

## Firma üst yönetimlerinde cinsiyet eşitsizliğinin bölgesel ve sektörel boyutları: Bir sınıflandırma ağacı ve rassal orman analizi

### The regional and industrial dimensions of gender inequality in firm top management positions: A classification tree and random forest analysis

Mehmet Güney Celbiş<sup>1,2</sup> 

<sup>1</sup> Dr. Öğretim Üyesi, Yeditepe Üniversitesi İİBF - Ekonomi Bölümü, İstanbul, Türkiye, [celbis@merit.unu.edu](mailto:celbis@merit.unu.edu)

ORCID: 0000-0002-2790-6035

<sup>2</sup> Bağlı Araştırmacı, Maastricht Economic and Social Research Institute on Innovation and Technology, Maastricht, Hollanda

#### Öz

İş gücü piyasasında cinsiyet eşitsizliği olgusu ve sürdürülebilir kalkınma ile ilişkisi Dünya’da ve Türkiye’de dikkat çeken bir konudur. Cinsiyet Eşitsizliğinin özel bir yansıması üst yönetici pozisyonlarında görülmektedir. Bu çalışmada Türkiye’de üst yönetici pozisyonları bağlamında cinsiyet eşitsizliğinin altında yatan etkenler irdelenmiş ve eşitsiz dağılıma yol açan durumların tespiti amaçlanmıştır. Ağaç-temelli makine öğrenmesi yöntemlerinden hareketle Türkiye’de firma üst yönetimlerinde cinsiyet eşitsizliği ile ilişkili kurumsal, sektörel ve bölgesel özellikler belirlenmiştir. Birçok başka sonucun yanı sıra, yolsuzluk, suç, arazi ve ruhsat sorunlarının yaşandığı firmalarda ve Kuzeydoğu ve Güneydoğu Anadolu bölgelerinde bulunan şirketlerde üst düzey yöneticilerin erkek olduğu görülmüştür. Bununla birlikte perakende, yiyecek-konaklama, tekstil, giysi ve imalat sektörlerinde ancak bazı bölgesel ve kurumsal şartlar altında kadın yöneticiler olduğu gözlemlenmiştir. Söz konusu bulgular üst yönetici pozisyonlarda kadın-erkek dengesizliğinin etkili olduğu sektör ve bölgeleri belirlemenin yanı sıra bu eşitsizliklere yol açan kurumsal faktörleri de gözlemlemektedir. Elde edilen bulgular spesifik belirleyicilerin analizi konusunda yeni araştırma kulvarları açmanın yanı sıra soruna yönelik yeni politikalar geliştirmede de etkili olma potansiyeline sahiptirler. Çalışmada sunulan modellerde Girişim Anketi-2019 verilerinden alınan 162 firma ve 254 değişkenin tümü algoritmik modeller ile analiz edilmiştir. Bulguların içeriksel önemine ek olarak bu çalışmada araştırma konusunun irdelenmesi yönünde ağaç bazlı makine öğrenmesi yaklaşımları yeni yöntemsel olanaklar olarak sunulmuştur.

**Anahtar Kelimeler:** Ekonomik Coğrafya, Cinsiyet Eşitsizliği, Firmalar, Üst Yönetim, Makine Öğrenmesi

**JeI Kodları:** R10, M0, L2, J7, C40

**Başvuru/Submitted:** 8/02/2021

**Revizyon/Revised:** 18/03/2021

**Kabul/Accepted:** 30/03/2021

**Yayın/Online Published:** 25/06/2021

**Atıf/Citation:** Celbiş, M.G., Firma üst yönetimlerinde cinsiyet eşitsizliğinin bölgesel ve sektörel boyutları: Bir sınıflandırma ağacı ve rassal orman analizi, bmij (2021) 9 (2): 439-455, doi: <https://doi.org/10.15295/bmij.v9i2.1777>

#### Abstract

The issue of gender inequality in the labour market is a topic attracting attention in Turkey and globally. One specific manifestation of gender inequality is observed in relation to top management positions. In the present study, the underlying effects that create gender inequality in top management positions are examined to identify the conditions that lead to an unbalanced distribution. Using tree-based machine learning methods, the present study identifies the institutional, industrial, and regional attributes related to gender inequality in top management positions of firms in Turkey. Alongside numerous other findings, we observe that the top managers in firms facing problems related to crime, corruption, access to land and licenses, and establishments located in Northeast and Southeast Anatolia tend to be male. In addition, we also observe that under certain regional and institutional conditions, top female managers are more common in the retail, restaurant and hotel, textiles, clothing, and manufacturing industries. Together with identifying the regions and industries with the highest female-male inequality, the findings also highlight the institutional factors that lead to these inequalities. The results, alongside opening new research paths regarding the analysis of specific determinants, also have the potential of helping the development of new policies to tackle the problem. In the models presented in this study, data on 162 firms and 254 variables presented in the Enterprise Survey-2019 are analyzed using algorithmic models. In addition to the content-related relevance of the findings, the present study presents tree-based machine learning methods as new methodological alternatives concerning assessing the research question.

**Keywords:** Economic Geography, Gender Inequality, Firms, Top Management, Machine Learning

**JeI Codes:** R10, M0, L2, J7, C40

## Extended Abstract

### The regional and industrial dimensions of gender inequality in firm top management positions: A classification tree and random forest analysis

#### Literature

##### Research subject

The subject of this research is the exploration of the underlying effects that create gender inequality in top management positions. These effects are examined to identify the conditions which lead to an unbalanced distribution. Using tree-based machine learning methods, the present study identifies the institutional, industrial, and regional attributes related to gender inequality in top management positions of firms in Turkey.

The article benefits from previous research primarily done in the areas of (i) sociology of the labour markets, (ii) the economic analysis of regional labour markets, (iii) innovation, (iv) management, and (v) organizational behaviour. Furthermore, a considerable portion of the literature is from machine learning as the methodology in this study is based on tree-based machine learning models. Consequently, the article brings together theoretical views and empirical findings from a wide range of areas in an interdisciplinary manner.

Due to this interdisciplinary nature, the literature review is not confined to a specific section but distributed across all paper sections. As a result, the study references more than sixty original research articles published either as journal articles, book chapters, reports, or conference proceedings

##### Research purpose and importance

Gender inequality in the labour market is a topic that has been attracting attention in Turkey and on a global scale. One specific manifestation of gender inequality is observed concerning top management positions. It is mainly of policy interest to understand the institutional and regional influences that create these inequalities. This study addresses this question through novel and state-of-the-art statistical learning techniques.

##### Contribution of the article to the literature

Alongside numerous other findings, we observe that the top managers in firms facing problems related to crime, corruption, access to land and licenses, and establishments located in Northeast and Southeast Anatolia tend to be male. In addition, we also observe that under certain regional and institutional conditions, top female managers are more common in the retail, restaurant and hotel, textiles, clothing, and manufacturing industries. Together with identifying the regions and industries with the highest female-male inequality, the findings also highlight the institutional factors that lead to these inequalities. The results, alongside opening new research paths regarding the analysis of specific determinants, also have the potential of helping the development of new policies to tackle the problem. In addition to the content-related relevance of the findings, the present study presents tree-based machine learning methods as new methodological alternatives with the assessment of the research question.

#### Design and method

##### Research type

This study, while being firmly grounded in established theoretical frameworks, adopts a data-driven quantitative empirical approach. The quantitative techniques used are heavily computerized and are established in the modern statistical and machine learning literature.

##### Research problems

The research aims to address the main problem is the differentiation and subsequent classification – through algorithmic prediction – of the firms based on whether they have a top female manager or not. On a second level, the aim is to understand what factors allow the successful classification of these firms. In other words, to explore the firm-level and regional features that are associated with the outcome classes.

##### Data collection method

In the models presented in this study, data on 162 firms and 254 variables presented in the Enterprise Survey-2019 are analyzed using algorithmic models. Therefore, the study uses secondary data. There has not been an effort directed towards the collection of primary data.

##### Quantitative/qualitative analysis

The study is strictly quantitative in nature. The quantitative methods are tree-based machine learning models.

##### Research model

The study uses two complementary models. (i) Classification tree is a component of the Classification and Regression Trees framework strongly established in machine learning. (ii) Based on the classification tree approach, a random forest analysis that includes many unpruned (unrestricted) classification trees is used. In order to investigate in detail the *ceteris paribus* effects of features, further diagnostic tools have been used. These are partial dependence plots and individual conditional expectation plots.

##### Research hypotheses

The research hypothesizes that firm-level institutional attributes, alongside regional determinants, play a significant role in influencing the presence of top female managers in firms.

## **Findings and discussion**

### **Findings as a result of analysis**

The analysis results find that the top managers in firms facing problems related to crime, corruption, access to land and licenses, and establishments located in Northeast and Southeast Anatolia tend to be male. Moreover, the findings also suggest that top female managers are more common in the retail, restaurant and hotel, textiles, clothing, and manufacturing industries under certain regional and institutional conditions. In addition to identifying the regions and industries with the highest female-male inequality, the findings also highlight the institutional factors that lead to these inequalities.

### **Hypothesis test results**

The hypotheses are tested through primarily variable importance scoring, individual conditional expectations and partial dependence values. Each variable in the dataset poses its hypothesis. While some are rejected and never selected by the algorithmic selection procedures, features related to crime, corruption, access to land and licenses, and is located in Northeast and Southeast Anatolia stand out as significant test results regarding the hypotheses on the roles of these features.

### **Discussing the findings with the literature**

The results, together with opening new research paths regarding the analysis of specific determinants, also have the potential of helping the development of new policies to tackle the problem. Alongside the content-related relevance of the findings, the present study presents tree-based machine learning methods as new methodological alternatives related to the existing research in the area.

### **Conclusion, recommendation and limitations**

#### **Results of the article**

In this article

- 1- it is observed that the top managers in firms facing problems related to crime, corruption, access to land and licenses, and establishments located in Northeast and Southeast Anatolia tend to be male.
- 2- it is observed that under certain regional and institutional conditions, top female managers are more common in the retail, restaurant and hotel, textiles, clothing, and manufacturing industries.
- 3- the regions and industries with the highest female-male inequality are identified
- 4- the findings underline the institutional factors that lead to these inequalities.
- 5- The results open new research paths regarding the analysis of specific determinants,
- 6- the findings have the potential of helping the development of new policies to tackle the problem.
- 7- tree-based machine learning is presented as new methodological alternatives that can be used in the research area.

#### **Suggestions based on results**

The results suggest that specific attention needs to be directed towards certain regions in relation to the presence of firms with top female managers. Regional effects may also be due to unmeasured influences such as travel times and distances to work. Moreover, the results suggest that we can expect to see more female top managers in a firm if problems such as crime, corruption, and disputed activities often involve illegal activities regarding land and license allocation.

#### **Limitations of the article**

The article relies on secondary data. This in itself poses limitations as essential features that are unmeasured in the data source cannot be assessed in our results. Machine learning methods perform exceptionally well when the data has high explanatory power. This being said, the author of this article does not have control regarding the data collection process. Therefore, the explanatory power of the analysis relies solely on the correct usage of the machine learning algorithms.

