

TÜRKİYE'DE TOPLUMSAL CİNSİYETİN VERİ MADENİLİĞİ SINIFLAMA MODELLERİ İLE ANALİZİ

Analisis Of Gender In Turkey Using Data Mining Classification Model

Dr. Gülsüm Merve GÖKÇİN¹

Dr. Asila KOÇAK²

Cite As: Gökçin, G.M. & Koçak, A. (2021). "Türkiye'de Toplumsal Cinsiyetin Veri Madencilik Sınıflama Modelleri İle Analizi", International Journal of Disciplines Economics & Administrative Sciences Studies, (e-ISSN:2587-2168), Vol:7, Issue:31; pp:480-490.

ÖZET

Kadın ile ilgili istatistikler, toplumsal cinsiyet eşitliğinin sağlanması, kadının toplum içindeki yerinin belirlenmesi ve sosyo-ekonomik kalkınma içinde, kadınların katkısının güçlendirilmesi ve olumsuz göstergelerin tespit edilerek sosyal ve ekonomik konularının iyileştirilmesi için oldukça önemlidir. Bu çalışmada veri madencilik sınıflama (denetleyici öğrenme) modellerinden k-ortalamalar algoritması ve CHAID karar ağacı algoritmaları kullanılarak, 81 ilin 2019 yılına ait nüfus, doğurganlık, sağlık, evlenme, aile yaşamı, eğitim, siyasal yaşam, iş hayatı konuları kapsamında, toplumsal cinsiyet bakış açısıyla seçilen toplumsal yapı ve toplumsal cinsiyet göstergelerine göre sınıflandırılması ve bu sınıflandırmada etkili olan kuralların ve değişkenlerin belirlenmesi amaçlanmaktadır. Cinsiyet eşitliğini sağlamaya yönelik önlemlerin ve iyileştirmelerin bölgesel düzeyde farklılıklarla ele alınması önemlidir. Bu çalışma sonuçlarının aynı zamanda bölgesel düzeyde cinsiyete dayalı eşitsizliğe ilişkin göstergelerin ölçülmesi, takip edilmesi, karşılaştırılması ve toplumsal cinsiyet eşitliğine yönelik bölgesel düzeyde farkındalık oluşmasına katkı sunması amaçlanmaktadır.

Çalışmadaki ifade edilen görüşler tamamen yazarlara aittir ve hiçbir şekilde Türkiye İstatistik Kurumu'nu bağlamaz.

Anahtar Kelimeler: Toplumsal cinsiyet, Veri madencilik, CHAID analizi, K-ortalamalar yöntemi

JEL classification: J12, J16,C10

ABSTRACT

Statistics on women are very important for achieving gender equality, determining the place of women in society and strengthening the contribution of women in socio-economic development and improving their social and economic status by identifying negative indicators. In this study, using the k-means algorithm and CHAID decision tree algorithms, one of the data mining classification (supervisory learning) models, the gender perspective in 81 provinces within the scope of 2019 population, fertility, health, marriage, family life, education, political life, business life. It is aimed to classify them according to the selected social structure and gender indicators and to determine the rules and variables that are effective in this classification. It is important that measures and improvements aimed at achieving gender equality are addressed with regional differences. The results of this study are also aimed to measure, monitor and compare indicators of gender-based inequality at the regional level and contribute to raising awareness on gender equality at the regional level.

Key words: Keywords: Gender, Data mining, CHAID analysis, K-means method

JEL classification: J12, J16,C10


1. GİRİŞ

Makalenizi yazım kurallarına göre buraya yapıştırabilirsiniz. Sosyolojide cinsiyet, genellikle bedenin erkek ya da dişi olarak tanımlanmasına neden olan anatomik ve biyolojik farklılıkları ifade etmek için kullanılır. Toplumsal cinsiyet ise erkekler ve dişiler arasındaki toplumsal ve kültürel farklılıklarla ilgilidir.


Toplumsal cinsiyette eşitlik kavramı, kişilerin kaynak ve fırsatlara sahipliği ile bu kaynak ve fırsatların kullanımındaki eşitliği ifade eder. Bu kaynak ve fırsatlar hem evin içinde, hem ev dışında, yaşamın her alanında söz konusu olabilir. Buna göre toplumsal cinsiyet ayrımcılığı sosyal, ekonomik, siyasi ve hukuki yaşamın her alanında kaynaklara ve fırsatlara sahiplikte ve bunların kullanımında eşitsizliklerle karşı karşıya kalmaktır.

Toplumsal cinsiyet biyolojik cinsiyetten farklı olarak toplumsal ve kültürel olarak belirlenen ve dolayısıyla içeriği yere ve zamana göre değişebilen "cinsiyet konumu" ya da "cins kimliği"dir. Yalnızca cinsiyet farklılığını belirtmekle kalmaz aynı zamanda cinsler arasındaki eşitsiz güç ilişkilerini de belirler. (Berktaş vd., 2011)

¹ Türkiye İstatistik Kurumu (TÜİK), Ankara/TÜRKİYE

 0000-0002-2644-1942

² Türkiye İstatistik Kurumu (TÜİK), Ankara/TÜRKİYE

 0000-0001-5048-3424

Biyolojik özelliklerin toplumsal, kültürel, politik, ideolojik ve ekonomik yapılarla etkileşimi sonucu farklılaşarak, biyolojik cinsiyetten başka bir hal alışı ve iki farklı yaşam alanının oluşması eşitsizliklere neden olmaktadır. (Sarp, 2008)

Tüm dünyada kadınların durumunda görülen iyileşmelere rağmen, toplumsal cinsiyet farklılıkları toplumsal eşitsizlikler için temel teşkil etmeye devam etmektedir.

Bugüne kadar dünyada ve Türkiye’ de kadın hakları alanında elde edilen gibi birçok olumlu gelişmeye rağmen kadınların toplumsal hayata katılımları hala yeterince iyi durumda değildir. Bir toplumda kadınların iyi durumda olması her alanda iyi koşullara sahip olması ile mümkündür.

2019 tarihli Küresel Toplumsal Cinsiyet Uçurumu Raporu’na (WEF) göre Türkiye 153 ülke arasında 130. Sırada yer almaktadır (WEF, 2019). Aynı zamanda 2019 yılı OECD raporuna göre de Türkiye, kadın istihdam oranı ve kadın bürokrat sayısı açısından birçok OECD ülkesinin gerisinde yer almaktadır (OECD, 2019).

