

# TÜRKİYE’DE GİRİŐİMCİLİK VE SÜRDÜRÜLEBİLİR BÖLGESEL KALKINMA: MAKİNE ÖĞRENMESİ YAKLAŐIMLARINDAN ELDE EDİLEN BULGULAR

## Entrepreneurship and Sustainable Regional Development in Turkey: Findings Obtained from Machine Learning Approaches

Ebru Tomris AYDOĞAN\* & Esra E. KARADENİZ\*\* & Mehmet Güney CELBİŐ\*\*\*

### Öz

**Anahtar Kelimeler:**  
Ekonomik Kalkınma, Giriőimcilik, Bölgesel Analiz, Makine Öğrenmesi.

**JEL Kodları:**  
O10, O18, M13, C88

**Keywords:**  
Economic Development, Regional Analysis, Entrepreneurship, Machine Learning.

**JEL Codes:**  
O10, O18, M13, C88

BirleŐmiş Milletler Kalkınma Programı (UNDP), Sürdürülebilir Kalkınma'nın (SK) sağlanmasında girişimciliđi önemli bir faktör olarak tanımlamaktadır. Bu çalışmanın amacı, sosyal bilimlerde nispeten yeni bir yöntem olan makine öğrenmesini uygulayarak ve Küresel Giriőimcilik Monitörü (GEM) verilerini kullanarak Türkiye'deki girişimcilik faaliyetlerini SK açısından etkileyen faktörleri analiz etmektir. Rastal Orman model bulgularına göre çalışma durumu, başka bir girişimci tanımak, kişinin kendi bilgisi ve becerisi hakkındaki görüşü, ihtiyaç/mecburiyet ve yaş, girişimci olmak için önemli faktörlerdir. Ayrıca, hanehalkı geliri ve nüfusu, kişinin çalışma durumu, cinsiyeti, eğitim düzeyi ve anket zamanı itibariyle şirket içi girişimcilik aktivitesinde bulunması da önemli faktörler olarak tespit edilmişlerdir. Stokastik Gradyan Artırma sonuçlarına göre Türkiye’de bölgesel kalkınmanın ve koşulların bireylerin girişimci olma olasılıklarını etkilediđi gözlenmiştir. Bu araştırma, Özdemir ve Karadeniz (2011) ile Karadeniz ve Özçam'ın (2018) Türkiye bağlamında bireysel girişimci kazanımlarından buldukları önceki sonuçları ilk kez makine öğrenmesi algoritmaları uygulaması aracılığıyla analiz etmekte ve genişletmektedir. Giriőimcilik faaliyetleri sürdürülebilir kalkınmayı destekleyen politikalar çerçevesinde planlanmalıdır.

### Abstract

The United Nations Development Programme (UNDP) defines entrepreneurship as an important factor in achieving Sustainable Development (SD). The aim of this study is to analyze the factors affecting entrepreneurship activities in Turkey in terms of SD by applying machine learning, which is a relatively new method in social sciences, and using Global Entrepreneurship Monitor (GEM) data. According to the results of the Random Forest model, working status, knowing another entrepreneur, one's opinion about one's own knowledge and skills, need/obligation and age are important factors for being an entrepreneur. Also, household income and population, the employment status, gender, education level, and the engagement in intra-company entrepreneurial activity of the person at the time of the survey were also determined as important factors. According to the results of the Stochastic Gradient Boosting, it is observed that regional development and conditions in Turkey affect the probability of individuals to become entrepreneurs. This research analyzes and extends the previous results of Özdemir and Karadeniz (2011) and Karadeniz and Özçam (2018) from individual entrepreneur gains in the Turkish context for the first time through the application of machine learning algorithms. Entrepreneurial activities should be planned within the framework of policies that support sustainable development.

\* Sorumlu Yazar, Dr. Öğr. Üyesi, Yeditepe Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonomi Bölümü, taydogan@yeditepe.edu.tr, ORCID: 0000-0002-1529-5484

\*\* Prof. Dr., Yeditepe Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonomi Bölümü, GEM Projesi Türkiye Koordinatörü, ekaradeniz@yeditepe.edu.tr, ORCID: 0000-0003-3655-9087

\*\*\* Dr. Öğr. Üyesi, Yeditepe Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonomi Bölümü celbis@merit.unu.edu, ORCID: 0000-0002-2790-6035

## 1. Giriş

Girişimcilik ekonomilere yenilikler getiren, değişim yaratarak rekabeti artıran, ekonomik kalkınmaya katkıda bulunan en önemli faktörlerden biridir. Girişimcilik faaliyeti, Birleşmiş Milletler Kalkınma Programı (United Nations Development Program [UNDP], 2021) Sürdürülebilir Kalkınma Amaçları (SKA) 1, 4, 8 ve 10 numaralı hedeflere ulaşmak için çok önemlidir. Bu hedefler; “SKA1-Yoksulluğa Son: Her türlü yoksulluğu her yerde sona erdirmek”, “SKA4-Nitelikli Eğitim: Herkes için kapsayıcı ve eşitlikçi nitelikli eğitim sağlanması ve ömür boyu eğitim fırsatlarının teşvik edilmesi”, “SKA8-İnsana Yakışır İş ve Ekonomik Büyüme: İstikrarlı, kapsayıcı ve sürdürülebilir ekonomik büyümeyi, tam ve üretken istihdamı ve herkes için insana yakışır işleri desteklemek” ve “SKH10-Eşitsizliklerin Azaltılması: Ülkelerin içinde ve arasındaki eşitsizlikleri azaltmak” tır. Birleşmiş Milletler Ekonomik ve Sosyal İşler Dairesi’nin (Department of Economic and Social Affairs [UN DESA], 2015) 2015 yılında öne sürdüğü SKA ile günümüz dünyasında girişimciliğin ve girişimcilik eko-siteminin geliştirilmesi ekonomik politikaların vazgeçilmez bir parçası olmuştur. Topluma girişimcilik ruhunun aşılması ve gelecekteki girişimci sayısının artırılması yürütülen çabaların başında gelmektedir. Ancak, uygulanan programlarda girişimci sayısını artırmak kadar, kurulan işletmelerin geliştirilmesi ve kalıcılığının sağlanması da amaçlanmaktadır. Temel amaç yenilikçi ve uluslararası düzeyde rekabet gücü yüksek, verimli ve küresel gelişmelere uyum sağlayan işletmeler yaratmaktır. Girişimciliğin ekonomik büyümeye ve kalkınmaya olan katkısından dolayı, bireyleri girişimci olmaya teşvik eden veya caydırıcı olan faktörlerin tespiti gerek akademisyenlerin gerekse politika yapımcılarının ilgilendikleri bir konu olmuştur.

Bu araştırmanın amacı, Türkiye’de girişimcilik faaliyetlerini SK açısından etkileyen faktörlerin neler olduğunun incelenmesidir. Araştırmanın literatüre katkısı, Küresel Girişimcilik Monitörü’nün (Global Entrepreneurship Monitor [GEM], 2018) verilerini kullanarak makine öğrenmesi (machine learning) ile girişimcilik endeksini etkileyen faktörleri analiz etmesidir. Karadeniz ve Özçam (2018) Türkiye’de girişimcilik faaliyet düzeyinde bölgelerarasında farklılık olduğunu ilk kez göstermişler ve ayrıca, girişimcilik davranışlarından başarısızlık korkusu, fırsatları görmek, bilgi ve beceriye sahip olmak ve eğitim faktörlerinin kişinin girişimci olma olasılığını etkilediğini bulmuşlardır.<sup>1</sup> Karadeniz ve Özçam (2018) tarafından yapılan lojistik regresyon yöntemi ile girişimcilik endeksini etkileyen faktörler üzerindeki araştırmalar makine öğrenmesi kullanılarak yapılan bu araştırma sonuçları ile karşılaştırılacaktır.

Bu makalenin kalan kısmı, 2. Teorik Altyapı, 3. Veri Seti, 4. Sınıflandırma Ağacı ve Rassal Orman Bulguları, 5. Stokastik ve Stokastik Olmayan Gradyan Artırma Makineleri Bulguları ve 6. Sonuç bölümlerinden oluşmaktadır.

## 2. Teorik Altyapı

Girişimciliği etkileyen faktörlerin tartışılması tek bir disiplinle sınırlanamaz (Verheul, Wennekers, Audretsch ve Thurik, 2001). Psikoloji alanındaki çalışmalar genellikle girişimcilerin

---

<sup>1</sup> Türkiye’deki coğrafi bölgeler arasında birçok açıdan önemli eşitsizlikler bulunmaktadır. Örneğin OECD ülkeleri arasında GSYH açısından bölgesel düzeydeki eşitsizliğin en yüksek olduğu ülke Türkiye (Organization of Economic Cooperation and Development [OECD], 2020). Firmaların ve çalışanların belirli şehirlerde yoğunlaşması, bölgeler arasında ekonomik ve de özellikle de verimlilik farklarına neden olmaktadır (Özgüzel, 2020).

güdüleri ve karakter özelliklerine; sosyoloji alanındaki çalışmalar girişimcilerin toplumdaki yerlerini, etkilerini ve geçmişlerini arařtırmaya; ekonomik çalışmalar, demografik yapının, inovasyon ve teknolojik gelişmelerin ve fırsat maliyetlerini kapsayacak şekilde ekonomik kalkınmanın girişimcilięi etkileyen faktörlerine odaklanmıştır.

SK'nın ekonomik boyutunu ifade eden ve ekonomik büyüme, beşeri sermaye, insani gelişmişlik ve bölgesel kalkınma faktörlerini kapsayan ekonomik kalkınma sürdürülebilir kalkınmanın temelini oluşturur (UNDP, 2021). Ekonomik kalkınma; yoksullar ve zenginler arasındaki gelir uçurumunu kaldıracak daha adil bir gelir dağılımı, herkes için eğitime ve sağlığa erişimde eşitlik ve cinsiyet eşitlięi faktörleri ile açıklanır. Bu faktörler bölgesel kalkınma için önem arz eder. SK'nın ekonomik, sosyal ve çevresel yönetişimi içeren 3 boyutu vardır ve girişimcilik faaliyetleri bu boyutların tümüne olumlu katkı sağlar (Filser, Kraus, Roig-Tierno, Kailer ve Fischer, 2019). SK'yı açıklamakta kullanılan faktörlerden olan gelir, eğitim ve cinsiyet faktörleri girişimcilik faaliyetlerini etkiler. Bu nedenle, girişimcilik faaliyetlerinin yapılabilmesi ve SK'ya ulaşılabilmesi için gelir adaleti ile eğitim ve cinsiyet eşitliklerinin sağlanması gerekir.

Birleşmiş Milletler Genel Kurulu (2016), girişimcilięi, inovasyonla birlikte Sürdürülebilir Kalkınma (SK)'nın itici gücü olarak tanımlamıştır. SKA4 (Nitelikli Eğitim)'in ve SKA8 (İnsana Yakıřır İş ve Ekonomik Büyüme)'nin hedefleri arasında yer alan girişimcilik ile eğitim, iş yaratma, yaratıcılık ve inovasyon arasında doğrudan bir ilişki vardır. Bununla birlikte, SKA'dan SKA7 (Erişilebilir ve Temiz Enerji), SKA9 (Sanayi, Yenilikçilik ve Altyapı), SKA11 (Sürdürülebilir Şehirler ve Topluluklar), SKA12 (Sorumlu Üretim ve Tüketim) ve SKA13 (İklim Eylemi) gibi girişimcilik ile doğrudan ilişkili olmayan hedefler de girişimciler için fırsat sunmaktadır. Tablo 1, SK'nın sağlanmasında girişimcilięin rolünü ve katkısını SK'nın ekonomik, toplumsal ve çevresel boyutları açısından göstermektedir (Filser vd., 2019).

**Tablo 1. Sürdürülebilir Kalkınma (SK)'nın Sağlanmasında Giriřimcilięin Rolü**

SK Boyutu	Giriřimcilik Katkısı
Ekonomik	“(...) Giriřimcilik, istihdam yaratarak, insana yakıřır işi, sürdürülebilir tarımı ve inovasyonu teşvik ederek ekonomik büyümeyi yönlendirir”.
Sosyal	“Giriřimcilik sosyal uyumu teşvik etmek, eşitsizlikleri azaltmak ve kadınlar, genç insanlar, engelliler ve en savunmasız insanlar da dahil olmak üzere herkes için fırsatları genişletmekte olumlu katkı sağlayabilir”.
Sosyal	“Giriřimcilik sosyal uyumu teşvik etmek, eşitsizlikleri azaltmak ve kadınlar, genç insanlar, engelliler ve en savunmasız insanlar da dahil olmak üzere herkes için fırsatları genişletmekte olumlu katkı sağlayabilir”.
Çevresel	“(...) girişimcilik, yeni iklim deęişiklięini azaltma ve adaptasyon teknolojileri ve dayanıklılık önlemlerinin getirilmesinin yanı sıra çevresel olarak sürdürülebilir uygulamaları ve tüketim kalıplarını teşvik ederek çevresel zorlukların üstesinden gelmek için yardımcı olabilir”.

**Kaynak:** (Filser vd., 2019, s. 3).

Belirli bir ülkedeki girişimcilik seviyesi, girişimcilięin arz ve talep tarafı arasında bir ayrım yapılarak açıklanabilir. Arz ve talep, bireysel düzeyde alınan girişimcilik kararı için koşullar yaratır. Talep, mal ve hizmetlere yönelik piyasa talebi yoluyla girişimcilik fırsatları yaratırken, arz ise fırsatlar doğrutusunda hareket edebilecek potansiyel girişimcilerin oluşmasını sağlar (Verheul vd., 2001). Yazara göre, girişimcilięe olan talep, ülkenin veya bölgenin ekonomik gelişmişlięini ve pazar talebini (çeşitlilięi) etkiler ve girişimcilik için fırsatlara yol açar. Giriřimcilik arzı, ülkenin yaş, küreselleşme ve teknolojik gelişme düzeyi gibi bir dizi faktör

tarafından tanımlanır. Bu faktörler, endüstriyel yapısı, nüfus yoğunluğu ve kentleşme oranı, göçmen sayısı ve nüfustaki veya işgücü piyasasındaki kadınların oranı, eğitim düzeyi ve kişilerin gelir düzeyi olmak üzere nüfusun büyüklüğüne ve bileşimine göre belirlenir.

Bu çalışmada girişimciliği belirleyen ve etkileyen faktörler GEM modeline göre analiz edilecektir. GEM (2018) girişimciliği “birey veya bireylerin veya işletmelerin yeni iş ya da yeni girişim yaratması, kendi işini kurması veya var olan işi genişletme faaliyetinde bulunması” olarak tanımlamıştır. GEM (2018) modeline göre, kişilerin cinsiyeti, yaşı, eğitim durumu, geliri, girişimcilik konusundaki görüşleri ve algı ve sezgileri girişimcilik arzını belirleyen faktörlerdir. Bölgelerde mal ve hizmetlere yönelik piyasa talebi yoluyla yaratılan girişimcilik fırsatları girişimcilik talebini oluşturur.

