ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

GİZLİLİK-TABANLI YÜZEY EĞİLİMİ ANALİZİ

Salih DEMİR

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA
2022

ÖZET

Doktora Tezi

GİZLİLİK-TABANLI YÜZEY EĞİLİMİ ANALİZİ

Salih DEMİR

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Dr. Öğretim Üyesi Bülent TUĞRUL

Genel olarak, Coğrafi Bilgi Sistemleri (CBS) karar vermek için matematiksel ya da istatistikler yaklaşımlar kullanmaktadır. Bu tezde CBS'lerin matematiksel yaklaşımlarından biri olan konumsal enterpolasyon yöntemleri, ölçülmemiş koordinatlar için tahmin değerleri üretmek amacıyla kullanılmıştır. Konumsal enterpolasyon yöntemlerinin güvenilirliği ve doğruluğu, çalışma bölgesinde toplanan örnek miktarına bağlıdır. Bu tür verilerin elde edilmesi çok fazla zaman ve bütçe gerektirmektedir. Bazı bölgelerde birden fazla kuruluş benzer faaliyetler içerisinde bulunabilir. Bu gibi durumlarda, kuruluşlar kendi verilerini kullanarak çalışmalar yapabilecekleri gibi daha doğru tahmin modelleri ortaya koyabilmek için diğer kuruluşlara ait verilerden faydalanabilir. Böylelikle zaman ve bütçe tasarrufu sağlanabilir. Ancak gizlilik endişesi nedeniyle bu tür değerli bilgiler diğer kuruluşlarla paylaşılmak istenmeyebilir. Ayrıca bazı ülkelerde kişisel/tüzel verilerin korunması kapsamında özel bilgilerin paylaşılmasına karşı yasalar vardır. Bununla birlikte bu yasaları ihlal etmeden gizlilik koruma çözümü ile kuruluşlara ait bilgiler gizlenerek diğer kuruluşun kullanımına sunulabilir. Böylelikle kuruluşlar arası işbirliği teşvik edilirken, verilerin güvenliği de sağlanmış olmaktadır.

Bu tezde verilerin gizliliği, güvenliği ve ortak bir şekilde işlenmesi üzerine çözüm modelleri sunulmaktadır. İlgili modellerde istemci-kuruluş, kuruluş-kuruluş ve bulut mimarisinde veri gizliliğini sağlayacak çözümler sunularak hem istemcilerin hem de veri sahiplerinin mahremiyetini sağlayan özel bir Yüzey Eğilim Analizi gerçekleştirilmiştir. Önerilen modeller sayesinde gizlilikler ihlal edilmeden doğru ve güvenilir tahminlerin elde edilmesi hedeflenmiştir. Önerilen modellerde verilerin şifrelenmesi ve işlenmesi, veri iletimi, depolama ve hesaplama maliyetlerini etkilemektedir. Her ne kadar bu maliyetler artsa da tahmin modellerinin genel performansı üzerine etkisi olmadığı gösterilmiştir.

Temmuz 2022, 70 sayfa

Anahtar Kelimeler: Konumsal enterpolasyon, Yüzey Eğilim Analizi, gizlilik, gizlilik-korumalı veri madenciliği

ABSTRACT

Ph.D. Thesis

PRIVACY-PRESEVING TREND SURFACE ANALYSIS

Salih DEMİR

Ankara University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Asst. Prof. Dr. Bülent TUĞRUL

Geographic Information Systems (GIS) generally use mathematical or statistical approaches to make decisions. In this thesis, spatial interpolation methods, one of the mathematical approaches of GIS, are used to generate predictive values for unmeasured coordinates. The reliability and robustness of spatial interpolation methods depend on the data collected in the relevant region. Obtaining such data requires a lot of time and budget. In some regions, more than one organization may engage in similar activities. In such cases, organizations can use their data and data from other organizations to produce more accurate forecasting models. Thus, time and budget savings can be achieved. However, due to privacy concerns, such valuable information may not be shared with other organizations. Also, some countries have laws against sharing private information within the scope of personal/legal data protection. However, with the privacy-preserving solution, the data belonging to the organizations can be obscured and made available to the other organization without violating these laws. Thus, while promoting cooperation between organizations, data privacy is also ensured.

This thesis presents solution models for privacy, security and joint data processing. In these models, a unique Trend Surface Analysis was carried out to ensure the privacy of both clients and data owners by providing solutions to ensure data privacy in client-server, server-server and cloud architecture. Thanks to the proposed models, it is aimed to obtain accurate and reliable estimates without violating confidentiality. Encryption and data processing in the proposed models affect data transmission, storage and computation costs. It has been shown that cost increases do not affect the overall performance of the forecasting models.

July 2022, 70 pages

Key Words: Spatial interpolation, Trend Surface Analysis, privacy, privacy-preserving data mining

TEŞEKKÜR

Bu çalışmanın ortaya çıkmasında deneyimleri, bilgileri ve yardımlarıyla bana yol gösteren, her zaman bana moral veren danışman hocam Dr. Öğretim Üyesi Bülent TUĞRUL'a sonsuz teşekkür ederim.

Akademik bilgileri ve destekleri ile rehberlik eden, moral veren çok değerli hocalarım Prof. Dr. Aysun COŞKUN'a ve Doç. Dr. Recep ERYİĞİT'e çok teşekkür ederim.

Çalışmamın her anında yanımda olan, tüm deneyimlerini benimle paylaşan, kendi teziymiş gibi gece-gündüz desteğini esirgemeyen arkadaşım Ufuk TANYERİ'ne çok teşekkür ederim.

Çalışmamı titizlikle inceleyip değerli katkıları ile tezime destek olan arkadaşlarım Mahmut KILIÇASLAN'a ve Ayhan AYDIN'a çok teşekkür ederim.

Her zaman yanımda olduğunu bildiğim, tezime zaman ayırmam için yardımcı olan arkadaşım Mesut SEVİNDİK'e çok teşekkür ederim.

Her zaman olduğu gibi destekleriyle hep yanımda hissettiğim canım anneme, babama, ağabeyime ve kız kardeşime teşekkür ederim.

Salih DEMİR
Ankara, Temmuz 2022

İÇİNDEKİLER

KISALTMALAR DİZİNİ

B-GKYEA	Bulut sunucularda Gizlilik-Korumalı Yüzey Eğilimi Analizi
BHS	Bulut Hizmeti Sağlayıcısı
CBS	Coğrafi Bilgi Sistemleri
D-GKYEA	Dağıtılmış verilerde Gizlilik-Korumalı Yüzey Eğilimi Analizi
EBOB	En Büyük Ortak Bölen
FHE	Tam Homomorfik Şifreleme (*Fully Homomorphic Encryption*)
GKVM	Gizlilik korumalı veri madenciliği
HE	Homomorfik Şifreleme (*Homomorphic Encryption*)
k-NN	*k-Nearest Neighbors*
KOKH	Kök Ortalama Kare Hatası
OMH	Ortalama Mutlak Hata
PHE	Kısmi Homomorfik Şifreleme (*Partially Homomorphic Encryption*)
SHE	Sınırlı Homomorfik Şifreleme (*Somewhat Homomorphic Encryption*)
SMC	Güvenli Çok taraflı Hesaplama (*Secure Multiparty Computation*)
SMIP	Güvenli Matris Tersini Hesaplama Protokolünü (*Secure Matrix Inversion Protocol*)
SMMP	Güvenli Matris Çarpma Protokolü (*Secure Matrix Multiplication Protocol*)
TMA	Ters Mesafe Ağırlıklandırma
YEA	Yüzey Eğilim Analizi

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

1. GİRİŞ

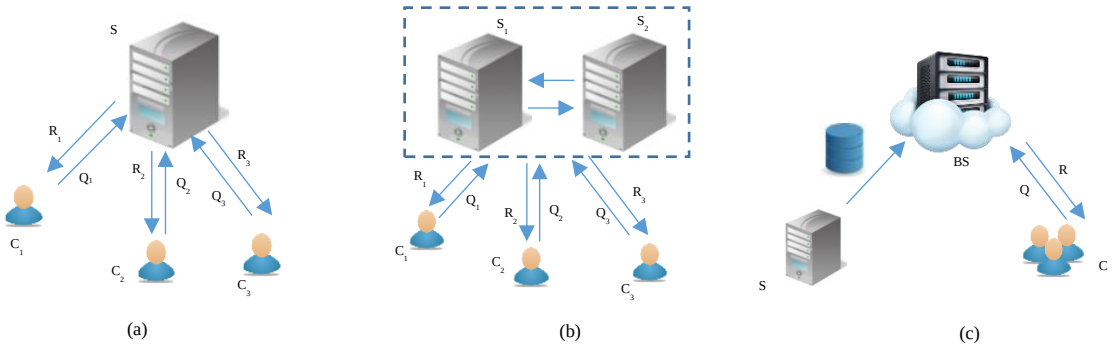
Bilgisayar teknolojisindeki gelişmeler nedeniyle Coğrafi Bilgi Sistemleri (CBS) uygulamaları popüler hale gelmiştir. CBS ile konumsal bilgiler belirli bir amaca yönelik olarak toplanabilir, yönetilebilir, analiz edilebilir ve görüntülenebilir. Ayrıca, planlama, lojistik ve yönetim ile ilgili birçok işlem CBS ile sağlanabilmektedir. Öte yandan, zaman ve maliyetler sebebiyle çalışma alanındaki tüm noktalarda ölçüm yapmak mümkün olmayabilir.

