Journal of the Faculty of Engineering and Architecture of Gazi University 32:4 (2017) 1301-1310

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University

Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Kişisel verilerin korunmasında öznitelik tabanlı gizlilik etki değerlendirmesi yöntemi

Hidayet Takçı[1], Pelin Canbay[2]*

[1]Cumhuriyet Üniversitesi, Bilgisayar Mühendisliği Bölümü, Sivas, 58800, Türkiye
[2]Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü, Ankara, 06800, Türkiye

ÖNEÇIKANLAR

- Kişisel verilerin korunmasına yönelik Gizlilik Etki Değerlendirmesi yöntemi
- Gizliliği koruyan teknolojiler
- Veri homojenliği ile hassas veri ölçümü

Makale Bilgileri
Geliş: 17.08.2016
Kabul: 27.05.2017

DOI:
10.17341/gazimmfd.369733

Anahtar Kelimeler:
Gizlilik,
homojenlik,
hassasiyet,
gruplama

ÖZET

Kişisel veriler öncelikli olarak korunması gereken hassas bilgi varlıklarıdır. Bugüne kadar kişisel verilerin korunabilmesi için gizliliği koruyan kurallar, rehberler ve tasarımlar geliştirilmiştir. Özellikle son zamanlarda Gizlilik Etki Değerlendirmesi yöntemleri Avrupa ülkelerinde büyüyen bir ilgiyle geliştirilmektedir. Bununla birlikte gelişen bilgi teknolojileri bu düzenlemeleri yetersiz bırakmaktadır. Bu çalışmada kişisel verilerin korunması amacıyla öznitelik tabanlı yeni bir Gizlilik Etki Değerlendirmesi yöntemi önerilmektedir. Çalışma, kişisel verilerin korunması alanında genel yaklaşım olan verilerin bütününü değerlendirmek yerine öznitelik bazında veri setinin gizlilik etkisini değerlendirmeye dayalıdır. Öznitelik bazında hesaplamalar ile kişisel verilerin daha hassas ve gizli kalması gereken bölümleri belirlenebilecek ve gizlenebilecektir. Gizlilik etki değerlendirme hesaplamaları için veri homojenliği yöntemi tercih edilmiştir. Çalışmanın çıktısı gizlilik etkisine göre gruplanmış veri öğeleridir. Önerimize göre daha homojen veri daha hassas veridir ve gizliliği daha önemlidir. Önerilen yöntem iki farklı veri kümesi üzerinde test edilmiş ve elde edilen sonuçlar analiz edilmiştir. Çalışmamızın en önemli bulgusu gizli görünmeyen niteliklerin nitelik birleştirme sonrası gizli olabilmesidir.

Attribute based privacy impact assessment method for the protection of personal data

HIGHLIGHTS

- Privacy Impact Assessment methods for the protection of personal data
- Technologies protecting privacy
- Sensitive data measurement with data homogeneity

Article Info
Received: 17.08.2016
Accepted: 27.05.2017

DOI:
10.17341/gazimmfd.369733

Keywords:
Privacy,
homogeneity,
sensibility,
grouping

ABSTRACT

Personal data is sensitive information asset primarily needed to be protected. In order to protect personal data, privacy-protected rules, designs, guidelines, and legal arrangements have been developed so far. Especially, Privacy Impact Assessment methods have been developed with a growing interest in European countries. However, developing information technologies leave these studies insufficient. In this work, a new feature based Privacy Impact Assessment method is proposed for the purpose of protection of personal data. This study focuses on evaluating the privacy impact of data set at attribute level instead of evaluating all of the data which is a general approach to protect data. With the help of calculation at feature level, more sensitive and private personal data parts can be defined and hidden. Data homogeneity method is preferred for privacy impact evaluation calculations. The outcome of this work is data items grouped by privacy impact. According to our proposal, more homogeneous data is more sensitive and its privacy is important. The proposed method is tested on two different data set and the obtained results are analyzed. The most important finding of our work is that attributes that do not appear to be private can be private after combining attributes.

*Sorumlu Yazar/Corresponding Author: pelin@cs.hacettepe.edu.tr / Tel: +90 312 780 7527

1. GİRİŞ (INTRODUCTION)

Gizlilik, bireysel bir hak olan "kişiliğin korunması" ile ilgili olup "bireyin yalnız kalma özgürlüğü" olarak tanımlanmıştır [1]. ABD yasalarına göre iki türde ele alınan gizliliğin ilk türü anayasal gizliliktir ve konusu kişinin kararlarını başkasının etkisi altında kalmadan verebilmesiyle ilgilidir. Diğer tür gizlilik ise kişilerin kendilerine özel bilgilere erişimi başkalarına kısıtlaması ile ilgilidir [2]. Kişisel veriler sadece isim, soy isim veya kimlik numarası değil aynı zamanda e-posta adresi, kişisel bilgisayarların IP numarası, kişilerin alışveriş veya eğlence örüntüleri vb. bilgilerdir. Kişisel verilerin son dönemde bir ticari mal olarak görülmesi dolayısıyla korunmasına eskisinden daha fazla ihtiyaç vardır. Gizliliğin korunması başta yasal düzenlemeler olmak üzere birçok açıdan yıllardır araştırmacılar tarafından ele alınmaktadır [3]. Yasal düzenlemeler toplanan verilerin işlenme amacının belirtilmesi, kullanımın sınırları, toplanan verilerin araştırma sonrası tutulup tutulmayacağı gibi bilgilerin net olmasını zorunlu kılmaktadır. Gelişen bilgi teknolojilerinin bir sonucu olarak kişisel bilgiler; zarar görme, eşitsizlik, ayrımcılık ve özerklik açılarından tehdit altındadır. Örneğin, bulut bilişimle ilgili son gelişmeler gizlilik sorunlarına farklı yaklaşımlar getirmiştir [4]. Bilgi teknolojilerindeki artan gizlilik problemlerinin nedeni bu teknolojilerin geliştirilmesi aşamasında güvenlik ve gizlilik konusunun yeterince planlanmamış olmasıdır. Bu durumun bir sonucu olarak, IT servisleri IT ürünlerinden çok gizlilik sorunlarıyla meşguldür [5]. Sorunun bir parçası olan bilgi teknolojileri çözüm için de olanaklar sunmaktadır. Bu bağlamda "değer duyarlı tasarım" teknoloji tasarımında insan haklarına duyarlı tasarımı destekler [6]. Değer 'gizlilik' olabilir ve değer duyarlı tasarım gizlilik dostu IT sistemlerinin tasarımı için bir metot olarak kullanılabilir. Gizlilik duyarlı tasarım yaklaşımları özellikle kurumlar arası işbirliklerinde sorumluluk ve güçlü iş uygulamaları gelişimi bakımından ilgiyle ele alınmaktadır [7]. Gizliliğe duyarlı olabilmek için eldeki verilerin veya çalışmanın gizliliğe etkisini ölçebilmek değerlendirebilmek ve çözüm alternatifleri sunabilmek gerekmektedir. Öncelikli olarak kişisel verilerin gizliliğinin korunması kaçınılmaz bir ihtiyaç olmakla beraber bu verilerin topluma faydalı üretimler için kullanılması ve kurumlar arası paylaşılması da bir o kadar ihtiyaç dâhilindedir. Bu kapsamda çalışmamızda söz konusu kişisel verilerin gizlilik etki değerinin hesaplanması ve hesaba dayalı olarak paylaşılabilecek ve paylaşılamayacak verilerin tanımlanması üzerinde durulmuştur.