Girişimciliği etkileyen önemli faktörlerden biri *cinsiyettir*. Cinsiyet farklılıkları açısından bakıldığında gerek dünya genelinde gerekse Türkiye’de erkek girişimci oranının kadın girişimci oranından daha yüksek olduğu görülmektedir (Bosma vd., 2020; Karadeniz ve Özçam, 2018; Özdemir ve Karadeniz, 2011). Bununla birlikte ulusal ve uluslararası yazın Türkiye’de cinsiyet temelli eşitsizlikleri ortaya koymaktadır (Türk, 2020). Avrupa Komisyonu’nun (2019) yaptığı açıklamaya göre, kadınlar toplam Avrupa nüfusunun %52’sini oluştururken ancak serbest meslek sahiplerinin yalnızca %34.4’ünü ve yeni kurulan girişimcilerin %30’unu temsil ederler.

Özdemir ve Karadeniz’in (2011), Rastgele Örneklem Yöntemi kullanarak 2006, 2007, 2008 ve 2010 yıllarında GEM Yetişkin Nüfus Anketi’ne (APS) göre toplanan 9601 birleşik veri seti kullanarak yapılan lojistik regresyon sonucunda, Türkiye’de gelir ve eğitim düzeyi yüksek, kendine güvenen, iş fırsatlarını belirleme becerisine sahip ve çevresi olan erkeklerin girişimci olma olasılıklarının daha fazla olduğunu tespit etmişlerdir. Yazarlar, ayrıca Türkiye’de erkeklerin girişimcilik faaliyetlerine katılma olasılıklarının kadınlardan iki kat daha fazla olduğunu belirtmişlerdir.

Özçam ve Karadeniz’in (2018), Türkiye’de genel girişimcilik endeksi ve erkek-kadın girişimci endeksi arasındaki farkın zaman içindeki değişimini Yatay Kesit ve Zaman Verileri Regresyon Analizi kullanarak araştırmışlardır. GEM projesi tarafından, 2006-2015 yılları arasında Türkiye için toplanan 56.109 görüşmeden oluşan temsili veri seti kullanarak yapılan analiz sonucunda Türkiye’de 2011 yılından itibaren girişimcilik faaliyetlerinde kayda değer bir artış kaydetmiştir. Aynı zamanda, Türkiye’de kadın girişimci sayısında da artış görülmüş, fakat kadın-erkek girişimcilik oranındaki fark %4 civarında sabit kalmıştır.

*Yaşın*, bireylerin girişimcilik kariyerlerini etkilediği, yapılan birçok çalışmada ortaya konulmuştur. Araştırmalar, bireylerin girişimcilik kariyerlerine çoğunlukla 25-45 yaşları arasında başladıklarını daha sonraki yaşlarda girişimci olan birey sayısında azalma olduğunu göstermektedir (Bosma vd., 2020; Karadeniz ve Özçam, 2018; Özdemir ve Karadeniz, 2011; Reynolds, Hay, Bygrave, Camp ve Autio, 2000). Vodă, Butnaru ve Butnaru’nin (2020) 18 Avrupa Ülkesini kapsayan çalışmalarında GEM’in 2007 ve 2014 yıllarını kapsayan Yetişkin Nüfus Anketi veri tabanı kullanarak yaptıkları lojistik regresyon analizi sonucunda yaşın girişimciliği etkileyen faktör olduğun ve gençlerin girişimcilik faaliyetlerinde bulunma olasılıklarının yaşlılara göre daha fazla olduğunu tespit etmişlerdir. Çetindamar, Gupta, Karadeniz ve Eğrican (2012) tarafından Türkiye’de 18-64 nüfus aralığını temsil eden 2417 örnekleme kullanarak yapılan lojistik regresyon sonucunda 25-34 yaş grubunda girişimci olanların sayısının daha yüksek, ancak 45 yaş ve üstü yaş grubunda daha düşük olduğunu tespit etmişlerdir.

*Toplumun giriřimcilik konusundaki grřleri ve algısı* giriřimcilik faaliyetlerini etkileyen nemli bir n kořuldur. GEM projesi, giriřimcilięin toplum tarafından desteklenip desteklenmedięini; yeni bir giriřimcilik faaliyetine bařlamanın iyi bir kariyer seęimi olup olmadıęına, yeni bir iř kurup bařarılı olanların yksek stat ve saygınlık sahibi olup olmadıęına ve ayrıca basında, yeni kurulan iřletmelerin bařarılarıyla ilgili haberlerin sıkça grlp grlmedięine bakarak deęerlendirir (Bosma vd., 2020).

GEM modeline gre kiřilerin yeni bir iř kurma kararı ile o *kiřilerin sezgileri ve evreyi algılamaları* arasında sıkı bir iliřki vardır (Arenius ve Minnit, 2005). Birok arařtırma kiřilerin yeni bir iř kurma kararı ile bireylerin evrelerindeki fırsatların farkına varmaları, yeterli bilgi ve yeteneęe sahip olmaları, giriřim faaliyetiyle uęrařan kiřileri tanımaları ve bařarısızlık korkusu arasındaki iliřkiyi saptamıřtır (Karadeniz ve zam, 2018; zdemir ve Karadeniz, 2011; Vod vd., 2020). Ayrıca, giriřimci ve giriřimci olamayan kiřiler arasında da kiřisel sezgiler bakımından farklar olduęu saptanmıřtır (zdemir ve Karadeniz, 2009).

Santos vd. (2017) tarafından 151,400 giriřimciden oluřan veri seti kullanarak boylamsal alıřma analizi sonucunda Portekiz, İtalya, Yunanistan, İspanya, İsve, Finlandiya ve Norve’te fırsat algısının olumlu bir etkisi olduęunu tespit etmiřlerdir. Ayrıca bu arařtırma kiřilerin giriřimcilik konusunda yeterli bilgi ve yeteneęe sahip olmaları kiřilerin giriřimci olma niyetlerini, davranıřlarını olumlu ynde etkileyen kritik bir deęiřken olduęunu kanıtlanmıřtır.

SK aısından blgesel kalkınmada giriřimcilięin geliřtirilmesi, dolayısı ile blgenin sahip olduęu potansiyellerin kalkınma srecine entegre edilmesi son derece nemlidir. Blgelerde giriřimcilik faaliyetlerini artırarak blgeler arası geliřmiřlik farkları azaltılabilir. Giriřimcilik srecinde lkedeki blgesel farklılıklar, toplumsal tutumlar ve algılamalar, giriřimci bir kltr yaratmada nemli rol oynamaktadır (Celbiř, 2021). Giriřimcilik faaliyetleri, belirli kltrel ve sosyal kořullarda yařayan insanlar tarafından gerekleřtirilir ve toplumun giriřimcilikle ilgili olumlu veya olumsuz algıları, potansiyel ve mevcut giriřimcilerin giriřimci istekleri zerinde etkilidir. *Eęitim ve iř deneyimi*, giriřimcilik iin bireylerin beřer sermayeleri aısından nemlidir (Kim, Aldrich ve Keister, 2006). Bireylerin eęitim seviyeleri arttıka beřer sermayeleri, bilgi ve becerileri ve buna baęlı olarak giriřimci olma istekleri de artar (Ployhart ve Moliterno, 2011). Wennekers, van Stel, Thurik ve Reynolds (2005) ekonomik kalkınma seviyeleri daha yksek olan lkeler iin reel gelir ve serbest meslek arasındaki negatif iliřkinin dzeldięini gstermiřlerdir.

Giriřimcilik SK’yı ve blgesel kalkınmayı arttırır. Bu nedenle, giriřimcilik faaliyetlerinin kalkınma zerinde nemli bir rol vardır. Siffert ve Guimares (2020) ekolojik bozulma ve sosyal eřiřsizliklerin kresel bazda arttıęı 21. yzyılda SK’nın nemini vurgulayıp; SK’nın ekonomik boyutu ile sosyal ve evresel boyutlarını da ieren bir kalkınma modeline ihtiya olduęunu ve giriřimcilięin blgesel yařam kořullarını iyileřtirerek blgesel kalkınmaya fayda saęladıęını belirtmiřlerdir.

### 3. Veri

Bu alıřmada, Kresel Giriřimcilik Monitr tarafından saęlanan ikincil veriler kullanılmıřtır. GEM, giriřimcilięin ekonomik byme srecindeki nemini arařtırmak amacıyla, 1997 yılında Babson College ve London Business School tarafından bařlatılmıř giriřimcilik konusundaki kapsamlı bir akademik arařtırmadır.

Veriler standart GEM anketi kullanılarak rassal yöntemle ülke nüfusunu temsil edecek minimum 2000 örnekleme temel alarak yetişkin nüfusla telefon anketi veya yüz yüze görüşme yapılarak toplanır. GEM, bireylerin girişimci niteliklerini, girişimcilikle ilgili tutum, algı ve yaklaşımlarını inceleyecek veriler ile birlikte ayrıca, yeni girişimlerin kurucuları ve kuruluş süreçlerinin farklı aşamaları hakkında veriler toplamaktadır. Böylece, GEM, girişimcilik hakkında standartlaştırılmış veri kümeleri sağlar (Bosma vd., 2020). GEM Projesi, rastgele seçilen büyük bir örnekleme içerdiğinden, bulguların tüm bir popülasyona veya bir alt popülasyona genellemesi muhtemel olduğundan, nicel metodoloji kullanır (Carr, 1994).

GEM (2018) veri setinin makine öğrenmesi modelleri tarafından sorunsuz olarak kullanılabilmesi amacıyla veri seti ile ilgili bir dizi düzenleme yapılmıştır. Öncelikle aynı göstergelyi ölçen değişkenlerden bir kısmı hariç tutulmuştur. Örneğin “şu anda bir iş kurmakta mısınız” sorusunu temsil eden değişken (GEM verisindeki adıyla bjobst) gibi kişinin girişimci olup olmadığını belirten ve dolayısıyla bağımlı değişken olan Erken Evre Girişimcilik Faaliyeti (TEA) ile hemen hemen aynı değerlere sahip olan diğer değişkenler analize katılmamıştır. Benzer sebeplerden dolayı mükerrer olarak yaş, eğitim, gelir, bölge gibi göstergeleri farklı değişkenler ile ölçen veya farklı kodlayan değişkenler de hariç tutulmuştur. Bu ilk veri temizleme aşaması sonucunda 2,424 gözlem ve 288 değişkenden oluşan veri seti 2,057 gözlem ve 216 değişken boyutuna dönüşmüştür. İkinci olarak, bir sonraki bölümde daha detaylı tanımladığımız bağımlı değişken değeri olmayan gözlemler analizde kullanılamayacakları için hariç tutulmuş ve bu sebepten dolayı değişken sayısı 72’ye düşmüştür.

Makine öğrenmesi algoritmaları çoğu ekonometrik analiz paketinin aksine eksik değeri olan gözlemleri otomatik olarak düşürüp analize devam edememektedirler. Dolayısıyla eksik gözlemlerin en aza indirgenmesi gerekmektedir. Bu aşamada 2,057 gözlem arasında sadece 209’unun tüm değişkenler için eksiksiz gözlem noktasına sahip olduğu görülmüştür. En çok gözlemin düşmesinden sorumlu olan değişken veya değişken kombinasyonları algoritmik olarak taranmış ve elenmiştir. Bu üçüncü aşamanın sonucunda verinin tüm analizlerde kullanılan 1,563 gözlem ve 44 değişkenden oluşan boyutuna ulaşılmıştır. Karşılaştırma kolaylığı sağlamak amacıyla çalışmada değişken isimleri GEM (2018) veri setinde olduğu şekilde kullanılmıştır. Etik kurul izni ve/veya yasal/özel izin alınmasına gerek olmayan bu çalışmada araştırma ve yayım etiğine uyulmuştur.

#### **4. Yöntem ve Bulgular**

Türkiye’deki girişimcilik ve sürdürülebilir bölgesel kalkınma ilişkisini analiz eden bu araştırma; makine öğrenmesi yöntemleri ile elde edilen bulguları bu bölümde (1) Sınıflandırma ağacı ve rassal orman yöntemi ve bulguları ve (2) Stokastik ve stokastik olmayan gradyan artırma makineleri yöntemi ve bulguları olarak adlandırılan iki başlık altında incelemektedir.

##### **4.1. Sınıflandırma Ağacı ve Rassal Orman Yöntemi ve Bulguları**

Makine öğrenmesi, verinin sunduğu varyasyonlardan öğrenmek suretiyle örüntüler keşfetmeyi amaçlayan algoritmik yöntemlere verilen addır (Aldrich ve Auret, 2013). Başka bir deyişle, istatistiksel teori üzerine kurulmuş olan makine öğrenmesi, yapay zekanın itici gücüdür (Alpaydın, 2016). Öte yandan, ağaç temelli makine öğrenmesi yönteminin önde gelen bilim insanı Leo Breiman günümüzden yirmi yıl kadar önce istatistiksel analiz alanında iki farklı kültürün

hakim olduđunu ve bunlardan birinin modelleme bazlı diđerinin ise algoritmik yaklařım olduđunu belirtmiřtir (Breiman, 2001b). Bununla birlikte nedensellik iliřkilerine ađırlık verilen ekonomik analizlerde bařlarda yavař da olsa makine öğrenmesi algoritmalarına son yıllarda verilen deđer artmıř ve bu yöntemler daha yaygın olarak kullanılmaya bařlanmıřtır (Imbens ve Athey, 2021). Buna ek olarak makine öğrenmesi araçları modelleme kararlarını kiřisel sezgiler yerine verinin kendisine dayandırarak hali hazırda var olan ekonometrik yöntemlerin ötesine geçmektedirler (Hersh ve Harding, 2018).

“Sınıflandırma ve Regresyon Ađaçları” çerçevesi içerisinde bulunan sınıflandırma veya karar ađaçları birçođ makine öğrenmesi topluluk modelinin (enseble models) temelini oluřturmaktadır (Breiman, Friedman, Olshen ve Stone, 1984). Makine öğrenmesi yöntemleri arasında ađaç bazlı yaklařımlar deđiřkenler arasında yüksek derecede etkileřimlere izin veren ve dođrusal olmayan karmařık iliřkilerin modellenmesine olanak sađlayan yaklařımlar olarak öne çıkmaktadırlar (Mullainathan ve Spiess, 2017). Bu yöntemler veriyi katmanlařtırarak ve bölerek öngörüler yapmayı hedeflemektedirler (James, Witten, Hastie ve Tibshirani, 2013). Bu yöntemler özünde basit olmakla beraber öngörü performansı ađısından kayda deđer biçimde başarılıdırlar (Friedman, Hastie ve Tibshirani, 2001). Sınıflandırma ađaçları verinin eđitme kısmını (training data) rekürsif (özçađrılı) bir biçimde ikili ve ařamalı olarak böler. Makine öğrenmesi algoritmalarını uygulayabilmek için ilk olarak GEM verisi içerisinde bulunan bireylerin yüzde 70’i eđitme verisi, yüzde 30’u ise test verisine dahil olacak řekilde rastgele ayrılmıřtır.

Kullanılan makine öğrenmesi algoritmalarına bir temel oluřturabilmek için öncelikle kiřinin Erken Evre Giriřimcilik faaliyetinde bulunup bulunmadıđının EVET-HAYIR kategorileri ile belirtildiđi bir sınıflandırma ađacını Friedman vd. (2001), James vd. (2013) ve Breiman vd. (1984) tarafından belirtilen yaklařım ile oluřturulmuřtur. Bu yöntemin tanımı ve bu çalışmada irdelenen arařtırma sorusuna uygulanıřı řu řekilde açıklanabilir: Bir sınıflandırma ađacındaki her veri bölme durađında ( $d$ ) algoritma çeřitlilik derecesini (impurity) azaltmayı hedefler. Eđer bu ikili bölme iřlemi, çeřitlilik derecesinde bir azaltma gerçeđleřtirmiyorsa; söz konusu iřlem verinin  $d$  durađına (node) tekabül eden kısmına uygulanmaz.