Modernleşme ve beraberinde sanayileşmeyle birlikte, doğuştan beklenen yaşam süresinin uzaması, anne ve çocuk ölümlerinin azalması gibi kadın sağlığı ve kadının işgücüne katılımı başta üzere birçok alanda olumlu gelişmeler görülmeye başlandı. Bu durum aynı zamanda, kadının eş ve anne olmak dışında hayatın her alanda eşit koşullara sahip olmalarına duyulan ihtiyacın artmasına da neden oldu. Ancak ne yazık ki ataerkil sistem ve erkek egemen yapı, kadınların yaşamında olumlu görünen bu gelişmelere rağmen sınırlayıcı, engelleyici, zorlayıcı önyargılarla devam etmektedir.

Kadınların maruz kaldığı eşitsizlikler, kadınların aleyhine işlemekte daha çok mücadele gerektiren alanlar olarak karşılırlarına çıkmaktadır.

Kadının evlilik dışında ekonomik alternatiflere sahip olması, çocuk, hasta, yaşlı bakımı ve ev işi gibi ev içi faaliyetler dışında, kamusal alanda eşit imkanlara sahip olarak varlığını sürdürmesi gelişmişlik düzeyi ve sürdürülebilir kalkınma için de önemli bir belirleyicidir.

Toplumsal cinsiyetin toplumsal yapıyı oluşturan sosyal, kültürel ekonomik dinamiklerle birlikte ele alınması çok boyutlu değerlendirmeler sağlamaktadır. Bu kapsamda kadının eğitime devam etme ya da etmeme, ev dışında çalışma ya da çalışmama, medeni durumunu seçme, anne olma ya da ne zaman kaç çocuk sahibi olacağına karar vermelerini etkileyen ya da belirleyen sosyal, kültürel etkenlerin yanısıra kentleşme, işgücü piyasası gibi ekonomik etkenler de toplumsal yapının dinamiklerini oluşturmakta ve kadınlar için caydırıcı ya da cesaretlendirici olabilmektedir.

Toplumsal yapının ve toplumsal cinsiyetin birlikte ele alınması, toplumsal cinsiyet eşitliğinin sağlanması, sosyo-ekonomik kalkınma içinde, kadınların katkısının güçlendirilmesi ve olumsuz göstergelerin tespit edilerek iyileştirilmesi açısından oldukça önemlidir.

Bu çalışmanın amacı da ülkemizde bölgesel düzeyde toplumsal yapının ve toplumsal cinsiyetin mevcut veriler ışığında bilimsel yöntemlerle ortaya koyulmasıdır. Bu amaç doğrultusunda veri madenciliği sınıflama modellerinden k-ortalamalar algoritması ve CHAID karar ağacı algoritmaları kullanılarak, 81 ilin 2019 yılına ait nüfus, doğurganlık, sağlık, evlenme, aile yaşamı, eğitim, siyasal yaşam, iş hayatı konuları kapsamında, toplumsal cinsiyet bakış açısıyla seçilen toplumsal yapı ve toplumsal cinsiyet göstergelerine göre sınıflandırılması gerçekleştirilmiş ve bu sınıflandırmada illerin kümelenmesinde etkili olan kurallar ve değişkenler tespit edilmiştir.

2. VERİ SETİ

Bu çalışmada 2019 yılına ilişkin toplumsal cinsiyet göstergelerine göre 81 ilde yaşayan kadınların değerlendirilebilmesi amacıyla 17 gösterge belirlenmiştir.

TÜİK tarafından yayınlanan 81 il düzeyinde toplumsal cinsiyet göstergelerine, kadının siyasi gücünü yansıtmak amacıyla “kadın milletvekili oranı” ve kadının iş hayatı durumunu yansıtmak amacıyla SGK kayıtlarından “4a/b/c kapsamında olan çalışan ve zorunlu sigortaya sahip kadın oranı hesaplanarak veri setine eklenmiştir.

Kümeleme analizinde, veriler arasında birim farklılıkları olması ve bazı verilerinin sayısal olarak diğer verilerden büyük olması nedeniyle (örneğin yabancı nüfuslu kadın sayısı) oluşacak sakıncaları ortadan kaldırmak için veriler oran düzeyinde standart hale getirilerek analiz edilmiştir. Tüm değişkenler kadın düzeyinde olup, göstergelerin oran hesaplamaları olarak Tablo 1’ de belirtilmiştir.

Tablo1. 81 İl Toplumsal Cinsiyet Veri Setinde Yer Alan Göstergeler

Gösterge	Kapsamı	Hesap	Kaynak
Kadın intihar oranı %	Sağlık	kadın intihar sayısı (İl) / toplam intihar sayısı(TR) *100	TÜİK
Kadın milletvekili oranı %	Siyasi Güç	kadın milletvekili sayısı (İl) / toplam milletvekili sayısı (İl) *100	TBMM
15+okuryazar olmayan kadın oranı %	Eğitim	15+okuryazar olmayan kadın (İl) / eğitim alan genel toplam kadın sayısı(İl) *100	TÜİK
Yüksekokul mezunu oranı %	Eğitim	Yüksekokul mezunu kadın (İl) / eğitim alan genel toplam kadın sayısı (İl) *100	TÜİK
Doğurganlık hızı	Doğurganlık	-	TÜİK
Evlilik süresi 1 yıldan az, doğum oranı %	Doğurganlık	Evlilik süresi 1 yıldan az doğum sayısı (İl) /toplam doğum sayısı (İl) *100	TÜİK
Doğurganlık	Doğurganlık	Evlilik süresi 1 yıl doğum sayısı (İl) /toplam doğum sayısı (İl) *100	TÜİK
Evlilik süresi 20 yıl, doğum oranı %	Doğurganlık	Evlilik süresi 20 + doğum sayısı (İl) /toplam doğum sayısı (İl) *100	TÜİK
Hiç evlenmeyen anne oranı%	Evlenme	Medeni durumu "hiç evlenmeyen" anne (İl) / toplam doğum sayısı (İl) *100	TÜİK
16-19 yaş evlenme oranı %	Evlenme	16-19 yaş evlenme sayısı (il) / toplam evlenme sayısı (İl) *100	TÜİK
Erkek 11 yaş büyük oranı%	Evlenme, Aile yaşamı	erkek 11 yaş büyük evlenme sayısı (İl) / toplam evlenme sayısı (İl)	TÜİK
Tek Çekirdek aile oranı %	Aile yaşamı	Tek çekirdekli aile sayısı (İl) / toplam hanehalkı sayısı (İl) *100	TÜİK
En az bir çekirdekli aile oranı %	Aile yaşamı	En az bir çekirdekli aile sayısı (İl) / toplam hanehalkı sayısı (İl) *100	TÜİK
Yabancı nüfuslu kadın oranı %	Nüfus	Yabancı nüfuslu kadın (İl) / toplam kadın nüfusu (İl) *100	TÜİK
Ortalama evlilik yaşı	Evlenme	-	TÜİK
Akraba evliliği oranı %	Evlenme, Aile yaşamı	Avraba evliliği sayısı (İl) / toplam evlenme sayısı (İl) *100	TÜİK
4A/b/c zorunlu sigortalı kadın oranı %	İş Hayatı	4A/b/c zorunlu sigortalı çalışan kadın sayısı (İl) / toplam kadın nüfusu (İl) *100	SGK