Çalışmamız kişisel verilerin korunması alanında yapılan birçok çalışma gibi verilerin bütününü değerlendirmek yerine öznitelik bazında veri setinin gizlilik etkisini değerlendirmeye dayalıdır. Böylece kişisel verilerin daha hassas ve gizli kalması gereken bölümleri belirlenebilecek ve değerlendirilebilecektir. Böylece fiziksel anlamda olmasa dahi uygulama anlamında gizliliğin belirlenmesi ve gizli kalması gereken veri parçalarının korunması imkânı sağlanacaktır. Ele aldığımız probleme yönelik geçmiş çalışmalar ve çalışmamız ile bağlantısı Bölüm 2'de ele alınmaktadır. Önermiş olduğumuz kişisel verilerin korunmasında öznitelik tabanlı gizlilik etki değerlendirmesi yöntemi Bölüm 3'te ayrıntılı olarak açıklanmaktadır. İki farklı veri kümesi kullanılarak yapılan uygulamalar Bölüm 4'te detaylandırılmış ve sonuçlar Bölüm 5'te analiz edilmiştir.

2. GİZLİLİK KORUMA TEKNOLOJİLERİ (PRIVACY PROTECTION TECHNOLOGIES)

Gizliliğe dayalı tasarım, sistemlerin kurulumunda gizliliğe dayalı problemleri oluşmadan engellemeyi amaçlar. Uygulamaların gizliliği üzerindeki potansiyel etkileri değerlendirmek için Gizlilik Etki Değerlendirmesi yaklaşımları artan bir ilgiyle gelişmektedir [8]. Gizlilik Etki Değerlendirmesi, bir proje, politika, sistem, program, servis, ürün veya diğer girişimlerin gizliliğe etkisini değerlendirmek ve paydaşlara danışarak olumsuz etkilerden kaçınma veya bu etkileri minimum düzeye indirmede gerektiğinde iyileştirici adımlar atmak için geliştirilen bir yöntembilimi olarak tanımlanmaktadır [9]. Gizlilik etki değerlendirmesi meydana gelebilecek gizlilik problemlerinin tespit edilmesinde ve problemler ortaya çıkmadan gerekli tedbirlerin alınmasında yeni bir ürün, servis veya teknoloji geliştiricilere bir rehber niteliğindedir. Sadece kişisel veya kurumsal gizliliğe değil aynı zamanda ahlaki değerlere etkilerin de değerlendirildiği çalışmalara olan ilgi ve ihtiyaç gün geçtikçe artmaktadır [10]. Son zamanlarda hızla yaygınlaşan akıllı şehir uygulamalarında da gizlilik etki değerlendirmesi çalışmaları kullanıcıların kişisel gizliliğini önemseyen kurumlar tarafından ön planda tutulmaktadır [11]. Önemine binaen gizlilik etki değerlendirmesi çalışmaları birçok ülkede ülke çapında standartlaşıp geliştirilme yolunda ilerlemektedir [12].

Çalışmamızda gizlilik etki değerlendirmesi kişisel veriler üzerinden ele alınmakta olup söz konusu verilerin gizlilik etki değerleri hesaplanarak korunma gereksinimleri üzerine odaklanılmıştır. Gizliliği koruyan bilişim sistemleri tasarlamak için çeşitli endüstri rehberleri de oluşturulmuştur. Bu rehberler bankacılık ve perakende sektörü gibi gizlilik ve güvenlik duyarlı sistemler için kurallar sağlar. Adil Bilgi Uygulamaları tabanlı Avrupa Birliği veri koruma direktifinin sağladığı prensipler [13] tasarım prensipleri olarak varsayılabilir. Kurallar ve rehberler ile tasarlanan sistemler prensipte Avrupa birliği gizlilik yasaları ile uyumlu olmalı ve kullanıcıların gizliliğine saygılı olmalıdır. Gizlilik bağlamında ortaya konan rehber ve standartlar kapsayıcı olmasına rağmen dinamik bir doğası olan bilişim sistemlerinde yetersiz kalabilmektedir.

Bir proje, politika, sistem, program, servis, ürün veya diğer girişimlerin uygulanmasında verilerden faydalanılmaktadır. Hedef kitle bireyler olduğundan bilimsel ve ticari çalışmaların gelişimi için kişisel veriler kullanılmaktadır. Gizlilik koruma yaklaşımları verilerin gizliliğini korumayı hedeflerken, aynı zamanda verilerden alınacak faydayı da maksimum düzeyde tutmaya çalışmaktadır. Çünkü veri

faydası ve veri gizliliği arasında zıt bir ilişki söz konusudur. Örneğin; bir veri kümesini herhangi bir gizlilik koruma işlemine tabi tutmadan olduğu gibi kullanmak o veri kümesinden alınacak faydayı maksimum yapacaktır, fakat veri kümesi saldırılara ve her türlü gizlilik ifşasına açık hale gelecektir. Aynı durumun tersi de söz konusudur. Yani bir veri kümesi tamamen gizlenir ve hiçbir bilgi kullanılmaz ise o veri kümesinden herhangi bir tespit veya ifşa işlemi yapılamayacak, dolayısı ile güvenlik ve gizlilik maksimum düzeyde korunurken veri faydası minimum düzeye inecektir. Veri faydası ve veri gizliliği arasındaki dengeyi ideal seviyede tutmak amacı ile veri kümelerinin anonimleştirilmesi üzerine odaklanılmıştır [14, 15]. Kişisel veri kümelerinde özniteliklerin belirli değer aralıklarında genelleştirilmesi ile verinin ait olduğu kişi anonimleştirilerek gizliliği sağlanmaktadır, örneğin yaş bilgisi 37 olan bir kişinin veri kümesindeki yaş bilgisi 30-40 aralığında genelleştirilerek kişi anonimleştirilmeye çalışılır. Veri gizliliğini veri faydası gözetilerek arttırabilmek adına çeşitli veri madenciliği tekniklerinden de faydalanılmıştır [16, 17]. Paylaşılacak verilerin veya belgelerin gizliliğini koruma anlamında başta şifreleme [18] olmak üzere, kimlik yönetimi [19], yetki denetimi [20] gibi birçok çalışma hala yürütülmektedir. Fakat paylaşılacak kişisel verilerin gizliliğini koruma anlamında geleneksel yöntemlerde olduğu gibi verinin tamamını gizlemek veya şifrelemek yerine veri setindeki gizli kalması gereken kısımların gizlenmesi ile veriden elde edilecek faydanın arttırılması çalışmamızın amacıdır. Nitelik bazında gizleme ham veri seviyesinde olamasa bile uygulama seviyesinde mümkün olabilecektir. Anonimleştirme ve veri gizliliği çalışmalarındaki temel sorun hangi özniteliklerin ne kadar gizlenmesi gerektiğidir. Çalışmamız bu sorunun çözümüne odaklanmaktadır. Kişisel veriler içeren bir veri kümesinde hangi niteliklerin daha ayırt edici olduğu ve dolayısı ile hangilerinin gizlenmeye diğerlerinden daha çok gereksinim duyduğu yapılan çalışmalar sonucu belirlenebilmektedir. Çalışmamızda kişisel veriler içeren iki ayrı veri kümesi üzerinden işlemler yapılarak hangi özniteliklerin daha gizli olması gerektiği, hangilerinin gizli olmasa da olabileceği ve hangilerinin de gizli kalsa daha iyi olacağı tarzından bilgiler elde edilecektir.