CBS'nin temel bileşenlerinden biri konumsal enterpolasyon yöntemidir. Bahsi geçen yaklaşımlar aynı bölgede gözlenen değerlerden ölçüm yapılmamış koordinatlar için tahmin modelleri üretmede kullanılmaktadır. Bu nedenle çalışma alanından örnekler alınmaktadır. Örnekleme stratejisi ve örnek sayısı enterpolasyon modellerinin doğruluğunu etkilemektedir. Konumsal enterpolasyon yöntemleri, Tobler (1970) tarafından önerilen Birinci Coğrafya Yasası olarak bilinen "her şey diğer her şeyle ilgilidir, ama yakın şeyler uzak şeylerden daha fazla ilişkilidir" temel prensibine dayanmaktadır. Ancak bu ilkeden çok önce başlamış olan çalışmalar da bulunmaktadır.

Konumsal enterpolasyon yöntemleri mühendislik, finansal ve birçok disiplin için önemli bir rol oynamaktadır. Li ve Heap (2014), yaygın olarak uygulanan 25 yöntemi özelliklerine göre sınıflandırmaktadır. Tüm konumsal enterpolasyon yöntemleri üç farklı sette ele alınmaktadır: i) jeositatistik olmayan, ii) jeoistatistik ve iii) hibrit yöntemler. Çalışma bölgesinin istatistik özelliklerini kullanan bir yöntem olan Kriging (Krige, 1951), en popüler yöntemlerden birisidir. Yanı sıra Ters Mesafe Ağırlıklandırma (TMA, Inverse Distance Weighting) (Shepard, 1968), En Yakın Komşular ve Yüzey Eğilim Analizi (YEA, Trend Surface Analysis) yöntemleri birçok mühendislik çalışmasında oldukça yaygın bir şekilde kullanılmaktadır. TMA yerel ve kesin bir enterpolatördür. TMA'da ölçüm yapılan bir noktanın ağırlığı, ölçüm yapılmamış noktaya olan uzaklığıyla ters orantılıdır. Uzaklığın gücü belirlenerek komşu noktaların etkileri hesaplanabilmektedir. YEA, konumsal verilere uygun en küçük karesel hataya göre en iyi polinomu arayarak bir enterpolasyon modeli oluşturmaktadır (Oldham & Sutherland, 1955). YEA, yerel ve global olmak üzere iki şekilde ele alınmaktadır. Yerel

YEA yalnızca bir tahminin istendiği komşu noktaları, global YEA ise tüm veri kümesini kullanmaktadır.

Geleneksel konumsal enterpolasyon şemasında iki taraf vardır. Birinci taraf, konumsal verileri toplayan, depolayan ve yöneten "veri sahibi" olarak; ikinci taraf ise "müşteri" ya da "istemci" olarak ifade edilmektedir. İstemciler, veri sahibinin konumsal verilere sahip olduğu bölgede ticari yatırımlar veya analizler yapabilmek için belirli bir alanda tahmin değerine ihtiyaç duymaktadır. Geleneksel tek sunuculu şema Şekil 1.1a'da tasvir edilmektedir. Veri sahibi tarafından oluşturulan konumsal enterpolasyon yöntemlerinin doğruluğu ve güvenilirliği, ölçülen konumların sayısına bağlıdır. Yetersiz miktarda veri kullanılarak yapılmış bir modelin çıktısı tartışmalı olmaktadır. Veri sahipleri, zaman ve bütçe kısıtlamaları nedeniyle yeterli miktarda veri toplamayabilir. Bu nedenle daha doğru modeller oluşturmak için farklı veri sahipleri tarafından toplanan verilerin birleştirilmesi gerekir. Birden çok veri sahibinin verileri birlikte kullanılarak elde edilen tahmin modeli, veri sahiplerinin tek başlarına yaptıkları modelden daha iyi sonuçlar üretecektir. Ancak toplanan ölçümler gizli tutulduğundan, veri sahipleri verilerini başkalarıyla paylaşma konusunda isteksiz davranmaktadır. İki veri sahipli şema şekil 1.1b'de tasvir edilmektedir. Şekil 1.1c'de ise verilerin bulut sunuculu mimaride toplandığı görülmektedir.



Şekil 1.1 Veri dağıtım senaryoları: a) Tek sunuculu b) iki sunuculu c) bulut sunuculu

Karmaşık ve büyük miktarda veriyi analiz etmek, çok fazla depolama ve bilgi işlem kapasitesi gerektirmektedir. Bulut bilişim, alt yapı modeliyle kullanıcılara kendilerine adanmış sunucular sunmaktadır (Şekil 1.1c). Veri sahipleri verilerini, düşük yatırım ve

maliyet, esneklik ve erişilebilirlik nedeniyle bulut sunuculara taşımaktadır. Veri sahipleri için hizmet aldığı sunucularda tutulan verilerinin gizliliği dikkate alınmalıdır. Hükümetler, vatandaşlarının mahremiyetini kötüye kullanımdan korumak adına yasal düzenlemeler yapmaktadır. Verilerin nasıl ve ne kadar süre saklanabileceği, kimlerle paylaşılabileceği sınırlandırılmaya çalışılmaktadır. Bu nedenle özel veriler, veri sahibi dışındaki hizmet sağlayıcı tarafından kontrol edilebilen bulut sunucularında açık olarak saklanmamalıdır.

Veri şifreleme, gizlilik ve güvenlik gerektiren hassas durumlarda kullanılan bir yöntemdir. Ancak şifrelenmiş veriler üzerinde işlem yapabilmek için öncelikle ilgili kişiler/kurumlar tarafından şifrenin çözülmesi ve daha sonra tekrar şifrelenmesi gerekmektedir. Fakat ilgili süreç, özellikle büyük verilerle çalışırken, işlem miktarının artmasına ve güvenlik açıklarının ortaya çıkmasına neden olabilmektedir. Veri sahipleri, verilerini hem istemcilerinden, hem işbirliği yapacakları diğer veri sahiplerinden hem de bulut sunucularından gizlemek istemektedir. İstemciler de para ve zaman harcayacakları koordinatları açık etmeden bir tahmin değeri almak istemektedirler. Gerçek hayatta taraflar kötü niyetli davranabilmektedir. Bu nedenle konumsal enterpolasyon analizi için tarafların mahremiyetini dikkate alan bir çözüm hayatidir.

Bu tezde, veri sahipleri ve istemci verilerinin gizliliğini sağlamak için merkezi, dağıtık ve bulut-sunucu veri dağıtım senaryolarında gizlilik korumalı YEA enterpolasyon yöntemleri önerilmektedir. Söz konusu metotlarda veri gizliliği ise homomorfik şifreleme algoritmaları ve güvenli çok taraflı bilgi işlem çözümleri ile sağlanmaktadır. Ayrıca her bir çözümün gizlilik ve ek maliyet analizleri yapılarak geleneksel yöntemin değerleri ile karşılaştırmalar yapılmaktadır. Önerilen çözümlerin yeterli doğru tahmin değerleri ürettiğini göstermek için gerçek veriler kullanılarak çeşitli deneyler yapılmış ve sonuçlar tartışılmıştır.

2. KURAMSAL TEMELLER

2.1 Konumsal Enterpolasyon

Konumsal veriler, çevresel yönetim planlaması, risk değerlendirmesi ve karar vermede önemli bir rol oynamaktadır. Bununla birlikte, coğrafi veriler genellikle kolay temin edilemezler. Özellikle söz konusu girdilere ulaşmak, dağlık bölgeler veya derin denizler gibi yerlerde oldukça maliyetlidir. Saha araştırmalarında toplanan verilere tipik olarak noktasal kaynaklardan ulaşılmaktadır. Bilim insanları etkili, güvenilir ve mantıklı yorumlar yapabilmek için bir bölge genelinde konumsal verilerden yararlanır. Coğrafi bilgi sistemleri ve modelleme teknikleri, doğal kaynakların yönetimi için güçlü araçlardır. Doğal olarak çevresel değişkenlerin konumsal verilerine olan ihtiyacın artacağı açıktır. Bu nedenle örneklenmemiş noktalardaki özniteliklerin değerlerinin tahmin edilmesi gerekmektedir. Dolayısıyla konumsal enterpolasyonun kullanılması önemlidir (Li ve Heap 2014).

Konumsal enterpolasyon, mekânsal verilerle ilgilenmektedir. Enterpolasyon yöntemlerinin amacı ölçülmemiş bir konum için tahmini bir değer üretmektir. Konumsal enterpolasyon için pek çok yöntem bulunmaktadır. Myers (1994) ilgili yaklaşımları deterministik ve stokastik olarak kategorize etmektedir. Ek olarak, Li ve Heap (2008) diğer araştırmacılardan farklı olarak; i) jeoistatistiksel ii) jeoistatistiksel olmayan ve iii) hibrit metotlar olarak gruplandırmaktadır. Webster ve Oliver (2007) tüm konumsal enterpolasyon tekniklerinin tahminlerini örneklenmiş verinin ağırlık ortalaması olarak ifade etmektedirler. Yukarıda bahsi geçen kategoriler içerisinde; Ters Mesafe Ağırlıklandırma (TMA) (Shepard, 1968), Kriging (Krige, 1951) ve Yüzey Eğilim Analizi (YEA) (Oldham & Sutherland, 1955) en bilinen metotlardır.