Çeřitlilik çođunlukla bir Gini sayısı ile ölçülür. Bađımlı deđiřkenimiz sadece iki sınıftan (EVET-HAYIR) oluřsa da  $c$  adet sınıf için  $d$  durađının bölünmesi sonucu çıkacak Gini sayısı  $\Gamma_d$  řu řekilde hesaplanabilir:

$$\Gamma_d = \sum_{c=1}^c s_{dc}(1 - s_{dc}) \quad (1)$$

Denklem 1’de yer alan  $s_{dc}$  ifadesi sınıflandırma ađacının  $d$  durađında  $c$  sınıfına  $c = 1, \dots, C$  düşen kiři oranını vermektedir:

$$s_{dc} = \frac{1}{N_d} \sum_{i \in M_d} \mathbf{1}(y_i = c) \quad (2)$$

Daha ađık řekilde olarak;  $s_{dc}$ ’yi hesaplamak için verinin  $M_d$  alt kümesindeki  $i$  bireyleri arasından bađımlı deđiřken ( $y_i$ ) deđerleri  $c$  sınıfına eřit olanlar sayılır ve bu toplam sayı  $d$  durađındaki gözlem sayısına ( $N_d$ ) bölünür. Dolayısıyla, Denklem 1’deki  $\Gamma_d$  sayısını hesaplamak için  $d$  durađındaki  $c$  sınıfının oranı ile o sınıfın dıřında kalan tüm diđer sınıflara düşen gözlem sayısının oranı çarpılır. İřlem tüm  $c$  sınıfları için tekrar edildikten sonra sonuçlar toplanır ve  $d$  durađı için Gini deđeri  $\Gamma_d$  hesaplanır. Böylece düşük bir  $\Gamma_d$  sayısı  $d$  ađaç durađında toplanmıř

bireylerin Erken Evre Girişimcilik durumları itibariyle benzeştikleri anlamına gelir. Başka bir deyişle  $d$  durağında EVET veya HAYIR kategorilerinden birinin oransal olarak dominant olduğunu gösterir (Géron, 2019; James vd., 2013). Çalışmanın gözlem birimlerinin Erken Evre Girişimcilik aktivitesinde bulunan ve bulunmayan kişiler olarak iki kategoride sınıflandırılmaları sebebiyle  $\Gamma_d$  değeri  $2s_{d1}(1 - s_{d1})$  şeklinde ifade edilebilir. Neticede, modellerin temelini oluşturan bu rekürsif ikili ayrıştırma (recursive binary partitioning) algoritması her bir ayrıştırma aşamasında test verisini  $d_1$  ve  $d_2$  alt duraklarına (child nodes) tekabül eden kişi sayısının toplam gözlem sayısı ( $N$ ) içindeki oranlarıyla ağırlıklandırılmış çeşitlilik değerlerini toplamda en aza indirme amacıyla öznelik uzayından ( $k = 1, \dots, K$ ) bir ayrıştırıcı değişken  $x_k$  ve onun ayrıştırma değeri (split value)  $v$ 'yi bulur ve veriyi bu değeri kullanarak ikiye ayırır (Breiman vd., 1984; Friedman vd., 2001; James vd., 2013):

$$\min_{k,v} \left[ \frac{N_{d_1}(k, v)}{N} \Gamma_{d_1}(k, v) + \frac{N_{d_2}(k, v)}{N} \Gamma_{d_2}(k, v) \right] \quad (3)$$

Doğal olarak bu özçâğrılı ikili ayrıştırma işleminin veri boyutunun mümkün kıldığı kadar devam etmesine izin vermek eğitime verisinin neredeyse tam bir kesinlikle ve başarıyla tahmin edilmesine yol açacaktır. Bu ilk bakışta olumlu bir durum gibi gözükse bile sıra test verisinin tahmin edilmesine geldiğinde aşırı uyumluluk (overfitting) sorun yaratacaktır ve modelin kullandığı örneklemin dışında kalan başka herhangi bir gözlem kümesinin tahmini de büyük ölçüde başarısız olacaktır. James vd. (2013), Sutton (2005) ve Friedman vd. (2001) tarafından tanımlandığı biçimde bu sorunu şu şekilde azaltmak mümkündür: ikili ayrıştırma algoritması “10 katlamalı çapraz doğrulama” (10-fold cross validation) yöntemi ile kısıtlanır ve bu işleme “budama” (pruning) adı verilir. Kısıtlamanın derecesi  $\tau$  ile temsil edilen kompleksite parametresi (complexity parameter) ile temsil edilmiştir. Doğrulama algoritması eğitime verisinin (training data) rastgele örneklem alınarak belirlenmiş her 10 dilimi için  $\tau$ 'nin farklı değerlerine karşılık gelen  $r = (1, \dots, R)$  alt ağaçları üretir. Her bir alt ağaca ait tüm nihai duraklar ( $\bar{d} = 1, \dots, |r|$ ) dahilinde çoğunlukta olan kategoriye  $c^*$  ile temsil edersek, bir  $M_{\bar{d}}$  gözlem kümesi için yapılan tahmin hatası  $\varepsilon_{\bar{d}}$  Denklem 4’te gösterilen şekilde belirtilir:

$$\varepsilon_{\bar{d}} = \frac{1}{N_{\bar{d}}} \sum_{i \in M_{\bar{d}}} \mathbf{1}(y_i \neq c^*) \quad (4)$$

En uygun  $\tau$  değerini belirleyebilmek amacıyla eğitime verisinin (training data) önceden belirlenmiş 10 kesiti için Denklem 5’teki ifadeyi ortalamada en düşük şekilde verecek olan  $\tau$  değeri aranır. Yukarıda belirtildiği gibi, bunu yapabilmek için her biri farklı bir  $\tau$  değeriyle hesaplanan alt ağaçlar üretilir. Başka bir deyişle, bir alt ağacın nihai duraklarındaki toplam ağırlıklı hata oranını (Denklem 5’in ilk terimini) minimize eden; ancak aynı zamanda, ağacın karmaşıklığını (nihai durak sayısı  $|r|$ 'yi) en düşük düzeyde tutacak bir denge aranır. Örneğin,  $\tau = 0$  bulunması karmaşıklık derecesine bir sınır konulmadığı ve sınıflandırma ağacının rekürsif ikili ayrıştırma adımlarına ayrıştırılacak gözlem kalmayana kadar devam edeceğini ifade eder.  $\tau > 0$  ise büyük bir  $|r|$  sayısının uygun olmadığı ve aşırı uyumluluğa karşı budanmış (kısıtlanmış) bir ağacın kullanılacağını belirtir:

$$\sum_{\bar{d}}^{|r|} \varepsilon_{\bar{d}} \frac{N_{\bar{d}}}{N} + \tau |r| \quad (5)$$



Sınıflandırma ağacının eğitme verisi dışında tahminlere yönelik olarak aşırı uyumluluktan kaynaklanabilecek başarısızlık ihtimalini azaltmak için ağacın karmaşıklığını  $\tau$ 'ya ek olarak her  $M_{\bar{d}}$  kümesine düşecek minimum gözlem miktarı ve herhangi bir nihai durak  $\bar{d}$ 'nin ulaşabileceği derinlik (ağacın kökünden herhangi bir son durağa erişene kadar uygulanan ikili ayrıştırma veya durak sayısı) parametreleri de belirlenir. 2-20 aralığında maksimum derinlik ve minimum gözlem için 361 kombinasyonu ve bu kombinasyonların her biri için elde edilen çapraz doğrulama hata seviyesine karşılık gelen  $\tau$  değeri tarama yoluyla (grid search) belirlenmiştir. Bu taramanın sonucunda  $\tau$  değeri 0.0107, maksimum derinliği 8 ve minimum son durak gözlem sayısı 9 olarak belirlenen sınıflandırma ağacı Şekil 1'de görselleştirilmiştir. Oluşturulan bu ağacın en baştan beri dışarıda bırakılan test verisini başarılı tahmin etme oranı (accuracy) %87'dir.

Herhangi bir rekürsif ayırım algoritması gerekmeyecek şekilde, test verisinde yer alan tüm kişiler girişimci değil (HAYIR) olarak tahmin edilseydi başarı oranı 0.85 olacaktı. Üstelik bu yaklaşımın doğal olarak herhangi bir açıklayıcılığı da olmayacaktı. Dolayısıyla, sınıflandırma ağacı uygulaması tahmin yapmada başarısız olmamakla birlikte girişimcilik çıktılarının altında yatan ilişkileri haritalandırabilmektedir. Öte yandan tekil bir sınıflandırma ağacı birçok açıdan yetersiz kalabilir. Aynı eğitme verisinden alınsa dahi farklı bir örneklem ile oluşturulan bir ağaç oldukça farklı bir şekilde ortaya çıkabilir. Seçilen değişkenler, ikili ayrıştırmalar, ağacın derinliği ve genel karmaşıklığı farklı olabilir. Üstelik eğer bazı açıklayıcı değişkenlerin arasında yüksek korelasyon var ise tekil sınıflandırma ağacı yüksek önem arz eden değişkenleri analize katmayabilir (Athey ve Imbens, 2019; James vd., 2013). Bu sebeplerden dolayı, tekil sınıflandırma ağacından farklı olan ve çok sayıda ağacın birleşmesinden oluşan toplum modelleri (Rassal Orman ve Stokastik Gradyan Artırma algoritmaları) bu çalışmanın ana metotlarını oluşturmaktadırlar.

Şekil 1'de sunulan tekil sınıflandırma ağacı 44 değişkenin içerisinde başka girişimciler tanıma (*knowent*), ihtiyaç ve işsizlik (*subnec18*, *occuump*, *occufull*), şirket içi girişimcilik aktivitesinde bulunmuş olmak (*ipactld\_all*), girişimciliğin toplum ve medya tarafından nasıl karşılandığı (*nbstatus*, *nbmedia*), başarısızlık korkusu (*fearfail*), yaş ve bölge (*age*, *TRregion*) ile ilgili özellikleri seçmiştir. Algoritmik olarak sınıflandırma ağacı tarafından seçilmiş bu değişkenler Tablo 2'de tanımlanmıştır.

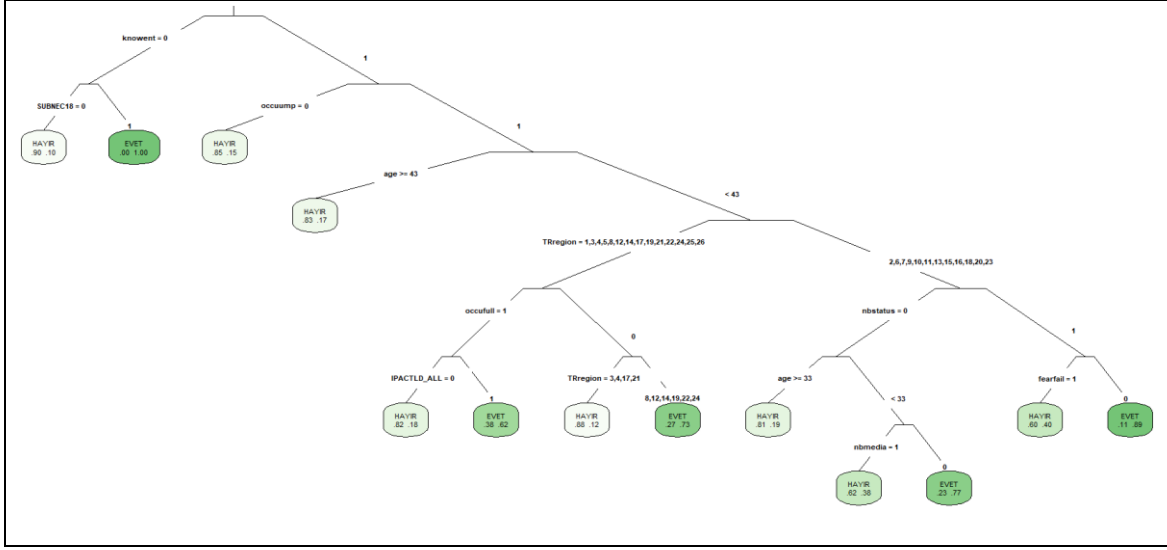
Şekil 1'deki sınıflandırma ağacının terminal (nihai) duraklarında o kümeye düşen kişilerin girişimci olarak tahmin edilip edilmediği EVET-HAYIR sınıflarına gruplanarak belirtilmektedir. Son duraklarda sırasıyla girişimci olmayan ve olan kişilerin nihai kümedeki oranı hem sayısal hem de renk tonları ile temsil edilmektedir. Koyu renk tonları girişimci olduğu tahmin edilen kişilerin oranının küme içerisinde yüksek olduğu anlamına gelmektedir. Tersini ise açık renkler ile temsil edilmektedir.

**Tablo 2. Tekil Sınıflandırma Ağacı Değişken Tanımları**

Değişken	Tanım	Değerler
<i>age</i>	Yıl olarak kişinin yaşı.	Sayısal değişken.
<i>fearfail</i>	Başarısız olma korkusunun kişiyi bir iş kurmaktan caydıracağı yönündeki görüşü.	Kategorik değişken (Evet-Hayır)
<i>IPACTLD_ALL</i>	Son üç yıl içerisinde aktif veya lider olarak şirket içi girişimcilik aktivitesinde bulunmuş olmak.	Kategorik değişken (Evet-Hayır)
<i>knowent</i>	Son iki sene içerisinde bir şirket kurmuş birisini tanıma durumu	Kategorik değişken (Evet-Hayır)
<i>nbmedia</i>	Kişinin yeni kurulan işletmelerin ülke medyasında çokça yer bulma veya bulamamalarına yönelik görüşü.	Kategorik değişken (Evet-Hayır)
<i>nbstatus</i>	Bir iş kurmayı başaran kişilerin ülkede daha iyi bir statü ve saygınlık düzeyine sahip oldukları konusunda kişinin görüşü.	Kategorik değişken (Evet-Hayır)
<i>occufull</i>	Kişinin tam zamanlı olarak istihdam edilme durumu.	Kategorik değişken (Evet-Hayır)
<i>occuump</i>	Kişinin işsiz olma durumu.	Kategorik değişken (Evet-Hayır)
<i>TRregion</i>	İşletmenin bulunduğu bölge, TÜİK İBBS-2 Düzey (26 Bölge) sınıflandırması	1: Adana, 2: Ağrı, 3: Ankara, 4: Antalya, 5: Aydın, 6: Balıkesir, 7: Bursa, 8: Erzurum, 9: Gaziantep, 10: Hatay, 11: İstanbul, 12: İzmir, 13: Kastamonu, 14: Kayseri, 15: Kırıkkale, 16: Kocaeli, 17: Konya, 18: Malatya, 19: Manisa, 20: Mardin, 21: Samsun, 22: Şanlıurfa, 23: Tekirdağ, 24: Trabzon, 25: Van, 26: Zonguldak.
<i>SUBNEC18</i>	Kişinin başlangıç düzeyinde ve ihtiyaç nedeniyle girişimcilik faaliyetinde olma durumu.	Kategorik değişken (Evet-Hayır)

**Kaynak:** GEM, 2018.

Birinci ikili ayrıştırma işleminin son iki yıl içerisinde şirket kurmuş birisini tanıma veya tanımama durumuna göre (*knowent*) yapıldığını görüyoruz. Ancak, şirket kurmuş birisini tanımamalarına rağmen girişimci olan kişilerin ihtiyaç/mecburiyet dolayısıyla girişimci olma yolunu seçenler olduklarını görüyoruz. Öte yandan sınıflandırma ağacı böyle bir ihtiyaç durumu mevcut olmayan ve şirket kurmuş olan birisini tanımayanların Erken Evre Girişimcilik faaliyetinde bulunmadıklarını tahmin etmektedir. Eğitim verisindeki girişimci tanıdığı olan kişilerin 43 ve üstü bir yaşta ise (*age*) veya işsiz değil ise (*occuump*) yine erken evre girişimci olmaya teşebbüs etmedikleri görülmektedir (Şekil 1).