3. ÖZNİTELİK TABANLI GİZLİLİK ETKİ DEĞERLENDİRMESİ YÖNTEMİ
(ATTRIBUTE-BASED PRIVACY IMPACT ASSESSMENT METHOD)

Önerdiğimiz gizlilik etki değerlendirmesi modeli kişisel verilerin gizliliğini öznitelik bazında hesaplamaya dayalıdır. Hesaplama işleminde veri homojenliği bilgisi kullanılacaktır. İşlem sonunda her bir nitelik gizlilik derecesine uygun olarak etiketlenecek ve veriyi korumakla yükümlü kişilere etiketlere dayalı olarak verilerin korunması konusunda öneri yapma şansı olabilecektir. Kişisel veriler çoğu zaman veri tabanlarında ve ilişkisel modele uygun olarak tutulurlar. Bu nedenle önerimizin detayları ilişkisel veri modeli kavramları yardımıyla anlatılacaktır. İlişkisel veri modeli iki boyutlu tablolar ile tarif edilir. Her bir tablo bir ilişki ve tabloda yer alan her bir satır da tuple olarak isimlendirilir. Bu arada her bir satır sütunlardan meydana gelir ve bu sütunlara da öznitelik adı verilir. Özniteliklerin yer aldığı sütunlar nitelik değerleriyle doldurulur. Nitelik değerlerinin alındığı veri havuzlarına etki alanı adı verilir. m, bir ilişkideki öznitelik sayısını vermek üzere A_1, A_2, ..., A_m o ilişkideki nitelikleri sunar. R ile bir ilişkiyi sunmak istediğimizde $R(A_1, A_2, ..., A_m)$ şeklinde sunabiliriz. Veri modeline başlık yerine içerik açısından baktığımızda bir ilişki satırların bir koleksiyonundan meydana gelir. Bu türden tanımlamada; n adet kayıt içeren bir ilişki $R=\{t_1, t_2,\ldots,t_n\}$ şeklinde sunulur. Her tuple t, m adet nitelik değerinden oluşan sıralı bir listedir ve $t=<v_1,v_2, \ldots,v_m>$ gibi gösterilir. $dom(A_i)$ i. nci niteliğin veri havuzu olmak üzere her v_i bu havuzun bir elemanıdır ya da boş bir değerdir. Etki alanlarına dayalı olarak bir ilişki Eş. 1'de gösterilen Kartezyen çarpım yardımıyla sunulabilir:

$$R=(dom(A_1)Xdom(A_2)X\ldots dom(A_m)),\quad 1 <= |\, dom(A_i)\,| <= n \tag{1}$$

Her bir etki alanı $dom(A_i)$ farklı sayıda değer içerir. Örneğin, İSİM niteliği için olası değer sayısı ile SOYİSİM niteliği için olası değer sayısı aynı değildir. Bir niteliğin daha fazla farklı içerikte değer almasına etki alanı zenginliği diyecek olursak; her bir nitelik için etki alanı zenginliği aynı değildir. Etki alanı bakımından zengin olan nitelikler hemen hemen her farklı nesne için farklı bir değerle eşleşir ve bu da o niteliği daha gizlilik açısından daha hassas hale getirir. Etki alanı bakımından zayıf olan nitelikler ise birden fazla nesne için tek bir değerle eşleşir ve bu durum o niteliğin daha az gizli olması gerektiği anlamına gelir, cinsiyet bilgisinde olduğu gibi. Dolayısıyla bir niteliği oluşturan değerlerin dağılımından o niteliğin gizlenmesi gerekip gerekmediği anlaşılabilir. Örneğin, bir vatandaşa ait kişisel bilgiler Tablo 1'deki gibi olsun. Biz bu kişiye ait bilgilerden sadece kimlik numarası bilgisi (TCKNO) ile o kişiyi doğrudan bulabiliriz, çünkü kimlik numarası tekildir fakat "Ebru" ismiyle veya "Ebru Yılmaz" {isim, soyisim} ikilisi ile ilgili şahsı bulmamız kimi zaman mümkün değildir çünkü aynı isimde birçok kişi olabilir. TCKNO her bir kayıt için farklıdır ve kişiyi eşsiz olarak bulmada kullanılabilmektedir yani gizlilik derecesi yüksek bir niteliktir. İSİM ve SOY İSİM ise birden çok kişi ile eşleşebilir, yani eşsiz değildir ve bundan dolayı da gizlilik derecesi daha düşüktür. Az sayıda kayıtta rastlanan veriler daha fazla gizlenmeli çok sayıda kayıtta rastlanan veriler ise daha az gizlenmelidir. Seyrek rastlanan verinin daha gizli sık rastlanan verinin daha az gizli olmasından yola çıkılarak bu çalışmada etki alanı zenginliği ölçümü ortaya konmuştur. Bu ölçüm ile nitelikler "Gizlenmelidir" veya "Gizlenmesine gerek yoktur" şeklinde

Tablo 1. Kişisel kayıt örneği (Sample of the personal record)

TCKNO	İSİM	SOYİSİM	YAŞ	CİNSİYET	ŞEHİR	MHAL
5472157002	Ebru	Yılmaz	23	Kadın	Bolu	Bekâr

etiketlenebilecektir. Bununla birlikte bazen de sık rastlanan ve dolayısıyla önerimize göre gizli görünmeyen niteliklerin ikili veya üçlü grupları bir araya gelerek gizli veriyi meydana getirebilirler. Çalışmamız her iki durum için de deneysel çalışmalar içermektedir.

3.1. Etki Alanı Zenginliği (The Richness of The Domain)

Etki alanı zenginliği Tablo 2'deki notasyona uygun olarak tanımlanacaktır. Etki alanı zenginliği hesaplamanın ilk aşaması her bir nitelikte birbirinden ayrık şekilde ortaya çıkan nitelik değerlerinin bütün veri setine oranının bulunmasıdır. Bu değere kullanım oranı adı vermek gerekirse; i.nci nitelikteki j.nci değer için kullanım oranı değeri $K_{ij} = |v_{ij}|/n$ şeklinde hesap edilebilecektir. Bu bilgi nitelik değerlerinin basit olasılığı gibi de görülebilir. Durumu örneklemek gerekirse; i.nci nitelik İSİM niteliği j.nci değer ise "Ali" olmak üzere K_{ij} değeri veri setinde İSİM niteliğinde "Ali" bilgisi bulunan kayıtlarının oranını verecektir. Aynı şekilde birbirinden farklı d_i adet isim için de aynı hesap işlemi yapılacaktır. Bu hesaplama bütün nitelik değerleri için elde edildikten sonra bu değerlerin homojenlik dağılımından K_i değeri elde edilir. K_i değeri artık kullanım oranı değil etki alanı zenginliği değeridir. Bu değerin elde edilmesinde homojenlik ölçümlerinden [22] Gini tercih edilmiştir. Gini yöntemine göre A_i niteliği için homojenlik değeri veya etki alanı zenginliği değeri Eş. 2 formülü ile bulunur.

$$K_i = 1 - \sum_{s=1}^{d_i} (K_{ij})^2 \quad (2)$$

K_i değerleri bütün nitelikler için elde edildikten sonra K_i değerleri üzerinde ayrıklaştırma işlemi yapılacaktır. Burada ayrıklaştırma işlemi basit olarak yerine getirilmiş ve sayısal değer kategorik değerlere dönüştürülmüştür. Önce maksimum K_i (K_{max}) ve minimum K_i (K_{min}) değerleri elde edilmiş ardından bu iki değer arasındaki fark (K_r) elde edilmiştir. Elde edilen bu bilgiler yardımıyla bütün nitelikler gizlilik derecelerine göre kategorik olarak etiketlenebilmiştir. Etiketleme algoritması aşağıda verilmektedir. Etki alanı zenginliğine göre etiketleme algoritması

-if ($K_{min} <= K_i < K_{min} + K_r/3$) then {Ai "Gizlenmesine gerek yoktur"}
-else if ($K_{min} + K_r/3 <= K_i < K_{min} + 2K_r/3$) then {Ai "Düşük seviyeli gizleme"}
-else {Ai "Gizlenmelidir"}