3. YÖNTEM

Veri madenciliğinin genel amacı; daha önce hiç keşfedilmemiş ya da tahmin edilemeyen ilişkileri ortaya çıkarmak ve ele alınan veriyi anlaşılır hale getirerek özetlemektir. Bu çalışmada da veri madenciliği sınıflama (denetleyici öğrenme) modellerinden k-ortalamlar algoritması ve CHAID karar ağacı algoritmaları kullanılarak, 81 ilin toplumsal cinsiyet göstergelerine göre sınıflandırılması ve bu sınıflandırmada etkili olan kuralların ve değişkenlerin belirlenmesi amaçlanmaktadır.

3.1 K-ortalamlar algoritması

K-ortalamlar algoritması, genellikle büyük veri kümeleri için en hızlı kümeleme yöntemidir. Başlangıçta verinin ait olduğu grupların ne olduğu bilinmemektedir ve veri kümesini farklı gruplara ayırmak için kullanılabilir. K-ortalamlar modelleri bir hedef değişkenine ihtiyaç duymaz. Hedef değişkeni olmayan bu tür öğrenmeye denetimsiz öğrenme denir. K-ortalamlar, bir sonucu tahmin etmeye çalışmak yerine, girdi alanları kümesindeki kalıpları ortaya çıkarmaya çalışır. Kayıtlar, bir grup veya küme içindeki kayıtlar birbirine benzeyecek, ancak farklı gruplardaki kayıtların birbirinden farklı olacak şekilde gruplanır.

K-ortalamlar algoritması dögüsel bir algoritmadır. İlk adımı verilerden türetilen bir dizi başlangıç küme merkezi tanımlamaktır. 2.adımda her kaydı, kaydın giriş alanı değerlerine göre en benzer olduğu kümeye atar. Kümenin benzerliği kümedeki objelerin ortalama değeriyle (mean value) ölçülür. Tüm veriler atandıktan sonra, küme merkezleri, her kümeye atanan yeni kayıt kümesini yansıtacak şekilde güncellenir. Kayıtlar daha sonra farklı bir kümeye yeniden atanmaları gerekip gerekmediğini görmek için tekrar kontrol edilir ve kayıt atama / küme yineleme süreci, maksimum yineleme sayısına ulaşıncaya veya bir yineleme ile sonraki arasındaki değişiklik, belirtilen eşik değerine ulaşana kadar yinelenir. Algoritma için modelin durdurma kriterleri belirlenebilir.

Çalışmada kullanılan IBM SPSS Modeler programında varsayılan kriterler 20 yineleme ve bir yineleme ile sonraki arasındaki değişiklik değeri için eşik değerinin 0,000001 değerinden küçük olmasıdır.

3.2 Karar ağaçları

Karar ağaçları, bir grup karar kuralına dayanarak gelecek gözlemleri sınıflayan ya da tahmin eden sınıflama sistemlerini geliştirmeye izin veren veri madenciliği modellerinden biridir. Bu modeller ile eğer ele alınan veri sınıflara ayrılabilirse, maksimum doğruluk ile eski ve yeni değerleri sınıflamak için kurallar belirlenebilir. Karar ağaçları bilgi keşfi sırasında pek çok test gerçekleştirerek, hedefi tahmin etmede en iyi sırayı bulmaya çalışır. Bu sayede belirlenen karar verme adımları ile büyük miktarlardaki kayıtlar, çok küçük kayıt gruplarına bölünebilmektedir. Karar ağaçlarının sıklıkla uygulamalarda tercih edilmesinin birçok sebebi vardır. Bunlar karar ağaçlarının uygulaması ve anlaması kolay modeller olması, sürekli ve kategorik verilerle çalışabilmesi, veri hazırlığına çok ihtiyaç duymamasıdır.

En yaygın kullanılan karar ağacı algoritmaları CHAID, C&R, C5 ve QUEST algoritmalarıdır. Algoritma seçimi, hedef değişkenin tipine göre değişkenlik gösterir. Karar ağaçlarında en sık kullanılan algoritmalar; kategorik değişkenler için Entropi, Gini, Sınıflandırma Hatası; sürekli değişkenler için ise En Küçük Kareler yöntemi şeklindedir.

Bu çalışmada kümelerin dağılımının hangi değişkenlerle ilgili olabileceğini tespit edebilmek için CHAID ağaçları kullanılmıştır. Hedef değişken k-ortalamlar yöntemine göre belirlenen küme gruplarıdır.

3.2.1 CHAID ağaçları

CHAID ağaçları, 1980 yılında Kaas tarafından geliştirilmiştir ve bir popülasyonu, bağımlı değişkendeki varyasyonu gruplar içi minimum, gruplar arası maksimum olacak şekilde farklı alt gruplara veya bölümlere tekrarlı olarak ayıran ve değişkenler arasındaki etkileşim veya kombinasyonları bulan bir yöntemdir (Gökçin,2019).