Şekil 1. Tekil Sınıflandırma Ağacı

Kaynak: Yazar tarafından hazırlanmıştır.

Sınıflandırma ağacına göre eğitim verisindeki 43 yaşından genç gözlemler için ise şöyle bir durum görülmektedir: bu kişiler arasında Adana, Ankara, Antalya, Aydın, Erzurum, İzmir, Kayseri, Konya, Manisa, Samsun, Şanlıurfa, Trabzon, Van veya Zonguldak İBBS-2 bölgelerinde bulunan grubun içerisinde tam zamanlı olarak çalışmakta olanlar (*occufull*) ve onların arasından son üç sene içerisinde şirket içi girişimcilik faaliyetlerinde aktif rol almış olanların (*IPACTLD\_ALL*) Erken Evre Girişimcilik faaliyetinde de bulunacakları ön görülmektedir. Aynı grup içerisinde bu yönde şirket içi aktivitede bulunmayanların ise girişimci olmayacakları yönünde bulgu mevcuttur. Yine bu grup arasından tam zamanlı bir işte çalışmayan kişiler Ankara, Antalya, Konya, veya Samsun bölgelerinde bulunuyorlarsa girişimci olmadıkları tahmin edilmektedir. Ancak, tam zamanlı çalışmayan bu kişiler eğer Erzurum, İzmir, Kayseri, Manisa, Şanlıurfa veya Trabzon'da iseler erken evre girişim faaliyetinde bulduklarını tahmin etmekteyiz.

Rekürsif ikili ayrıştırma modeline göre 43 yaş veya daha altı yaşlardaki kişiler arasından Ağrı, Balıkesir, Bursa, Gaziantep, Hatay, İstanbul, Kastamonu, Kırıkkale, Kocaeli, Malatya, Mardin veya Tekirdağ İBBS-2 bölgelerinde bulunanlar için şöyle bir çıkarım yapılmaktadır: girişimcilik faaliyetlerinin Türkiye'de bir insanı daha iyi bir statü ve saygınlık düzeyine erdirdiğini düşünmeyen (*nbstatus*) ve 33 yaşında veya daha büyük olan kişilerin erken evre girişimci olmadığını tahmin etmekteyiz. Bu grup arasından 33 yaşından küçük olup erken evre girişimci olarak tahmin edilen kişilerden girişimcilik faaliyetinde bulunanların ülke medyasında çokça yer bulabildiğini düşünmediklerini görmekteyiz. Tersisi durum ise girişimci olmayan kişiler için gözlenmektedir. Ancak, bu tahminin çeşitlilik seviyesi yüzde dağılımlarından görülebileceği gibi yüksektir, dolayısıyla tahmin çok "güçlü" değildir. Her halükarda *nbstatus* ve *nbmedia* değişkenleri tarafından ölçülen sosyal statü ve tanınırlığın bu kesim için önem arz ettiğini görmekteyiz. Belirtilen bölgelerde bulunup, statü ve saygınlık olarak girişimciliğin kişiyi daha yüksek bir noktaya taşıdığını düşünenler için ise başarısız olma korkusu (*fearfail*) önemli görülmektedir. İş kurma konusunda başarısız olma endişesine sahip olanların, olmayanların aksine, erken evre girişimci olmadığı sınıflandırma ağacı tarafından tahmin edilmektedir.

Yukarıda altı çizilen eksikliklerine rağmen tekil sınıflandırma ağacı, girişimcilik ile ilgili doğrusal olmayan, interaktif ve yüksek etkileşimli mekanizmaların yapısına yönelik tüyolar vermektedir. Aynı zamanda Şekil 1’de görselleştirilen rekürsif ikili ayrıştırma adımları bu çalışmada kullanılan ana modeller olan Rassal Orman ve Gradyan Artırma algoritmalarının bireysel yapı taşlarının bir örneği olarak görülebilir. Tekil ağaç yerine birçok ağacın kullanıldığı topluluk algoritmaları ise birçok avantaj sunmaktadır. Topluluk algoritmalarının önde gelen uygulamaları çok sayıda ağaçtan oluşan Rassal Orman (Random Forest) ve ardışık ağaçlardan oluşan Gradyan Artırma’dır (Gradient Boosting). Bu yaklaşımlar arasında ilk kullandığımız yöntem olan Rassal Orman algoritması Breiman (2001a) tarafından daha önce geliştirilmiş olan rastgele örneklem toplama metodunun (Breiman, 1996) daha da rassal bir versiyonu olarak görülebilir ve neticede iki metot da birbirine devamlılık açısından bağlıdır. Bu algoritma rassallığı  $L$  sayıda sınıflandırma ağacı üretirken ( $l = 1, \dots, L$ ) iki çeşit rassallaştırma yapmaktadır. Öncelikle  $L$  ağaçların üretildiği her bir  $l$  iterasyonda  $N$  kadar gözlem yerine koyarak örnekleme usulüyle veri setinin %70’lik eğitime kısmından çekilir ve  $l$ ’inci ağaç bu örneklem kullanılarak üretilir. İkinci olarak Rassal Orman her  $l$ ’inci ağacın ikili ayrıştırma adımlarında tüm  $K$  değişkenler yerine rastgele belirlenmiş  $\sqrt{K}$  büyüklüğünde bir alt küme içerisindeki değişkenleri hesaba katar (Breiman 2001a; Friedman vd., 2001). İki farklı şekilde sağlanan bu stokastisite sayesinde hem sınıflandırma ağaçlarının arasındaki korelasyon azaltılmakta (Friedman, 2001; James vd., 2013) hem de birbirleri arasında yüksek korelasyon olan değişkenlerin bazılarının hatalı şekilde analiz dışında bırakılma durumu en aza indirgenmektedir (Athey ve Imbens, 2019; James vd., 2013). En son aşamada  $i$  kişinin girişimci olup olmadığı üretilen  $L$  ağacın oy çoğunluğu ile tahmin edilir.

Yukarıda detaylandırıldığı üzere Breiman (2001a) tarafından geliştirilen Rassal Orman algoritmasının her bir iterasyonunda farklı örneklemler ve değişken alt kümeleri kullanılarak üretilen çok sayıda ağaç içermektedir. Dolayısıyla görsel açıdan tekil ağaç örneğindeki gibi örüntüleri detaylı biçimde temsil etmek mümkün değildir. Sonuçta, öznelik uzayında bulunan herhangi bir  $k$  değişkeni farklı iterasyonlara değişik fonksiyonel yapılarda veya etkileşimler ile dahil edilmiş veya iterasyonda oluşturulan ağacın dışında kalmış olabilir (Breiman, 2001a; Friedman vd., 2001; James vd., 2013). Dolayısıyla örüntüler tüm algoritma için temsil edilemeye bile genel bir “öznelik önemi” (feature importance) çıkarımı oldukça kullanışlıdır (Friedman vd., 2001; James vd., 2013). Öznelik önemini ölçmek için bir değişkenin algoritmanın kestirim performansına yaptığı katkı ölçülür (Friedman vd., 2001; James vd., 2013).

Denklemler 3’te belirtildiği gibi bir sınıflandırma ağacındaki her bir ikili ayrıştırma aşaması çeşitlilik seviyesini (impurity) azaltabilmeyi amaçlar (Breiman, 2001a; Friedman vd., 2001; James vd., 2013). Eğer çeşitlilik yani benzeşmezlik seviyesi hiçbir değişken tarafından azaltılamıyor ise zaten söz konusu durakta ikili ayrıştırma işlemi yapılmaz ve bu durak terminal yani nihai statüde olur (Breiman, 2001a; Friedman vd., 2001; James vd., 2013). Rassal Orman içerisindeki tüm  $L$  iterasyonda (ağaçta) bir  $x_k$  değişkeninin  $I$  ile temsil ettiğimiz çeşitlilik derecesini ne denli azaltabildiğini ölçebilmek amacıyla  $d$  durağının  $d_1$  ve  $d_2$  duraklarına ayrıştırılmasından önce olan Gini değeri  $\Gamma_d$  ile ayrıştırma sonrası oluşan iki durak için elde edilen Gini değerleri  $\Gamma_{d_1}$  ve  $\Gamma_{d_2}$ ’nin ağırlıklı toplamı ile farkı yani  $\Delta I$  hesaplanır (Breiman, 2001a; Friedman vd., 2001; James vd., 2013):

$$\Delta I = \Gamma_d - \left[ \frac{N_{d_1}}{N} \Gamma_{d_1} + \frac{N_{d_2}}{N} \Gamma_{d_2} \right] \quad (6)$$

Sonrasında  $x_k$  'nın tüm ayrıştırma aşamalarında ve ağaçlarda kazandırdığı  $\Delta I$  değerlerinin toplamı 1-100 sayıları arasına ölçeklenir ve  $x_k$  'nın öznitelik önemi hesaplanır (James vd., 2013). 500 ağaçtan oluşan ve tahminde %88 oranında başarılı olan Rassal Orman modeli tarafından en yüksek öznitelik önemine sahip olduğu belirlenen ilk 20 değişken Şekil 2'de gösterilmiş ve aralarından halihazırda Tablo 2'de tanımlanmayanlar Tablo 3'te tanımlanmışlardır. Denklemlerde kullanılan tüm işaret ve simgeler Ek'te sunulan Tablo 5'te açıklanmıştır.

**Tablo 3. Rassal Orman Değişken Tanımları (Tekil ağaç modelinin seçtiği değişkenler hariç)**

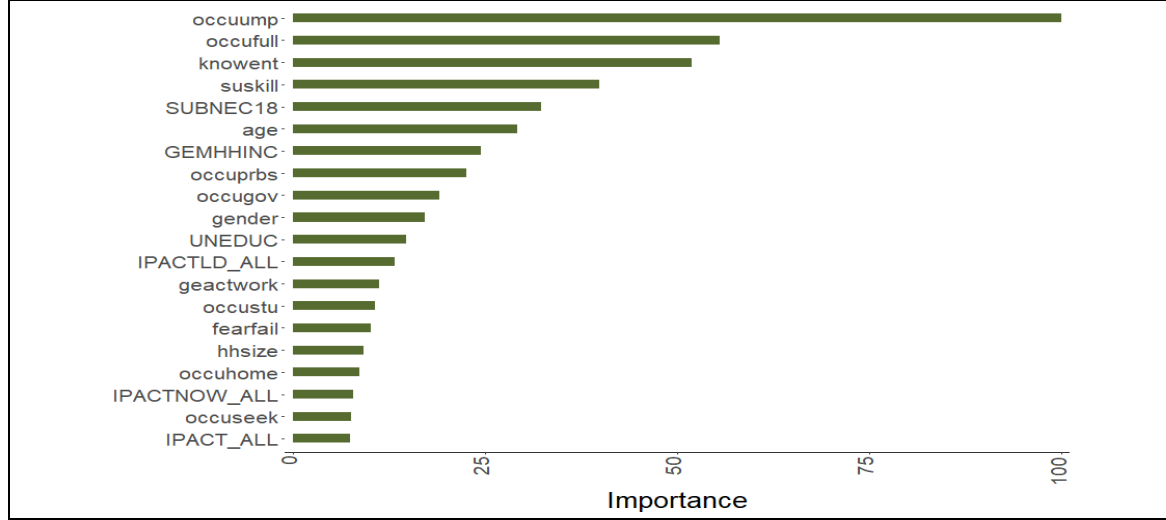
Değişken	Tanım	Değerler
<i>GEMHHINC</i>	Hanehalkı gelir grubu.	Sıralı değişken. 1: ilk %33'lük dilim, 2: ikinci %33'lük dilim, 3: üçüncü %33'lük dilim.
<i>gender</i>	Cinsiyet.	Kategorik değişken (Evet-Hayır)
<i>IPACTNOW_ALL</i>	Şirket içi girişimcilik faaliyetinde bulunuyor olmak.	Kategorik değişken (Evet-Hayır)
<i>IPACT_ALL</i>	Son üç yıl içerisinde liderlik seviyesinde olmasa da şirket içi girişimcilik faaliyetinde aktif olmak.	Kategorik değişken (Evet-Hayır)
<i>hhsiz</i>	Kişinin kendisi dahil hane halkı nüfusu	1 – 25 arası sayısal değişken.
<i>occugov</i>	Kişinin devlet bünyesinde istihdam edilme durumu.	Kategorik değişken (Evet-Hayır)
<i>occuhome</i>	Kişinin tam zamanlı olarak kendi ev işlerinde çalışması durumu.	Kategorik değişken (Evet-Hayır)
<i>occuprbs</i>	Kişinin özel bir işletmede istihdam edilme durumu.	Kategorik değişken (Evet-Hayır)
<i>occuseek</i>	İş arıyor olma durumu.	Kategorik değişken (Evet-Hayır)
<i>occustu</i>	Öğrenci olma durumu.	Kategorik değişken (Evet-Hayır)
<i>suskill</i>	Bir iş kurmak için gerekli bilgi ve becerilere sahip olma konusunda kişinin görüşü.	Kategorik değişken (Evet-Hayır)
<i>UNEDUC</i>	Eğitim durumu.	Sıralı değişken. 0: Okul öncesi, 1: İlköğretim, 2: Alt ilköğretim (5-8 sınıflar), 3: Üst ilköğretim (lise), 4: Lise sonrası üniversite öncesi eğitim, 5: Önlisans eğitimi, 6: Üniversite.

**Kaynak:** GEM, 2018.

Rassal Orman modeli ile sınıflandırma ağacı bulguları arasında paralellikler bulunmasının yanı sıra beklenildiği üzere Rassal Orman yaklaşımı yeni bilgiler edinmemizi sağlamaktadır. Kullanılan ikili ayrıştırma iterasyonlarının çokluğundan ve her birinde üretilen ağaçların heterojenliğinden dolayı değişkenlerin kısmi bağımlılık değerlerini incelemeyen girişimcilik ile aralarındaki ilişkinin yönünü gözlemlemek mümkün değildir. Ancak, öznitelik önemleri grafiği ilk aşamada yukarıda belirtildiği üzere tahminleri yapmada hangi değişkenlerin en çok kullanıldığını özetler.

Şekil 2'de özetlenen bulgulara göre iş durumu değişkenleri (*occuump*, *occufull*) en önemli iki değişken olarak bulunmuştur. Bir başka deyişle bu değişkenlerin kazandırdığı toplam  $\Delta I$  değerleri diğer değişkenlerden çok daha fazladır. Sınıflandırma ağacında da gözlemlendiği üzere başka bir girişimci tanımak da (*knowent*) yüksek önem arz etmektedir. İlk altı değişken arasında

sınıflandırma ağacının **gözlemleyemediği** tek değişken olan kişinin kendi bilgisi ve becerisi hakkındaki girişimcilik bağlamındaki görüşü (*suskill*) dördüncü sırada gelmektedir ve sırasıyla ihtiyaç/mecburiyet ve yaş faktörleri onu izlemektedir.