2.1.1 Ters Mesafe Ağırlıklı Enterpolasyon Yöntemi

TMA, ölçüm yapılmamış bir noktada çevresindeki ölçümleri kullanarak tahmin değeri elde etmeyi sağlayan bir enterpolasyon tekniğidir. Bu işlemi yaparken tahmin değeri istenen konuma yakın olan ölçümler, uzaktakilerden daha fazla ağırlığa sahiptir. Başka bir deyişle ölçüm yapılan konumların etkileri, tahmini istenen konuma olan uzaklıklarıyla ters orantılıdır. Söz konusu etki *m*'ye bağlıdır. *m*=0 ise, ağırlıklar mesafeye bağlı değildir. $m \geq 1$ olduğunda veya mesafe arttığında ölçüm yapılan noktaların etkisinin azalacağı açıktır. TMA yaklaşımı kullanılarak i konumundaki ölçüm (P_i) denklem 2.1'deki gibi hesaplanmaktadır (Armstrong, 1998):

$$P_i = \frac{\sum_{j=1}^{n}\left[P_j/\left(D_{ij}\right)^m\right]}{\sum_{j=1}^{n}\frac{1}{(D_{ij})^m}} \tag{2.1}$$

n, komşu noktaların sayısını, P_j j noktasındaki ölçümü ve D_{ij}, i ve j noktaları arasındaki mesafeyi temsil etmektedir. D_{ij}, bir koordinat sistemi olan bir uzayda Öklid uzaklığı metriği kullanılarak hesaplanmaktadır. Öklid uzaklığı herhangi iki nokta arasındaki düz çizgi mesafesini göstermektedir. Düz çizgilerin olmadığı daha büyük alanlarda D_{ij}'yi hesaplamak için jeodezik mesafe kullanılmaktadır (Johnson vd., 2001).

TMA yöntemi varsayım olarak alttaki yüzeyi pürüzsüz alanlarda kullanılmaktadır. Çıktı olarak ızgaralı bir yüzey ya da konturlar elde edilmektedir. TMA, tek değişkenli, deterministik ve yerel olarak çalıştırılan bir enterpolasyon tekniğidir. İlgili yöntemde, hata değerlendirmesi yapılamamaktadır. Arama alanının boyutu ve mesafe ağırlıklandırma parametresinin seçiminin TMA'nın sonuçlarına etkisi büyüktür. Kötü seçilmiş arama alanı boyutu yüksek veri yoğunluklarında yapaylık vermektedir. Seyrek veri yapısında ise hızlı çalışmaktadır ve hesaplama yükü düşüktür (Li ve Heap 2008, Li ve Heap 2014).

2.1.2 Kriging Enterpolasyon Yöntemi

Kriging tekniği, ilk olarak Güney Afrikalı maden mühendisi Daniel Krige tarafından ortaya atılmıştır. Jeo-istatistik alanında geliştirilen bir enterpolasyon tekniğidir (Krige, 1951). Başlarda maden kaynaklarını belirlemede kullanılsa da örnek noktalarını kullanarak doğru tahminler üretebilme yeteneğiyle birçok mühendislik alanında yaygın olarak kullanılmaktadır. Kriging, doğrusal olmayan ve çok modlu problemlerde iyi performans göstermektedir (El. Saad & Dong, 2015). Jeo-istatistiksel bir yöntem olarak Kriging'in kullanıldığı alanlara yağış ve toprak özelliği tahmini, yeraltı suyu istatistikleri ve dijital arazi modellemesi örnek verilebilir. Kriging enterpolasyonu, çalışma alanındaki bölgeselleştirilmiş değişkenlerin optimal tahmini için variyogram teorisine ve yapısal analize dayanmaktadır. Yöntem temel olarak denklem 2.2 gibi ifade edilmektedir.

$$\hat{Z}(P_0) = \sum_{i=1}^{n} \lambda_i Z(P_i) \tag{2.2}$$

Denklem 2.2'de, P_0 tahmin değeri hesaplanacak konumu, $Z(P_i)$, i. konumda ölçülen değeri, λ_i ölçülen değer için bilinmeyen bir ağırlığı ve son olarak n ise ölçüm sayısını ifade etmektedir. İlaveten λ_i ağırlıkları, tahminin yansız olduğu ve tahmin hatasının varyansının minimum olduğu şekilde seçilmektedir. λ_i, variogram modeline, tahmin konumuna olan uzaklığa ve tahmin değeri istenen noktanın etrafında gözlenen değerlerin mekânsal ilişkisine bağlıdır. λ_i'lerin toplamı 1'e eşittir.

Kriging yönteminde varsayım olarak yüzey pürüzsüz kabul edilirse çıktı veri yapısı ızgaralı bir yüzey olmaktadır. Tek değişkenli ya da çok değişkenli bir yapısı bulunmaktadır. Ayrıca TMA'nın aksine stokastik olup, tahmin modelinin oluşturulmasında girdi verilerin bazıları kesin olarak bilinse de bazı önemli değişkenlerin değerleri karar alınana kadar bilinmemektedir. Kriging, ilgili bilinmeyen değişkenlere dair olasılıkları kullanarak belirsizliği modele katmaktadır. Yerel çalışan

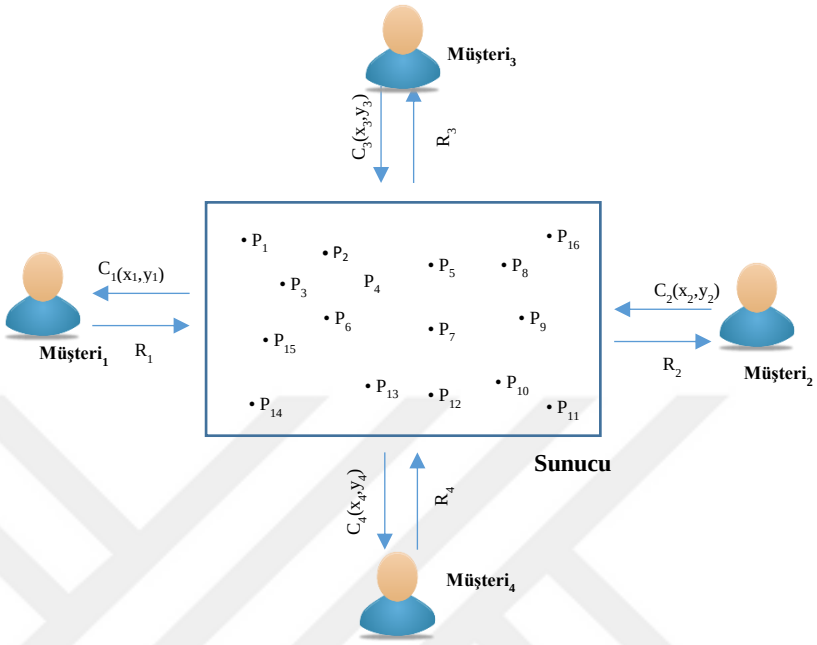
bir tekniktir. Kriging'in sınırlaması, hata değerlendirmesi, variograma, veri noktalarının dağılımına ve enterpolasyonlu blokların boyutuna bağlıdır. Variogramları hesaplamak için yeterli veri olduğunda, Kriging seyrek veriler için iyi bir enterpolatördür (Li ve Heap 2008, Li ve Heap 2014).

Kriging yöntemleri, i) Sıradan (Ordinary) Kriging, ii) Basit (Simple) Kriging ve iii) Evrensel (Universal) Kriging olarak bilinmektedir. Sıradan kriging, doğrusal denklemler sistemini çözmenin $O(n^3)$ zaman karmaşıklığından dolayı, büyük kayan noktalar kümesinde hesaplama açısından maliyetlidir. Sürekli değişen uzay-zamansal veri akışları üzerinde gerçek zamanlı Kriging enterpolasyonu yapmak bu nedenle özellikle zor olmaktadır (Zhong vd., 2016). Basit Kriging yöntemi, yatay ve dikey variogramların analizinden elde edilen bilgilere dayandırılarak doğrudan uygulanmaktadır. Basit Kriging, matematiksel olarak en basit enterpolasyon tekniği olmasına rağmen, yaygın değildir. Rastgele alan beklentisinin bilindiğini varsayar ve bir kovaryans fonksiyonuna dayanmaktadır. Bununla birlikte, çoğu uygulamada kovaryans önceden bilinmemektedir (Cho vd., 2020). Son olarak, global Kriging, yarı variogramları veya kovaryansları ifade etmek için kullanılan matematiksel formlardır. Dönüşümleri kullanabilir ve ölçüm hatasına izin verebilir (Aumond vd., 2018).

2.1.3 Yüzey Eğilim Analizi Enterpolasyon Yöntemi

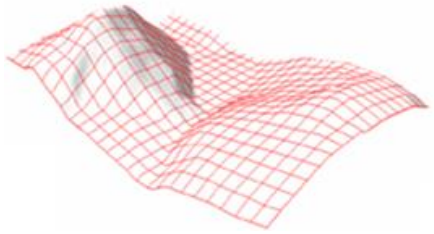
YEA yöntemi arazi üzerindeki bütün ölçüm noktalarını kullanarak arazinin genel eğilimini ortaya koyan yaklaşık bir yöntemdir. YEA, ölçülen değerler ile modellenmiş yüzey arasındaki karelerin toplamını en aza indiren en küçük kareler yöntemiyle polinomların oluşturulmasına dayanmaktadır. YEA lokal veya global olmak üzere iki farklı şekilde uygulanabilir. Lokal YEA tahmin yapılacak noktanın etrafında ölçüm değerleri bilinen ve belirli kriterleri sağlayan komşu noktaları kullanarak arazinin eğilimini bulmaya çalışmaktadır. Buna karşın global YEA arazideki bilinen bütün ölçüm noktalarını kullanarak arazinin genel eğilimini ortaya çıkarır. Global YEA'da bütün ölçüm noktalarının üzerinden geçmek zorunlu değildir. Ortaya çıkan hataların

toplamını minimize edecek optimum bir model kurulmaktadır (Burrough vd., 2015). Şekil 2.1 geleneksel bir YEA modelini şematize etmektedir.