3.2. Nitelik Çiftleri için Gizlilik Hesabı (Privacy Assessment for The Attribute Pairs)

Önerdiğimiz gizlilik değerlendirme yöntemine göre nitelikler; "Gizlenmesine gerek yoktur", "Düşük seviyeli gizleme" ve "Gizlenmelidir" şeklinde etiketlenecektir. Bununla birlikte gruplandırmada hata yapılıp yapılmadığının kontrolü için daha hassas analiz yapılması bir ihtiyaçtır. Daha hassas analizin yolu "Düşük seviyeli gizleme" ve "Gizlenmesine gerek yoktur" şeklinde etiketlenen niteliklerin ikili olarak yeniden ele alınmasıdır. Dolayısıyla ilk test sonucuna göre elde edilen etiketler yardımıyla bazı nitelikler ikili olarak gruplanır ve test işlemine alınır. Nitelik birleştirme yapılırken tek bir niteliğin gizlilik değerine değil her iki niteliğin gizlilik değerine bakılır. Yani, bir nitelik değerinin bireysel olarak meydana gelme sayısı yerine nitelik çiftlerinin birlikte meydana gelme olasılıklarına göre işlem yapılır.

3.2.1. Nitelik çiftlerinin birliktelik kuralları madenciliği ile elde edilmesi (Obtaining of the attribute pairs using association rule mining)

Nitelik birleştirmede hangi niteliklerin eşleştirileceği konusu özellikle çok boyutlu veri setleri için karmaşık bir problemdir. Bu nedenle nitelik çiftlerinin tespitinde birliktelik kuralları madenciliğinden faydalanılacaktır. Önce gizlenmesine gerek olmayan niteliklerden tek elemanlı nitelikler (1-items) bulunacak ve ardından birliktelik kuralları madenciliğine göre iki elemanlı gizli olmasına gerek olmayan nitelikler (2-items) elde edilecektir. 2-items için destek değerleri $|v_{ij}|$ olarak kullanılacaktır. Böylece, gizli olmasına gerek olmayan niteliklerin bir araya gelmesi ile gizli olması gereken bir nitelik çifti meydana gelip gelmediği görülecektir.

3.2.2. Birliktelik kuralları madenciliği (Association rule mining)

Birliktelik kuralları madenciliği iki aşamadan oluşur. Bu aşamalardan ilki eleman kümesi üretimi diğeri ise kural üretimidir. Çalışmamızda bu aşamalardan sadece ilki kullanılacaktır. Bir sepet mantığında gizli olmayan ve gizli olabilecek nitelikler sepete atılacak ve ardından aralarındaki eşleşmeler elde edilecektir. Bir veri setinde yer alabilecek olası bütün eleman kümeleri için Apriori algoritması aşağıda görülmektedir.

Apriori algoritması
-k=1

Tablo 2. Etki alanı zenginliği ölçümü için notasyon (Notation for the measurement of the richness of the domain)

Parametre	Açıklama
A_i	i.nci nitelik
$dom(A_i)$	i.nci niteliğin değer aldığı etki alanı
d_i	i.nci nitelik için olası birbirinden farklı nitelik değeri adedi
J	i.nci nitelik için olası birbirinden farklı nitelik değerlerinin indisi $\{1,2,\ldots,d_i\}$
v_{ij}	i.nci nitelik için $dom(A_i)$ etki alanından alınan j.nci nitelik değeri
$\|v_{ij}\|$	i.nci nitelikteki, j.nci farklı değerin tekrar adedi
N	Veri seti örnek adedi

-Boyutu 1 olan eleman kümeleri üret
-Yeni hiçbir sık eleman kümesi bulunmayana kadar tekrar et
-k boyutlu sık eleman kümelerinden (k+1) boyutlu aday eleman kümeleri üret
-Aday eleman kümeler eğer yeteri kadar sık değilse sistemden çıkar
-Veri setini tarayarak her bir aday eleman kümesi için destek değerini hesapla
-Sık olmayan adayları ele ve sadece sık olanlarla devam et

3.3. Önerdiğimiz Yöntem ve Gizlilik Etki Değerlendirmesi (Our Proposed Method and Privacy Impact Assessment)

Gizlilik etki değerlemesi gizliliğe dayalı tasarımla ilgili geniş bir çerçeveye işaret etmektedir. Ayrıca gizliliğe dayalı risklerin ortaya çıkarılması ve riskli durumun ortadan kaldırılması konularıyla yoğun olarak ilgilenmektedir. Her ne kadar gizlilik etki değerlemesi düzenlemeler ve kurallar bütünü olarak karşımıza çıksa da gizlilik etki değerlemesinde önemli olan nesnelerin değerinin elde edilmesi ve bu değere dayalı olarak bir tasarım ortaya konulmasıdır. Çalışmamızda ortaya koymaya gayret ettiğimiz konu da bu yönüyle gizlilik etki değerlemesine hizmet etmektedir. Diğer yöntemlerin daha büyük parçalar üzerinde yaptığı çalışmalar önerimizde daha küçük parçalar için ele alınmaktadır. Çalışmamız verinin bütünü üzerinde değil de parçaları üzerinde yapılacak bir gizlilik derecelendirmesini desteklemekte ve böylece daha hassas şekilde kişisel verinin korunmasına yardımcı olmaktadır. Dolayısıyla, çalışmamızın nitelik seviyesinde gizlilik değeri elde etme özelliği gizlilik etki değerlemesine nitelik detayında dahi değerli bilginin elde edilmesi imkânını verecektir. Bunun doğal bir sonucu da diğer yöntemlerle gözden kaçma ihtimali olan hassasiyette çalışma imkânı olacaktır. Bir diğer konu ise şudur. Önerimiz bilişim sistemi yöneticilerine tavsiye getirmek üzerine ortaya konmuştur. Nitelikler gizlilik derecesine göre hesap edildikten sonra bilgi işlem yöneticilerine bu bilgiler sağlanacak ve onların uygulama seviyesinde gizleme yapmalarına olanak sağlanacaktır. Hatta konu veritabanı yöneticisi ve veritabanı programcılarıyla da ilgilidir. Gizli kalması gereken nitelik ortaya çıkarıldıktan sonra bu niteliğin sistem seviyesinde anonimleştirilmesi, şifrelenmesi ve benzer bir teknik düşünülebilir. Önerimiz gizli kalması gereken niteliğin ne olduğunu tespitte görev yapacak olup diğer işlemler başka tekniklerle yerine getirilecektir. Ayrıca, sistem seviyesinde gizleme yapılmasa dahi uygulama seviyesinde çözüm düşünülebilir. Yıllardır güvenlik maksadıyla kullanılan veritabanının görünümü ile çalışma burada da düşünülebilir. Gizli kalması gereken veriler halka açık uygulamalarda sunulmayarak da gizliliğe uygun çalışma yapılmış olacaktır. Çalışmamız gizlilik etki değerlendirmesi yöntemine bir alternatif olmaktan çok onu destekleyen ve onu daha hassas şekilde uygulamaya izin veren bir yöntem sunmaktadır.

4. DENEYSEL ÇALIŞMA (EXPERIMENTAL STUDY)

Öznitelik bazında gizlilik değerlendirme iki farklı veri seti üzerinde test edilecektir. Bu veri setlerinden birisi TÜİK istatistiklerine dayalı olarak türetilmiş personel verileri diğeri ise University of California tarafından sağlanan yetişkinler veri setidir. Bu iki veri seti de kişisel veriler içermektedir. Deneysel çalışmanın çıktısı her iki veri seti için de niteliklere ait gizlilik değerleridir. Niteliklerin gizlilik değerlerini bulmada yeni bir yaklaşım olarak etki alanı zenginliği kullanılmış olup etki alanı zenginliği veri homojenliği ile yakından ilgilidir.