## Giriş

Karar verme mekanizmalarına kadınların aktif katılımı sürdürülebilir kalkınmanın belirleyici etkenlerinden biri olarak görülmektedir (Şahin, 2011; Ayyıldız Ünnü, Baybars ve Kesken, 2014). Buna paralel olarak ülkelerin kalkınmalarında kadın katılımının önemi Birleşmiş Milletler sürdürülebilir kalkınma hedeflerinde belirtilmiştir (Birleşmiş Milletler, 2019). Öte yandan firma üst yönetimlerinde kadın katılımının firma işleyişine farklı perspektifler getirdiği ve bundan doğan bilgi ve tecrübe çeşitliliğinin kalkınma ve inovasyonu canlandırdığı öne sürülmüştür (Barczak, Lassk ve Mulki, 2010; Ritter-Hayashi, Vermeulen ve Knobens, 2016; Becic ve Vojinic 2018). Ancak çoğu ülkede ve Türkiye’de kadınların üst yönetim pozisyonlarında bulunma düzeyleri son dönemlerde artmakla birlikte düşüktür (Smith vd., 2006; Örucü vd., 2007). Bu durumun bölgesel gelişmişlik ile ilgili bir boyutu olduğu da gözlemlenmiştir (Karatepe ve Arıbaş, 2017).

Kadınların iş hayatında karşılaştıkları bariyeri sembolize eden “cam tavan” kavramı literatürde sıkça dile getirilmektedir (Cotter vd., 2001). Bununla birlikte terfi süreçlerinde de erkek çalışanlara nazaran kadınların daha çok engel ile karşılaştıkları gözlemlenmiştir (Taşkın, 2012; Karcioğlu ve Leblebici, 2014). Başka bir bakış açısından ise cam tavanı aşarak yönetici konumuna gelen kadınların bir kısmının şirketin performansının kötüleşmesi sonucu bu pozisyonlara getirildiği öne sürülmüştür (Yıldız vd., 2016; Cook ve Glass, 2013). Bu tür gözlemlerden hareketle kadınların iş yaşamında bulunma ve yükselme olanakları, erkek çalışanlara kıyasla kazançları, iş yükleri ve daha birçok benzer konu uzun yıllardır yoğunlukla araştırılmaktadır (Brenner ve Levi, 1987; Jacobs, 1995; Boden, 1999; Adams ve Funk, 2012). Bu araştırmalar çoğunlukla işletme, örgütsel yönetim ve sosyoloji temelli bakış açıları ile yapılmaktadır ve genellikle kişi odaklı bir yapıya sahiptirler. Örneğin kadın yöneticilerin sorunlarına bir çalışma sosyolojisi bakış açısıyla yaklaşan Cohen ve Huffman (2007), belli hiyerarşik seviyelerdeki kadın-erkek dağılımının kadınların işe alım ve terfilerinde önemli rol oynadığını gözlemlenmişlerdir. Organizasyon kültürü bağlamında kadın yöneticilerin tabi olduğu engelleri inceleyen Rutherford, (2001) ise özellikle erkek odaklı kurumsal kültürün hâkim olduğu kurumlarda kadın yöneticilerin önünde büyük zorluklar olduğunu gözlemlenmiştir. Ekonomi odaklı araştırmalar göz önüne alındığında ise kadın yöneticilerin varlığının kurum finansal performansına olan etkisi öne çıkan bir konudur (Shrader vd., 1997; Smith vd., 2006; Campbell ve Minguez-Vera, 2007). Etkili çalışmalardan seçilmiş bu örneklerde görüldüğü üzere farklı disiplinler ve yaklaşımlar ile işlenen yöneticilik seviyesindeki cinsiyet eşitsizliği ve kadın katılımı konusunun işlenişine ilgili çalışmanın teorik kurgusuna bağlı olarak farklılık göstermektedir. Bu çalışmanın içerik ve yönetsel katkıları iki kısımda değerlendirilebilir. Çalışmanın hipotezi kurumsal ve yerel etkenlerin bir firmada kadın yönetici olma durumunu güçlü biçimde belirlediğini varsaymaktadır. Dolayısıyla ilk olarak, bu hipotezi test etme noktasında analiz biriminin kişiden ziyade şirket olması kurumsal özelliklerin etkilerinin değerlendirilmesine izin vermektedir. Bölge etkileri de katıldığında çalışma Türkiye bağlamında çalışma yeni bir bakış açısı getirmektedir ve istihdam ve girişimcilik konularında yapılmış çalışmalara ek olarak yöneticilik kademelerindeki eşitsizlikleri analiz etmek yolu ile de literatüre katkı sunmaktadır. Çalışmanın ikinci önemli katkısı ise söz konusu araştırma sorusunun cevaplanmasına yönelik kayda değer bir yönetsel yenilik sunmaktır. Geniş bir anket verisinden elde edilmiş değişkenlerin algoritmik seçimi ve değerlendirilmesi sayesinde bağımlı değişken kısıtlı modeller ile değil, tüm veriyi sistematik bir biçimde işleme almak yoluyla açıklanmaktadır. Bu yaklaşımı mümkün kılan makine öğrenmesi yöntemleri firma üst yönetiminde gözlemlenen cinsiyet eşitsizliği konusuna yönelik olarak ilk defa bu çalışmada kullanılmışlardır. Bu yöntemlerden hareketle firmaların kurumsal özelliklerine ve bölgelere göre üst yöneticinin kadın olma ihtimali sınıflandırma modelleri ile araştırılmıştır.

Çalışmada giriş bölümünün ardından, ikinci bölümde sosyoekonomik araştırma sorularının makine öğrenmesi yöntemleriyle irdelenmesinin faydaları tartışılmaktadır. Üçüncü ve dördüncü bölümler sırasıyla veri setini ve kullanılan analitik yöntemler ile sonuçlarını sunmaktadır. Beşinci bölümde ise çalışmanın sonuçları değerlendirilmiştir.

## Sosyoekonomik veri analizinde makine öğrenmesi

Literatürde bölgesel farklılıklar, cinsiyet eşitsizliği ve firma performansı konuları neredeyse her zaman klasik yöntemlerle irdelenmektedir. Bu geleneksel yöntemler ekonometri gibi sayısal yaklaşımları ve vaka analizi gibi kalitatif analiz yöntemlerini kapsamaktadır. Öte yandan algoritmik ve esnek yaklaşımlar karmaşık ve doğrusal olmayan sosyoekonomik etkileri araştırmada çok yardımcı olabilirler. Bu bağlamda makine öğrenmesi metodları kadınların üst yöneticilikteki mevcudiyetleri gibi karmaşık etkilere bağlı konuları anlamada ve yeni bilgileri günışığına çıkarmada avantajlar sunmaktadırlar (Varian, 2014; Mullainathan ve Spiess, 2017; Harding ve Hersh, 2018; Athey, 2018).

Hem bölgesel çalışmalar hem de cinsiyet eşitsizliği araştırmalarında makine öğrenmesi metodlarının kullanımı artmaya başlamakla birlikte halen çok nadirdir. Yakın zamanda, bölgesel bir bakış açısıyla bölgelerarası eşitsizlik ve yakınsama gibi konular üzerine yaptıkları çalışmalarda Chang vd. (2014); Bang vd. (2017); Nyman ve Ormerod (2017) ve Nyman ve Ormerod (2020) regresyon ağaçları, yapay sinir ağları, rastgele örnekleme agregasyonu, gradyan artırma ve rassal orman yöntemlerini kullanmışlardır. Cinsiyet eşitsizliği konusunda ise makine öğrenmesi uygulamaları daha da nadirdir. Briel ve Töpfer (2020) kadın ve erkeklerin kazanç farklılıklarını tahmin edebilmek için En Az Mutlak Büzülme Seçici Operatörü metodunu kullanmış ve çalışılan iş kolunun özelliklerinin önemli belirleyiciler arasında olduğunu gözlemiştir. Verma vd. (2018) ise Avrupa'da öğretmenleri cinsiyetlerine göre sınıflandırmada rassal orman ve Bayes ağlarını kapsayan yöntemler kullanmıştır.

Cinsiyet eşitsizliği veya bölgesel çalışmalar alanlarının yanı sıra, genel olarak sosyoekonomik konuları araştırmada makine öğrenmesi metodlarının kullanımı özellikle 2017 öncesi dönemde çok az olmuştur. İşsizlik oranlarının tahmini için yapay sinir ağları modellerini kullanan Xu vd. (2013) ve Çin vatandaşlarının seyahat tercihlerini araştıran Zhang ve Zhang, (2014) tarafından yapılan çalışmalar 2017 öncesi için bazı örnekler olarak verilebilir. Büyük veri edinme olanaklarının yaygın hale gelmesi sosyal bilimlerde makine öğrenmesi kullanım fırsatlarının artmasını sağlamıştır. Örneğin uydu verilerinin kullanımı buna örnek olarak verilebilir. Glaeser vd. (2018) Google uydu verilerini kullanarak destek vektör makineleri (support vector machines) aracılığıyla New York şehri için gelir seviyelerini tahmin etmiştir. Robinson vd. (2017) ise uydu verilerini yapay sinir ağları metoduyla analiz ederek ABD için nüfus tahminleri yapmıştır. Öte yandan Peru'daki çocuk işçiliğinin ve Darüsselam şehrinde gönüllü kayıt dışı istihdamın yaygınlıklarını Lopez-Yucra vd. (2018) ve Kaiser (2018) sırasıyla yapay sinir ağları ve denetimsiz Bayes makine öğrenme tekniklerini kullanarak araştırmışlardır. Makine öğrenmesi yaklaşımları göç alanında da kullanılmıştır. Liu (2019) ve Iman ve Tarasyev (2018) işgücü piyasası ve göç araştırmalarında denetimsiz makine öğrenmesi ve "En Yakın K-Komşu" (KNN) yaklaşımlarını kullanmışlardır. Göç, işsizlik ve nüfus gibi konulara ek olarak suç araştırmalarında da makine öğrenmesi metodları kullanılmaktadır. Örneğin Mittal vd. (2019) Hindistan'da suç oranları tahmininde yapay sinir ağları, rassal orman ve karar ağacı metodlarını kullanmışlardır.

## Veri seti

Bu çalışmada, Dünya Bankası, Avrupa İmar ve Kalkınma Bankası ve Avrupa Yatırım Bankası'nın Türk firmaları üzerine ortaklaşa olarak yaptığı "Girişim Anketi 2019" (Enterprise Survey 2019) verisi kullanılmıştır. Birçok başka ölçümün yanı sıra söz konusu ankette firma temsilcilerine firma üst yöneticisinin kadın olup olmadığı sorulmuştur. Anket 1,663 firmayı kapsamaktadır ve ham haliyle sorulara verilen cevapları temsilen 433 değişkenden oluşmaktadır. Doğal olarak bu değişkenlerin önemli bir kısmı idari bilgileri temsil etmektedir ve bu tür değişkenler çalışmada yapılan analizlerin dışında tutulmuştur (örneğin anketin yapıldığı gün, saat, anketi yapan kişinin kodu vb.). Buna ek olarak, makine öğrenmesi algoritmalarında kullanılmak üzere maksimum firma sayısını muhafaza etmek amaçlanmıştır.