Değişkenler arasındaki ilişki doğrusal yapıdan daha karmaşık ise veride gizli olan bu ilişkiyi bulmak için verinin belli kısımlarını eleme yöntemi olan CHAID kullanılır. “Ki-kare” ismini almasının nedeni algoritmasında birçok çapraz çizelgenin kullanılması ve istatistiksel önem oranları ile çalışmasıdır (Pehlivan, 2006; Ekici, 2012).

CHAID algoritması, tahmin edici değişkenin tüm değerlerini kullanarak analiz yapmaktadır. Yapılacak testler hedef değişkenin türüne göre değişmektedir. Eğer değişken sürekli bir değişkense F testi, kategorik bir değişkense χ^2 testi kullanılır. Eğer kategorik bir hedef değişkeni varsa, CHAID algoritması, her bir girdi ve hedef değişkeni arasındaki çapraz tablo yardımıyla Eş. 1.1’deki χ^2 bağımsızlık testini kullanarak anlamlılığı test eder. Burada kurulacak sıfır hipotezi “değişkenler bağımsızdır” şeklindedir.

$$\chi^2 = \sum_{i=1}^n \frac{(X_i - E_i)^2}{E_i} \quad (1.1)$$

Burada X_i , i değişkeninin gözlem değerini, E_i değeri, i değişkeninin beklenen değerini ve n değişken sayısını belirtmektedir. Büyük χ^2 değerleri gözlem verisinin aykırı değer olduğunu belirtmektedir (Chandola, Banerjee, Kumar, 2009).

Eğer bu ilişkilerden birden fazlası istatistiksel olarak anlamlıysa, CHAID en anlamlı girdi değişkenini seçer. Eğer girdi değişkeni ikiden fazla kategoriye sahipse, bunlar karşılaştırılır ve hedef değişkeninde fark göstermeyen kategoriler birlikte ele alınır. Bu süreç en küçük anlamlı farkı gösteren kategori çiftlerini birleştirerek devam eder. Kategori birleştirme süreci, geriye kalan kategorilerin hepsinin tanımlanan test seviyesinde fark yarattığında durur (Oktay,2013).

Çalışmada kullanılan IBM SPSS Modeler programında maksimum ağaç derinliği, kök işlemcinin maksimum bölünme sayısıdır. Varsayılan bölünme sayısı 5’tir. Bagging ve çok büyük veriler şıkları için hedef değişkeni kategorik olduğunda kurallar, en yüksek olabilirlik, en yüksek ortalama ya da oy verme (voting) ile birleştirilir. Burada oy verme, kategorik hedef değişkeni için tahmin değerlerini toplulaştırma değildir.

4. BULGULAR

Çalışmada yer alan 17 göstergeye göre (Tablo 1) illerin hangi kümede yoğunlaştığını görebilmek için k-ortalamlar kümeleme yöntemi ve bu kümelerin hangi göstergelerin etkisi nedeniyle farklılaştığını tespit edebilmek için ise karar ağacı analizi kullanılmıştır. K-ortalamlar yöntemine göre öncelikle farklı küme sayıları seçilmiş ve küme bazında en iyi dağılım 5 sayısında elde edilmiştir. K-ortalamlar yöntemine ilişkin

elde edilen kümelemenin kalitesini belirten ortalama Silhouette ölçüm değeri 0.3 olup orta düzeyde olduğu tespit edilmiştir.

Silhouette ölçü ortalamaları, tüm kayıtlar üzerinden $(B - A) / \max(A, B)$ olarak hesaplanır. Burada A değeri kaydın küme merkezine uzaklığı ve B değeri kaydın ait olmadığı en yakın küme merkezine uzaklığıdır. 1'lik bir Silhouette katsayısı, tüm vakaların doğrudan kendi küme merkezlerinde bulunduğu anlamına ; -1 değeri ise tüm vakaların başka bir kümenin küme merkezlerinde yer aldığı anlamına gelmektedir.

Hangi ilin hangi kümede olduğu Şekil 1.'de Türkiye haritası üzerinde görselleştirilmiştir. Buna göre 5 kümenin illere göre dağılıma bakıldığında 81 ilden 26'sı 1. kümede; 15'i 2. kümede; 13'ü 3. Kümede; 17'si 4. Kümede ve 10'u 5.kümede yer almaktadır.

Tablo 2' de kümelerin seçilen toplumsal cinsiyet göstergelerine göre ortalamaları bulunmaktadır. Buna göre; bazı gösterge değerlerinin birbirinden oldukça farklı olduğu görülmektedir. Tablo 2' de Toplumsal cinsiyet göstergelerinin, en avantajlı küme değeri kırmızı koyu renk ile en avantajsız küme değeri koyu siyah renk ile gösterilmiştir.

Bu tabloda "1 yıllık evlilik süresine göre doğumlar", ortalama evlilik yaşının küçük olduğu durumlarda avantajsız olarak değerlendirilebilir. Ayrıca "evlilik süresi 20+ yıl doğum oranı" ise, çocuk sahibi sayısının fazla olması ve ortalama evlilik yaşının küçük olması ile birlikte düşünüldüğünde toplumsal cinsiyet açısından avantajsız olarak değerlendirilebilir. Bu sebeple Tablo2' de koyu renk ile belirtilmeden nötr bırakılmıştır. Ortalama evlenme yaşının küçük olması ise, kadının eğitime devam etme ve iş hayatına katılma kararı üzerinde yol ayrımına sebep olmaktadır. Bu sebeple ortalama evlenme yaşının küçük olması toplumsal cinsiyet açısından avantajsız sonuçlar doğurmaktadır.

Tablo2' de "kadın intihar oranı" bakımından en avantajlı küme 0,393 ortalama ile 2. Küme'dir ve genellikle Karadeniz bölgesi illerini içine almaktadır. En avantajsız küme ise büyük kentleri içine alan nüfus yoğunluğunun fazla olduğu illeri kapsamaktadır (küme 5). Bu küme de yer alan İstanbul, 81 il için de 12,69 ile en yüksek intihar oranına sahip ildir. İstanbul' u sırasıyla 5,51 ve 4,61 oranlarıyla yine bu küme de yer alan Ankara ve İzmir takip etmektedir.