4.1. Veri Seti (Dataset)

Modelimizi test etmek için kullandığımız veri setlerinden ilki TÜİK istatistikleri kullanılarak elde edilen veri setidir. Bu veri seti, temsil edici nitelikte olup popülasyonu sunacak karakteristiklere sahiptir. Üretilen veri setinin ortalaması ve standart sapması popülasyon ortalaması ve standart sapmasına oldukça yakındır. Nüfus kayıt veri seti detayları Tablo 3'de kısmi olarak sunulmuştur. Tablo 3'de verilen veri setinde; Türkiye'de en çok kullanılan isim ve soy isimler daha sık yer almakta, ayrıca; cinsiyet ve medeni hal dağılımı gibi bilgiler aslına uygun olarak yer almaktadır. Veri üretimi sırasında Türkiye'de yaşayan kişilerin yaş aralıkları da dikkate alınmıştır [21]. Ancak T.C. Kimlik numarası, gizlilik dikkate alınarak 11 haneli olduğu halde, 10 haneli olarak ele alınmıştır. Veri seti olarak temsil ediciliği yüksek, 300 adet veri üretilmiştir. Veri üretiminde veriyi oluşturan her bir nitelik için en uygun istatistiksel dağılım kullanılmış ve bu dağılımların parametreleri popülasyon verilerinden elde edilmiştir. Veri üretiminden sonra tekrar doğrulama yapılmış ve elde edilen verilerin popülasyona uygunluğu test edilmiştir. Örnek verilerden bir kısmı Tablo 4'de görülmektedir. Deneylerde kullanılan diğer bir veri seti UCI makine öğrenmesi veri deposundan [22] elde edilen yetişkinler veri setidir. Yetişkinler veri setinde; meslek, eğitim, gelir, ülke, cinsiyet, ırk, yaş gibi alanları içeren yirmi binin üzerinde kayıt mevcuttur. Yetişkinler veri setinde çok sayıda boş değer olduğu için deneyler öncesinde önişlem yapılmıştır. Kayıp değerlerle mücadelede kullanılan yöntemlerden; kayıp değerlerin bulunduğu satırların veri setinden ayrılması kullanılarak veri temizleme yapılmıştır. Tablo 5 yetişkinler veri seti detaylarını göstermektedir. Bu gösterimde verilerin aldığı örnek değerler de yer almaktadır.

Tablo 3. Nüfus kayıt veri seti özellikleri (Population registration dataset properties)

Alan Adı	Veri Tipi	Veri
TCKİMLİKNO	Sürekli	11 rakamdan oluşan milyonlarca bilgi
İSİM	Kategorik	Ahmet, Fatma, Mehmet, Zeynep, …
SOYİSİM	Kategorik	Demir, Yıldırım, Kaya, Çelik, …
YAŞ	Sürekli	18, 23, 27, 32, 41, 48, 55, 38, 29, …
CİNSİYET	Kategorik	Kadın, Erkek
MEDENİDURUM	Kategorik	Bekâr, Evli, Dul
MEMLEKET	Kategorik	İstanbul, Ankara, İzmir, Rize, …

Tablo 4. Türetilen veri setinden bazı örnek veriler (Some sample data derived from the dataset)

TCKİMLİKNO	İSİM	SOYİSİM	YAŞ	CİNSİYET	MEMLEKET	MEDENİDURUM
4789511209	Ahmet	Demir	18	Erkek	İstanbul	Bekar
2009412019	Serap	Demir	27	Kadın	İzmir	Evli
6434863079	Mehmet	Altan	48	Erkek	Bursa	Evli
7821901254	Metin	Tunca	23	Erkek	Trabzon	Bekar
6921125437	Serpil	Çelik	38	Kadın	Ankara	Evli
…	…	…	…	…	…	…
5329741020	Ahmet	Yıldırım	55	Erkek	Urfa	Evli

Tablo 5. Yetişkinler veri seti öznitelikleri (Adult dataset attributes)

Alan Adı	Veri Tipi	Veri
Yaş	Sürekli	10, 14, 27, 34, 50
Çalışma Alanı	Kategorik	Özel Sektör, Serbest Meslek, Kamu
Son Ağırlık Değeri	Sürekli	77516, 83311, 338409
Eğitim Seviyesi	Kategorik	Lise, Lisans, Yüksek Lisans, Doktora
Okul Sınıfı	Sürekli	13, 9, 8, 11
Medeni Durumu	Kategorik	Bekâr, Evli, Dul
Meslek Bilgisi	Kategorik	Satıcı, Öğretmen, Mühendis, Doktor
Akrabalık Derecesi	Kategorik	Eş, Öz çocuk, Aileden değil, Anne
Irk	Kategorik	Uzak Doğulu, Siyah, Eskimo
Cinsiyet	Kategorik	Erkek, Kadın
Gelir Bilgisi	Sürekli	1200, 2450, 300, 678
Gider Bilgisi	Sürekli	0, 100, 230, 900
Haftalık Çalışma Saati	Sürekli	40, 13, 36
Ülke Bilgisi	Kategorik	Türkiye, İngiltere, Almanya
Sınıf Bilgisi	Kategorik	Sınıf 1, Sınıf 2

4.2. Deneysel Tasarım (Experimental Design)

Çalışmamızda, kişisel verilerin gizliliği kişisel verileri meydana getiren niteliklerin gizliliği üzerinden elde edilmiştir. Bunun için veri seti önce niteliklere ayrılmış ve ardından her bir nitelik için etki alanı zenginliği değeri hesap edilmiştir. Karar ağacı algoritmalarında bilgi kazancı metriği olarak da kullanılan Gini homojenlik ölçümü, çalışmamızda nitelik değerlerinin ne kadar homojen dağıldığını bulmada kullanılmıştır. Her bir niteliğin bir düğüm, birbirinden farklı her bir nitelik değerinin bir sınıf, o nitelik değerine sahip eleman sayısının ise o sınıftaki eleman sayısını verdiği bir kurguda düğüm için elde edilen Gini index değeri etki alanı zenginliği bilgisini verecektir. Veri setindeki bütün nitelikler için etki alanı zenginliği bilgisi hesap edildikten sonra maksimum ve minimum etki alanı zenginliği değeri bulunmuş ve aradaki farktan etki alanı zenginliği için bir aralık bulunmuştur. Maksimum, minimum ve aralık bilgileri kullanılarak her bir nitelik; "Gizlenmeli", "Düşük seviye gizleme" ve "Gizlenmesine gerek yoktur" şeklinde etiketlenmiştir. Bu çalışmanın ardından "Gizlenmesine gerek yoktur" şeklinde etiketlenen nitelikler kendi aralarında eşleştirilerek ikililer için gizlilik değeri ölçülmüş ve ilk bakışta gizli olarak görülmeyen niteliklerin gizli olup olamayacağı test edilmiştir.

4.3. Deneysel Sonuçlar (Experimental Results)

Deneysel tasarımda ayrıntıları verilen etki alanı zenginliğine dayalı etiketleme işlemi önce bireysel nitelikler için ardından nitelik çiftleri için yerine getirilmiştir. Bu sayede hem niteliklerin tek başına gizli olabileceği hem de nitelik çiftleri şeklinde ele alındıklarında gizli olabilecekleri gösterilmek istenmiştir.