Şekil 2. Rassel Orman Öznitelik Önemleri

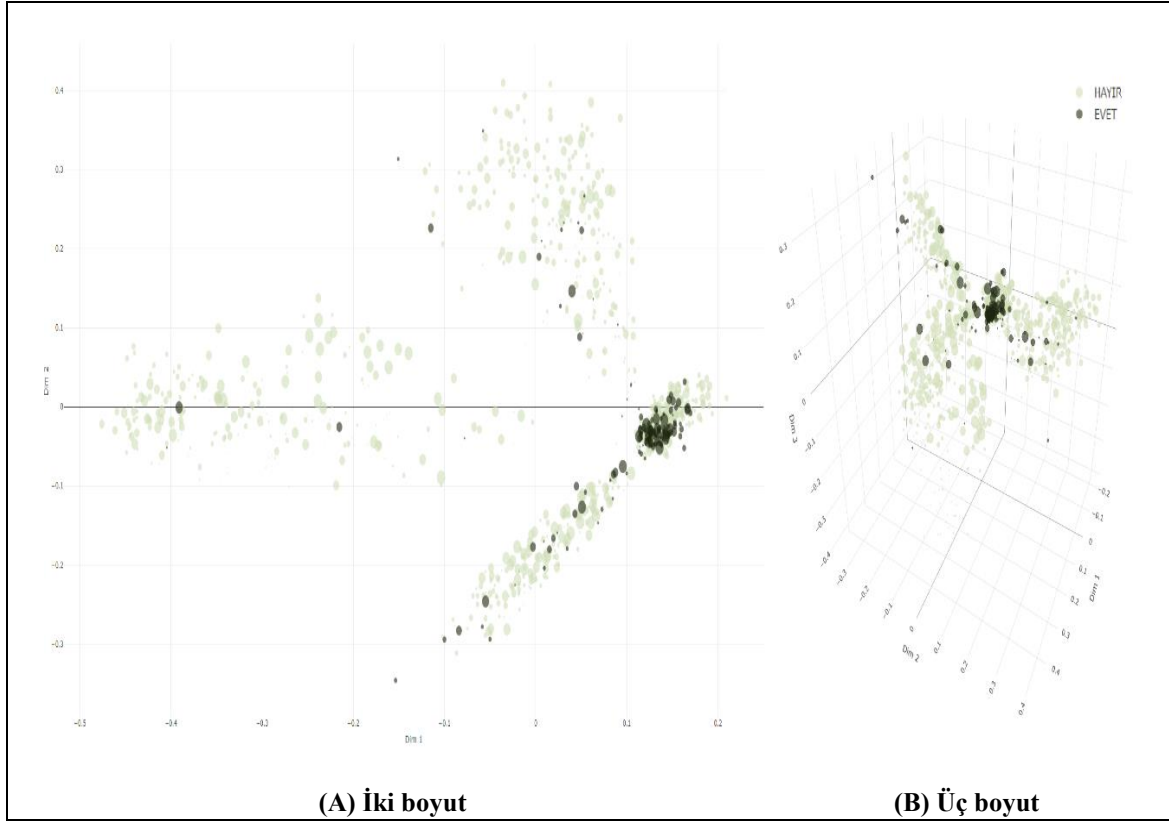
**Kaynak:** Yazar tarafından hazırlanmıştır.

Budanmamış ağaçlardan oluşan Rassel Orman, sınıflandırma ağacının kullandığı değişkenlere ek olarak ve daha düşük önem seviyeleri ile hanehalkı geliri (*GEMHHINC*), kişinin özel sektör, kamu sektörü, kendi ev işlerinde çalışması, öğrenci olması veya iş aramakta olması (sırasıyla *occuprbs*, *occugov*, *occuhome*, *occustu*, *occuseek*), cinsiyeti (*gender*), eğitim düzeyi (*UNEDUC*), hanehalkı nüfusu (*hhsize*), anket zamanı itibariyle şirket içi girişimcilik aktivitesinde bulunuyor olması veya son üç yılda liderlik seviyesinde olmasa bile bu konuda şirket içerisinde aktif olması (*IPACTNOW\_ALL*, *IPACT\_ALL*) değişkenlerini kullanmıştır.

En çok kullanılan değişkenler açısından sınıflandırma ağacı sonuçları ile tutarlı olmakla beraber, Rassel Orman ve bir sonraki bölümde sunulacak Gradyan Artırma modelleri yeni değişkenler seçmiş ve bunları göreceli olarak önem açısından tartmışlardır. Değişkenlerin girişimcilik çıktısı ile ilişkisini daha iyi anlayabilmek için bir sonraki bölümde kısmi bağımlılık ve koşullu bireysel beklenti metot ve bulgularını her iki model için karşılaştırmalı olarak irdelemektediriz. Ancak, öncesinde Rassel Orman algoritmasının başarısına bakmak faydalı olacaktır. Rassel Orman her ardışık işlemde yeni bir örneklem çektiği için eğitim veri seti içerisindeki her bir *i* gözlemi (kişisi) için o gözlemin hariç tutulduğu iterasyonlar uygulanmaktadır. Bu iterasyonlara denk gelen tekil sınıflandırma ağaçları vasıtasıyla tüm *i* gözlemlerinin girişimci olup olmama durumu onları dışarıda bırakan, bir başka deyişle “torba dışı” (out of bag, OOB) kaldıkları sınıflandırma ağaçları tarafından tahmin edilebilir ve bunun sonucunda test veri seti için hata tahmini yapılabilir. Çapraz doğrulama yaklaşımı gibi bu metot da test hatasını kestirme amacı taşır; ancak, gözlemler sistematik (dilimli) şekilde değil, rassal bir biçimde her iterasyonda hariç tutularak ayrılır (James vd., 2013).

Rassel Orman algoritmasının girişimci ve girişimci olmayan olarak tahmin edilen gözlemleri bu iki kategori arasında ne kadar belirgin bir şekilde ayırttığı konusunda fikir sahibi olabilmek için Rassel Orman yakınlık grafiği (Random Forest proximity plot) kullanışlı bir

görselleřtirme sunmaktadır (Breiman ve Cutler, 2020; Friedman, 2001). Yukarıda belirtildiđi řekilde gözlemlerin kendilerinin hariç tutulduđu sınıflandırma ağaçları tarafından tahmin edilmesi işlemine dayanarak bu işlem boyunca gözlemlerin birbirleri ile hangi sıklıkta aynı terminal durađa düřtükleri kayıt altına alınır. İki gözlem aynı nihai durakta bulunduđu her bir iterasyonda bu iki gözlem arasındaki yakınlık puanına 1 eklenir. Sonuç olarak her bir gözlem arasındaki yakınlık derecesinin puanlandıktan sonra sınıflandırma ağacı sayısına bölünüp normalize edilmesiyle bir  $N \times N$  matris elde edilir. Metrik Çok Boyutlu Ölçekleme kullanılarak bu matrisi iki veya üç boyuta indirgenerek gözlemler arası benzerlik görselleřtirilir (Breiman ve Cutler, 2020; Friedman, 2001).



Şekil 3. Rassal Orman Yakınlık Grafiđi  
Kaynak: Yazar tarafından hazırlanmıştır.

Şekil 3'te sunulan iki ve üç boyutlu yakınlık grafiklerine göre Rassal Orman Metodu sınıfları ayrıştırmada oldukça başarılı olmuştur. Her iki grafikte de girişimci olarak tahmin edilen gözlemlerin (EVET) kendi aralarında kümelenildiđi gözükmektedir.

#### 4.2. Stokastik ve Stokastik Olmayan Gradyan Artırma Makineleri Yöntemi ve Bulguları

Bu bölümde sınıflandırma ağacı ve Rassal Orman tekniklerine ek olarak ardışık ağaçlardan oluşan topluluk algoritmaları ile sonuçlarımıza alternatif bir bakışla yaklaşmaktayız. Stokastik Gradyan Artırma algoritması (stochastic gradient boosting machine) stokastik olmayan Gradyan Artırma modelinin rassal bir versiyonudur (Friedman, 2001; Friedman, 2002). Bu iki model

içerisindeki ağaçlar ardışık olarak bir önceki ağacın hatalarından öğrenme ve önceki yanlış tahminleri düzeltme amacı taşırlar. Bir başka deyişle, girişimci olma ve olmama kategorilerine sınıflandırılması bir önceki ağaç tarafından hatalı biçimde yapılmış gözlemlere ağırlık verilerek bir sonraki ağaç oluşturulur.

Stokastik olmayan Gradyan Artırma algoritması öncelikle bir başlangıç tahmini yapar. Bu tahmin eğitim verisindeki bir kişinin girişimci olmasının oranının logaritmasıdır (Friedman, 2001; Friedman vd., 2001; Friedman, 2002). İkinci adımda ilk tahminden elde edilen kalıntı terimleri (residuals) yeni bir regresyon ağacı vasıtasıyla kestirilir.  $L$  sayıda ardışık ağaç arasında yer alan bir  $l$  ağacından önce gelen  $l - 1$ 'inci ağacın ( $l = 1, \dots, L$ ) tüm  $|r|$  nihai duraklarında ( $M_{\bar{d},l-1}$ ) bulunan kişilere denk gelen  $\varepsilon_{\bar{d},l-1}$  kalıntı terimleri elde edilir. Ancak, bu kalıntı terimleri  $i$  kişisi için yapılan tahminini ( $\hat{y}_{i,l-1}$ ) revize etmek için tümünden kullanılmaz. Neticede  $l - 1$  ağacının yaptığı “düzeltme” teşebbüsünün kendisi hatalı olabilir. Dolayısıyla  $\varepsilon_{\bar{d},l-1}$  terimleri sadece kısmen kullanılır. Bir başka deyişle, ağaçlar kendilerinden önce gelen ağaçtan “öğrenirken” bu öğrenme hızı yavaşlatılır. Bu yavaşlatma işlemi “öğrenme hızı” (learning rate) parametresi  $\gamma$  tarafından belirlenir ( $0 < \gamma < 1$ ). Ardışık ağaçlar yoluyla seri halinde yapılan bu özçağrılı (rekürsif) düzeltme adımları ancak seriye yeni eklenen sınıflandırma ağaçları kestirimleri artık daha doğru yapmadığı zaman durdurulur (Friedman, 2001; Friedman vd., 2001; Friedman, 2002). Öğrenme hızı yavaşlatıldıkça, Gradyan Artırma algoritması önceki iterasyonda yapılan hataların kısmen düzeltilmesi için çok sayıda ve farklı biçimlerde oluşturulmuş sınıflandırma ağacının seriye eklenmesini sağlar (James vd., 2013). Bu rekürsif yapı şu şekilde özetlenebilir:

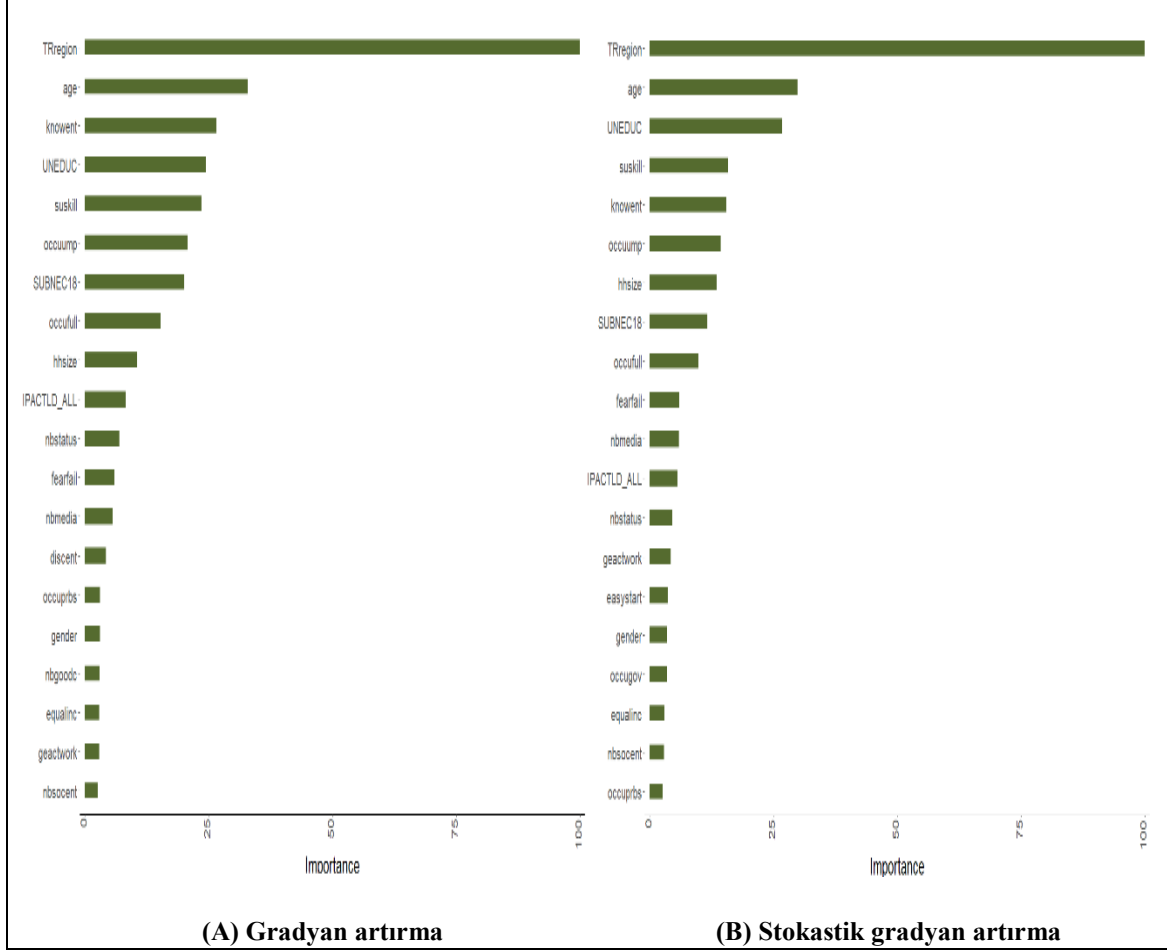
$$\hat{y}_{il} = \hat{y}_{i,l-1} + \gamma \varepsilon_{\bar{d},l-1} \mathbf{1}(i \in M_{\bar{d},l-1}) \quad (7)$$

Yukarıdaki denklemde  $\hat{y}_{il}$  terimi  $i$  kişisi için tahmin edilen girişimci olma veya olmama durumunu temsil eden değerdir.  $\mathbf{1}$  operatörü ise denklemdeki son terimin ancak  $i$  gözleminin  $M_{\bar{d},l-1}$  içerisinde yer alması durumunda sifıra eşit olmayan bir değer aldığı, aksi takdirde 0 olduğuna işaret eden karakteristik fonksiyonu (indicator function) belirtir. Öğrenme hızı parametresine ek olarak stokastik ve stokastik olmayan Gradyan Artırma algoritmalarını çalıştırmadan önce iterasyon sayısı (sınıflandırma ağacı miktarı) ve maksimum ikili ayrıştırma sayısı belirlenmelidir. Bu sayılar sırasıyla 4 ve 8 olarak belirlendiğinde kestirim performansının yüksek olduğu gözlemlenmiştir (Friedman, 2001; Friedman vd., 2001).