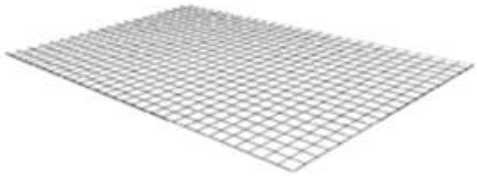


Şekil 2.1 Geleneksel YEA Modeli

Koordinatları (x, y) bilinen bir arazi için $P(x, y) = a_0 + a_1 x + a_2 y$ olmak üzere 1. dereceden bir düzlem ile tanımlanır. Arazi şartlarına bağlı olarak 2. dereceden ve daha yüksek dereceden modeller kurulabilir. 2. dereceden denklem $P(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2$, 3. dereceden ise $P(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2 + a_6 x^3 + a_7 x^2 y + a_8 xy^2 + a_9 y^3$ şeklinde oluşturulur. 2. dereceden denklemi oluşturmak için 6 adet, 3. dereceden bir denklem için ise 10 adet katsayının belirlenmesi gerekmektedir. Arazinin yapısı ne kadar çok değişiklik gösteriyorsa kurulacak olan model de o kadar yüksek dereceye sahip olmaktadır. Küçük dereceli modeller daha basit olmasına rağmen hata miktarı çoktur. Aksine yüksek dereceli modeller, daha karmaşık olup hata miktarı azdır. Şekil 2.2 gerçek bir yüzeyin 3 farklı dereceye göre YEA modellemesini göstermektedir.

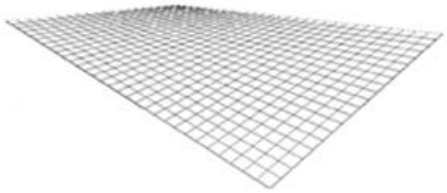


a) Gerçek yükseklik dağılımı



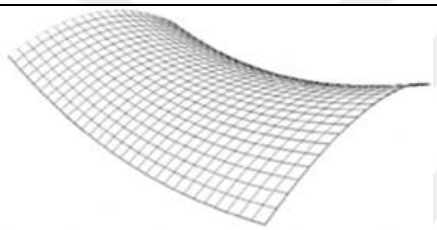
b) 1. derece doğrusal yüzey eğilimi

$$P(x,y) = a_0 + a_1x + a_2y$$



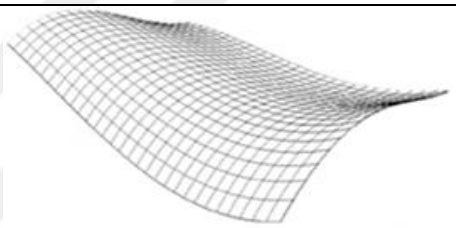
c) 1. derece çift doğrusal yüzey eğilimi

$$P(x,y) = a_0 + a_1x + a_2y + a_3xy$$



d) 2. derece yüzey eğilimi

$$P(x,y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy$$

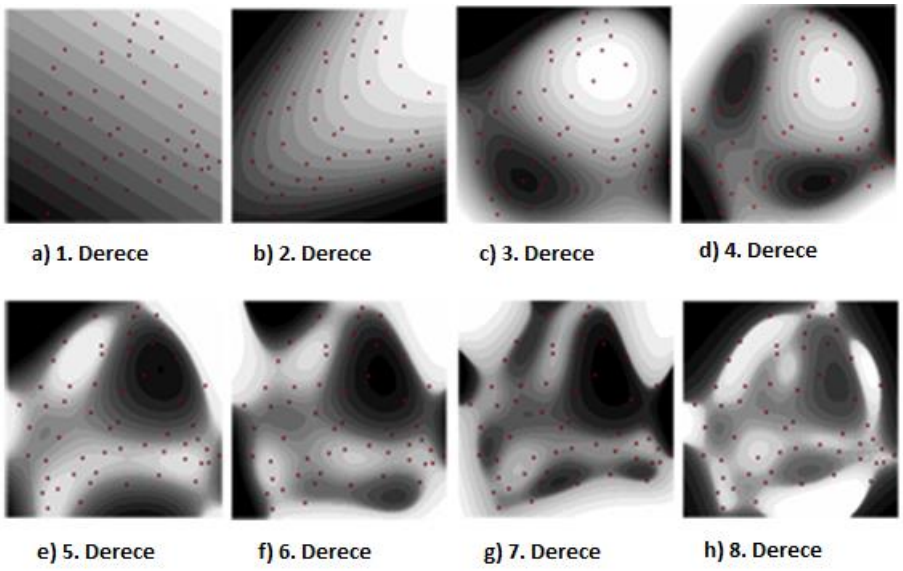


e) 3. derece yüzey eğilimi

$$P(x,y) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy + a_6x^3 + a_7x^2y + a_8xy^2 + a_9y^3$$

Şekil 2.2 Farklı derecelerde YEA modellemesi: (a) gerçek yükseklik (b) 1. derece doğrusal (c) 1. derece çift doğrusal (d) 2. derece (e) 3. derece yüzey eğilimi (Anonymous, 2012)

Şekil 2.3 ise 1. dereceden 8. dereceye kadar İsviçre Sarine bölgesindeki 55 yerleşim yerinin nüfus artışındaki normalleştirilmiş farkın konumsal dağılımını YEA ile modellenmesini göstermektedir.



Şekil 2.3 İsviçre Sarine bölgesindeki 55 yerleşim yeri için nüfus artışındaki normalleştirilmiş farkın konumsal dağılımı (Anonymous, 2012)

Denklem 2.3 YEA'nın 2. dereceden bir tahmin modelini ifade etmektedir ve söz konusu modeldeki sabitler denklem 2.4'te verilen metot ile hesaplanmaktadır.

$$z = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy \tag{2.3}$$

Daha karmaşık verileri temsil etmek için daha yüksek derece denklemler kullanılabilir. Denklem 2.3'te verilen katsayılar $(a_0, a_1, a_2, a_3, a_4, a_5)$, regresyon analizinde olduğu gibi hesaplanmaktadır. Şen (2009), bahsi geçen katsayıları bulmak için gereken hesaplamayı denklem 2.4'teki gibi sunmuştur:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 1 & \bar{x} & \bar{y} & \overline{x^2} & \overline{xy} & \overline{y^2} \\ \bar{x} & \overline{x^2} & \overline{xy} & \overline{x^3} & \overline{x^2y} & \overline{xy^2} \\ \bar{y} & \overline{xy} & \overline{y^2} & \overline{x^2y} & \overline{xy^2} & \overline{y^3} \\ \overline{x^2} & \overline{x^3} & \overline{x^2y} & \overline{x^4} & \overline{x^3y} & \overline{x^2y^2} \\ \overline{xy} & \overline{x^2y} & \overline{xy^2} & \overline{x^3y} & \overline{x^2y^2} & \overline{xy^3} \\ \overline{y^2} & \overline{xy^2} & \overline{y^3} & \overline{x^2y^2} & \overline{xy^3} & \overline{y^4} \end{bmatrix}^{-1} \begin{bmatrix} \bar{z} \\ \overline{xz} \\ \overline{yz} \\ \overline{x^2z} \\ \overline{xyz} \\ \overline{y^2z} \end{bmatrix} \tag{2.4}$$

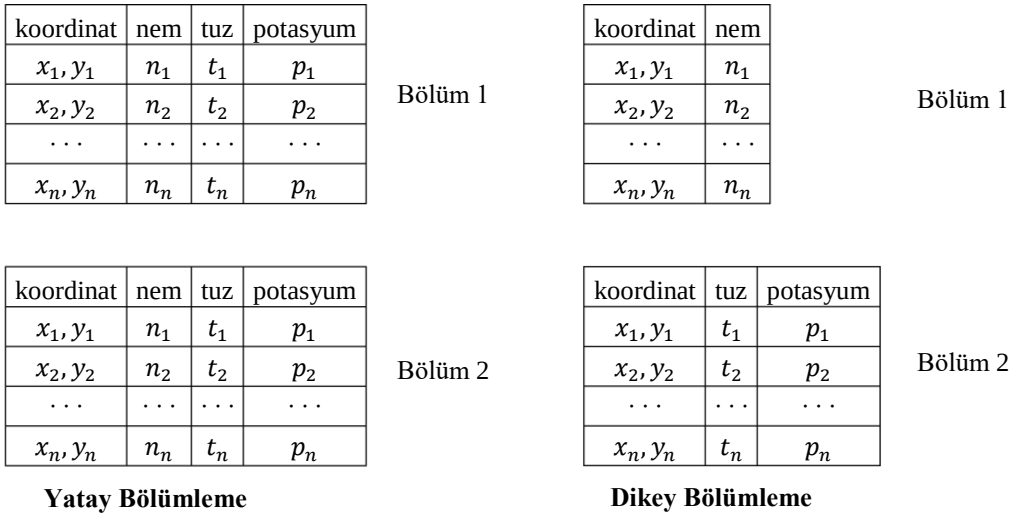
Denklem 2.4'ün sol tarafı eğilim yüzey denkleminin katsayılarına karşılık gelmekte ve ilgili denklemin sağ tarafı, veri kümesindeki değişkenlerin ortalamalarından oluşmaktadır.