4.3.1. Bireysel nitelikler için gizliliğe dayalı etiketleme (Privacy-based tagging for individual attributes)

Gini tabanlı yönteme uygun olarak elde edilen etki alanı zenginliği bilgisi bireysel nitelikler için Tablo 6 ve Tablo 7'de sunulmuştur. Özellikle TÜİK verisinden elde edilen nitelikler için bazı değerler birbirine yakın olsa dahi gizlilik derecelendirmesi bakımından doğru sonuçlara ulaşılabilmiştir. Bu durum da önerimizi doğrulamak anlamında değerli bir bilgi sunmaktadır.

Tablo 6. Nüfus veri seti için birikimli etki alanı zenginlikleri
(Cumulative domain riches for census dataset)

N	Nitelikler (A_i)	K_i
1	TCKİMLİKNO	0,996
2	İSİM	0,980
3	SOYİSİM	0,984
4	YAŞ	0,970
5	CİNSİYET	0,499
6	MEMLEKET	0,963
7	MEDENİDURUM	0,503

Tablo 7. Yetişkinler veri seti için birikimli etki alanı zenginlikleri (Cumulative domain riches for adult dataset)

N	Nitelikler (A_i)	K_i
1	Age	0,978
2	Workclass	0,421
3	Fnlwgt	0,999
4	Education	0,806
5	education_num	0,806
6	marital_status	0,661
7	Occupation	0,888
8	Relationship	0,733
9	Race	0,256
10	Sex	0,443
11	capital_gain	0,157
12	capital_loss	0,092
13	hours_per_week	0,762
14	native_country	0,165

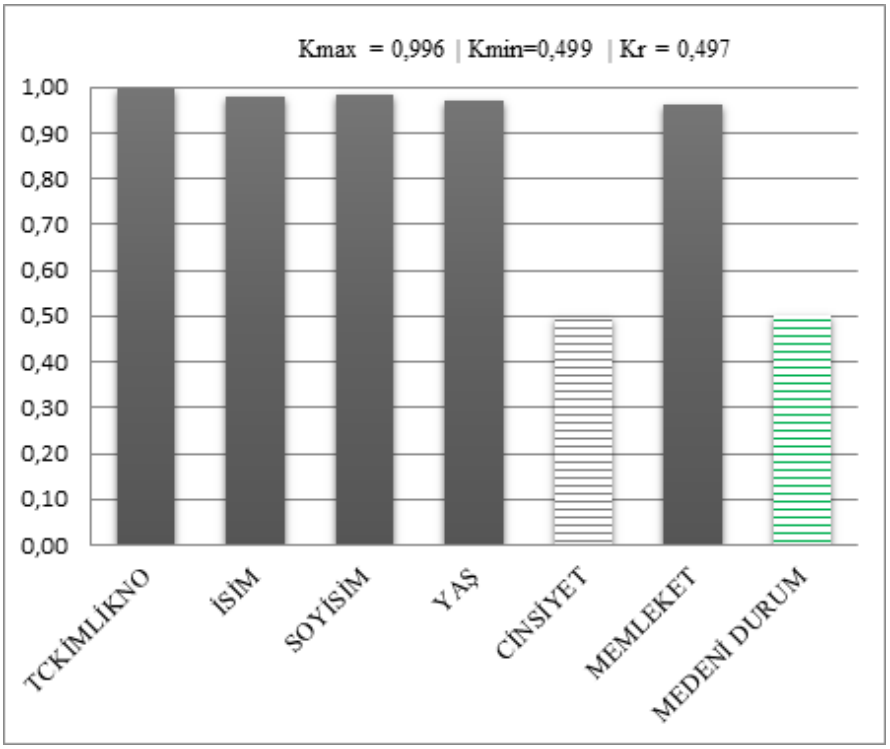
Daha sonra K_i değerleri yardımıyla maksimum etki alanı zenginliği (K_{max}), minimum etki alanı zenginliği (K_{min}) ve aralık (K_r) değerleri bulunmuştur. Bu değerler; K_{max}=0,996, K_{min}=0,499 ve K_r=0,497 olup sınırlar şu şekilde ortaya çıkmıştır.

0,499<=K_i<0,665 değerleri için "Gizlenmesine gerek yoktur",
0,665<=K_i<0,830 değerleri için "Düşük seviyeli gizleme" ve
0,830<=K_i<=0,996 değerleri için "Gizlenmelidir"

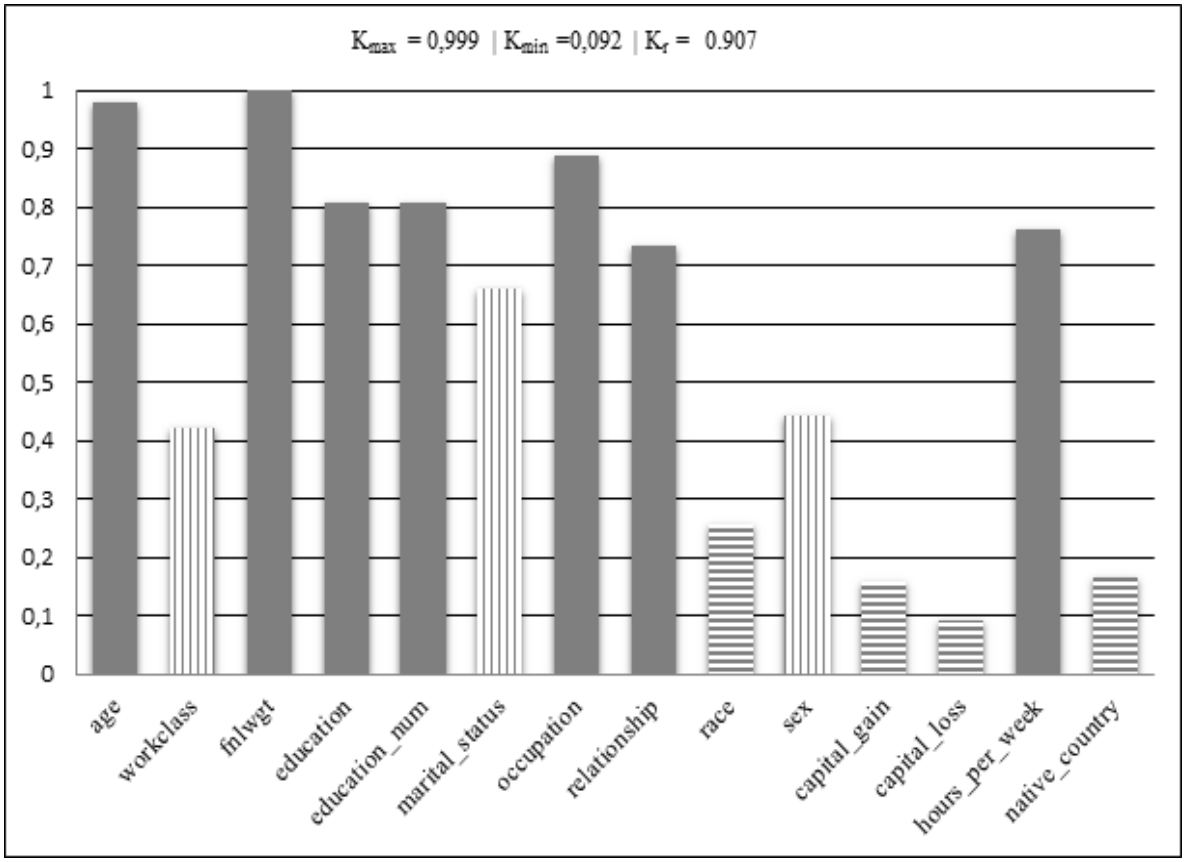
Şekil 1'de gizli kalması gereken nitelikler düz dolgu ile gizli olmayan nitelikler yatay çizgilerle sunulmuştur. Birinci veri seti için yapılan deneyler sonrasında; TCKİMLİKNO, İSİM, SOYİSİM, YAŞ, MEMLEKET gibi bilgiler "Gizlenmelidir"; CİNSİYET ve MEDENİDURUM bilgileri "Gizlenmesine gerek yoktur" şeklinde elde edilmiştir. TÜİK kaynaklı veriler için olduğu gibi yetişkinler veri seti için de aynı işlem yapıldığında aşağıdaki değerler elde edilmiştir.