Yüksek değişken sayısı dolayısıyla eksik gözlemi olan değişkenler ve bu değişkenlerin kombinasyonları veri setinden büyük miktarda gözlemin kaybolmasına yol açmaktadır. Hatta ham haliyle veri setinde tüm değişkenler için gözlemlerin tam olduğu tek bir firma bile yoktur. Bu duruma karşı bir önlem olarak gözlemlerinin %10'undan fazlasının eksik olduğu değişkenler analizlerden hariç tutulmuştur. Ancak bu noktada başka bir sorun daha bulunmaktadır. Çalışmanın birinci bölümünde tartışıldığı üzere kadın yönetici oranı oldukça düşüktür. Yukarıda belirtilen adımlardan sonra veri setinde kalan firma sayısı 1,165 olmasına rağmen bu şirketlerin sadece 81'inin üst yöneticisi kadındır. Makine öğrenmesi modelleri öngörü yapmada maksimum verimi tercih ederler. Veri setindeki kayda değer dengesizlik yüzünden tüm şirketlerin yöneticisinin erkek olduğunu tahmin eden bir model bile %93'lük bir başarıya ulaşacaktır ancak bu modellerin açıklayıcılığı düşük olacaktır. Bu dengesizliği gidermek üzere üst yöneticisi erkek olan 1084 firmadan 81 şirketten oluşan bir basit rastgele örnekleme çekilmiştir. Veri düzenlemesinde son adım olarak çok sayıda kategoriden oluşan değişkenlerdeki her bir kategori 0 ve 1 değerlerinden oluşan ikili kod şekline dönüştürülmüştür. Bu değişkenler sırasıyla işletmenin Uluslararası Standart Sanayi Sınıflaması (ISIC) kodu, firmanın ana ürünü/hizmeti ve firmanın yöneticilerinin işletmenin işleyişinin önündeki en büyük engeli olduğunu düşündüğü sorunu belirten değişkendir. Sonuç olarak kadın veya erkek üst yöneticisi olan firma sayısının iki durum için de 81 olduğu 162 firma ve 254 değişkenden oluşan veri seti makine öğrenmesi modellerinde kullanıma hazır hale getirilmiştir.

## Sınıflandırma ağacı ve rassal orman uygulamaları ve bulguları

Sınıflandırma ağaçları, diğer adıyla karar ağaçları, Breiman vd. (1984) tarafından geliştirilen ve “Sınıflandırma ve Regresyon Ağaçları” olarak bilinen algoritmik yöntemlerden biridir. Bu algoritma eğitim verisi üzerine uygulanan özçagrılı (rekürsif) ikili bölünme işlemi üzerine temellidir.<sup>1</sup> Firma üst yönetimlerinde kadın yönetici olup olmadığını öngörebilmek için veri öncelikle eğitim ve test verisi olarak rastgele biçimde ikiye bölünmüştür. Eğitim verisi tam veri setinin rastgele seçilmiş %70’i olarak belirlenirken diğer %30 ise test verisi olarak ayrılmıştır.<sup>2,3</sup>

Bağımlı değişkenin firma üst yöneticisinin kadın olup olmadığını “Evet” ve “Hayır” şeklinde iki kategori ile temsil ettiği sınıflandırma ağacı Breiman vd. (1984), Friedman vd. (2001) ve James vd. (2013) tarafından ortaya konulan yöntemsel çerçeveye uygun olarak oluşturulmuştur. Bu çerçeveye göre bir ağaçtaki her  $m$  ağaç durağının (node) çeşitlilik (impurity) ölçütü bir Gini endeksi ( $G_m$ ) biçiminde tanımlanmıştır:<sup>4</sup>

$$G_m = \sum_{k=1}^K p_{mk}(1 - p_{mk})$$

öyle ki,

$$p_{mk} = \frac{1}{N_m} \sum_{i \in D_m} \mathbf{1}(y_i = k) \quad (1)$$

Yukarıdaki denklemlerde  $p_{mk}$  terimi  $m$ 'inci ağaç durağında  $i$  firması için gözlenen çıktının ( $y_i$ ) “Evet” ve “Hayır” sınıflarından birisi olan  $k$  sınıfına ( $k = 1, \dots, K$ ) denk gelme oranını vermektedir;  $D_m$  terimi  $m$ 'inci durağa tekabül eden firma kümesi ve  $N_m$  bu firmaların sayısıdır. Bir başka deyişle, eğer  $i$  firması sınıflandırma ağacı tarafından  $D_m$  kümesine yerleştirildiyse ve  $k$  sınıfına giriyorsa bu kümenin bulunduğu ağaç durağındaki aynı sınıftaki firmalar ile birlikte duraktaki  $k$  sınıfına giren firma sayısının duraktaki toplam firma sayısına oranını hesaplamada kullanılır.  $G_m$  değeri ise  $m$  durağında her bir  $k$  sınıfı için  $k$  sınıfına ait gözlemler ve bu sınıfa ait olmayan gözlemlerin oranlarının çarpımlarının toplamından oluşur. Dolayısıyla  $G_m$  ne kadar düşük ise  $m$  durağındaki gözlemler o kadar benzeşmektedir (James vd., 2013; Géron, 2019). Bir başka deyişle aynı sınıfa denk gelen gözlemler çoğunlukta. Bu çalışmada kadın üst yöneticiye sahip olan ve olmayan firmalar “Evet” ve “Hayır” şeklinde iki sınıfta gruplandırıldığı için  $G_m$  terimi  $2p_{m1}(1 - p_{m1})$  olarak hesaplanabilir. Sonuç olarak algoritma, özçagrılı ikili bölünme adımlarının her birinde veriyi alt duraklar  $m_1$  ve  $m_2$ 'nin bu duraklara düşen firma sayısının toplam firma sayısı  $N$ 'ye oranlarıyla ağırlıklandırılmış çeşitlilik ölçütlerinin toplamını minimize edecek şekilde öznelik uzayından ( $j = 1, \dots, J$ ) bir bölme değişkeni  $z_j$  ve bu değişkenin bölme değeri  $s_j$ 'yi seçerek ve böler (Breiman vd., 1984; Friedman vd., 2001; James vd., 2013):

$$\min_{j,s} \left[ \frac{N_{m_1}(j,s)}{N} G_{m_1}(j,s) + \frac{N_{m_2}(j,s)}{N} G_{m_2}(j,s) \right] \quad (2)$$

Eğitim verisini yukarıda özetlenen biçimde bölmeye devam etmek ve öngörülerini buna göre yapmak aşırı uyumluluk sorununa yol açacaktır. Bir başka deyişle test verisindeki bilgiyi kullanarak ikili bölünme işlemini duraklara düşecek gözlem kalmayana kadar devam ettirmek eldeki veri içerisindeki gözlemleri çok yüksek bir başarı ile tahmin edebilecek, ancak örneklem dışı veri bağlamında performans düşük olacaktır. Aşırı uyumluluk sorununu gidermek için sınıflandırma ve regresyon

<sup>1</sup> Bu çalışmada makine öğrenmesi terimlerinin Türkçe karşılıkları için Ethem Alpaydın'ın öne sürdüğü önerilerden yararlanılmıştır (Alpaydın, 2011).

<sup>2</sup> Bu çalışmada R yazılım dili kullanılmıştır. Analizleri uygulamada kullanılan R paketleri şunlardır: sınıflandırma ağaçları için Atkinson & Therneau (2000) tarafından geliştirilen “rpart”, bu ağaçların görselleştirilmesi için “train” (Kuhn, 2008), rassal orman uygulaması için “randomForest” (Liaw & Wiener, 2002) ve kısmi bağımlılık ve bireysel koşullu beklenti göreselleri için “pdp” (Greenwell, 2017).

<sup>3</sup> Bu çalışmada kullanılan veri seti 2019 yılında yayınlanmıştır. Veri toplama zamanı 1 Ocak 2020'den önce olmasından dolayı bu çalışma için etik kurul izin belgesi gerekmemektedir.

<sup>4</sup> Bazı kaynaklarda “node” terimi için Türkçe “düğüm” karşılığı, “impurity” için ise “katışıklık” sözcüğü kullanılmıştır. Türkçe makine öğrenimi çalışmalarının - özellikle sosyal bilimler alanında - kılığını göze alarak, bu çalışmada anlaşılması daha kolay karşılıklar kullanabilme amacıyla “node” için “durak” ve “impurity” için “çeşitlilik” karşılıkları kullanılmıştır. Bu seçimlerin birincisinin sebebi “node” adı verilen noktalarda herhangi bir düğümleme aktivitesi olmamasıdır. “Tree node” veya “terminal node” adı verilen noktalar verinin bölündükten sonra gruplandırıldığı veri bölgelerine ağaç temsiline karşılık gelen kümelerle verilen isimdir. Dolayısıyla bu veri bölgeleri ikili bölünme işleminin durakları olarak görülebilirler. “Impurity” ise belirli bir durakta oluşan kümelerin ne derecede aynı sınıfa ait olmadığını belirten ölçümdür. “Saf olmama” veya “tek çeşit olmama” derecesi olarak görülebilir. Bu yüzden bu ölçüme karşılık “çeşitlilik” sözcüğü kullanılmıştır.

ağaçları 10 katmanlı çapraz doğrulama (10-fold cross-validation) yöntemi ile kısıtlanır. Budama (pruning) adı verilen bu işlemin ana hatları Friedman vd. (2001), Sutton (2005) ve James vd. (2013) tarafından belirtildiği üzere şu şekilde tanımlanır: bir  $l$  alt ağacının her bir son durağı  $\bar{m}$  içerisinde ( $\bar{m} = 1, \dots, |l|$  ve  $l = 1, \dots, L$ ) çoğunluğa sahip olan sınıf ( $k^*$ ) ise  $D_{\bar{m}}$  kümesindeki hatalı tahmin oranı

$$u_{\bar{m}} = \frac{1}{N_{\bar{m}}} \sum_{i \in D_{\bar{m}}} \mathbf{1}(y_i \neq k^*) \quad (3)$$

olarak ifade edilebilir. Bu durumda eğitim verisi 10 katmana bölünüp her bir katman için  $\lambda$  olarak tanımlanan karmaşıklık parametresine denk gelecek bir dizi alt ağaç oluşturulduktan sonra en düşük hata oranını veren  $\lambda$  değeri bulunur. Diğer bir deyişle denklem 4'te belirtildiği gibi hata oranını mümkün olduğunca düşüren ancak bunu yaparken ağacın toplam son durak sayısını da minimize eden karmaşıklık parametresi belirlenir. Örneğin eğer  $\lambda$  sıfır ise ağacın karmaşıklığının dikkate alınmadığı ve sadece toplama işaretinden önceki terimin minimize edildiği, yani ağacı hiç budanmadan ikili bölme işleminin en son aşamaya kadar devam etmesine izin verildiği anlamına gelir. Daha yüksek  $\lambda$  değerleri ise ağaç büyüklüğünün cezalandırıldığı anlamına gelir ve aşırı uyumu engellemek için belirlenen  $\lambda$  değerine denk gelecek şekilde sınıflandırma ağacının büyümesinin durdurulması gerektiğini saptar:

$$\sum_{\bar{m}}^{|l|} u_{\bar{m}} \frac{N_{\bar{m}}}{N} + \lambda |l| \quad (4)$$