"Kadın milletvekili oranı", "15+ okuryazar olmayan kadın oranı", "yüksekokul mezunu kadın oranı", "4/a/b/c zorunlu sigortalı kadın oranı" bakımından en avantajlı küme İstanbul Ankara ve İzmir illerini içine alan küme 5' tir. Diğer taraftan Küme2, "Kadın milletvekili oranı" bakımından en avantajsız illerin görüldüğü kümedir.

"15+ okuryazar olmayan kadın oranı" ve "yüksekokul mezunu kadın oranı" bakımından en avantajsız küme önemli farkla, Doğu ve Güneydoğu Anadolu Bölgeleri illerini kapsayan küme 3' tür. Bu küme "ortalama evlilik yaşının" en düşük olduğu, "16-19 yaş evlenen kadın oranının" en yüksek olduğu, "4/a/b/c zorunlu sigortalı kadın oranı" en düşük olduğu ve "akraba evliliğinin" en yüksek olduğu illerden oluşan küme olarak da dikkat çekmektedir. Küme 3' ün buna göre toplumsal cinsiyet karnesinin parlak olmadığı görülmektedir.

Ayrıca Tablo2'de ANOVA analiziyle, değişkenlerin, kümeler arası farklılık gösterip göstermediğinin incelenmesi amaçlanmaktadır. Değişkenlerin kümeler itibariyle farklılık göstermesi beklenmektedir. Çünkü kümeleme analizi, bu farklılığı kendi oluşturur ve kümeler arası farkı (uzaklığı) maksimum yapar.

SPSS Modeler programında birbirinden bağımsız gruplar arası kare ortalamalarının testi (ANOVA) analiz aşamasında hesaplanan p değeri 0.90'dan düşük olanlar kümeler itibariyle farklı çıkmayan, 0.90'dan yüksek olanlar kümeler itibariyle farklılık yaratanlar olarak belirlenmektedir. Ayrıca tabloda hesaplanan F değeri 1'e yaklaştıkça gruplar arası ortalamaların birbirinden farksız olduğu söylenebilmektedir. F değeri arttıkça gruplar arası ortalamalar da birbirinden uzaklaşmakta ve fark artmaktadır. Elde edilen sonuçlara göre tüm göstergeler için kümeler arasında anlamlı farklılık bulunmaktadır.

81 İl Toplumsal Cinsiyet Küme Haritası



Şekil 1. 81 İl Toplumsal Cinsiyet Küme Haritası

Tablo 2. 81 İl Toplumsal Cinsiyet Göstergelerinin Küme Ortalamalarına Göre En Avantajlı-En Avantajsız Olduğu Kümeler

Gösterge	küme-1*	küme-2*	küme-3*	küme-4*	küme-5*	F-Test	sd	önemlilik
Kadın intihar oranı (%)	1,44	0,393	1,321	0,649	2,846	4,126	4,76	0,996
Kadın Milletvekili oranı (%)	16,247	2,222	23,15	5,852	22,996	10,641	4,76	1
15+okuryazar olmayan kadın oranı(%)	4,304	8,387	13,297	8,069	3,223	74,031	4,76	1
Yüksekökl mezunu kadın oranı(%)	13,345	13,01	9,204	11,416	17,486	35,621	4,76	1
Doğurganlık hızı	1,681	1,585	2,909	2,034	1,508	62,656	4,76	1
1 yıldan az evlilik süresine göre doğum oranı (%)	9,739	8,992	13,89	10,05	9,07	51,314	4,76	1
1 yıllık evlilik süresine göre doğum oranı (%)	15,715	16,587	11,426	14,488	15,862	40,779	4,76	1
20 yıldan fazla evlilik süresine göre doğum oranı (%)	1,039	1,229	1,043	1,295	0,859	5,163	4,76	0,999
Hiç evlenmeyen doğum yapan kadın oranı (%)	1,253	0,713	3,08	1,572	1,418	16,086	4,76	1
16_19 yaş evlenen kadın oranı(%)	12,656	8,893	27,148	20,341	8,967	43,504	4,76	1
Erkek 11 yaş büyük evlilik oranı(%)	6,595	6,703	8,169	7,872	7,233	5,549	4,76	0,999
Tek Çekirdek aile_oranı(%)	65,206	59,596	64,456	66,279	62,013	9,619	4,76	1
En az bir Çekirdek_oranı(%)	14,879	15,141	21,709	15,262	13,222	22,797	4,76	1
Yabancı nüfuslu kadın oranı(%)	1,241	1,144	0,272	1,372	3,217	10,535	4,76	1
Ortalama evlilik yaşı	24,668	25,515	23,585	23,875	25,671	19,147	4,76	1
Akraba evliliği oranı (%)	2,194	2,467	11,102	4,745	1,61	86,101	4,76	1
4A/b/c zorunlu sigortalı kadın oranı (%)	14,167	11,736	6,166	9,356	17,953	57,484	4,76	1

Kümelerin oluşmasında hangi göstergelerin önemli olduğunun ve ayrışma sebeplerinin belirlenebilmesi için bu çalışmada CHAID karar ağacı yöntemi seçilmiştir. İlgili yöntemde bağımlı yani hedef değişken illerin ayrıştığı kümelerdir. CHAID karar ağacına ilişkin oluşturulan modelin derinliği 3'tür ve Pearson Ki-Kare yöntemine göre ağacın dalları oluşturulmuştur. Elde edilen model sonuçlarına göre 17 göstergeden 5'i en önemli fark yaratan, kuralların oluşmasını sağlayan değişkenler olarak belirlenmiştir. İstatistiksel olarak önem sırasına göre bu değişkenler sırasıyla "4/a/b/c çalışan zorunlu sigortalı kadın oranı", "16-19 yaş evlenen kadın oranı", "1 yıllık evlilik oranına göre doğum oranı %", "Erkek 11 yaş büyük evlilik oranı %" ve "Ortalama Evlilik Yaşı" değişkenleridir, CHAID karar ağacı algoritmasında kuralların oluşmasını sağlayan en önemli değişken olarak belirlenmiştir.

Modelin sınıflandırma performansına bakıldığında; doğru sınıflandırma yüzdesi % 82,72; yanlış sınıflandırma yüzdesi % 17,21 olarak tespit edilmiştir. Buna göre bu modelin oldukça yüksek doğrulukta bir sınıflandırma yaptığını söyleyebiliriz.