0,092<=K_i<0,394 değerleri için "Gizlenmesine gerek yoktur",
0,394<=K_i<0,696 değerleri için "Düşük seviyeli gizleme" ve
0,696<=Ki<=0,999 değerleri için "Gizlenmelidir"

Bu değerler esas alınarak Şekil 2'deki çizim elde edilmiştir. Diğerinden farklı olarak dikey çizgiler "Düşük seviyeli gizleme" derecesindeki gizlilikleri sunmada kullanılmıştır. Yetişkinler veri seti ile TÜİK veri seti arasında nitelikler açısından benzerlik bulunmaktadır. O benzerliğe sebep olan nitelik çiftleri şöyledir; {(YAŞ, age), (MEDENİDURUM, relationship), (CİNSİYET, sex), (MEMLEKET, native country)}. Bu niteliklerden YAŞ ve age nitelik çifti her ikisi de gizli olarak etiketlenmiştir. MEDENİDURUM ve relationship nitelikleri ise benzer olmalarına rağmen birisi "Gizlenmesine gerek yoktur" diğeri ise "Gizlenmelidir" olarak etiketlenmiştir. Benzer şekilde CİNSİYET niteliği birinci veri setinde "Gizlenmesine gerek yoktur", ikinci veri setinde "Düşük seviyeli gizleme" şeklinde yer etiketlenmiştir. MEMLEKET ile "native country" nitelikleri de aynı içeriklere sahip olmasa bile benzer içeriğe sahiptir, biri illeri biri ise ülke isimlerini içerir. Bazı nitelikler her iki veri seti için de aynı gizlilik etiketi almıştır.



Şekil 1. TÜİK kaynaklı veri seti için nitelikler ve gizlilik dereceleri (Attributes and the privacy degrees for TÜİK-sourced)



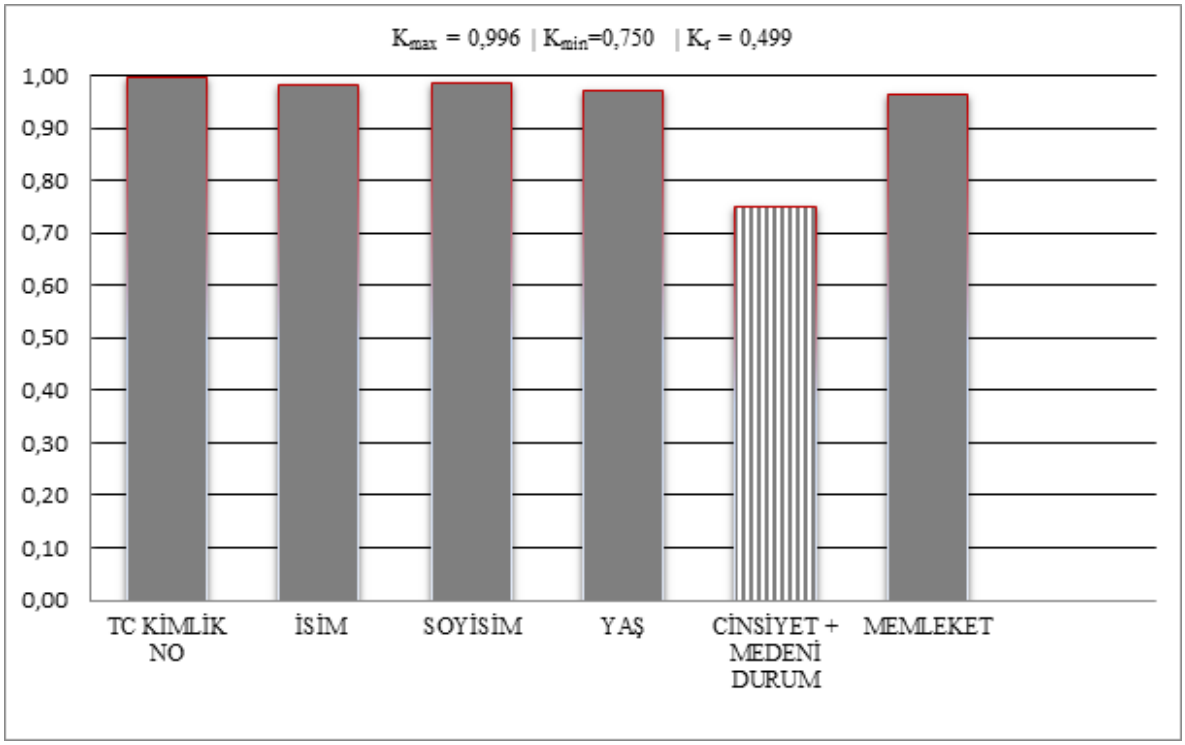
Şekil 2. UCI yetişkinler veri seti için nitelikler ve gizlilik dereceleri (Attributes and the privacy degrees for adult dataset)

Bu durum önerimiz açısından istenen durumdur. Fakat bazı nitelikler için gizlilik değerleri farklıdır. Bu durumun farklı sebepleri vardır. Öncelikle, veri setleri boyut açısından oldukça farklıdır, 300 ile 20.000, ikincisi veri setlerinden birisi Türkiye kaynaklı diğeri ise yabancı kaynaklıdır. Üçüncüsü, her iki veri setinde birbirinden farklı nitelikler yer aldığı için bu niteliklerden elde edilen; K_{max} ve K_{min} değerleri farklı ve doğal olarak aralık değerleri farklıdır. Bu durumda etiketlemeyi ve sonuçları etkilemektedir.

4.3.2. Nitelik çiftleri için gizliliğe dayalı etiketleme
(Privacy-based tagging for attribute pairs)

Konuyla ilgili bir başka çalışma nitelik çiftleri için gizlilik değerlerinin bulunmasıdır. Tek başına ele alındığında gizli olmayan CİNSİYET ve MEDENİ DURUM nitelikleri nitelik çifti olarak birlikte ele alındıklarında gizlilik değeri; KMEDENİDURUM, CİNSİYET = 0,750 olmuştur. Bu değer; cinsiyet ve medeni durum niteliklerinin bireysel gizlilik değerlerinden daha yüksek bir değerdir. Dolayısıyla gizli olmayan niteliklerin bir araya gelmesiyle gizli niteliklerin ortaya çıkabileceği görülmüştür. Nitelik birleştirmesi sonucu elde edilen sonuçlar Şekil 3'te gösterilmektedir. Şekilden de görüleceği üzere gizli olmayan iki nitelik bir araya getirildiğinde ilgili satır için gizlilik derecesi artmıştır. Benzer şekilde yetişkinler veri setindeki tüm A_i'ler (i = 1, … m) için birikimli kullanım oranları (K_i)'ler hesaplanmıştır. Buna göre, Şekil 2'de "Gizlenmesine gerek yoktur" şeklinde etiketlenen race, capital_gain, capital_loss ve native_country niteliklerinin çiftleri, birlikte incelendiğinde; $K_{race,capital_gain}$ = 0,378, $K_{capital_gain,capital_loss}$ = 0,241, $K_{race,capital_loss}$ = 0,327, $K_{capital_gain,native_country}$ = 0,298, $K_{race,native_country}$ = 0,352, $K_{capital_loss,native_country}$ = 0,243 değerleri elde edilmiştir. Değerler dikkatle incelendiğinde görülecektir ki nitelik birleştirmeleri her zaman gizlilik değerini tekli gizlilik değerlerine oranla artırmaktadır. Nitelik sayısı arttıkça doğrusal olarak gizlilik derecesi de yükselmektedir. Bu her iki veri seti içinde bu durum doğrulanmıştır. Yapılan deneylerden de görüleceği üzere cinsiyet ve medeni durum nitelikleri görünüşte gizli olmayan nitelikler olmasına rağmen veri setinin durumuna göre kimi yerde bir kişiyi tanımlamada kullanılabilir bir bilgi meydana getirmektedir.