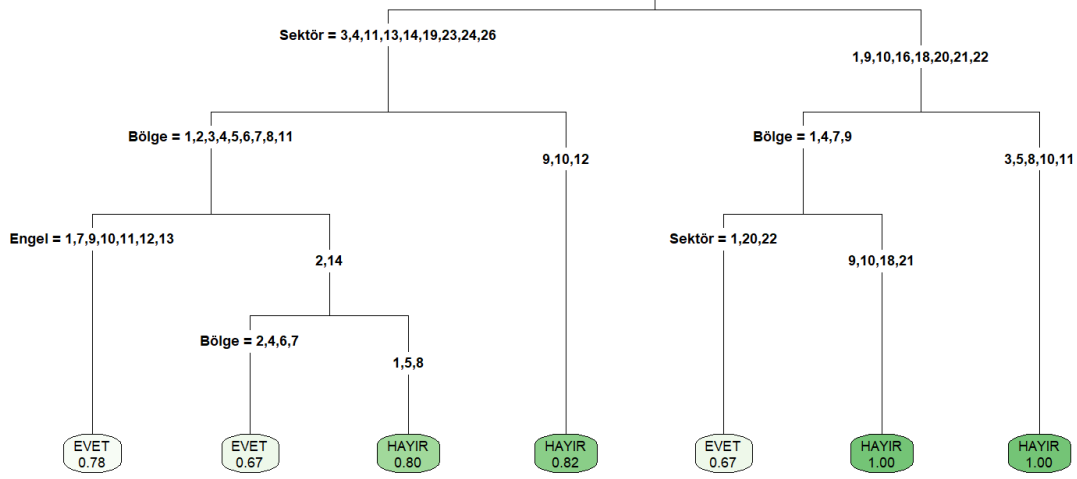
Örneklem dışı tahmin başarısını destekleme amacıyla aşırı uyumluluğu sınırlandırmak için  $\lambda$ 'nın yanı sıra her ağaç durağında bulunabilecek minimum gözlem sayısına ve de ağacın azami derinliğine sınırlamalar getirilir. Bu iki parametrenin değerlerinin en uygun kombinasyonunu belirlemek amacıyla 2 ve 20 sayıları arasında bir tarama gerçekleştirilmiştir ("grid search" yöntemi). Sonuç olarak eğitim verisindeki gözlemlerin aşırı doğru tahminini hedeflemekten ziyade test verisindeki gözlemlere genelleştirme amaçlanmıştır. Sonuç olarak sınıflandırma ağacı belirlenen şu üç parametreye tabi olarak Şekil 1'de görülebileceği şekilde ortaya çıkmıştır:  $\lambda = 0.01$ , ikili bölme için gerekli asgari gözlem sayısı = 5, azami ağaç derinliği = 8.

Şekil 1'deki sınıflandırma ağacının 254 değişken arasından seçtiği değişkenlerin tanımları Tablo 1'de verilmiştir. Şekilde gösterilen ağaçtaki her bir son durakta durağın içerisine düşen gözlemlerin EVET veya HAYIR sınıflarından hangisine ait olarak tahmin edildikleri ve o sınıfın duraktaki oranı belirtilmiştir. Bu bilgiler aynı zamanda renkler ile de temsil edilmiştir: daha açık renkler EVET sınıfının daha yoğun olduğu kümeleri, daha koyu renkler ise HAYIR sınıfının daha yoğun olduğu kümeleri göstermektedir.

Öncelikle, sadece tek bir sınıflandırma ağacının birçok eksikliğe sahip olduğunu hatırlamak önem arz etmektedir. Örneğin, farklı firmalardan oluşan bir veri seti çok farklı bir ağaç ortaya çıkabilir. Aynı zamanda bu algoritma değişkenler arası korelasyonun yüksek olduğu durumlarda önemli değişkenleri analizin dışında bırakabilir (James vd., 2013; Athey ve Imbens, 2019). Dolayısıyla Şekil 1'de sunulan sınıflandırma ağacı, aşağıda sunulacak rassal orman analizinin temelini oluşturan bireysel ağaçlara bir örnek olarak görülebilir.

Şekil 1'deki sınıflandırma ağacı 254 değişken arasından sektör, bölge ve firmaların karşılaştıkları engelleri belirten değişkenleri seçmiştir. Ancak unutulmamalıdır ki, yukarıda belirtildiği üzere bu tek ağaç örneğinin ampirik açıdan yetersizlikleri vardır ve bu yüzden bir topluluk modeli (ensemble model) olan rassal orman algoritması bu çalışmada uygulanmıştır. Buna rağmen sınıflandırma ağacı, bağımlı değişken ve geri kalan değişkenler arasında doğrusal olmayan bir ilişki olduğunu gözler önüne sermektedir. Ortaya çıkan sınıflandırma ağacı ilk ikili bölmeyi Sektör değişkenini kullanarak yapmış ve Tablo 1'de tanımlanan sektör kodlarına göre veriyi ayırmıştır. Her iki alt durakta ise bölünme Bölge değişkenine göre yapılmıştır. Ağacın sol kısmına bakıldığında Doğu Karadeniz, Kuzeydoğu Anadolu ve Güneydoğu Anadolu'da yer alan tekstil, giysi, mineral, fabrikasyon metaller, makine ve ekipman, geri dönüşüm, perakende, otel ve restoran ve bilgi işlem sektörlerinde iş yapan firmaların üst yöneticilerinin erkek olduğunun tahmin edildiği görülmektedir. İstanbul, Batı Anadolu ve Batı Karadeniz'de ise vergi oranları ve araziye erişim engelleri ile karşı karşıya olan firmaların erkek üst yöneticilere sahip olduğu öngörülmektedir. Aynı engellere tabi olan firmalar Batı ve Doğu Marmara, Akdeniz ve Orta Anadolu'da iseler kadın yöneticiye sahip oldukları tahmin edilmiştir. Engel durumuna bakılmaksızın yukarıda listelenen sektörlerde iş yapan geri kalan firmaların üst yöneticilerinin kadın olduğu öngörülmüştür.

Sınıflandırma ağacına göre bir firmanın üst yöneticisinin kadın olma ihtimali gıda, kimyasal ürünler, plastik ve kauçuk, hassas aletler, mobilya, inşaat, servis ve motorlu araçlar ve toptan satış sektörlerinde çok daha düşüktür; sadece İstanbul, Doğu Marmara, Orta Anadolu ve Doğu Karadeniz’de gıda, inşaat ve toptan satış sektörlerinde - daha az bir kesinlikle - kadın üst yöneticilerin varlığı öngörülmektedir. Bundan hareketle sektörel etkenlerin yanı sıra Türkiye’de firmaların kadın üst yöneticiler tarafından yönetilme durumunun belirgin bir bölgesel boyuta sahip olabileceğine yönelik bir işaret gözlemlenmektedir. Bu bulgular aşağıda doğruluk oranı daha yüksek olan rassal orman algoritması ile kapsamlı olarak test edilmektedir. Yukarıda özetlenen eksikliklere tabi olan bu sınıflandırma ağacının test verisi üzerindeki başarı oranı (accuracy) %54’tür.



Şekil 1: Sınıflandırma Ağacı

Kaynak: Yazar tarafından üretilmiştir.

Tablo 1: Değişken Tanımları: Sınıflandırma Ağacı

Değişken	Tanım	Değerler
Sektör	Firmanın ait olduğu sektör.	Kategorik değişken 1: Gıda, 2: Tütün, 3: Tekstil, 4: Giysi, 5: Deri, 6: Ahşap, 7: Kâğıt, 8: Yayın/baskı/kayıt, 9: Kimyasal ürünler, 10: Plastik ve Kauçuk, 11: Mineral, 12: Temel metaller, 13: Fabrikasyon metaller, 14: Makine ve ekipman, 15: Elektronik, 16: Hassas aletler, 17: Taşımacılık makineleri, 18: Mobilya, 19: Geri dönüşüm, 20: İnşaat, 21: Servis ve motorlu araçlar, 22: Toptan satış, 23: Perakende, 24: Otel ve restoran, 25: Taşımacılık, 26: Bilgi işlem.
Bölge	İşletmenin bulunduğu bölge, TÜİK İBBS-1 Düzey sınıflandırması.	Kategorik değişken. 1: İstanbul, 2: Batı Marmara, 3: Ege, 4: Doğu Marmara, 5: Batı Anadolu, 6: Akdeniz, 7: Orta Anadolu, 8: Batı Karadeniz, 9: Doğu Karadeniz, 10: Kuzeydoğu Anadolu, 11: Orta Doğu Anadolu, 12: Güneydoğu Anadolu.
Engel	İşletmenin işleyişine en büyük engeli oluşturduğu belirtilen kategori.	Kategorik değişken. 1: Finansal kaynaklara erişim, 2: Arazi (mekânsal) erişim, 3: Ruhsat/izin belgesine erişim, 4: Yolsuzluk, 5: Mahkemeler, 6: Suç, hırsızlık ve asayiş 7: Gümrük ve ticaret regülasyonları, 8: Elektrik, 9: Düşük eğitilmiş işgücü, 10: İşgücü piyasası düzenlemeleri, 11: Siyasi istikrarsızlık, 12: Kayıt dışı rekabet, 13: Vergi yönetimi, 14: Vergi Oranları, 15: Taşımacılık.

Kaynak: Yazar tarafından üretilmiştir.

Yukarıda bir örneği oluşturulan sınıflandırma ağacının çok sayıda farklı versiyonunu oluşturup bunların toplamından hareketle tahminler yapan algoritmalara topluluk modelleri denir. Breiman, (2001) tarafından geliştirilen rassal orman yaklaşımı daha eski olan rasgele örnekleme agregasyonu (bootstrap aggregation) modeline dayanır (Breiman, 1996). Rassal orman, sınıflandırma veya regresyon



ağaçlarına stokastisite katarak  $T$  sayıda ağacın her biri için eğitim verisinden  $N$  büyüklüğünde rastgele örneklem alır ( $t = 1, \dots, T$ ). Ancak yukarıda tek bir ağaç için görüldüğü gibi, bu durumda da değişkenler arasında yüksek korelasyonlar bazı önemli değişkenlerin analiz dışında kalmasına neden olabilir (Athey ve Imbens, 2019; James vd., 2013). Buna ek olarak, tüm sınıflandırma ağaçları aynı değişkenlerden seçim yaparlarsa kendileri arasında da korelasyon olabilir (Friedman, 2001; James vd., 2013). Sınıflandırma ağaçları arasında oluşacak korelasyonu engellemek için (dekorelasyon) rassal orman analizi her ikili bölme noktasında tüm  $Z$  değişken seti içerisinde sadece  $\sqrt{Z}$  kadarını seçer ve böylece değişkenlerin bir kısmını bu ikili bölme işlemi gerçekleştirmek için kullanılan ve Denklem 2’de belirtilen minimizasyon sırasında dikkate almaz (Breiman 2001; Friedman vd., 2001). Son olarak  $i$  firmasının üst yöneticisinin kadın olup olmadığı ortaya çıkan tüm  $T$  sınıflandırma ağaçlarının çoğunluk oyu ile tahmin edilir.