YEA, normal dağılımlı veriler için kullanılırsa çıktının veri yapısı sürekli ve ızgaralı yüzeye sahip olmaktadır. YEA, Kriging yönteminde olduğu gibi stokastik ve global olarak çalışan bir entepolasyon yöntemidir. İlgili metotta aykırı değerler ve kenar etkileri yüzeyi bozabilir. YEA yaklaşımının hesaplama maliyeti düşük olup, hızlı değerlendirme yapılacak sistemlerde tercih edilmektedir (Li ve Heap 2008, Li ve Heap 2014).

2.2 Veri Dağıtım Senaryoları

Enterpolasyon için ölçüm değerleri ve bunların konum bilgileri genellikle tek bir kurum tarafından tutulmaktadır. Veri sahibi, sorgu sahiplerine mevcut verilerini kullanarak talep ettikleri konuma ait tahminler sunmaktadır. Bu tür veri dağıtım senaryosuna merkezi tabanlı denilmektedir.

Farklı kurumlar aynı alanda benzer çalışmalar yapabilmektedir. Kurumlar kendi verilerini kullanarak tahminler üretse de güvenilir ve doğru olmayabilir. Yapılacak tahminlerin güvenirliği ölçüm yapılan noktaların sayısına ve çalışma alanındaki dağılımına bağlıdır. Kurumların çalışma alanında yeni ölçüm yapması oldukça maliyetli olup, zamansal olarak mümkün olamamaktadır. Veri sahibi kurumlar, verilerini birleştirerek ortak veri seti üzerinden tahminler üretmeye karar verebilirler. Çünkü veri sahiplerinin oluşturdukları yeni veri setinden daha güvenilir ve kesin tahminler üretme olasılığı bulunmaktadır. Böyle bir durum, bölümlenmiş veri olarak adlandırılır. Benzer şekilde veriler ikiden fazla tarafa dağıtıldığında söz konusu durum dağıtılmış veri olarak adlandırılmaktadır. Dağıtılmış veriler yatay veya dikey olarak doğrudan bölümlenebilir. Ancak, dikey bölümlemenin coğrafi istatistik yöntemlerinde uygulanması pratik değildir. Şekil 2.4 yatay ve dikey bölümlenmiş dağıtık veri senaryosunu göstermektedir.

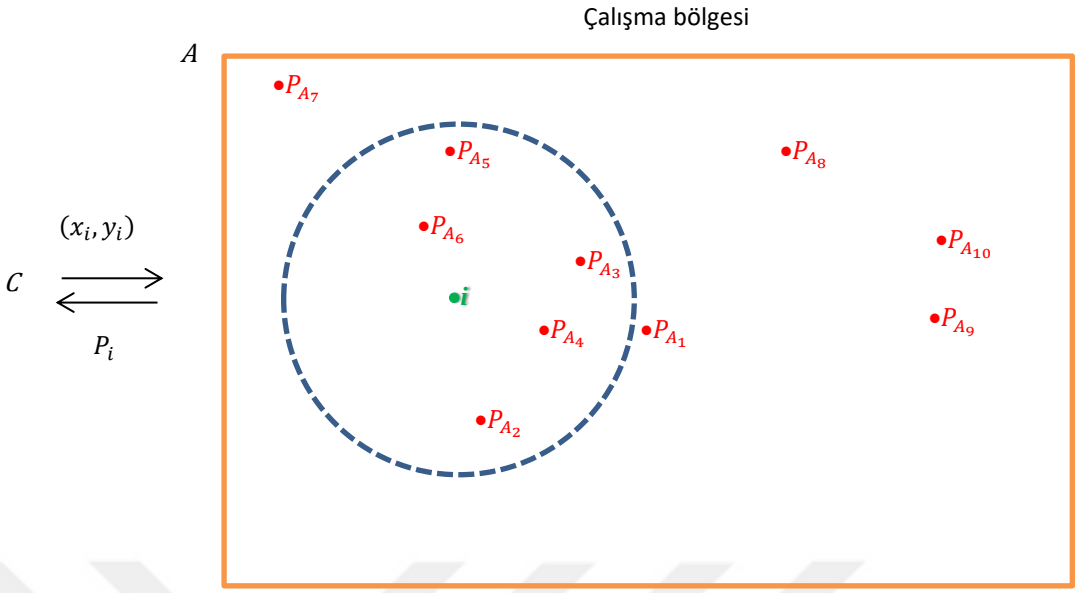


Şekil 2.4 Yatay ve dikey bölümlenmiş dağıtık veri senaryosu

Son olarak veri sahipleri, verilerini saklamak ve istemcilere sunmak için her zaman çevrimiçi olan sunuculara ihtiyaç duymaktadır. Kesintisiz hizmet verebilmek için sunucuların her an ulaşılabilir olması ve güvenliğinin sağlanması gibi görevlerin de yerine getirilmesi gerekmektedir. Veri sahipleri performans, ölçeklenebilirlik ve maliyet gibi avantajlarından dolayı verilerini bulut sunucular üzerinde saklayıp istemcilere bu sunucular üzerinden hizmet sunmaktadır. Bu durum ise bulut mimari tabanlı veri dağıtımı olarak ifade edilmektedir.

2.2.1 Merkezi veri dağıtımı

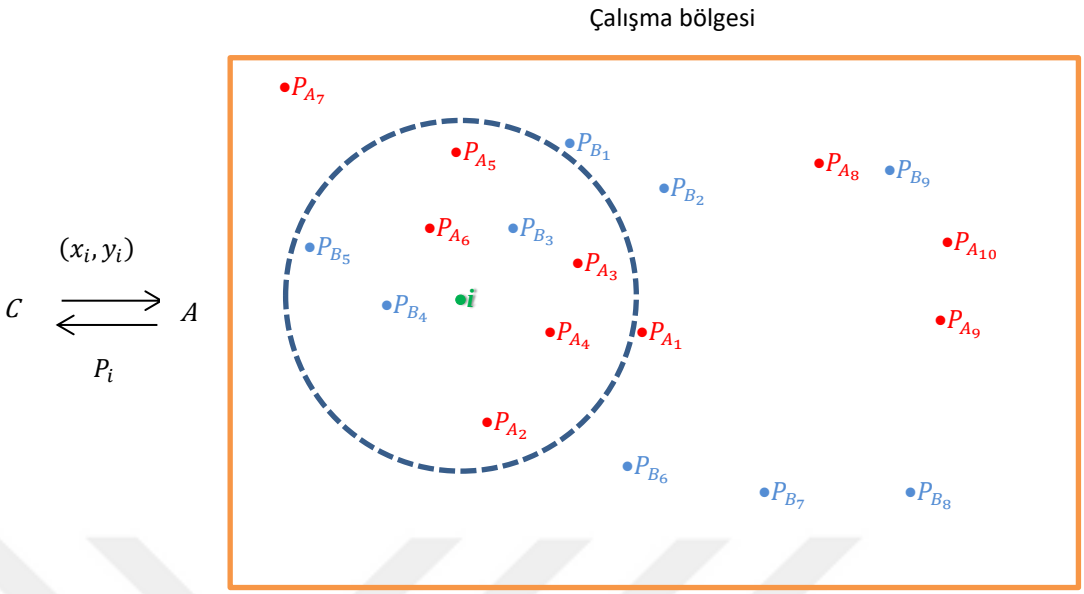
YEA ile tahmin değerleri üretebilmek için ölçüm yapılan koordinatlar ve ölçülen değerler kullanılarak bir model oluşturulması gerekmektedir. YEA ölçüm noktalarını lokal ya da global olarak tercih etmektedir. Lokal YEA modeli, tahmin değeri istenen konumun etrafında belirlenen uzaklıktaki ya da sayıdaki noktaları kullanmaktadır. İstemci, tahmin değerini istediği *i* konum bilgisini veri sahibine gönderir. Veri sahibi, tahmin modelini gelen konuma göre işleterek istemciye döndürür. Şekil 2.5'teki örnek şemada P_i olarak adlandırılan *i* konumundaki ölçümü tahmin etmek için o konumunun etrafındaki $(A_2, A_3, A_4, A_5, A_6)$ noktalar kullanılmaktadır. Benzer şekilde global YEA tahmin modelinde çalışma alanındaki tüm ölçüm yapılan konumlar $(A_1, A_2, A_3, A_4, A_5, A_6, A_7, A_8, A_9, A_{10})$ kullanılmaktadır.



Şekil 2.5 Merkezi veri dağılımı

2.2.2 İki taraflı veri dağılımı

Enterpolasyon amacıyla bir bölge üzerinde ölçümler yapmak mümkün olmayabilir. Bunun sebebi çalışma bölgesinde yeni ölçümlerin oldukça maliyeti olması ya da o bölgeye ait geçmiş zaman ölçümlerinin yapılamamasıdır. Enterpolasyon için toplanan ölçümlerin yetersizliğinden dolayı hesaplanan tahmin değeri tatmin edici olmayabilir. Tahmin değerleri üretmek için iki kurum aynı bölgeye ait farklı konumlardaki ölçümlere sahip olabilir. İki taraf arasında bölünmüş veri tabanlı enterpolasyon örneği şekil 2.6'da verilmektedir. İlgili şemada bazı ölçümlerin veri sahibi A tarafından $(P_{A_1}, P_{A_2} \dots, P_{A_{10}})$ ve kalan ölçümlerin veri sahibi B tarafından $(P_{B_1}, P_{B_2} \dots, P_{B_9})$ tutulduğu gösterilmektedir. Sorgulama yapan istemci C, i konumu için bir tahmin değeri istemektedir. C, i konumumun x_i, y_i koordinat bilgilerini hizmet sağlayıcı kurum A'ya gönderir. A, kendi verileriyle B'nin verilerini birlikte kullanarak P_i tahmin değerini hesaplar ve sonucu C'ye gönderir.