Çalışmamızda bir niteliğin gizli olup olmaması o niteliğin gizlilik sınırları ile elde edilmektedir. Eğer elde edilen kullanım oranı bilgisi için 0.75 gibi bir sınır verdiysek 0.75 ve üzeri bir kullanım oranı değeri veren nitelik gizlenmesi gereken niteliktir. Nitelik bazında gizlilik sağlanmalıdır. Önerimiz güvenlik yöneticileri için, verilerden yola çıkarak otomatik olarak gizlilik önerisi getirmeye yardımcı olabilecektir. Bununla birlikte kurumların güvenlik algısına bağlı olarak gizlilik sınırları belirlenebilir. Bu konuda yapılması gereken işlem gizlilik sınırlarını değiştirme yetkisinin güvenlik yöneticilerine bırakmaktır. Böylece sistemin kullanılabilirliği daha da artacaktır.



Şekil 3. Nitelik birleştirmesi sonucu değişen gizlilik durumu (Changing privacy with the result of attribute combination)

5. SONUÇLAR (CONCLUSIONS)

Son günlerin belki en fazla tehdide maruz kalan varlıkları kişisel verilerdir. Kişisel verilerin çalınması, ele geçirilmesi gibi tehditler bu konudaki riski artırmaktadır. İlgili riske karşı önlem olarak çeşitli politikalar geliştirilmiş ve yasalar çıkarılmıştır. Bununla birlikte öncelikli olarak korunması gereken bilgi varlıklarının neler olduğunun, hangi bilgi varlıklarının daha fazla risk taşıdığının, veri paylaşımı sırasında özellikle korunması gereken veri parçalarının neler olduğunun uygun bir şekilde ölçümlenmesi ve böylece korunması ihtiyacı vardır. Bu çalışmada gizliliğe dayalı bir etki derecelendirme yapılarak gizliliğe dayalı hassas bir veri koruma çalışması ortaya konmuştur. Verilerin gizliliği veri homojenliği yardımıyla hesap edilmiştir. Veri, nitelikler bazında ele alınmış ve önce niteliklerin bireysel gizlilik değerleri bulunmuş, ardından nitelik çiftleri için gizlilik değeri elde edilmiştir. Bu hesaplama sırasında nitelik birleştirmelerinin gizliliği artırdığı görülmüştür. Bu artış birikimli toplam yerine birikimli toplamlardan kesişimlerin çıkarılması şeklinde meydana gelmiştir. Çalışmalar esnasında gizlilik etiketi vermede maksimum ve minimum risk değerinin de etkili olduğu görülmüştür. Sistemin kullanılabilirliği açısından risk sınırlarının kullanıcı tanımlı olarak değiştirilebilmesi uygundur. Ayrıca, olası bütün veri setleri için maksimum risk değeri 1 minimum risk değeri ise 0 civarındadır. Dolayısıyla, risk değerlendirme yaparken [0,1] aralığında değerleri esas almak, daha doğru karşılaştırma yapabilmek için diğerlerinden daha etkili olabilecektir. Bu çalışmada ortaya konan etiketleme yöntemi gizlilik merkezli risk yönetimini çerçevesinde de ele alınabilecektir. Özellikle risk değerlendirme, gizlilik etki değerlendirme adımına katkı sağlayacak bu öneri bilgi varlıklarına değer biçme için de bir yöntem ortaya koymaktadır.

KAYNAKLAR (REFERENCES)

1. Warren S.D. ve Brandeis L.D., The right to privacy, Harvard Law Review, A.B.D., 2010.
2. Van Den Hoven J., Blaauw M., Pieters W. ve Warnier M. Privacy and Information Technology. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/spr2016/entries/it-privacy. Yayın tarihi Bahar, 2016. Erişim tarihi Şubat 13, 2017.
3. DeCew J.W., In pursuit of privacy: Law, ethics, and the rise of technology, Cornell University Press, 1997.
4. Svantesson D., Clarke R., Privacy and consumer risks in cloud computing, Computer Law & Security Review 26 (4), 391-397, 2010.
5. Pieters W., On thinging things and serving services: technological mediation and inseparable goods, Ethics and information technology, 15 (3), 195-208, 2013.
6. Friedman B., Kahn Jr P. H., Borning A., Value sensitive design and information systems, Early engagement and new technologies: Opening up the laboratory, Springer, 55-95, 2013.
7. Cavoukian A., Taylor S., Abrams M.E., Privacy by Design: essential for organizational accountability and strong business practices, Identity in the Information Society, 3 (2), 405-413, 2010.
8. Wright D., De Hert P., Introduction to privacy impact assessment, Privacy Impact Assessment, Springer Link, 3-32, 2012.

9. Wright D., The state of the art in privacy impact assessment, Computer Law & Security Review, 28 (1), 54-61, 2012.

10. Wright D., Mordini E., Privacy and ethical impact assessment, Privacy impact assessment, Springer Link, 397-418, 2012.

11. Seto Y., Application of Privacy Impact Assessment in the Smart City, Electronics and Communications in Japan, 98 (2), 52-61, 2015.

12. Warren A., Charlesworth A., Privacy impact assessment in the UK, Privacy Impact Assessment, Springer Link, 205-224, 2012.

13. Gellman R. Fair information practices: A basic history. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2415020. Yayın tarihi Mart 27, 2014. Erişim tarihi Şubat 23, 2017.

14. Canbay P., Sağlık Hizmetlerinde Anonimlik: Dağıtık Yapılar Için Ideal Bir Veri Paylaşım Modeli, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 2014.

15. Fung B., Wang K., Chen R., Yu P. S., Privacy-preserving data publishing: A survey of recent developments, ACM Computing Surveys (CSUR), 42 (4), 14, 2010.

16. Canbay P., Sever H., The Effect of Clustering on Data Privacy, IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami-Amerika, 277-282, 9-11 Aralık, 2015.

17. Aggarwal C.C., Philip S.Y., A general survey of privacy-preserving data mining models and algorithms, Privacy-Preserving Data Mining: Model and Algorithms, Springer, 11-52, 2008.

18. Clifton C., Kantarcioglu M., Vaidya J., Lin X., Zhu M. Y., Tools for privacy preserving distributed data mining. ACM Sigkdd Explorations Newsletter, 4 (2), 28-34, 2002.

19. Hansen M., Berlich P., Camenisch J., Claub S., Pfitzmann A., Waidner M., Privacy-enhancing identity management, Information Security Technical Report, 9 (1), 35-44, 2004.

20. Uğur A., Soğukpınar I., Sustainable Authorization in Enterprise Workflow and Authorized Digital Signature Model, Journal of the Faculty of Engineering and Architecture of Gazi University, 29 (3), 559-568, 2014.

21. Türkiye İstatistik Kurumu. Türkiye İstatistik Kurumu nüfus istatistikleri 2009 verileri. www.tuik.gov.tr. Erişim tarihi Haziran 11, 2016.

22. Frank A., Asuncion A. UCI Machine Learning Repository 2010. http://archive.ics.uci.edu/ml. Erişim tarihi Haziran 23, 2016.