Rassal orman analizinde test verisi öngörülürken oluşacak hatanın kestirimi, eğitim verisi içerisindeki her  $i$  firmasının o firmayı dışarıda bırakan örneklem ile oluşturulan sınıflandırma ağaçlarının toplamının yaptığı tahmine bakarak yapılır. Bir başka deyişle, rassal orman içerisindeki her ağaç veriden rasgele bir örneklem seçtiği için bazı firmalar belirli bir ağaç grubu tarafından hiç seçilmemiş olacaklardır. Bu gözlemlere torba dışı gözlem (out-of-bag observation) adı verilir. Bu yöntem yukarıda tartışılan çapraz doğrulama yöntemiyle benzer fonksiyona sahip olmak ile birlikte, gözlemleri veri dilimleri şeklinde dışarıda bırakarak değil, rassal orman algoritması sırasında kendiliğinden oluşan bu torba dışı gözlemleri kullanarak hatayı tahmin eder (James vd., 2013).

Sınıflandırma ağaçlarının sayısının çokluğundan dolayı tek bir ağaç modelinde olduğu gibi rassal ormandaki tüm ağaçları görselleştirmenin açıklayıcılığı yoktur. Ancak rassal orman algoritması oldukça kullanışlı olan “değişken önemi” (variable importance) adı verilen ölçümü yapabilir. Bir değişken rassal ormandaki farklı ağaçlarda farklı şekillerde kullanılmış, bazıları tarafından ise hiç kullanılmamış olabilir. Bununla birlikte bu değişkenin toplamda tahminlerin yapılabilmesine ne derecede katkıda bulunduğu ölçülebilir. Denklem 2’de gösterildiği üzere, her bir ikili bölme adımında amaç duraktaki çeşitliliği (impurity) ölçütünü azaltmaktır. Bir değişkenin çeşitlilik ölçütünü tüm  $T$  sınıflandırma ağaçları genelinde ne kadar azalttığını belirlemek için öncelikle  $m'$ inci ağaç durağının  $m_1$  ve  $m_2$  olarak iki yeni durağa bölünmesi sonucu çeşitlilik azalması şu şekilde ifade edilebilir:

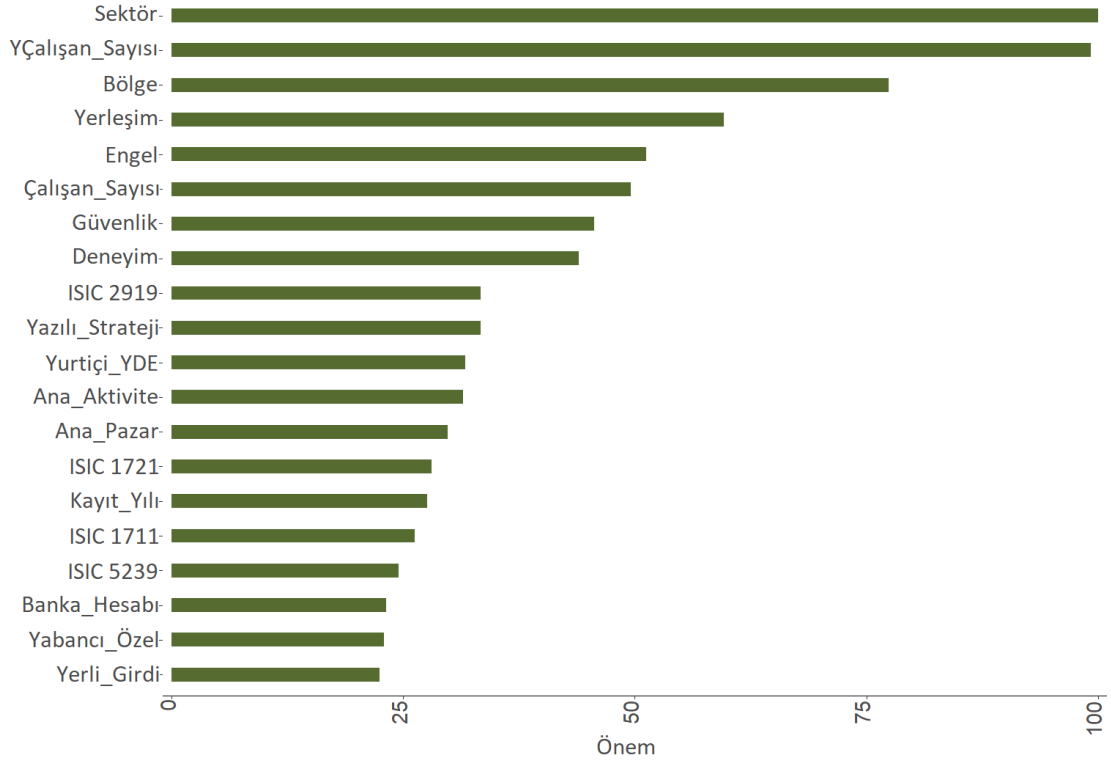
$$\Delta\text{Çeşitlilik} = G_m - \left[ \frac{N_{m_1}}{N} G_{m_1} + \frac{N_{m_2}}{N} G_{m_2} \right] \quad (5)$$

Yukarıdaki denklemde  $\Delta$  operatörü  $m'$ inci durağın  $m_1$  ve  $m_2$  duraklarına bölünme öncesi ve sonrası çeşitlik ölçütünde oluşan farkı temsil etmek amacıyla kullanılmıştır.  $\Delta\text{Çeşitlilik}$  değerini 1 ve 100 arasına ölçeklendirerek değişkenlerin tahmin başarısına katkılarını karşılaştırabilmek mümkün olmaktadır (James vd., 2013). Şekil 2’de rassal orman algoritmasının firma üst yöneticilerin cinsiyetini tahmin etmekte olan katkıları nedeniyle en çok önemi atfettiği ilk yirmi değişkenin önem sıralamalarını ve dereceleri görülmektedir. Uygulanan rassal orman algoritması ile 500 rassal ağaç oluşturulmuştur ve test verisi üzerindeki başarı oranı %60,4’tür.

Şekil 2’de görülebileceği gibi, daha önce oluşturulan sınıflandırma ağacına paralel olarak 500 rassal sınıflandırma ağacı da toplu olarak en çok Sektör değişkeninden faydalanmışlardır. Yani Denklem 5’de belirtilen ifadeyi en fazla bu değişken yükseltmiştir. Ancak daha önce üretilen tek sınıflandırma ağacının göstermediği yeni bilgiler rassal orman analizi sonucunda gözlemlenmektedir. Örneğin, Sektör değişkeninden sonra gelen ve tüm algoritmanın seçtiği bütün değişkenler ile birlikte Tablo 2’de tanımlanan *YÇalışan Sayısı* en yüksek ikinci öneme sahip değişken olarak belirlenmiştir. Değişken önemi çıktısında bu değişken ile firma üst yöneticisinin cinsiyeti arasında nasıl bir ilişki olduğu görülememektedir. Bu nedenden dolayı, analizin ilerleyen kısımlarında bu değişken ile birlikte diğer bazı seçilmiş değişkenlere daha derinden bir bakış sağlamak amacıyla kısmi bağımlılık ve de bireysel koşullu beklenti eğrileri incelenmektedir.

Rassal orman algoritması, yine tek sınıflandırma ağacı analizine benzer olarak Bölge değişkenini yoğun ve etkili biçimde kullanmıştır. Seçilmiş sayısal değişkenleri yukarıda sözü geçen araçlarla daha detaylı incelemek mümkündür. Kategorik bir değişken olan Bölge değişkenini de mekansal kısmi bağımlılık sonuçları aracılığıyla bu bölümün sonunda incelenmektedir. Geri kalan ve yüksek önem arz eden değişkenler Tablo 2’de tanımlanmışlardır ve sırasıyla yerleşim biriminin büyüklüğü, işletmenin önündeki engel cinsi, başlangıçtaki çalışan sayısı, güvenlik harcamaları, üst yönetiminin deneyim durumu, firma aktivitesinin genel amaçlı makine imalatı olması (ISIC 2919), firmanın yazılı bir stratejisi olup olmadığı, satışların yurtiçi payı, sektör içi ana aktivite dalı, firmanın ana pazarı, firmanın tekstil

sektöründe olması, kayıt yılı, perakende satış yapmak, banka hesabı olması, yerli girdi ve yabancı şahıslara aidiyet oranı olarak belirlenmişlerdir.



**Şekil 2:** Rassal Orman Değişken Önemleri

**Kaynak:** Yazar tarafından üretilmiştir.

**Tablo 2:** Değişken Tanımları: Rassal Orman

Değişken	Tanım	Değerler
<i>Ana_Aktivite</i>	İşletmenin sektör içi ana aktivite dalı.	Kategorik değişken. <b>1:</b> İmalat, <b>2:</b> Perakende, <b>3:</b> Toptancılık, <b>4:</b> İnşaat, <b>5:</b> Makine ve Ekipman, <b>6:</b> Otel/Restoran, <b>7:</b> İnşaat, <b>8:</b> Hizmet.
<i>Ana_Pazar</i>	İşletmenin ana Pazar tipi.	Kategorik değişken. <b>1:</b> Yerel, <b>2:</b> Ulusal, <b>3:</b> Uluslararası.
<i>Banka_Hesabı</i>	İşletmenin bir banka hesabı bulunması durumu.	Kategorik değişken (Evet-Hayır).
<i>Çalışan_Sayısı</i>	İşletmenin operasyona başladığı yıldaki çalışan sayısı.	1 - 2.200 arası sayısal değişken.
<i>Deneyim</i>	İşletmenin en üst yöneticisinin yıl bazında tecrübesi.	3 - 65 arası sayısal değişken.
<i>Güvenlik</i>	Son mali yıl içerisinde işletmenin güvenlik harcaması yapma durumu.	Kategorik değişken (Evet-Hayır).
<i>ISIC</i>	İşletmenin Uluslararası Standart Sanayi Sınıflaması (ISIC Rev. 3.1) kodu.	Her bir sınıflama için ikili değişken (0-1) tanımlanmıştır.
<i>Kayıt_Yılı</i>	İşletmenin kuruluş/kayıt yılı.	Sıralı değişken (yıl).
<i>Yabancı_Özel</i>	İşletmenin yabancı özel şahıslara aidiyet oranı.	Sayısal değişken (yüzde cinsinden).
<i>Yazılı_Strateji</i>	İşletmenin yazılı bir resmi iş strateji metnine sahip olma durumu.	Kategorik değişken (Evet-Hayır).
<i>YÇalışan_Sayısı</i>	Son mali yıl itibarıyla işletmenin çalışan sayısı.	2 - 4.100 arası sayısal değişken.
<i>Yerleşim</i>	İşletmenin bulunduğu yerleşim biriminin nüfus büyüklüğü.	Sıralı değişken. <b>1:</b> 50.000 - 250.000 arası, <b>2:</b> 250.000 - 1.000.000 arası, <b>3:</b> 1.000.000 ve üzeri.
<i>Yerli_Girdi</i>	Son mali yıl itibarıyla İşletmenin üretimine olan yerli girdi oranı.	Sayısal değişken (yüzde cinsinden).
<i>Yurtiçi_YDE</i>	İşletmenin satışlarının yurtiçi payı.	Sayısal değişken (yüzde cinsinden).