Şekil 2.6 İki ortaklı veri dağıtımı

P_i'yi tahmin etmek için tarafların işbirliği yapması gerekir. Lokal YEA tahmin modeli için çalışma alanının şekil 2.6'daki çemberle belirlenmiş alan olduğunu kabul edelim. Taraflar, i konumunda tahminlerini işbirliği yaparak hesapladıklarında, şekil 2.6'daki çemberde yer alan 8 nokta $(A_2, A_3, A_4, A_5, A_6, B_3, B_4, B_5)$ örnek noktalar olarak seçilmektedir. Bununla birlikte hizmet sağlayıcı A, tahmini kendi başına yapmak isterse, ölçümleri A tarafından tutulan $(A_2, A_3, A_4, A_5, A_6)$ 5 noktayı komşu olarak seçmektedir. Benzer şekilde, B de tahmini kendi başına hesaplamak isterse, ölçümleri B tarafından tutulan en yakın (B_3, B_4, B_5) 3 noktayı komşu olarak seçmektedir. Global YEA tahmin modelinde tarafların işbirliği yapması durumunda $(A_1 \dots A_{10})$ ve $(B_1 \dots B_9)$ olmak üzere 19 nokta kullanılmaktadır.

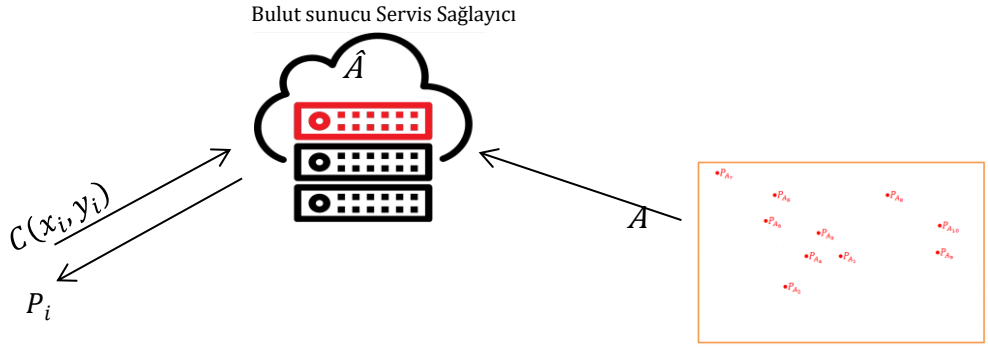
2.2.3 Bulut mimari

Geleneksel mimaride veri sahibi kurumlar verilerini kendi sunucularında saklamaktadırlar. Bilimsel hesaplamalar doğası gereği çok fazla zaman, depolama ve hesaplama gücü gerektirir. Son yıllarda ilgili yükleri bulut sunucularına aktarmak için yeni bir eğilim vardır. Bulut bilişim, bellek, işlemci, bant genişliği ve güç açısından düşük kapasiteli kullanıcıların verilerini ve gerekli hesaplama işlemlerini, zaman ve

para maliyetlerini azaltmak için yüksek kapasiteli sunuculara taşımaya olanak tanımaktadır.

Bulut tabanlı çözümler üç farklı hizmet modeli olarak sunulmaktadır: i) Yazılım hizmeti (Software as a Service-SaaS); ii) Platform hizmeti (Platform as a Service-PaaS); ve iii) Altyapı hizmeti (Infrastructure as a Service-IaaS) (Simmon, 2018). Ayrıca bulut hizmetinin genel, topluluk, karma ve özel olmak üzere kurulum modelleri bulunmaktadır. Genel bulut, Amazon AWS ve benzeri kurumlarda olduğu gibi kaynağı bir hizmet sağlayıcıdan genel kullanıcılara sunan bir yerleşim modelidir. Topluluk bulutu ise aynı eğilim ve kaygıları yüklenen bireysel topluluktaki türlü organizasyonlar arasındaki altyapının paylaşılması olarak ifade edilmektedir. Karma bulut da farklı amaçlardaki iki veya daha çok bulutun bir araya getirilmesidir. Son olarak özel bulut, üçüncü taraf aracılığıyla sadece tek bir organizasyona koşturulan bir yerleşim şeklidir (Mell & Grance, 2011).

Alt yapı ya da platform hizmeti alınan modelde, enterpolasyon için ölçümler yapmış veri sahibi A, verilerini bulut sunucuya yükler. Tahmin modelini sunucu üzerinde çalıştırarak, istemcilere hizmet sunmaktadır. A kendisi için ayrılan alanda verilerini güncelleme gibi işlemleri yaparken, sunucunun aktif olup olmaması tamamıyla bulut hizmeti veren kuruluşun sorumluluğundadır. Şekil 2.7 bulut mimari veri dağıtımını örneklemektedir.



Şekil 2.7 Bulut mimari veri dağıtımı

2.3 Coğrafi Gizlilik

Gizlilik, mahrem verilerin korunmasına ilişkin genel bir kavramdır. Söz konusu ifade genellikle gayri resmi olarak bir bireyin kişisel bilgilerini kontrol etme hakkıdır (Sun vd. 2020). Mahremiyetin diğer bir tanımı da bireyin, grubun veya kurumun neyi, ne zaman, nasıl ve ne ölçüde kendilerini ifşa edeceğine karar verme yeteneği olduğu iddiasına dayanmaktadır (Z. Sun vd. 2020). Gizlilik, özellikle sağlık ve suç verileri çalışmalarında önemlidir (Young vd. 2009). İlgili kavram ilerleyen bölümlerde coğrafi alanda ve merkezi, iki taraflı ve bulut dağıtım senaryolarında incelenmektedir.

Coğrafi gizlilik, haritalama işlemlerinde bireysel konum bilgilerinin korunmasını amaçlamaktadır (Kwan vd. 2004). Haritalar genellikle akademik yayınlarda önemlidir. Ancak bazı kritik durumlarda, gizlenmesi istenen bilgi örneğin HIV hastalarının veya aile içi şiddet mağdurlarının adresleri gibi hassas veriler içerebilir. Bu durumda, gizlilik amacını aşarak mahremiyet ihlallerine yol açabilmektedir. Araştırmacılar ayrıca coğrafi gizliliğin nasıl korunacağına odaklanmaktadır. Çünkü bireysel verilerin gizliliği sağlanırken sonuçların değişmemesi hayatidir. Leitner ve Curtis (2006) haritalardaki konumsal bilgilerin gizliliğini sağlayan genel bir çerçeve önermektedirler. Bu amaçla kişisel konumların gizliliğini koruyan coğrafi maskeleme yöntemlerinin nasıl tanımlanacağını incelemişlerdir. İlgili yazarlar, hastaların konum bilgilerini ve sağlık kayıtlarını, hastalıklar ve yer verileri arasındaki ilişkiyi gösteren analitik haritalar oluşturmak için kullanmışlardır.

Coğrafi bilgilerin gizliliğinin önemini vurgulayan pek çok çalışma vardır. Araştırmalar, sağlık veya suç gibi hassas bilgiler içeren haritaların yer aldığı 57 makalenin 2014 yılında yapılan bir incelemesinde, bu makalelerden 41'inin coğrafi maske kullanmayı ihmal ettiğini ve bunu yaparak 68.000'den fazla kişinin hassas ev konumlarının ortaya çıkarıldığını göstermektedir (Kounadi & Leitner, 2014). Bu bulgular, Haley vd.'nin (2016) yaptığı bir çalışma ile desteklenmektedir. İlgili çalışmada cinsel sağlık dergilerinde yayınlanan haritalar gözden geçirilmektedir. Haritaların %56'sının ev konumlarını yeterince maskelemediği açıkça görülmektedir. Yukarıda bahsi geçen araştırmalar hassas ve kritik veriler içeren konumsal bilgilerin gizliliğinin önemini ve de

verilerin gizliliği sağlanırken sonuçların doğruluğunun temin edilmesinin gerekliliğini vurgulamaktadır.

Araştırmacılar, gizliliği sağlamak için veri maskeleme, karıştırma, veri şifreleme gibi mahremiyet koruma yöntemleri geliştirmektedirler. Curtis vd. (2006), haritalarda yer alan konum bilgilerinin, tersine mühendislik yapılarak ev adreslerini belirlemede kullanılabileceğini göstermektedirler. Bu çalışmaları coğrafi maskelerin önemini ortaya koymaktadır. Gambs vd. (2013) ise coğrafi konum bilgisine sahip büyük verilerde gizlilik analizini verimli bir şekilde gerçekleştirmek adına MapReduce paradigmasının kullanılmasını önermektedirler. Exeter vd. (2014) kişilerin mahremiyetini tehlikeye atmadan coğrafi ve sağlık verilerinin nasıl yayınlanması gerektiğine dair bir coğrafi gizlilik erişim çerçevesi önermektedirler. Swanlund vd. (2020) ise kullanıcılardan ekstra bir veri gereksinimi duymadan uygulayabilecekleri sokak maskeleme yöntemini sunmaktadırlar.