Rassal Orman modeline benzer biçimde Gradyan Artırma modellerine de rassallık katılabilir. Rassallık sayesinde hem Rassal Orman kısmında belirtilen avantajlar sağlanır hem de modellerin standart bir bilgisayarda işleme hızı arttırılabilir. Rassallık, Gradyan Artırma modellerine öncelikle her bir iterasyonda kalıntı terimlerini tahmin ederken, yerine koymadan örnekleme yöntemiyle eğitim verisinden rasgele örneklem alarak, ikinci olarak da seri içerisindeki bir  $l$  ağacına ait her bir  $d$  durağında ikili ayrıştırmayı yapacak değişkenin seçimini öznitelik uzayındaki  $K$  sayıda değişken arasından rasgele seçilmiş bir alt kümeyle sınırlamak suretiyle katılır (Friedman, 2002). Gradyan Artırma modellerinin parametreleri (Friedman, 2002) tarafından gösterildiği üzere aşırı uyumluluk ihtimalini en aza indirgeyebilmek amacıyla şu şekilde belirlenmiştir: terminal durak derinliği 6, iterasyon rassal alt küme oranı 0.5 (sadece Stokastik Gradyan Artırma için). Buna ek olarak, tarama yaklaşımı ile (grid search) öğrenme hızı parametresi ( $\gamma$ ) 0.1, ikili ayrıştırma adımlarında öznitelik altkümesi oranı 0.9 ve bir  $M_d$ 'de bulunabilecek minimum gözlem sayısı 5 olarak belirlenmiştir. Bu parametreler ile işleme konan Gradyan Artırma algoritması test verisi göz önüne alındığında %88, Stokastik Gradyan Artırma algoritması ise %87 oranında doğruluk oranına sahiptir.



Gradyan Artırma ve Stokastik Gradyan Artırma öznitelik önemleri Şekil 4'ün sırasıyla A ve B panellerinde sunulmuştur. Her iki algoritmanın deęişken havuzundan seçtięi ve önceki tablolarda hali hazırda tanımlanmamış deęişkenlerin tanımları Tablo 4'te yapılmıştır.



Şekil 4. Gradyan Artırma ve Stokastik Gradyan Artırma Öznitelik Önemleri

Kaynak: Yazar tarafından hazırlanmıştır.

Her iki Gradyan Artırma modeli de sınıflandırma ağacına tutarlı olarak; ancak, Rassal Orman bulgularının aksine bölge deęişkenine yüksek önem atfetmiştir. Stokastisiteden bağımsız olarak çeşitlilik deęerini azaltmada açık bir ara ile en başarılı olan deęişken TRregion olmuştur. Sonrasında ise hem Rassal Orman hem de sınıflandırma ağacı bulguları ile paralel olarak kişinin yaşı, girişimci tanıdığı olması, eğitimi, girişimcilik bilgi ve becerileri, istihdam durumu ve ihtiyaç/mecburiyet faktörleri göreceli olarak dięer deęişkenlerden daha önemli olarak puanlanmıştır.

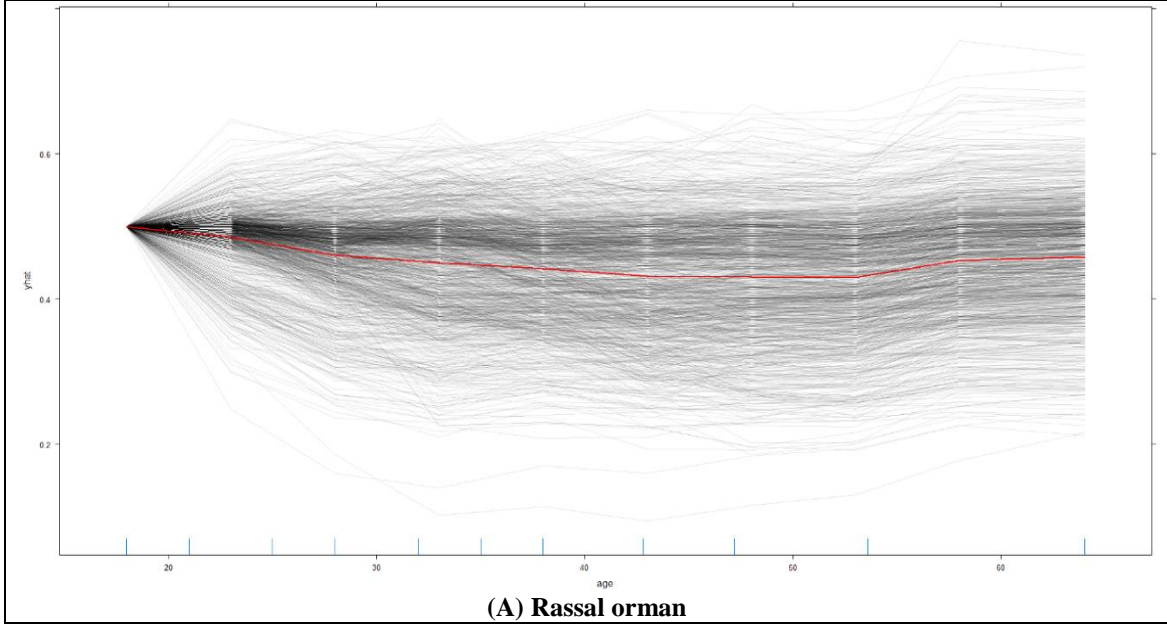
**Tablo 4. Gradyan Artırma ve Stokastik Gradyan Artırma Değişken Tanımları (Tekil ağaç ve rassal orman modellerinin seçtiği değişkenler hariç)**

Değişken	Tanım	Değerler
<i>discent</i>	Kişinin son 12 ay içerisinde sahip olduğu bir işletmeyi kapatmış olma durumu.	Kategorik değişken (Evet-Hayır)
<i>easystart</i>	Kişinin ülkede iş kurmanın kolay olduğu yönündeki görüşü.	Kategorik değişken (Evet-Hayır)
<i>equalinc</i>	Kişinin insanların yaşam standartlarında eşitlik olması gerektiğine dair düşüncenin ülkede hakim fikir olduğu veya olmadığı yönündeki görüşü.	Kategorik değişken (Evet-Hayır)
<i>nbgoodc</i>	Kişinin ülkede iş kurmanın istenen bir kariyer seçimi olduğu yönündeki görüşü.	Kategorik değişken (Evet-Hayır)
<i>nbsocent</i>	Kişinin ülkedeki sosyal sorunları çözmeye odaklı işletmelerin bulunduğuna yönelik görüşü.	Kategorik değişken (Evet-Hayır)

**Kaynak:** GEM, 2018.

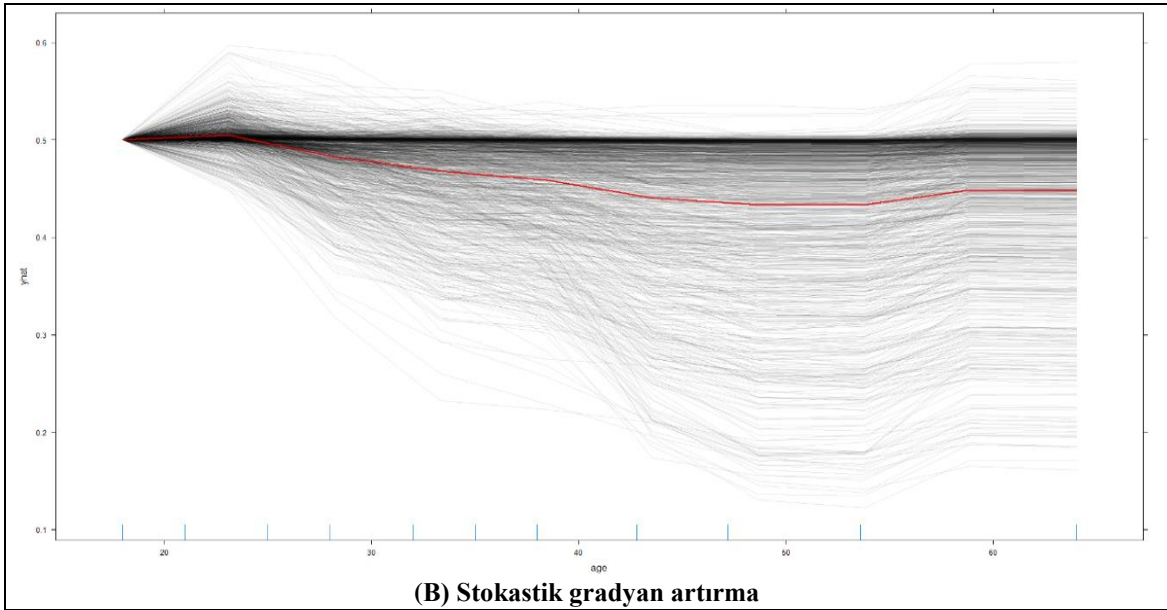
Önceki algoritmalara ek olarak Gradyan Artırma ve Stokastik Gradyan Artırma modelleri düşük önem seviyesinde olsalar dahi bazı yeni değişken seçimlerinde bulunmuşlardır. Bu seçimler kişinin son 12 ay içerisinde işyerini kapatmış olması (*discent*), Türkiye’de iş kurmanın kolaylığı, istenen bir kariyer seçimi olduğuna yönelik görüşü, sosyal sorunları çözmeye yönelik işletmelerin var olduğuna yönelik inancı, ve ülkede eşitliğe önem verilip verilmediği konusundaki pozisyonunu (sırasıyla *easystart*, *nbgoodc*, *nbsocent*, *equanlinc*) temsil eden değişkenlerden oluşmaktadır.

Gradyan Artırma ve Rassal Orman sonuçlarının belli değişkenler özelinde koşullu bireysel beklenti (“individual conditional expectation,” (Goldstein, Kapelner, Bleich ve Pitkin, 2015) ve kısmi bağımlılık (“partial dependence,” (Friedman, 2001)) grafikleri ile daha yakından irdelenmesi mümkündür. Kısmi bağımlılığı hesaplanan bir değişkenin kendisinin dışında kalan değişkenlerin sabit olması koşuluyla girişimci olup olmama olasılığına olan marjinal etkisi o değişkenin aldığı farklı değerler için hesaplanır. Sayısal ve kategorik değişkenler için bu etkilerin görselleştirilmesi farklıdır. Sayısal bir değişken olan yaş (*age*) değişkenini ele aldığımızda Şekil 5’te sunulan Rassal Orman ve Stokastik Gradyan Artırma koşullu bireysel beklenti grafiklerindeki kırmızı çizgi tüm *i* gözlemleri için ortalama bağımlılığı temsil etmektedir. Siyah çizgilerden her biri ise tek bir kişinin (gözlemin) diğer değişkenler sabit tutulurken, öncesinde oluşturulan topluluk modelini kullanarak yaş (*age*) değişkenine farklı değerler verip o kişi için yeniden tahmin yapılması sonucunda bulunan girişimci olma olasılıklarını göstermektedir.

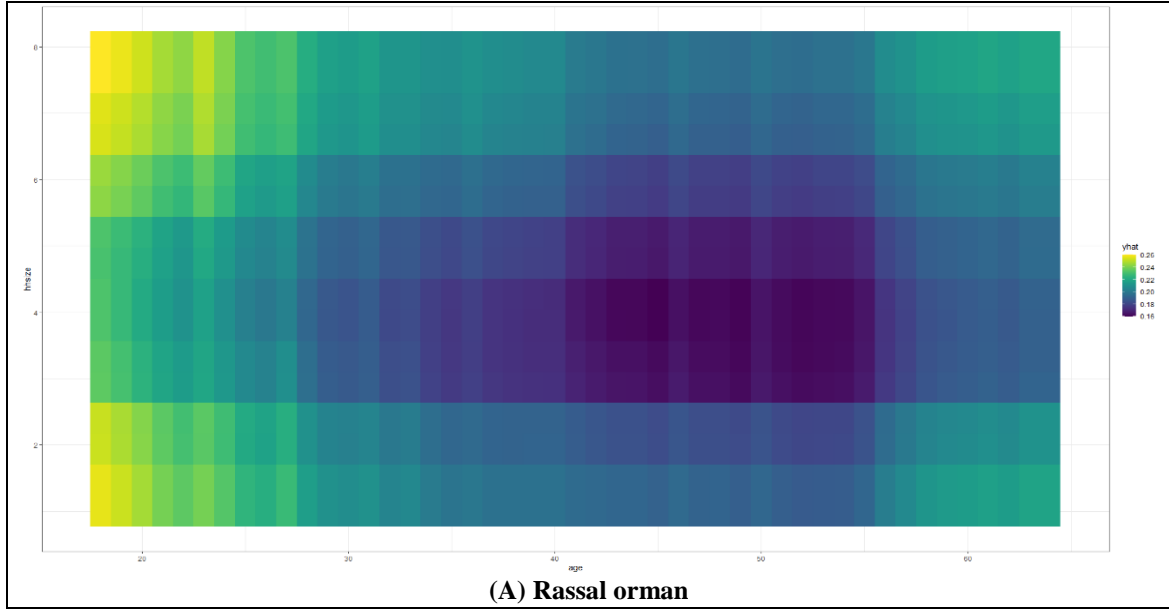


**Şekil 5. Bireysel Koşullu Beklenti Grafiđi (Ortalanmıř): Yař**  
**Kaynak:** Yazar tarafından hazırlanmıřtır.

Her iki model de yař arttıka, özellikle 20’li yařların ortalarından itibaren giriřimci olma olasılıđında bir dūřuř bařladıđını gōstermekte. Ortalama çizgisi olasılıkta kōçük bir ortalama deđiřimi gōsterse de özellikle Stokastik Gradyan Artırma modeline gōre birgok kiři için, siyah kořullu bireysel beklenti çizgilerine bakarak, giriřimci olma ihtimalinin %50’lerden %20’ler seviyesine kadar dūřtūđünü gōruyoruz. 50 yař ortalarından sonra hafif bir yūkselme gōzükse de yatay eksen de dikey çizgilerle temsil edilen onda birlik dilimlere bakılınca bu yař grubu için verinin gōreceli olarak seyrek olduđunu da gōrmekteyiz.