**Kaynak:** Yazar tarafından üretilmiştir.

Rassal orman algoritması tarafından seçilmiş değişkenlerin rolünü kısmi bağımlılık ve bireysel koşullu beklenti eğrilerini inceleyerek daha detaylı olarak görebiliriz. Tüm değişkenler için bu analizleri yapmak tek çalışmada mümkün olmayacağı için kategorik değişkenlerden *Bölge* ve *Engel*, sayısal değişkenlerden ise *YÇalışan Sayısı* ve *Deneyim* değişkenlerini kullanarak firmaların bölgesel ve

kurumsal özelliklerine bir bakış sunulmaktadır. Kısmi bağımlılık Friedman (2001) tarafından geliştirilmiş bir değerlendirme yöntemidir. Sayısal değişkenler için genelde seçilmiş bir veya iki değişkene odaklanılır ve tüm  $i$  gözlemler için diğer değişken değerleri olduğu gibi sabit tutulur. Ardından seçilmiş değişkenlere farklı değerler vererek modeldeki tüm ağaçların  $i$  gözlemi (firması) için tahminleri o değerlere göre yeniden hesaplanır. Böylece seçilmiş değişkenlerin farklı değerlerine göre  $i$  gözleminin çıktısı tahmin edilir bu ilişki tek bir fonksiyon olarak bir kısmi bağımlılık grafiğinde (Partial Dependence Plot) görselleştirilebilir (Friedman2001). Goldstein vd. (2015) kısmi bağımlılığın görselleştirilmesinde sadece genel veya ortalama değer değil, tüm gözlemler için kendi ayrı fonksiyonlarının temsil edildiği bir yöntemi öne sürmüştür ve bu görsele bireysel koşullu beklenti grafiği (Individual Conditional Expectation Plot) adı verilmiştir. Kategorik değişkenler için ise yine diğer değişkenler sabit tutularak tüm  $i$  gözlemlerinin kategorileri (örneğin bizim örneğimizde Bölge veya Engel değişkenleri) eşitlenir. Örneğin, Akdeniz bölgesi için bu değer hesaplanırken tüm firmaların Akdeniz bölgesinde olması durumunda ve diğer değişken değerleri sabit iken üst yöneticinin kadın olma ihtimali hesaplanır. Tüm kategoriler için aynı işlem yapıldıktan sonra değerleri karşılaştırmak mümkün olmaktadır. Çoğu yöntemde olduğu gibi bu görselleştirme tekniğinin de dezavantajları mevcuttur. Özellikle karmaşık ve doğrusal olmayan ilişkilerde eğer seçilmiş değişkenler ve sabit tutulan değişkenler arasında yüksek korelasyon varsa bulunan değerlerin doğruluk payı az olabilmektedir.

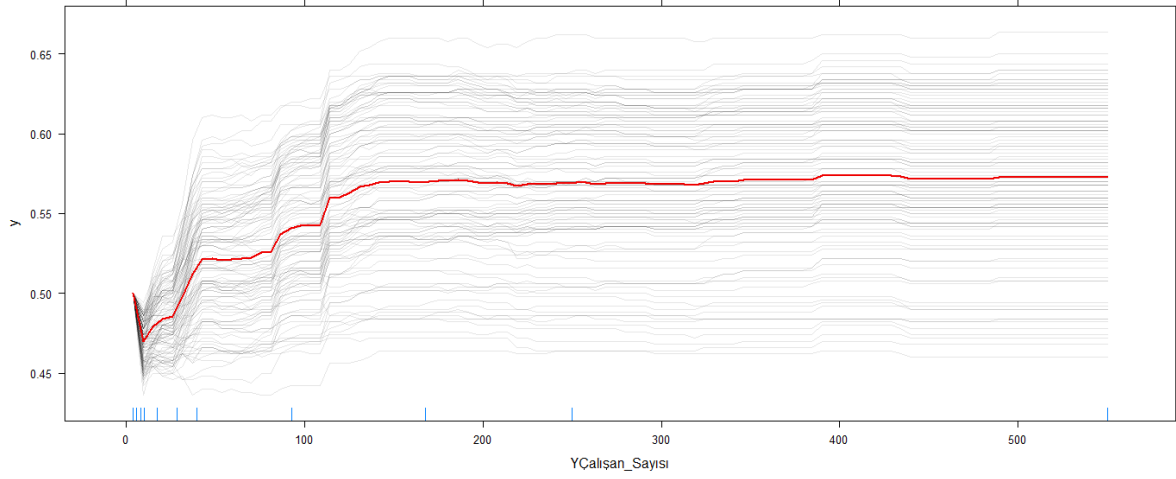
Bir firmanın üst yöneticisinin kadın olma ihtimalinin o firmanın çalışan sayısına (son mali yıl itibariyle) olan kısmi bağımlılığı Şekil 3'te gösterilmektedir. Y eksenini söz konusu olasılığı, X eksenini ise çalışan sayısını vermektedir. Bu eksen üzerindeki kısa mavi düşey çizgiler çalışan sayısındaki onda birlik kısımları ve minimum/maksimum değerleri göstermektedir ve dağılımda pozitif çarpıklığa işaret etmektedir. Tüm kısmi bağımlılık eğrilerinin ortalamasını temsil eden kırmızı eğri ve tüm firmaların bireysel eğrileri karşılaştırılabilirlik için 0,5 olasılığına ortalanmıştır. Bireysel eğrilerin dağılımı veride kayda değer bir heterojenlik olduğunu göstermektedir. Bununla birlikte ortalama eğrisinin de işaret ettiği gibi genel örüntü küçük (az çalışanlı) firmalarda kadın üst yönetici bulunma ihtimalinin düşük olduğunu göstermektedir. Kadın üst yöneticiler daha yüksek olasılık ile daha çok çalışan olan firmalarda bulunmaktadır. Bazı firmalar için bu olasılık %65 seviyelerine kadar çıkmaktadır. Bununla birlikte yaklaşık 150 çalışan sayısından sonra bu etki sabitlenmektedir. Dolayısıyla veri setindeki firmaların çalışan sayısı aşağı yukarı 100'den az olduğunda o firmalarda kadın üst yönetici bulunma olasılığının daha az olduğu gözlemlenmektedir.

Bulgularımıza göre önem arz eden *Engel* değişkeni kategorik olduğu için bu değişkene göre sunulan kısmi bağımlılık dereceleri sütun grafiği şeklinde gösterilebilir. Şekil 4'e göre araziye ve elektrige erişim, ruhsat alımı, yolsuzluk ve suç sorunları yaşayan şirketlerin üst yöneticilerinin erkek olma ihtimali daha fazladır. Bunun yanında kadın üst yöneticiler yönettikleri firmalarda daha çok siyasi istikrarsızlık, finansal kaynaklara erişim, işgücünün eğitimsizliği, kayıt dışı rekabet, vergi ve gümrük konuları ile ilgili zorluklar yaşamaktadırlar. Bu bulgular, suç, yolsuzluk, arazi ve ruhsat konuları gibi çatışmacı durumların yaşanabildiği alanlarda kadın üst yöneticilerin azınlıkta olduğu ile ilgili bir işaret vermektedir. Dolayısıyla kadın üst yöneticisi oranını artırmak için sektörlerdeki bu tür sorunların azaltılması dolaylı yoldan önemli bir olumlu etkiye yok açabilir.

Şekil 5'te ise iki yönlü bir kısmi bağımlılık grafiği görülmektedir. Y eksenini çalışan sayısını, X eksenini bulgularımızda öne çıkan diğer bir sayısal değişken olan üst yöneticinin deneyim süresini, sarı renge yakın açık tonların da yüksek olasılığını temsil etmektedir. Şekil 3 ile tutarlı olarak düşük çalışan sayısı koyu renklere tekabül etmektedir ve bu yöneticinin deneyiminden büyük ölçüde bağımsızdır. Ancak çalışan sayısı arttıkça şirket üst yöneticilerinin deneyimlerinin rolü daha belirgin hale gelmekte ve 20-30 yıl arası deneyime sahip olan ve yaklaşık 150 çalışan sayısından daha çok çalışana sahip olan firmaların üst yöneticilerinin kadın olma ihtimalinin diğer kombinasyonlara göre yüksek olduğu görülmektedir. Bu aralıktan daha az ve daha çok deneyime sahip olan üst yöneticilerin ise daha yüksek olasılıkla erkek oldukları gözlemlenmektedir.

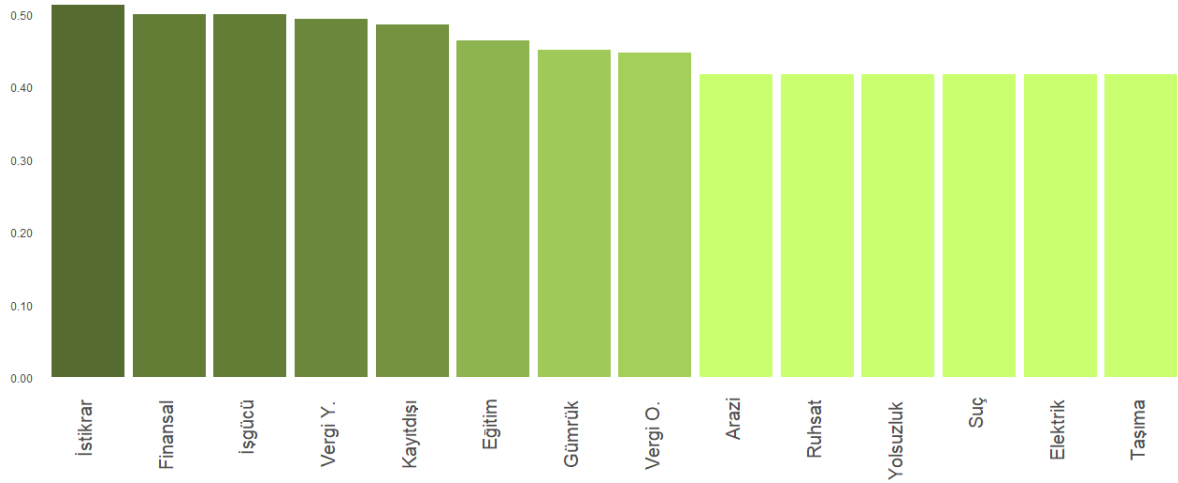
Hem sınıflandırma ağacı hem de rassal orman sonuçlarına göre bölge faktörü en çok öneme sahip değişkenlerden biridir. Şekil 6'da gösterilen haritada Girişim Anketi 2019'da kullanılan İstatistiki Bölge Birimleri Sınıflandırması 1. Düzey İBBS bölgelerine göre kadın üst yönetici kısmi bağımlılık değerleri gösterilmektedir ve koyu renkler daha yüksek olasılıkları temsil etmektedir. Türkiye'nin ekonomik merkezi olmasına karşın İstanbul en yüksek olasılığa sahip değildir. Firmalar eğer Doğu Marmara, Batı Marmara ve Akdeniz bölgelerinde bulunuyorlar ise kadın üst yöneticiye sahip olma olasılıkları daha yüksektir. Ankara'nın dahil olduğu Batı Anadolu bölgesinde de bu olasılık geri kalan bölgelere nazaran düşük değildir. Kadın üst yönetici bulunma olasılığının en düşük olduğu bölgeler ise Kuzeydoğu Anadolu ve Güneydoğu Anadolu bölgeleridir. Dolayısıyla çalışmanın bulgularından hareketle politika

yapıcılarının ve araştırmacıların bu bölgelere odaklı vaka çalışması türü detaylı bakışlar gerektiren incelemelerde bulunması tavsiye edilebilir. Bu bölgelerde ekonomik aktivite ve şirket sayısı düşük olsa bile yöneticilik pozisyonunda bulunan kadın-erkek dengesinin azlığı dikkate değer bir konudur.