Sosyal ağlar ve konum tabanlı hizmetler günümüzün popüler konuları arasındadır. Konum özellikli teknolojiler yaygınlaştıkça kullanıcıların konum bilgileri de birçok hizmet sağlayıcı ile paylaşılmaktadır. Örneğin Instagram, Facebook, Swarm ve Twitter gibi konum tabanlı hizmetlerin, kişilerin özel verilerini yayınladığı bilinen bir gerçektir. Bu bağlamda, Huang vd. (2019) ve Li ve Goodchild (2013) araştırmalarında Twitter kullanıcılarının ev ve iş adreslerinin konumlarının nasıl tahmin edileceğini incelemektedirler. Rizwan vd. (2020) ise çalışmalarında coğrafi konum tabanlı sosyal ağ uygulamalarının zaman-mekansal kişisel aktiviteleri tahmin etmede kullanılabileceğini göstermektedirler. Anlaşılacağı üzere yaygın olarak kullanılan konum tabanlı uygulamaların kritik verilerin gizliliğini sağlaması önemli bir rol oynamaktadır. Öte yandan gizliliği sağlarken yapılan işlemlerin sonuçları etkilememesi arzu edilen diğer bir durumdur.

Konumsal veriler merkezi ya da birden fazla alanda tutulabilmektedir. Bazen de hizmetin verimliliği, kullanılabilirliği ve maliyeti bakımından bulut ortamları tercih edilmektedir. Dolayısıyla konumsal veriler içeren coğrafi bilgilerin gizliliği bu yönüyle ele alınmalıdır.

2.3.1 Merkezi durumda gizlilik

Merkezi veri dağıtımından bölüm 2.2.1'de bahsedilmektedir. Bilindiği üzere merkezi durum söz konusu olduğunda veri, tek bir sunucuda tutulmaktadır. Veri toplama ve yayınlama için tipik bir senaryo şekil Şekil 1.1 a'da gösterilmektedir. Veri toplama aşamasında, veri sahibi veriyi ölçümler yaparak toplamaktadır. Veri yayımlama aşamasında ise veri sahibi, toplanan verileri bir sunucu üzerinden istemcilerinin erişimine açar.

Merkezi durum senaryosunda istemci sunucu arasındaki veri iletişimini üçüncü taraflardan gizlemek için çeşitli yöntemler kullanılmaktadır. İki taraf arasındaki iletişimin şifreli olarak gerçekleştirilmesi bunlardan biridir. Merkezi sunucunun, istemcinin verilerini açık bir şekilde öğrenmesi onlar için çeşitli tehlikeler oluşturmaktadır. Benzer şekilde istemcinin sunucu üzerinde kısıtlamasız isteklerde bulunarak veriler elde etmesi de veri sahipleri için tehlike oluşturmaktadır.

2.3.2 İki taraflı durumda gizlilik

Merkezi durumun dışında veriler birden fazla sunucuda dağıtılmış olarak tutulabilir. Veri toplama ve yayınlama için tipik bir senaryo şekil Şekil 1.1 b'de gösterilmektedir. İki taraflı dağıtılmış senaryolarda, verilerden elde edilecek bilgiler kurumların gelecekteki konumunu belirleyeceğinden dolayı veri analizi oldukça popülerdir. Elde edilen veriler üzerinde çeşitli veri madenciliği yöntemleri kullanılarak değerli bilgiler ortaya çıkartılabilmektedir. Modellerin doğruluğu ve güvenilirliğinde veri kalitesi ve miktarı önemli bir rol teşkil etmektedir. Taraflar arasında gizlilikle bölünmüş veriler üzerinde farklı hesaplamalar yapmak için çeşitli çalışmalar bulunmaktadır. Bu çalışmalar, iki veya daha fazla tarafın verilerinin birlikte kullanılmasıyla daha doğru ve güvenilir modellerin kurulabileceğini göstermektedir. (Kaleli, 2012; Silva ve Bellala, 2017). Özel verileri koruyan birçok çözüm vardır (Aggarwal ve Yu, 2008; Evfimievski vd., 2004; Vaidya ve Clifton, 2002). Araştırmacılar çözümlerini yatay, dikey ve hibrit olmak üzere üç farklı veri bölümleme şeması üzerinde kurmaktadırlar. Yatay bölümleme şemaları, veri kümesinin iki farklı veri tabanının satırlarına aynı öznitelik

kümesine katılarak oluşturulduğunu varsayar. Öte yandan dikey bölümlendirme şemasında iki farklı veri tabanında depolanan aynı nesneye ait veriler, ortak kimlik bilgisine sahip sütunlar aracılığıyla birleştirilmektedir. Hibrit şema oluşturmak için de farklı veri tabanı bölümleri seçilmektedir (Aggarwal ve Yu, 2008; Sheikhalishahi ve Martinelli, 2017). Abdel Wahab vd., (2014) yatay olarak bölünmüş veriler üzerinde önerdikleri gizliliği koruyan bir veri madenciliği modelinde dağıtılmış özel verileri korurken birliktelik kuralı sorguları geliştirmişlerdir. Sun vd. (2015), gizliliği koruyan bir proksimal destek vektörü makine çözümü sunmaktadırlar. Çözümleri, sınıflandırma amacıyla dikey olarak bölümlenmiş veriler üzerinde tasarlanmaktadır. Yang vd., (2015), bulutta dikey olarak bölümlenmiş veri tabanlarına dayalı bir tıbbi veri paylaşımını önermektedirler. Bringer vd., (2013) şifreleme teknikleri aracılığıyla gizliliği sağlayarak, güvenli iki taraflı hesaplamanın biyometrik tanımlamaya nasıl uygulanacağını göstermektedirler. Henecka ve Schneider (2013), daha az bellek kullanarak hızlı olan güvenli iki taraflı bir protokol önermektedir. Domadiya ve Rao (2019), dikey olarak bölünmüş sağlık verilerinde, RSA algoritması ile gizliliği koruyan kural tabanlı yeni bir yaklaşım önermektedirler.

2.3.3 Bulut ortamında gizlilik

Bulut bilişimin hızlı gelişimi ile giderek daha fazla kullanıcı verilerini ve uygulamalarını bulutta taşımaktadırlar. Sanallaştırma teknolojisi de bulut bilişimin temelini oluşturmaktadır. İlgi teknoloji, veri merkezinin bilgi işlem, ağ ve depolama donanımlarını bir araya getirip, yüksek verimlilik, düşük maliyet, hızlı uygulama ve yüksek kullanılabilirlik sunmaktadır. Bu nedenle birçok kuruluş verilerini ya da uygulamalarını bulut sunucular üzerinde barındırmaktadırlar. Ancak bulut sunucular performans ve maliyet konusunda avantaj sağlarken güvenlik konusunda bazı riskler de getirmektedir.

Şekil 1.1c'de gösterildiği gibi, veri sahipleri verilerini doğrudan bulut sunucusuna yüklemektedir. Artık veriler, Bulut Hizmeti Sağlayıcısının (BHS) kontrolündeki sunucularda barındırılmakta ve yönetilmektedir. Bu da verilerin sahipliği ve yönetiminin farklı taraflarda olduğu anlamına gelmektedir. BHS, kendi kontrolündeki

sunuculardaki depolanan verilere serbestçe erişebilir ve arama yapabilir. Saldırganlar, kullanıcının verilerini elde etmek için BHS sunucularına da saldırabilir. Yukarıda bahsi geçen durum, kullanıcılar için bilgi sızıntısı ve veri kaybı tehlikesi oluşturur. İlgili sorunların çözümü için genellikle erişim kısıtlamalarına veya veri şifrelemeye odaklanılmaktadır. Bu bağlamda şifreleme yöntemleri kullanılarak verilerin gizliliği sağlanabilmektedir. Ancak bu, verilerin hesaplanmasını, sorgulamasını engellemektedir. Bu nedenle gizliliğin yanı sıra şifreli veriler üzerinde sorgulama ve hesaplama yapmak, bulut bilişim alanında dikkat çekici konular arasındadır.

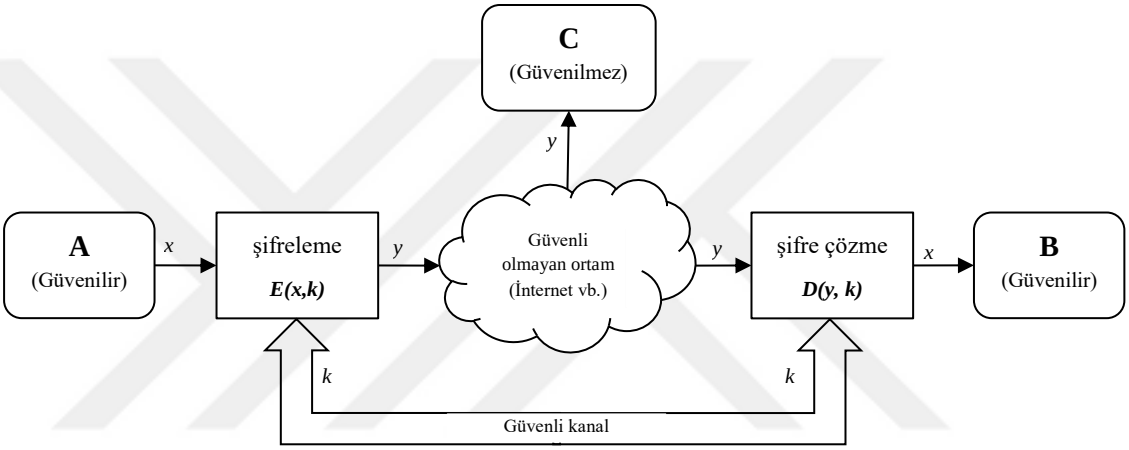
Yiu vd. (2012), hassas veriler için benzerlik arama tekniklerinin dış kaynak kullanımına nasıl sağlanacağını önermektedirler. Zhou vd. (2017) bulut sunucularında bulunan şifreli veriler üzerinde güvenli *k*-NN (k-Nearest Neighbors) sorgulaması yapmak için yeni bir çözüm önermektedirler. Çözümleri, veri sahiplerinin ve istemcilerin özel verilerini, bulut sunucusu sağlayıcılarından korumaktadır. Liu vd. (2020), veri madenciliğinde kullanılması adına homomorfik şifreleme yöntemi kullanarak güvenli *k*-NN sınıflandırıcısı olarak adlandırdıkları dış kaynaklı bir hesaplama araç seti geliştirmişlerdir. Kartheeban ve Murugan (2017), bulutta şifreli belgelerde arama yapmayı sağlayan dosyaların birden çok sunucuya dağıtık şekilde depolandığı bir yöntem sunmuşlardır.