1.Kümede yer alan 26 ilden 3'ünde "4/a/b/c zorunlu sigortalı kadın oranı" yaklaşık %8 ile %10,6 arasındadır. Bu iller "16-19 yaş evlilik oranı" değişkenine göre ayrışmaktadır; bu illerden Elazığ ilinde "16-19 yaş evlilik

oranı” %10 ile %11 arasındadır. Konya ve Tokat illerinde ise “16-19 yaş evlilik oranı” %11-27 arasındadır ve “erkek 11 yaş büyük evlilik oranı” %6 ‘dan küçüktür.

1.kümede yer alan 26 ilden 9’unda “4/a/b/c zorunlu sigortalı kadın oranı” yaklaşık %10.6 ile %13.7 arasındadır. Bu illerden sadece Afyonkarahisar ilinde evlilik yaşı 23’ten küçüktür. 8’inde ortalama evlilik yaşı 24’ten yüksektir. Bu iller Adana, Amasya, Burdur, Mersin, Kayseri, Kütahya, Samsun ve Bartın’dır.

1.kümede yer alan geriye kalan 14 ilde “4/a/b/c zorunlu sigortalı kadın oranı” yaklaşık %13,7’den yüksektir. Bu iller Aydın, Balıkesir, Bilecik, Bursa, Denizli, Isparta, Kırklareli, Kocaeli, Manisa, Sakarya, Tekirdağ, Uşak, Karaman, Düzce ‘dir.

“4a/b/c Sigortalı Kadın Oranına” ilişkin kümelerin ortalamaları dikkate alınarak iller gruplandırılmıştır ve hazırlanan Türkiye haritası Şekil 2’de belirtilmiştir.



Şekil 2. 4a/b/c Zorunlu Sigortalı Kadın Oranına Göre Türkiye Haritası

2.kümede yer alan 15 ilde “4/a/b/c zorunlu sigortalı kadın oranı” %7,7’den fazladır. Bu illerden Gümüşhane, Malatya ve Bayburt’ta bu oran %7,7 ile %10,6 arasındadır. Ayrıca bu 3 ilde “16-19 yaş evlilik oranı” %10’dan azdır. “16-19 Yaş Evlenen Kadın Oranına” ilişkin kümelerin ortalamaları dikkate alınarak iller gruplandırılmıştır ve hazırlanan Türkiye haritası Şekil 3’de belirtilmiştir.



Şekil 3. 16-19 Yaş Evlenen Kadın Oranına Göre Türkiye Haritası

2.kümede yer alan 12 ilde “4/a/b/c zorunlu sigortalı kadın oranı” %10,6-%13,7 arasındadır ve tamamında “ortalama evlilik yaşı” 24’den fazladır. Bu iller Artvin, Çankırı, Erzincan, Giresun, Kastamonu, Ordu, Rize, Sinop, Trabzon, Tunceli, Zonguldak, Karabük’dür.

Evlenme istatistiklerine göre; resmi olarak ilk evliliğini 2019 yılında yapmış olan kadınların ortalama evlenme yaşı 25,0 iken erkeklerin 27,9’dır. Ortalama ilk evlenme yaşının en yüksek olduğu il, kadınlarda 28,4 yaş, erkeklerde 31,1 yaş ile Tunceli’dir. Ortalama ilk evlenme yaşının en düşük olduğu iller ise kadınlarda 21,9 yaş ile Ağrı, erkeklerde 26,0 yaş ile Şanlıurfa olarak görülmektedir. “Ortalama Evlilik Yaşına” ilişkin kümelerin ortalamaları dikkate alınarak iller gruplandırılmıştır ve hazırlanan Türkiye haritası Şekil 4’de belirtilmiştir.



Şekil 4. Ortalama Evlilik Yaşına Göre Türkiye Haritası

3.kümede yer alan 13 ilin tamamında “4/a/b/c zorunlu sigortalı kadın oranı” %10,6’dan azdır. Bu iller Ağrı, Bitlis, diyarbakır, Hakkari, Mardin, Muş, Siirt, Şanlıurfa, Van, Batman, Şırnak, Iğdır’dır. Bu illerden Kars ili hariç diğerlerinde “4/a/b/c zorunlu sigortalı kadın oranı” %7’den azdır. Bu illerin tamamında “1 yıllık evlilik süresine göre doğum oranı” %13,2’den düşüktür. “1 yıllık Evlilik Süresine göre Doğum Oranına” ilişkin kümelerin ortalamaları dikkate alınarak iller gruplandırılmıştır ve hazırlanan Türkiye haritası Şekil 5’de belirtilmiştir.



Şekil 5. 1 yıllık Evlilik Süresine göre Doğum Oranı Türkiye Haritası

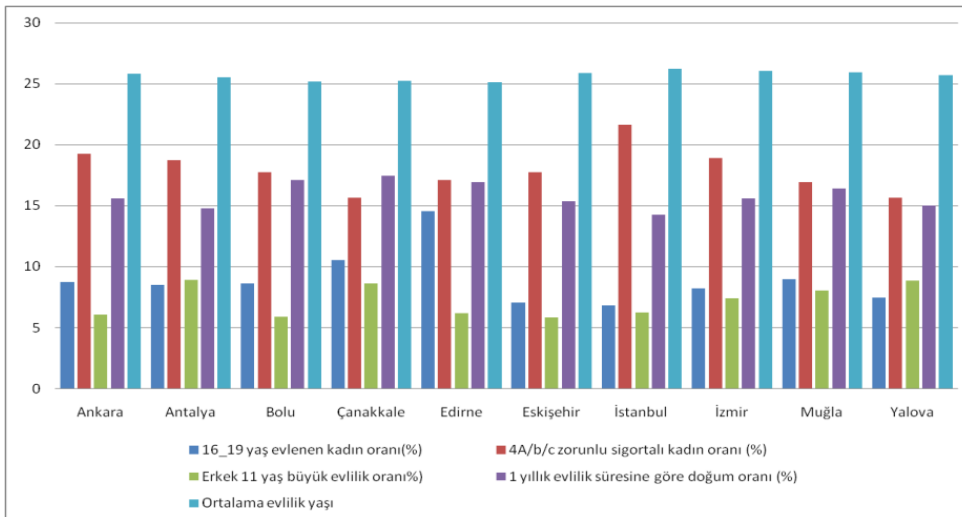
4.kümede yer alan 17 ilin tamamında “4/a/b/c zorunlu sigortalı kadın oranı” %13,7’den azdır. Bu illerden 2’sinde (Adıyaman ve Bingöl’de) “4/a/b/c zorunlu sigortalı kadın oranı” %7,7’den azdır ve bu illerde “1 yıllık evlilik süresine göre doğum oranı” %13,2’den fazladır. Nevşehir ili de “4/a/b/c zorunlu sigortalı kadın oranı” %10,6-13,7 arasında olması ve ortalama evlilik yaşının 23-24 arasında olması nedeniyle ayrı bir yerdedir. Geriye kalan 14 ilin tamamında “4/a/b/c zorunlu sigortalı kadın oranı” %7,8-10,6 arasındadır. Bu iller Kilis, Çorum, Erzurum, Gaziantep, Hatay, Kırşehir, Kahramanmaraş, Niğde, Sivas, Yozgat, Aksaray, Kırıkkale, Ardahan, Osmaniye illeridir. Bu illerden Kilis ilinde “16-19 yaş evlilik oranı” %27,2 olması nedeniyle ayrılmaktadır. Geriye kalan 13 ilin tamamında “16-19 yaş evlilik oranı” %11-27 arasındadır ve “erkek 11 yaş büyük olma oranı” %6’dan fazladır.