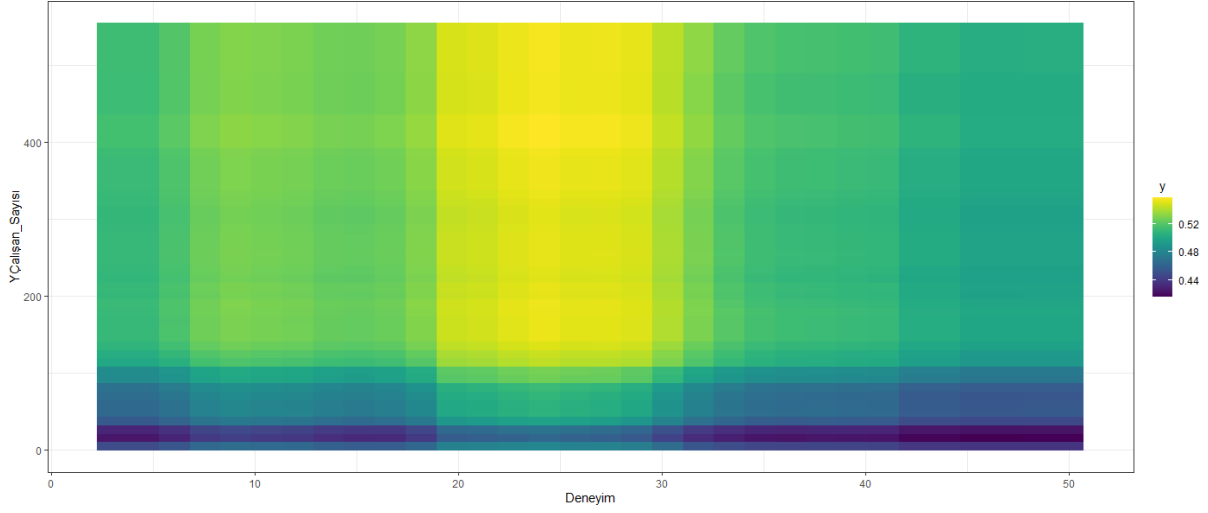
**Şekil 3:** Bireysel Koşullu Beklenti Grafiği (ortalananmış): Çalışan Sayısı

**Kaynak:** Yazar tarafından üretilmiştir.



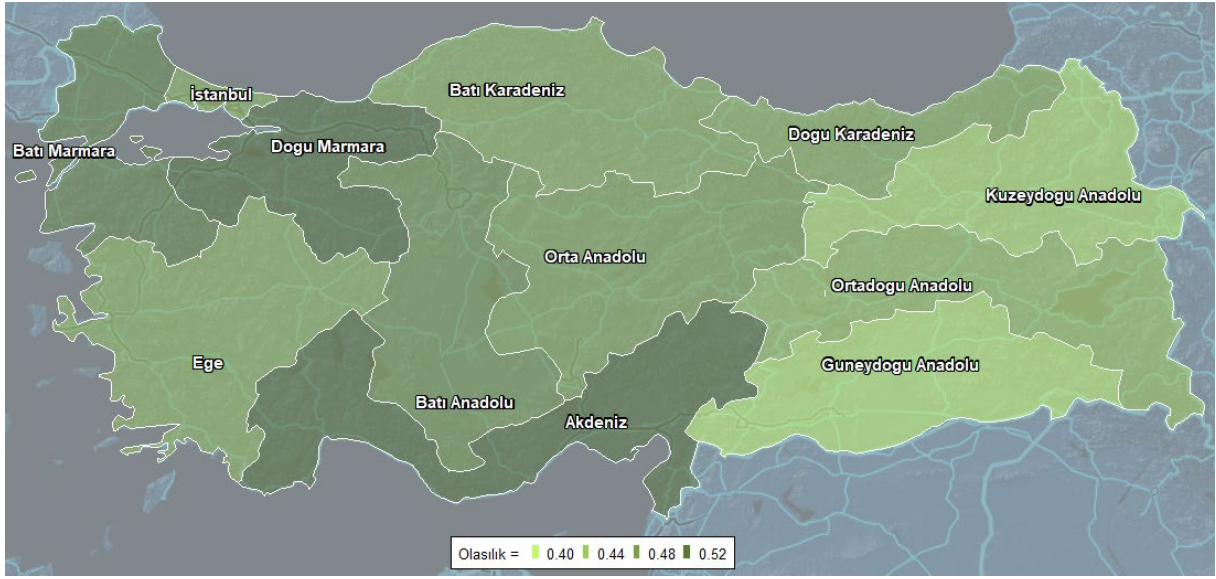
**Şekil 4:** Kısmî Bağımlılık Grafiği: Firmanın Karşılaştığı Engeller

**Kaynak:** Yazar tarafından üretilmiştir.



Şekil 5: Kısmî Bağımlılık Grafiği: Çalışan Sayısı ve Deneyim

Kaynak: Yazar tarafından üretilmiştir.



Şekil 6: Kısmî Bağımlılık Grafiği: Bölgeler

Kaynak: Yazar tarafından üretilmiştir.

## Sonuç

Bu çalışmadaki bulguların Türkiye’de kadın üst yönetici oranını artırmak için sunabileceği katkıları şu şekilde özetleyebiliriz: Sektörel olarak suç, yolsuzluk, arazi ve ruhsat konularının ön plana çıktığı, bölgesel olarak ise Kuzeydoğu ve Güneydoğu Anadolu içerisinde olan firmaların sorunlarına yönelik politikalar geliştirmek, ekonomik aktivite yönünden Türkiye’nin en aktif bölgesi olmasına rağmen kadın üst yönetici bulunma ihtimalininin yüksek olmadığı İstanbul için ise vaka çalışmalarında bulunmak ihtiyacı gözlemlenmiştir. Bununla birlikte çalışan sayısı bakımından görece küçük firmalarda kadın yönetici olma ihtimalininin daha büyük firmalara nazaran düşük olduğu da bulgular arasındadır. Bu gözlemden çıkarılabilecek bir sonuç kadın yöneticilerin firma kuruluş aşamasında değil, ancak firma belli bir ölçüğe eriştikten sonra kariyer basamaklarını tırmanarak bu pozisyonlara gelmeleri olabilir. Bu bulgu kadın girişimcilere daha çok destek verilmesi gerektiği anlamına gelmektedir. Türkiye Kadın Girişimcilik Endeksi 2019 araştırma raporuna göre kadın girişimcilerin perakende, toptan satış ve imalat sektörlerinde daha aktif olduğu gözlemlenmiştir (KADİGER vd., 2019). Ancak bulgularımızda bu ve benzeri gözlemlerin belirgin bir bölgesel boyutu olduğu açıktır.

Kadın yöneticiler üzerine yaptıkları çalışmada İslam ve Amin (2015) bulgularımız arasında yer alan suç etkeninin yanı sıra iş ve ev arası yolculuk zaman ve mesafesinin kadınların iş gününe katılımını olumsuz etkilediğine dikkat çekmiştir. Bu durum özellikle evli kadınlar için daha belirgin biçimde gözlemlenmiştir (Black vd., 2014). Yukarıda tartışılan bulgularda İstanbul Bölgesinin üst sıralarda olmamasının sebebi mesafe ve zaman faktörü olabilir. Öte yandan Türkiye’de kadın girişimciliği ve genel olarak istihdamının bölgelere göre farklılık arz ettiği birçok defa gözlemlenmiştir. Örneğin Berber ve Eser (2008) kadın istihdamının bölgesel boyutları bağlamında Güneydoğu Anadolu bölgesinin en düşük seviyede olduğunu belirlemiştir. Can ve Eyidiker (2019) ise bulgularımıza göre kadın yönetici bulunma olasılığının en düşük olduğu ikinci bir İİBS bölgesi olan Kuzeydoğu Anadolu’da kadın istihdamının azaldığı sonucunu elde etmişlerdir. Öte yandan, Dayoglu ve Kırdar (2010) ise Kuzeydoğu Anadolu Bölgesinde kırsal alanlarda kadın istihdamının yüksek olduğunu ancak kırsal nüfus oranında düşüş görüldüğünü belirtmiştir. Girişimcilik odaklı bir yaklaşım ile de benzer örüntülerin bulunduğu görülmektedir. Öztürk (2018) bulgularında 2009 ve 2016 yılları arasında Kuzeydoğu Anadolu Bölgesindeki kadın girişimcilerin oranının yaklaşık %4, Güneydoğu Anadolu Bölgesi için ise yaklaşık %6’lık bir düşüş gözlemlenmiştir. Girişimcilik ve istihdam ile ilgili bu bulguların çalışmamızın sonuçları ile paralellik göstermesi beklenen bir durumdur. Bu bağlamda girişimcilik ve istihdamdaki cinsiyet eşitsizliğinin üst yöneticilik pozisyonlarına güçlü bir yansıması olduğu görülmektedir. Dolayısıyla kadın üst yönetici varlığının yerel beşerî kaynak seviyeleri ile yakından ilgili olduğu görülmektedir. Literatürdeki bulgular ışığında ve ampirik sonuçlarımızdan hareketle nüfus yoğunluğunun yüksek olduğu bölgelerde ulaşım sorunlarını giderilmesine yönelik politikaların kadın istihdamının ve girişimciliğinin artmasına olumlu katkı yapacağı beklenebilir. Eğitim, ekonomik aktivite ve bölgelere özgü sorunların cinsiyet eşitsizliği bağlamında irdelenmesi ve bunlara yönelik politika çözümlerinin getirilmesi kadın yöneticilerin iş dünyasındaki varlığına olumlu etkiler olarak öne çıkacaklardır.