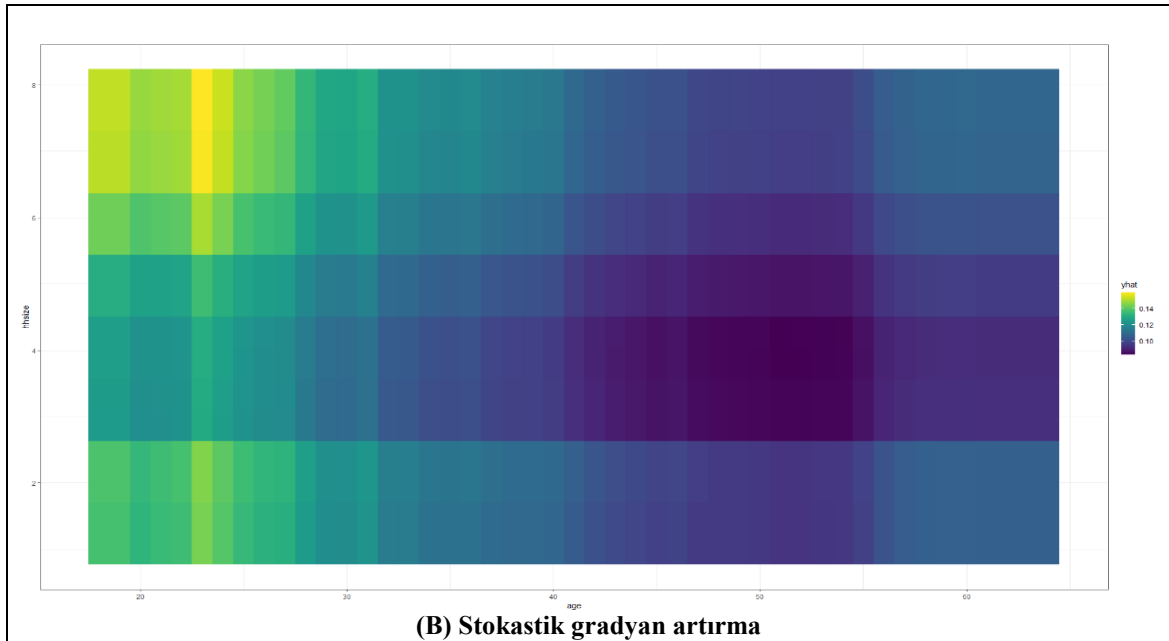
**Şekil 6. Bireysel Koşullu Beklenti Grafiđi (Ortalanmıř): Yař**  
**Kaynak:** Yazar tarafından hazırlanmıřtır.



Şekil 7. Kısmi Bağımlılık Grafiği: Yaş ve Hanehalkı Nüfusu

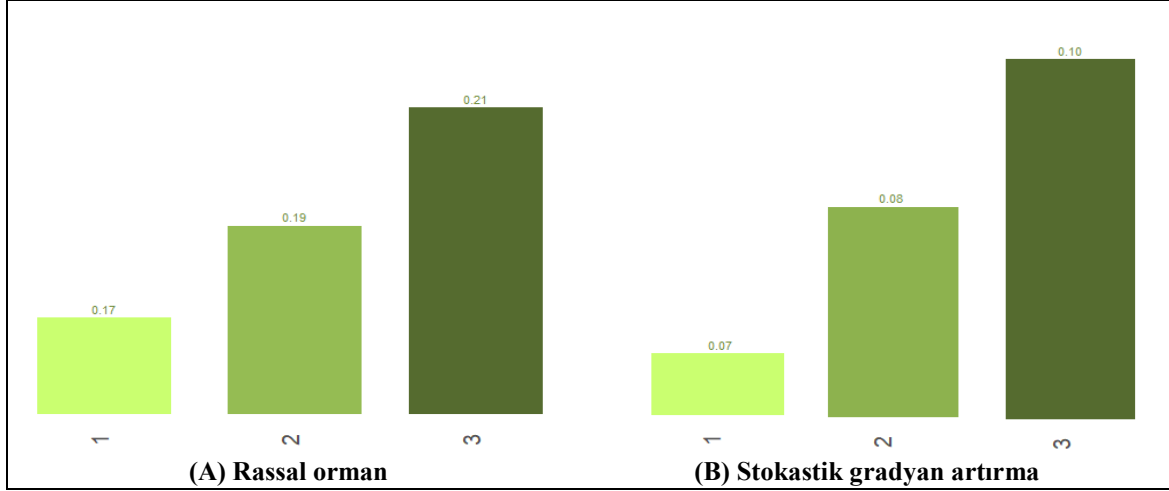
Kaynak: Yazar tarafından hazırlanmıştır.

Yaş değişkeni ile bulgularımızda nispeten yüksek önem arz ettiği görülen bir diğer sayısal değişken olan hanehalkı nüfusunun kısmi bağımlılıklarının birlikte görselleştirildiği Şekil 6’ya bakıldığında her iki modele göre de genç yaşlarda erken evre girişimci olma ihtimalinin daha yüksek olduğu; ancak, küçük nüfuslu ve yüksek nüfuslu hanelerde bu gözlemin daha öne çıktığı gözlemlenmektedir. 4-5 kişilik hanelerde yaşayan gençler ise aynı kategoride yer alan 40 yaş üstü kişilere göre daha yüksek bir girişimci olma olasılığına sahipken, belirtildiği üzere daha kalabalık veya daha küçük hanelerde yaşayan gençlere göre düşük bir olasılığa sahiplerdir.



Şekil 8. Kısmi Bağımlılık Grafiği: Yaş ve Hanehalkı Nüfusu

Kaynak: Yazar tarafından hazırlanmıştır.

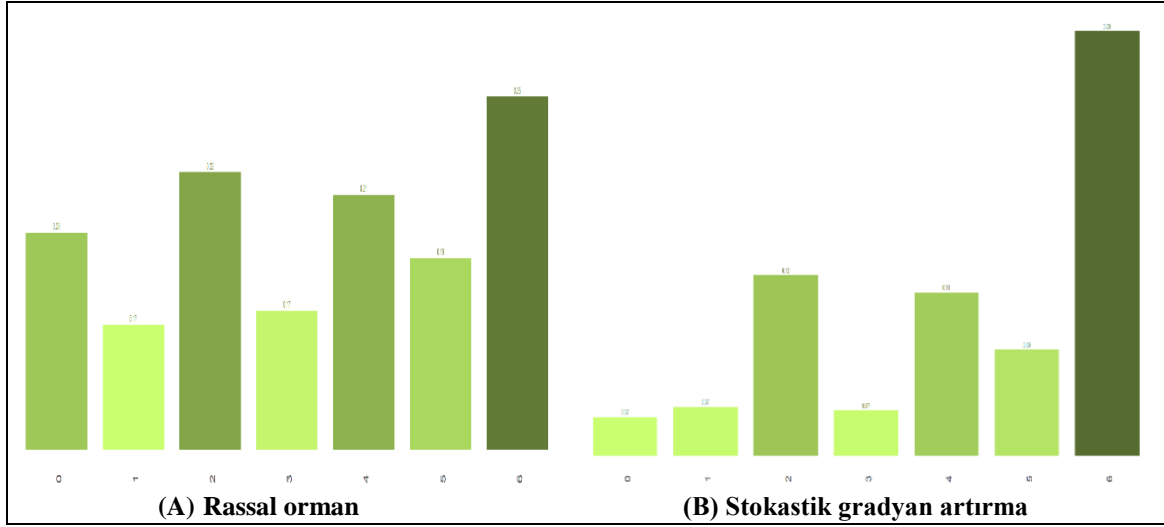


**Şekil 9. Kısmi Bağımlılık Grafiği: Gelir Grubu**  
**Kaynak:** Yazar tarafından hazırlanmıştır.

Kategorik olan deęişkenler söz konusu olduğunda girişimciliğin kısmi bağımlılığının hesaplandığı kategori için geri kalan tüm gözlemleri de sanki o kategorideymiş gibi ele alarak (örneğin işsizlik için tüm *i* gözlemlerini işsiz olarak kabul ederek veya Ankara bölgesine göre kısmi bağımlılığı hesaplarken tüm gözlemlerin Ankara'da bulunduğunu varsayarak) tahminler yeniden yapılır. Kategorilerin hepsi için bu işlem uygulandıktan sonra bağımlı deęişkenin dięer deęişkenler sabitken belirli bir kategoriye olan kısmi bağımlılığı hesaplanır. Bulgularımızda gelir grubu önemli bir faktör olarak öne çıkmıştır. Bu deęişkenin kategorik kısmi bağımlılık grafiği baktığımızda (Şekil 7) her iki modele göre ilk yüzde 33'lük dilimdeki hanehalklarında bulunan kişilerin en düşük, sırasıyla ikinci ve üçüncü dilimlerde bulunanların ise daha yüksek olasılıklara sahip olduğu görülmektedir. Olasılıklar arasındaki farklar küçük gözükse de dahi bu grafiğin sadece kısmi bağımlılıkları yani ortalama olasılıkları gösterdiğini hatırlamakta fayda vardır. Koşullu bireysel beklenti grafiklerinde gördüğümüz gibi ortalamanın çok altında iken üst gelir gruplarında olsalar ortalamanın çok daha üstünde bir olasılığa erişecek kişiler olduğuna yönelik bir varsayımı rahatlıkla yapabiliriz.



**Şekil 10. Kısmi Bağımlılık Grafiği: Çalışma Durumu**  
**Kaynak:** Yazar tarafından hazırlanmıştır.



Şekil 11. Kısmi Bağımlılık Grafiği: Eğitim Durumu

Kaynak: Yazar tarafından hazırlanmıştır.

Çalışma durumu ile ilgili değişkenler tutarlı biçimde bulgularımızda öne çıkmışlardır. Şekil 8’de başlıca kategorilere denk gelen ortalama olasılıklar karşılaştırılmaktadır. Girişimcilik yolunu seçmesi en olası olan grubun işsiz bireylerden oluştuğunu görmekteyiz. Bu grubu sırasıyla öğrenciler, özel sektör çalışanları ve kamu sektörü çalışanları takip etmektedir. Eğitim durumuna baktığımızda ise Tablo 3’te tanımlanan kategorilere denk gelen ortalama olasılıkların her iki model tarafından da yüksek eğitim düzeyine sahip kişiler için en yüksek olarak hesaplandığını görmekteyiz (Şekil 9). Bununla birlikte hem Rassal Orman hem de Stokastik Gradyan Artırma sonuçlarına göre geri kalan gruplar arasında ortaokul mezunu ve lise sonrası üniversite öncesi eğitim almış kişiler diğer gruplardan daha yüksek olasılığa ulaşmaktadırlar.

Sınıflandırma ağacı sonuçlarında gördüğümüz üzere bölgesel koşulların kişilerin erken evre girişimci olmalarındaki rolleri doğrusal olmayan ve diğer değişkenlerle etkileşimli bir yapıdadır. Topluluk modellerinden edinilen sonuçlara göre ise, bölgelerin kısmi bağımlılık dereceleri Şekil 10’da harita biçiminde sunulmuştur. Açık renkler düşük, koyu renkler ise yüksek olasılıkları temsil etmektedir. Ortalama olasılıklar arasında en yüksek ve en düşük bölge arasında %20 dolaylarında bir fark mevcuttur. Bir kişinin girişimci olma ihtimali eğer o kişinin Kastamonu İBBS-2 bölgesinde bulunduğu varsayıldığında göreceli olarak en düşük seviyededir. Kastamonu kadar düşük olmasa da Tekirdağ, Balıkesir, Bursa ve Erzurum İBBS-2 bölgelerinde de girişimci olma ihtimalinin herhangi bir kişi için ortalama daha az olacağı görülmektedir. Sınır bölgeleri arasında ise Van, Gaziantep ve Trabzon İBBS-2 bölgeleri nispeten yüksek değerlere sahiptir. Zonguldak bölgesi ise yüksek bir olasılıkla öne çıkmaktadır. Unutulmamalıdır ki, tekil sınıflandırma ağacı örneğinde görüldüğü gibi bu bölgesel etkiler başka birçok faktöre tabidir. Örneğin bir kişi Zonguldak veya Van bölgesinde bulunuyor ise sınıflandırma ağacına göre 43 yaşından genç, tam zamanlı olarak istihdam edilmiş; ancak, mecburiyet ve ihtiyaç dolayısıyla erken evre girişimciliği seçmiş bir kişi olması beklenmektedir. Kısmi bağımlılık grafiğinde en düşük çıkan Kastamonu bölgesinde bulunan bir kişi sınıflandırma ağacına göre eğer 33 yaşından küçük ve girişimciliğe toplum ve medya tarafından yeterli önem verildiğini düşünmeyen birisi ise bu kişi erken evre girişimci olarak tahmin edilmiştir (Şekil 10).



**Şekil 12. Bölgesel Kısmi Bağımlılık: Stokastik Gradyan Artırma**  
**Kaynak:** Yazar tarafından hazırlanmıştır.

## 5. Sonuçlar ve Politika Önerileri

Ülkelerin Sürdürülebilir Kalkınma (SK) ve ekonomik büyüme seviyeleri girişimcilik ve inovasyon ile artmaktadır. SKA4 (Nitelikli Eğitim)'in ve SKA8 (İnsana Yakışır İş ve Ekonomik Büyüme)'nin hedefleri arasında yer alan girişimcilik ile eğitim, iş yaratma, yaratıcılık ve inovasyon arasında doğrudan bir ilişki vardır (Filser vd., 2019). Yapılan arařtırmalara göre (Karadeniz ve Özçam, 2018; Özdemir ve Karadeniz, 2011; Vodã vd., 2020) fırsatları değerlendirme motivasyonu ile girişimcilik faaliyetlerinde bulunan girişimcilerin daha eğitilmiş ve yaratıcı oldukları ve ekonomik büyümeye katkıda buldukları gözlemlenmiştir. Bu arařtırmadaki sonuçlarımız arasında, Türkiye'de girişimcilerin fırsatları değerlendirme motivasyonundan çok; şirket kurmuş birisini tanımamalarına rağmen ihtiyaç ve mecburiyetten dolayı girişimci oldukları bulgusu da bulunmaktadır.

Bu çalışmada kullanılan Rastal Orman ve Gradyan Artırma modellerine göre; Türkiye'de genç yaşta bireylerin girişimci olma ihtimallerinin yüksek olduğu ve bu sonucun literatür tarafından desteklendiği görülmüştür (Vodã vd., 2020). Ayrıca, analiz sonuçları Türkiye'de yüksek gelir grubunda olanların girişimci olma olasılıklarının daha fazla olduğunu göstermiştir. Çalışma durumu ile ilgili değişkenlerin girişimciliği belirlemede önemli bir faktör olduğu tespit edilmiştir. Özellikle işsiz kişilerin mecburiyetten dolayı girişimci oldukları gözlemlenmiştir. Hem Rastal Orman hem de Gradyan Artırma sonuçlarına göre yüksek eğitimli bireylerin girişimci olma olasılıklarının daha yüksek olduğu görülmüştür. Stokastik Gradyan Artırma sonuçlarına göre Türkiye'de bölgesel kalkınmanın ve koşulların bireylerin girişimci olma olasılıklarını etkiledikleri ve bölgelerarasındaki en yüksek ve en düşük erken evre girişimcilik olasılıkları farkının %20 civarında olduğu tespit edilmiştir. Rastal Orman öznelik önem bulgularına göre çalışma durumu, başka bir girişimci tanımak, kişinin kendi bilgisi ve becerisi hakkındaki görüşü, ihtiyaç/mecburiyet ve yaş, girişimci olmak için önemli faktörlerdir. Rastal Orman sınıflandırma ağacının kullandığı değişkenlere ek olarak ve daha düşük önem seviyeleri ile hanehalkı geliri ve nüfusu, kişinin çalışma durumu, cinsiyeti, eğitim düzeyi, anket zamanı itibarıyla şirket içi girişimcilik aktivitesinde bulunuyor olması veya son üç yılda liderlik seviyesinde olmasa bile bu

konuda şirket içerisinde aktif olması girişimciliği etkileyen başlıca faktörler olarak tespit edilmiştir. Türkiye’de makine öğrenmesi ile GEM (2018) verileri kullanılarak yapılan bu araştırma sonuçları Özdemir ve Karadeniz (2011) ve Karadeniz ve Özçam (2018) tarafından yapılan araştırma sonuçları ile karşılaştırıldığında birbirlerini desteklemektedirler. Bu araştırmanın literatüre katkısı makine öğrenmesi kullanarak aynı sonuçlara ulaşılmış olmasıdır. Ayrıca, makine öğrenmesi metodu kullanılarak bu iki çalışmadan farklı olarak bu çalışma ile Türkiye’de, şirketlerde tam zamanlı çalışan ve şirket içi girişimcilik faaliyetlerinde aktif rol alan kişilerin daha sonra bağımsız girişimci oldukları tespit edilmiştir. Literatürdeki bulgulara katkı açısından sonuçlarımız yüksek derecede doğrusal olmayan ve etkileşimli ilişkilere ve belirli koşullara tabi olduğu için çalışmanın bulguları bireysel koşullu beklenti ve kısmi bağımlılık çerçevesi içerisinde detaylı olarak incelenmiştir.