“Erkek 11 Yaş Büyük Evlilik Oranına” ilişkin kümelerin ortalamaları dikkate alınarak iller gruplandırılmıştır ve hazırlanan Türkiye haritası Şekil 6’da belirtilmiştir.



Şekil 6. Erkek 11 Yaş Büyük Evlilik Oranı Türkiye Haritası

Bu kümelerin dışında kalan iller ise Küme 5’te toplanmıştır. Bu iller Ankara, Antalya, Bolu, Çanakkale, Edirne, Eskişehir, İstanbul, İzmir, Muğla ve Yalova olarak belirlenmiştir. Bu illerin tamamında “4/a/b/c zorunlu sigortalı kadın oranı” %15,7’den fazladır. Ortalama evlilik yaşı bu illerde 25-26 olup, “1 yıllık evlilik süresine göre doğum oranı” ortalama %15,6’dır. Bu illerden Edirne ve Çanakkale illeri hariç 16-19 yaş evlilik oranı ortalama %6-8 arasındayken, Edirne ilinde bu oran %14; Çanakkale ilinde bu oran %10,5’dir. CHAID analizinde kümelerin belirlenmesinde en önemli olarak seçilen göstergelere göre illerin durumu Şekil 7’de belirtilmiştir.



Şekil 7. Küme 5 ve Önemli Farklılık Yaratıcı Değişkenlere Göre İller

5. SONUÇ ve ÖNERİLER

Bu çalışmada veri madenciliği sınıflama modellerinden k-ortalama algoritması ve CHAID karar ağacı algoritmaları kullanılarak, 81 ilin 2019 yılına ait nüfus, doğurganlık, sağlık, evlenme, aile yaşamı, eğitim, siyasal yaşam, iş hayatı konuları kapsamında, toplumsal cinsiyet bakış açısıyla seçilen toplumsal yapı ve toplumsal cinsiyet göstergelerine göre sınıflandırılmış ve bu sınıflandırmada etkili olan kural ve değişkenler belirlenmiştir.

Buna göre k-ortalama yöntemine göre belirlenen 5 kümenin illere göre dağılıma bakıldığında 81 ilden 26'sı 1. kümede; 15'i 2. kümede; 13'ü 3. Kümede; 17'si 4. Kümede ve 10'u 5.kümede yer almaktadır. Bu kümelerin hangi değişkenlere göre ayrıştığı belirlenmesi amacıyla CHAID karar ağacı analizi yapılmıştır. Elde edilen model sonuçlarına göre, 17 göstergeden 5'i CHAID karar ağacı algoritmasında kuralların oluşmasını sağlayan ve kümelerin ayrışmasında en önemli farklılığı yaratan değişkenler olarak belirlenmiştir. İstatistiksel olarak önem sırasına göre bu değişkenler sırasıyla "4/a/b/c çalışan zorunlu sigortalı kadın oranı", "16-19 yaş evlenen kadın oranı", "1 yıllık evlilik oranına göre doğum oranı %", "Erkek 11 yaş büyük evlilik oranı %" ve "Ortalama Evlilik Yaşı" değişkenleridir.

Elde edilen kümeler incelendiğinde bazı bulgular dikkat çekicidir. "Kadın milletvekili oranı", "15+ okuryazar olmayan kadın oranı", "yüksekokul mezunu kadın oranı", "4/a/b/c zorunlu sigortalı kadın oranı" bakımından en avantajlı küme İstanbul, Ankara ve İzmir illerini de içine alan küme 5' tir.

"15+ okuryazar olmayan kadın oranı" ve "yüksekokul mezunu kadın oranı" bakımından en avantajsız küme önemli farkla, Doğu ve Güneydoğu Anadolu Bölgeleri illerini kapsayan Küme3' tür. Bu küme ortalama evlilik yaşının en düşük olduğu, 16-19 yaş evlenen kadın oranının en yüksek olduğu, 4/a/b/c zorunlu sigortalı kadın oranının en düşük olduğu ve Akraba evliliğinin en yüksek olduğu illerden oluşan küme olarak da dikkat çekmektedir. Bu bağlamda Küme 3' ün toplumsal cinsiyet karnesinin parlak olmadığı görülmektedir.

Toplumsal cinsiyet göstergelerinin detaylı yorumlanması için ile aile hayatı ve toplumsal yapının birlikte ele alınmasının önemini bir kez daha vurgulamak gerekir.

Adrese Dayalı Nüfus Kayıt Sistemi (ADNKS) sonuçlarına göre; 2019 yılında toplam resmi evlenmeler içerisindeki akraba evlilikleri illere göre incelendiğinde; 2019 yılında toplam evli bireyler içerisinde son evliliğinde akraba evliliği yapmış 16 ve üzeri yaştaki bireylerin oranının en fazla görüldüğü il, %20,8 ile Mardin'dir. Bu ili sırasıyla Şanlıurfa ve Diyarbakır izlemektedir. Akraba evliliği yapmış bireylerin oranının en az görüldüğü il ise %1,1 ile Edirne'dir. Bu ili sırasıyla Kırklareli ve Çanakkale izlemektedir.