Firmalarda kadın üst yönetici bulunma durumunun açıklamasının karmaşık mekanizmalara tabi olmasından dolayı bu çalışmada doğrusal olmayan algoritmik modeller kullanılmıştır. Sonuçlar geleneksel metotlardan ziyade en güncel ağaç temelli makine öğrenmesi modelleri ışığında gözlenmiş ve değişkenler arasındaki karmaşık ilişkiler ortaya çıkarılmıştır. Buna ek olarak kullanılan makine öğrenmesi uygulamaları ekonometrik ve diğer geleneksel yaklaşımların kullanamayacağı kadar çok değişkenden yararlanmışlardır. Dolayısıyla politika odaklı yeni sonuçlar ortaya çıkarmasının yanı sıra, bu çalışma makine öğrenmesi metotlarının Türk firmaları üzerine bir uygulama olması bakımından da yöntemsel bir örnek teşkil etmektedir.

#### **Hakem Değerlendirmesi / Peer-review:**

Dış bağımsız

*Externally peer-reviewed*

#### **Çıkar Çatışması / Conflict of interests:**

Yazar(lar) çıkar çatışması bildirmemiştir.

*The author(s) has (have) no conflict of interest to declare.*

#### **Finansal Destek / Grant Support:**

Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

*The author declared that this study has received no financial support.*

#### **Kaynakça / References**

- Adams, R. B., ve Funk, P. (2012). Beyond the glass ceiling: Does gender matter? *Management Science*, 58(2):219–235.
- Alpaydın, E. (2011). *Yapay öğrenme*. Boğaziçi Üniversitesi Yayınevi.

- Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, 507–547. University of Chicago Press.
- Athey, S., ve Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11.
- Atkinson, E. J., ve Therneau, T. M. (2000). An introduction to recursive partitioning using the RPART routines. Rochester: Mayo Foundation.
- Ayyıldız Ünnü, N. A., Baybars, M. ve Kesken, J. (2014). Türkiye’de kadınların üniversiteler bağlamında yetki ve karar verme mekanizmalarına katılımı. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (42):121–134.
- Bang, J., Sen, T., ve Basuchoudhary, A. (2017). New tools for predicting economic growth using machine learning: A guide for theory and policy.
- Barczak, G., Lassk, F., ve Mulki, J. (2010). Antecedents of team creativity: An examination of team emotional intelligence, team trust and collaborative culture. *Creativity and Innovation Management*, 19(4):332–345.
- Birleşmiş Milletler (2019), *The Sustainable Development Goals Report 2019*. New York.
- Becic, M. ve Vojinic, P. (2018). The role of female top manager in innovation activities: Case of Ceecs’ firms. *Proceedings of the 10th Economics & Finance Conference, Rome kitabı içinde*. International Institute of Social and Economic Sciences.
- Berber, M., ve Eser, B. Y. (2008). Türkiye’de kadın istihdam: ülke ve bölge düzeyinde sektörel analiz. *ISGUC The Journal of Industrial Relations and Human Resources*, 10(2):1–16.
- Black, D. A., Kolesnikova, N., ve Taylor, L. J. (2014). Why do so few women work in New York (and so many in Minneapolis)? Labor supply of married women across US cities. *Journal of Urban Economics*, 79:59–71.
- Boden, R. J. (1999). Gender inequality in wage earnings and female self-employment selection. *The Journal of Socio-Economics*, 28(3):351–364.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., ve Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.
- Brenner, S.-O., ve Levi, L. (1987). Long-term unemployment among women in Sweden. *Social Science & Medicine*, 25(2):153–161.
- Briel, S., ve Töpfer, M. (2020). The gender pay gap revisited: Does machine learning offer new insights? *LASER*.
- Campbell, K., ve Minguez-Vera, A. (2007). Gender diversity in the boardroom and firm financial performance. *Journal of Business Ethics*, 83(3):435–451.
- Can, Y., ve Eyidiker, U. (2019). Türkiye’de kadın istihdam profilinin istatistiksel bölgeler arasındaki farklılıkların incelenmesi. XIII. IBANESS İktisat, İşletme ve Yönetim Bilimleri Kongreler Serisi – Tekirdağ-Güz / Türkiye. 05-06 Ekim 2019
- Chang, H.-J., Shih, Y.-S., ve Su, T.-J. (2014). Split selection methods for regression tree on detecting regional economic convergence. *Journal of the Chinese Statistical Association*, 52(2):180–208.
- Cohen, P. N. and Huffman, M. L. (2007). Working for the woman? Female managers and the gender wage gap. *American Sociological Review*, 72(5):681–704.
- Cook, A., ve Glass, C. (2013). Glass cliffs and organizational saviors. *Social Problems*, 60(2):168–187.
- Cotter, D. A., Hermsen, J. M., Ovadia, S., ve Vanneman, R. (2001). The glass ceiling effect. *Social Forces*, 80(2):655–681.
- Dayoğlu, M., ve Kırdar, M. G. (2010). Türkiye’de kadınların işgücüne katılımında belirleyici etkenler ve eğilimler. DPT Yayınları, Ankara.
- Friedman, J., Hastie, T., ve Tibshirani, R. (2001). *The Elements of Statistical Learning*, volume 1. Springer Series in Statistics, New York.

- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 5:1189-1232.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.
- Glaeser, E. L., Kominers, S. D., Luca, M., ve Naik, N. (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry*, 56(1):114-137.
- Goldstein, A., Kapelner, A., Bleich, J., ve Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1):44-65.
- Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. *The R Journal*, 9(1):421-436.
- Harding, M. and Hersh, J. (2018). Big data in economics. IZA World of Labor, (451).
- Iman, H. S., ve Tarasyev, A. (2018). Machine learning methods in individual migration behavior. In Russian Regions in the Focus of Changes: Conference proceedings. Ekaterinburg, 2018, 72-81. LLC Publishing office EMC UPI.
- Islam, A., ve Amin, M. (2015). Women managers and the gender-based gap in access to education: Evidence from firm-level data in developing countries. *Feminist Economics*, 22(3):127-153.
- Jacobs, J. A. (1995). Gender Inequality at Work. ERIC.
- James, G., Witten, D., Hastie, T., ve Tibshirani, R. (2013). An Introduction to Statistical Learning, volume 112. Springer, New York.
- Kadın Girişimciler Derneği ve Türk Tuborg (2019). Türkiye Kadın Girişimcilik Endeksi-2019 Araştırma Raporu.
- Kaiser, M. (2018). Segmentation in urban labor markets: a machine learning application and a contracting perspective.
- Karatepe, S., ve Arıbaş, N. N. (2017). İş hayatında kadın yöneticilere ilişkin cinsiyet ayrımcılığı: Türkiye için bir değerlendirme. *Yasama Dergisi*, (31):7-23.
- Karcioğlu, F. ve Leblebici, Y. (2014). Kadın Yöneticilerde Kariyer Engelleri: "Cam Tavan Sendromu" Üzerine Bir Uygulama. *Atatürk University Journal of Economics & Administrative Sciences*, 28(4).
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5):1-26.
- Liaw, A., ve Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18-22.
- Liu, Y. (2019). Predicting labor market competition and employee mobility - a machine learning approach. PhD Dissertation, University of Iowa.
- Lopez-Yucra, K., Libaque-Saenz, C., Lazo, J., ve Bravo, E. (2018). Could machine learning improve the prediction of child labor in Peru? In Information Management and Big Data: 4th Annual International Symposium, SIMBig 2017, Lima, Peru, September 4-6, 2017, Revised Selected Papers, volume 795, 15. Springer.
- Mittal, M., Goyal, L. M., Sethi, J. K., ve Hemanth, D. J. (2019). Monitoring the impact of economic crisis on crime in India using machine learning. *Computational Economics*, 53(4):1467-1485.
- Mullainathan, S., ve Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87-106.
- Nyman, R., ve Ormerod, P. (2017). Predicting economic recessions using machine learning algorithms. arXiv preprint arXiv:1701.01428.
- Nyman, R., ve Ormerod, P. (2020). Understanding the great recession using machine learning algorithms. arXiv preprint arXiv:2001.02115.
- Örücü, E./ Kılıç R./ Kılıç T. (2007), "Cam tavan sendromu ve kadınların üst düzey yönetici pozisyonuna yükselmelerindeki engeller: Balıkesir ili örneği," *Celal Bayar Üniversitesi İ.İ.B.F. Yönetim ve Ekonomi Dergisi*, 14 /2: 117-135



- Öztürk, S. (2018). Türkiye’de kadın girişimciliğinin iktisadi ve kültürel normlar çerçevesinde değerlendirilmesi. *Girişimcilik ve Kalkınma Dergisi*, 13(1):1-11.
- Ritter-Hayashi, D., Vermeulen, P. A. M., ve Knoben, J. (2016). Gender diversity and innovation: The Role of Women’s Economic Opportunity in Developing Countries. (DFID çalışma tebliği). Radboud University Nijmegen.
- Robinson, C., Hohman, F., ve Dilkina, B. (2017). A deep learning approach for population estimation from satellite imagery. In Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities, 47-54.
- Rutherford, S. (2001). Organizational cultures, women managers and exclusion. *Women in Management Review*, 16(8):371-382.
- Şahin, F. (2011). Kadınların siyasal katılımlar çerçevesinde kadın meclislerinin yerel siyasetteki etkinlikleri ve üye profilleri. TC Başbakanlık Kadın Statüsü Genel Müdürlüğü, (Uzmanlık Tezi), Ankara.
- Shrader, C. B., Blackburn, V. B., ve Iles, P. (1997). Women in management and firm financial performance: An exploratory study. *Journal of managerial issues*, 355-372.
- Smith, N., Smith, V., ve Verner, M. (2006). Do women in top management affect firm performance? a panel study of 2, 500 Danish firms. *International Journal of Productivity and Performance Management*, 55(7):569-593.
- Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. In Handbook of Statistics, 303-329. Elsevier.
- Taşkın, E. (2012). Kadın yöneticilerin cam tavan algısının cam tavan aşma stratejilerine etkisi: Bursa örneği. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 1(33):19-34.
- The World Bank (2019) Turkey - Enterprise Survey (ES) 2019, Ref. TUR\_2019\_ES\_v01\_M. <https://www.enterprisesurveys.org/portal/login.aspx>.
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3-28.
- Verma, C., Tarawneh, A. S., Illes, Z., Stoffova, V., ve Dahiya, S. (2018). Gender prediction of the European school’s teachers using machine learning: Preliminary results. In 2018 IEEE 8th International Advance Computing Conference (IACC). IEEE.
- Xu, W., Li, Z., Cheng, C., ve Zheng, T. (2013). Data mining for unemployment rate prediction using search engine query data. volume 7, 33-42. Springer.
- Yıldız, S., Alhas, F., Sakal, Ö., ve Yıldız, H. (2016). Cam Uçurum: Kadın Yöneticiler Cam Tavanı Ne Zaman Aşar? *Ankara Üniversitesi SBF Dergisi*, 71 (4), 1119-1146.
- Zhang, C., ve Zhang, J. (2014). Analysing Chinese citizens’ intentions of outbound travel: a machine learning approach. *Current Issues in Tourism*, 17(7):592-609.