SK ve bölgesel kalkınma girişimcilik faaliyetlerinden olumlu etkilenmektedir. Sürdürülebilir kalkınma ve bölgesel kalkınmanın sağlanabilmesi için; girişimcilik faaliyetlerinin SK’nın ekonomik, sosyal ve çevresel boyutlarını da kapsayan bir kalkınma modeli (Siffert ve Guimarães, 2020) çerçevesinde desteklenmesi ve daha iyi yaşam koşulları sağlayacak olan politikalar uygulanması önemlidir.

#### **Araştırmacıların Katkı Oranı Beyanı**

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

#### **Çıkar Çatışması Beyanı**

Bu çalışmada herhangi bir potansiyel çıkar çatışması bulunmamaktadır.



## Kaynakça

- Aldrich, C. and Auret, L. (2013). *Unsupervised process monitoring and fault diagnosis with machine learning methods*. London: Springer.
- Alpaydın, E. (2016). *Machine learning: The new AI*. Cambridge: MIT press.
- Arenius, P. and Minniti, M. (2005). Perceptual variables and nascent entrepreneurship. *Small Business Economics*, 24(3), 233-247. doi:10.1007/s11187-005-1984-x
- Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685-725. <https://doi.org/10.1146/annurev-economics-080217-053433>
- Avrupa Komisyonu. (2019). *Female entrepreneurs*. Retrieved from [https://ec.europa.eu/growth/smes/promoting-entrepreneurship/we-work-for/women\\_en](https://ec.europa.eu/growth/smes/promoting-entrepreneurship/we-work-for/women_en)
- Birleşmiş Milletler Genel Kurulu. (2016). Sürdürülebilir kalkınma için girişimcilik: Genel kurul tarafından kabul edilen karar. Erişim adresi: <https://digitallibrary.un.org/record/855166>
- Bosma N., Hill, S., Ionescu-Somers, A., Kelley D., Levie, J. and Tarnawa, A. (2020). *Entrepreneurship monitor 2019/2020 global report*. Retrieved from <https://www.babson.edu/media/babson/assets/global-entrepreneurship-monitor/2019-2020-GEM-Global-Report.pdf>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140. <https://doi.org/10.1007/BF00058655>
- Breiman, L. (2001a). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199-231. doi:10.1214/ss/1009213726
- Breiman, L. and Cutler, A. (2021). *Random forests*. Retrieved from <https://www.stat.berkeley.edu/~breiman/RandomForests/>
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks.
- Carr, L. T. (1994). The strenghts and weaknesses of quantitative and qualitative research: What method for nursing? *Journal of Advanced Nursing*, 20(4), 716-721. <https://doi.org/10.1046/j.1365-2648.1994.20040716.x>
- Celbiş, M. G. (2021). A machine learning approach to rural entrepreneurship. *Papers in Regional Science*, 100(4), 1079-1104. <https://doi.org/10.1111/pirs.12595>
- Cetindamar, D., Gupta, V. K., Karadeniz, E. E. and Egrican, N. (2012). What the numbers tell: The impact of human, family and financial capital on women and men's entry into entrepreneurship in Turkey. *Entrepreneurship & Regional Development*, 24(1-2), 29-51. <https://doi.org/10.1080/08985626.2012.637348>
- Department of Economic and Social Affairs. (2015). Sustainable development. Retrieved from <https://sdgs.un.org/>
- Filser, M., Kraus, S., Roig-Tierno, N., Kailer, N. and Fischer, U. (2019). Entrepreneurship as catalyst for sustainable development: Opening the black box. *Sustainability*, 11(16), 4503. <https://doi.org/10.3390/su11164503>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 5, 1189-1232. doi:10.1214/aos/1013203451
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38(4), 367-378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Friedman, J., Hastie, T. and Tibshirani, R. (2001). *The elements of statistical learning*. Berlin: Springer.
- Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems*. USA: O'Reilly Media, Inc.

- Global Entrepreneurship Monitor. (2018). *The World's foremost study of entrepreneurship*. Retrieved from [www.gemconsortium.org](http://www.gemconsortium.org)
- Goldstein, A., Kapelner, A., Bleich, J. and 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. Retrieved from <http://www.jstor.org/>
- Hersh, J. and Harding, M. (2018). Big data in economics. *IZA World of Labor*, (451). doi:10.15185/izawol.451
- Imbens, G. and Athey, S. (2021). Breiman’s two cultures: A perspective from econometrics. *Observational Studies*, 7(1), 127-133. doi:10.1353/obs.2021.0028
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.
- Karadeniz, E. E. and Özçam, A. (2018). Regional disparities in entrepreneurship in Turkey with respect to gender using a regression of pooling cross sections: 2006-2015. In N. Faghieh and M. R. Zali (Eds.), *Entrepreneurship in the Middle East and North Africa (MENA)*. New York: Springer. <https://doi.org/10.1007/978-3-319-75913-5>
- Kim, P. H., Aldrich, H. E. and Keister, L. A. (2006). The impact of financial, human, and cultural capital on entrepreneurial entry in the United States. *Small Business Economics*, 27, 5-22. <https://doi.org/10.1007/s11187-006-0007-x>
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1-26. Doi: 10.18637/jss.v028.i05
- Liaw, A. ve Wiener, M. (2002). Classification and regression by randomforest. *R News*. 2(3), 18-22. Retrieved from <https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf>
- Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87-106. doi:10.1257/jep.31.2.87
- Organization of Economic Cooperation and Development. (2020). *OECD regions and cities at a glance 2020*. Retrieved from <https://www.oecd.org/regional/oecd-regions-and-cities-at-a-glance-26173212.htm>
- Özçam, A. and Karadeniz, E. (2018). How did women entrepreneurship change over time in Turkey (2006-2015): A regression of pooling cross sections across time. *Contemporary Research in Economics and Social Sciences*, 2(1), 43-69. Retrieved from <https://dergipark.org.tr/en/pub/conress/>
- Özdemir, Ö. and Karadeniz, E. E. (2009). *Differences between being an opportunity-driven and necessity-driven entrepreneur: Evidence from Turkey’s GEM data*. Paper presented at the 4th European Conference on Entrepreneurship and Innovation. University of Antwerp, Belgium. Retrieved from <https://avesis.metu.edu.tr/yayin/e7b50ba9-31c0-4d2d-bbfb-b6b4c6857442/differences-between-being-opportunity-driven-and-necessity-driven-entrepreneur-evidence-from-turkeys-gem-data>
- Özdemir, Ö. and Karadeniz, E. E. (2011). Investigating the factors affecting total entrepreneurial activities in Turkey. *METU Studies in Development*, 38, 275-290. Retrieved from <http://www2.feas.metu.edu.tr/metusd/ojs/>
- Özgüzel, C. (2020). Agglomeration effects in a developing economy: Evidence from Turkey (PSE Working Papers No. 1341). Retrieved from <https://erf.org.eg/app/uploads/2019/09/1341.pdf>
- Ployhart, R. E. and Moliterno, T. P. (2011). Emergence of the human capital resource: A multilevel model. *Academy of Management Review*, 36(1), 127-150. <https://doi.org/10.5465/AMR.2011.55662569>
- Reynolds, P. D., Hay, M., Bygrave, W. D., Camp, S. M. and Autio, E. (2000). *Global entrepreneurship monitor: 2000 executive report*. MO: Kauffman Foundation Kansas.
- Santos, S. C., Caetano, A., Spagnoli, P., Fernandes Costa, S. and Neumeyer, X. (2017). Predictors of entrepreneurial activity before and during the European economic crisis. *The International Entrepreneurship and Management Journal*, 13(4), 1263-1288. doi:10.1007/s11365-017-0453-8

- Siffert, P. V. and Guimarães, L. D. O. (2020). Entrepreneurial ecosystem and sustainability as catalysts for regional development: Proposition of a theoretical framework. *Interações (Campo Grande)*, 21, 739-752. <http://dx.doi.org/10.20435/inter.v21i4.2647>
- Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of Statistics*, 24, 303-329. [https://doi.org/10.1016/S0169-7161\(04\)24011-1](https://doi.org/10.1016/S0169-7161(04)24011-1)
- Türk, U. (2020). Gelir dağılımında fırsat eşitsizliği ve alt kırılımları: Türkiye üzerine bir araştırma. *Alternatif Politika*, 12(2), 311-335. Eriřim adresi: <https://alternatifpolitika.com>
- United Nations Development Program. (2021). *Sürdürülebilir kalkınma amaçları*. Eriřim adresi: <https://www.undp.org/sustainable-development-goals>
- Verheul, I., Wennekers, S., Audretsch, D. and Thurik, R. (2001). *An eclectic theory of entrepreneurship: Policies, institutions, and culture* (Tinbergen Institute No. 030/3). Retrieved from <https://papers.tinbergen.nl/01030.pdf>
- Vodã, A. I., Butnaru, G. I. and Butnaru, R. C. (2020). Enablers of entrepreneurial activity across the European Union: An analysis using GEM individual data. *Sustainability*, 12(3), 1022. <https://doi.org/10.3390/su12031022>
- Wennekers, S., van Stel, A., Thurik, R. and Reynolds, P. (2005). Nascent entrepreneurship and the level of economic development. *Small Business Economics*, 24, 293-309. <https://doi.org/10.1007/s11187-005-1994-8>

## EKLER

### Ek 1 – Notasyon Tablosu

İşaret	Tanım
$c$	Gözlem sınıfı.
$d$	İkili bölünme durağı (node).
$\Gamma$	Gini endeksi.
$s$	Ağaç durağına düşen gözlem oranı.
$N$	Gözlem sayısı.
$M$	Veri alt kümesi.
$i$	Gözlem birimi.
$y$	Bağımlı değişken
$x$	Öznitelik (feature)
$k$	Öznitelik endeksi (Toplam sayı: $K$ ).
$v$	Ayrıştırma değeri (Split value)
$\tau$	Kompleksite parametresi (Complexity parameter)
$r$	$\tau$ değerlerine karşılık gelen ağaç (Toplam ağaç sayısı: $R$ ).
$\bar{d}$	Nihai durak (Terminal node).
$ r $	$r$ ağacındaki nihai durak sayısı.
$c^*$	Bir nihai duraktaki çoğunlukta olan sınıf.
$\varepsilon$	Belirli bir alt küme için tahmin hatası.
$l$	Rassal orman içerisindeki bir iterasyon.
$l$	Bir başka deyişle $l$ iterasyonunda üretilen sınıflandırma ağacı (toplam sayı: $L$ ).
$\Delta I$	İkili bölünle sonrasında Gini değerinde oluşan fark.
$\hat{y}$	Öngörülen değer.

**Kaynak:** Yazar tarafından hazırlanmıştır.

## **ENTREPRENEURSHIP AND SUSTAINABLE REGIONAL DEVELOPMENT IN TURKEY: FINDINGS OBTAINED FROM MACHINE LEARNING APPROACHES**

### **EXTENDED SUMMARY**

#### **Aim of the Study**

The aim of this study is to analyze the factors affecting entrepreneurship activities in Turkey in terms of SD by applying machine learning, which is a relatively new method in social sciences, and using Global Entrepreneurship Monitor (GEM) data. This research analyzes and extends the previous results of Özdemir and Karadeniz (2011) and Karadeniz and Özçam (2018) from individual entrepreneur gains in the Turkish context for the first time through the application of machine learning algorithms.

#### **Literature**

Entrepreneurship is one of the most important factors that foster innovative activity and increase competition by creating change, and therefore, contribute to economic development. Entrepreneurship activity is crucial to achieving United Nations Development Programme (UNDP, 2021) Sustainable Development Goals (SDGs) 1, 4, 8 and 10. These goals are; “SKA1-No Poverty: Ending all forms of poverty everywhere”, “SKA4-Quality Education: Providing inclusive and equitable quality education for all and promoting lifetime educational opportunities”, “SKA8-Decent Work and Economic Growth: Supporting stable, inclusive and sustainable economic growth, full and productive employment, and decent jobs for all” and “SKH10-Reduced Inequalities: Reducing inequalities within and between countries”.

The United Nations General Assembly (2016) has defined entrepreneurship as the driving force of Sustainable Development (SD) along with innovation. There is a direct relationship between entrepreneurship and education, job creation, creativity and innovation which are among the objectives of SKA4 (Quality Education) and SKA8 (Decent Work and Economic Growth) (Filsler vd., 2019).

In this research, the factors that determine and affect entrepreneurship are analyzed according to the GEM (2018) model which defines entrepreneurship as “individuals or individuals or businesses creating new business or new ventures, establishing their own business or expanding existing business”. According to the GEM (2018) model, working status, knowing another entrepreneur, one’s opinion about one’s own knowledge and skills, need/obligation and age are important factors for being an entrepreneur. Also, with lower levels of importance household income and population, the employment status, gender, education level of the person and engagement in intra-company entrepreneurial activity as of the time of the survey affect entrepreneurship.

### **Methodology**

Based on the Random Forest and Gradient Boosting models used in this study; it is observed that young individuals in Turkey are more likely to become entrepreneurs, and this result is supported by literature (Vodã vd., 2020). In addition, the results of the analysis showed that those in the high-income group in Turkey were more likely to become entrepreneurs. Variables related to working status are found to be an important factor in determining entrepreneurship. It is also observed that unemployed people in particular are entrepreneurs due to necessity. According to both the Random Forest and Gradient Boosting results highly educated individuals are more likely to become entrepreneurs.

### **Findings**

The results of the Stochastic Gradient Boosting suggest that regional development and conditions in Turkey affect the likelihood of individuals becoming entrepreneurs and that the difference in the highest and lowest early-stage entrepreneurship probabilities among regions is around 20%. Based on the random forest variable importance scores, working status, knowing another entrepreneur, one's opinion on one's own knowledge and ability, need/necessity and age are important factors for becoming entrepreneurs. Variables with lower importance levels that are selected by the Random Forest classification tree, namely household income and population, working status, gender, education level, intra-entrepreneurial activity (as of the time of the survey) or being active in the company even if it is not at the leadership level in the last three years are identified as the main factors affecting entrepreneurship.

### **Conclusion**

This research is novel in the sense that machine learning models are applied on the GEM data for the first time in Turkey. The findings support the results of the research conducted by Özdemir and Karadeniz (2011) and Karadeniz and Özçam (2018). The contribution of this research to the literature is that it has yields results similar to those of the aforementioned studies and extends them using machine learning. For instance, unlike the earlier two studies, the present study finds that that people who work full-time in companies and take an active role in in-house entrepreneurship activities in Turkey, later became independent entrepreneurs. Since our results are highly nonlinear and subject to interactive relationships and conditions, the findings of the study were examined in detail within the framework of individual conditional expectations and partial dependences.