Resmi evlenmeler içerisindeki akraba evlilikleri illere göre incelendiğinde; 2019 yılında gerçekleşen toplam evlenmeler içindeki akraba evliliği oranının en yüksek olduğu ilin, %14,9 ile Şanlıurfa olduğu, bu ili, Siirt ve Mardin illerinin izlediği görülmektedir. En yüksek akraba evliliği oranına sahip illerin aynı kümede yer alarak, diğer toplumsal cinsiyet göstergeleri bakımından en avantajsız kümede yer alması bir tesadüf değildir.

ADNKS sonuçlarına göre; Türkiye'de ortalama hanehalkı büyüklüğünün 2019 yılında 3,35 kişi olduğu görülmektedir. İllere göre incelendiğinde; 2019 yılında ortalama hanehalkı büyüklüğünün en yüksek olduğu il, 6,11 kişi ile Şırnak iken Şırnak ilini 5,39 kişi ile Hakkari ve 5,38 kişi ile Şanlıurfa illeri izlemektedir. Benzer şekilde ortalama hane büyüklüğünün en fazla olduğu illerin aynı kümede yer alarak, diğer toplumsal cinsiyet göstergeleri bakımından en avantajsız kümede yer almaktadır.

2019 yılında geniş aile olarak tanımlanan ve en az bir çekirdek aile ve diğer kişilerden oluşan hanehalklarının oranının en yüksek olduğu il, Şırnak'tır. Şırnak ilini Hakkari ve Batman izlemektedir. Benzer şekilde en az bir çekirdek aile ve diğer kişilerden oluşan hanehalklarının en fazla olduğu iller aynı kümede yer alarak, diğer toplumsal cinsiyet göstergeleri bakımından en avantajsız kümede yer almaktadır.

Son yıllarda nüfusun yapısını belirleyen önemli tartışma konularından biri de yabancı nüfustur. Toplumsal yapı ve toplumsal cinsiyet ile ilgili elde edilen sonuçların yorumlanmasında ele alınması gereken bir göstergedir.

ADNKS sonuçlarına göre; yabancı kadın nüfus oranı benzerlik gösteren iller, 81 İl Toplumsal Cinsiyet Küme Haritasında aynı kümede yer almaktadır.

Sonuç olarak bu çalışma ile Türkiye'nin toplumsal cinsiyet açısından bölgesel farklılık gösterdiği istatistiksel olarak ortaya konulmaktadır. Kadına atfedilen roller, kadının toplumdaki algılanma biçimi, maruz ya da mahrum kaldığı alanlar 81 ilin benzerlik ve farklılıklarından yola çıkarak istatistiksel olarak ele alınmıştır.

Çalışmanın sonuçlarının, bölgesel düzeyde, toplumsal yapıya ve toplumsal cinsiyete dikkat çekmesi, toplumsal cinsiyet eşitliğine yönelik bölgesel farkındalık oluşmasına katkı sağlaması, karar vericiler ve politika yapıcılar için yararlı olması beklenmektedir. Ayrıca toplumsal yapı ve toplumsal cinsiyet için seçilen göstergelerin yıllar itibarıyla değişim gösterip göstermediğini ya da nasıl değişim gösterdiklerini konu alan benzer çalışmalar daha detaylı bir şekilde yapılabilecektir. İlerleyen çalışmalar verinin il-ilçe-mahalle düzeyinde elde edilebilmesi halinde il, ilçe hatta mahalle düzeyinde tespitler ve iyileştirilmeler yapılmasına olanak sağlayacaktır. Aynı zamanda mevcut veri setine 81 il düzeyinde ilave edilebilecek gösterge sayısı artırılarak detaylı çalışmalar yapılabilecektir.

KAYNAKÇA

Berktaş, F. (2011). Feminist teoride açılımlar, toplumsal cinsiyet çalışmaları. Eskişehir: Anadolu Üniversitesi Yayını, 2-23.

Chandola, V., Banerjee, A., Kumar, V., (2009). Anomaly Detection : A Survey, ACM Computing Surveys, University of Minnesota, 1, 72.

Ekici, E. (2012). Farklı sınıflandırma yöntemlerinin karşılaştırılması ve bir uygulama, Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ, 13-19.

Gökçin, G.Merve, (2019). Çimento ve Taş Kömürü örnek sektörlerinde stok miktarı tahmini için sınıflama tipi modeller ile çoklu regresyon modellerinin karşılaştırılması, Doktora Tezi, Gazi Ün. Sosyal Bilimler Enstitüsü, Ankara, 84-86.

OECD, (2019). Social Institutions and Gender Index 2019 [Veri Seti]. Erişim adresi: <https://stats.oecd.org/Index.aspx?DataSetCode=SIGI2019>

Oktay, G.Merve, (2013). Aylık sanayi üretim verisinin veri madenciliği ile incelenmesi ve model performanslarının değerlendirilmesi, Türkiye İstatistik Kurumu, Uzmanlık Tezi, Ankara, 58-59.

Pehlivan, G. (2006). CHAID analizi ve bir uygulama, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 17.

Sarp, Ü. (2008). Toplumsal cinsiyet eşitliği, T.C. Başbakanlık Kadının Statüsü Genel Müdürlüğü, Ankara, 6-8.

SGK, (2019). İşyeri, Sigortalı, Aylık ve Gelir Alanlara Ait İstatistikler [Veri seti]. Erişim adresi: <http://www.sgk.gov.tr/wps/portal/sgk/tr/kurumsal/istatistik>

TÜİK, (2019). Toplumsal yapı ve cinsiyet istatistikleri [Veri seti]. Erişim adresi: <https://data.tuik.gov.tr/Kategori/GetKategori?p=nufus-ve-demografi-109&dil=1>

TBMM, (2019). Türkiye Büyük Millet Meclisi Milletvekilleri Dağılımı [Veri seti]. Erişim adresi: https://www.tbmm.gov.tr/develop/owa/milletvekillerimiz_sd.dagilim

WEF, (2019). 2019 Küresel Toplumsal Cinsiyet Uçurumu Raporu. Erişim adresi: <https://www.weforum.org/reports/gender-gap-2020-report-100-years-pay-equality>