2.4 Şifreleme Algoritması ve Protokoller

Şifreleme yöntemleri, herhangi bir bilgiyi istenmeyen kişilerden gizlemek amacıyla yaygın olarak kullanılmaktadır. Söz konusu metotlar simetrik ve asimetrik olarak kategorize edilmektedir (Cominetti & Simplicio, 2020). Aynı özel anahtarın şifreleme ve deşifreleme görevleri adına kullanılması simetrik şifreleme olarak bilinir. Bahsi geçen yaklaşım, şifrelemede karışıklık ve yayılma işlemlerini kullanmaktadır. Asimetrik algoritmalarda ise kullanıcı simetrik şifrelemede olduğu gibi gizli ve açık (genel) bir anahtar çiftine sahiptir. İlgili teknik gizliliği sağlamak için karmaşık modüler aritmetikten yararlanmaktadır. Asimetrik algoritmalar, klasik veri şifrelemenin yanı sıra dijital imzalar ve anahtar oluşturma gibi uygulamalar için de kullanılabilmektedir (Paar

ve Pelzl 2010). İlaveten, bu algoritmalar tam sayılı çarpanlara ayırma, ayrık logaritma ve eliptik eğriler tabanlı yürütülmektedir.

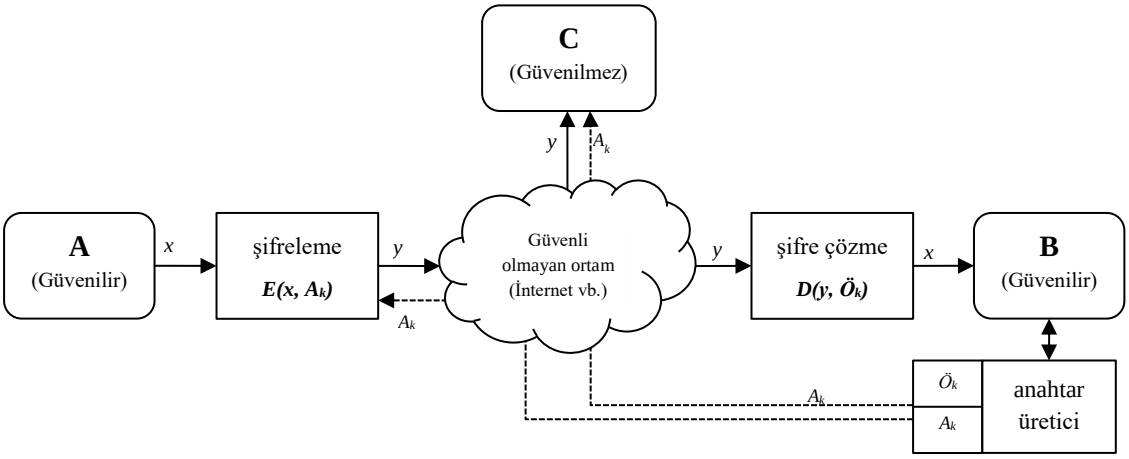
Bölüm 2.3'de farklı veri dağıtım senaryolarındaki coğrafi verilerin gizliliğinin gerekliliğinden bahsedilmektedir. İlgili senaryolardan bulut mimarideki durumun simetrik şifreleme uygulaması şekil 2.8, asimetrik biçimi ise şekil 2.9 ile gösterilmektedir. Şekil 2.8 ve şekil 2.9'da konumsal verilerde gizlilik sağlanırken güvenilir A, B ve güvenilmez C olmak üzere iki taraf bulunmaktadır.



Şekil 2.8 Simetrik şifreleme şeması

A ve B arasındaki iletişimi gizleyen şifreleme şemasında ***x*** düz metin, ***k*** anahtar, ***y*** şifreli metin, *E* şifreleme ve *D* şifre çözme fonksiyonudur (Şekil 2.8). C, özel anahtara sahip olmadığı için şifreyi çözemez ve gizli metni öğrenemez.

Simetrik algoritmalar, güvenli bir iletişim kurmak isteyen her iki kullanıcı tarafından paylaşılan yalnızca bir anahtar kullanmaktadır (Şekil 2.8). Ancak asimetrik şifreleme algoritmalarında her kullanıcının iki anahtarı vardır (Şekil 2.9).



Şekil 2.9 Asimetrik şifreleme şeması

Şekil 2.9'da ilk anahtar tüm kullanıcılarla paylaşılabilecek açık anahtar, A_k olarak bilinir. İkincisi, anahtarın sahibi tarafından gizli tutulması gereken özel anahtardır, $Ö_k$. Şifreli metin sadece $Ö_k$ ile çözülebilmektedir. C, şifreli metine ve açık anahtara sahip olsa bile açık anahtardan özel anahtarı tahmin edemez.

2.4.1 Homomorfik şifreleme

Bulut bilişim gibi şifrelenmiş veriler ile hesaplamalar gerekli olduğu durumlarda, geleneksel şifreleme teknikleri yetersiz kalmaktadır. Bu durumda kullanıcılar, hassas verilerinin şifresini çözdükten sonra hesaplama yapıp daha sonra verileri tekrar şifrelemekte ya da hesaplama yapmak için güvenilmeyen taraflara bazı gizli bilgileri açıklamak zorunda kalmaktadır. Homomorfik şifreleme (Homomorphic Encryption - HE), şifrelenmiş veriler üzerinde, deşifre etmeden işlem gerçekleştirme konusunda ayırt edici bir özelliğe sahiptir (Cominetti & Simplicio, 2020). Şifrelenmiş metinler üzerinde yapılan hesaplamaların sonuçları da şifrelidir. İlgili sonuç metnin şifresinin çözülmüş hali ile orijinal metinler üzerinde yürütülen aynı hesaplamaların sonucu birbirine eşittir (Sathya vd. 2018). Homomorfik şifreleme, matematiksel işlevleri hesaplamak için idealdir. Örneğin, HE uygulanmış büyük veri kümesinde ortalama veya ortanca değer gibi istatistikler hesaplanabilmektedir. Bankacılık, tıbbi veri analizi, veri madenciliği gibi uygulamalarda işlenen verilerin gizliliğini sağlamasından dolayı da bulut bilişimde kapsamlı bir şekilde kullanılmaktadır (Wang ve ark. 2019, Will ve Ko 2015).

Mevcut HE şemaları asimetrik ve simetrik olmak üzere iki sınıfa ayrılmaktadır. Gentry (Gentry, 2009), DGHV (van Dijk vd., 2010), RSA (Sathya vd., 2018), El Gamal (1985), Paillier (Paillier, 1999) gibi asimetrik şemalar ve Domingo Ferrer (Ferrer, 1996), MORE (Matrix Operation for Randomization and Encryption) ve PORE (Polynomial Operation for Randomization and Encryption) gibi simetrik şemalar vardır (Hariss vd., 2017).

Herhangi iki düz metin m_1, m_2 ve bir şifreleme algoritması E için

$$E(m_1) \cdot E(m_2) = E(m_1 + m_2) \tag{2.5}$$

eşitliği sağlanırsa, sistemin toplama işleminde homomorfik olduğunu söyleriz. El Gamal (1985), Paillier (1999) ve Benaloh (1994) şifreleme algoritmaları homomorfik toplamaya örnektir. Benzer şekilde

$$E(m_1) \cdot E(m_2) = E(m_1 \cdot m_2) \tag{2.6}$$

eşitliği sağlandığında, çarpma işleminde homomorfik demektir. El Gamal (1985) ve RSA şemaları (Rivest vd., 1983), homomorfik çarpma örnekleridir.

Temel olarak HE, Kısmi Homomorfik Şifreleme (Partially Homomorphic Encryption - PHE), Sınırlı Homomorfik Şifreleme (Somewhat Homomorphic Encryption-SHE) ve Tam Homomorfik Şifreleme (Fully Homomorphic Encryption-FHE) şeklindedir (Shrestha & Kim, 2019).

PHE'de şifreli mesaj üzerinde sınırsız sayıda sadece toplama veya çarpma işlemine izin verilmektedir. RSA (Rivest vd., 1983), Goldwasser-Micali (Goldwasser & Micali, 1982), Benaloh (J. C. Benaloh, 1987), ElGamal (El Gamal 1985) ve Paillier (Paillier, 1999) bilinen PHE yöntemleridir. Öte yandan, SHE'de hem toplama hem de çarpma işlemine izin verilmekte olup, işlem sayısında sınırlılıklar vardır. BGN bahsi geçen tekniğe bir örnektir (Boneh vd., 2005). Son olarak FHE ile şifreli mesaj üzerinde sınırsız sayıda işleme izin verilmektedir. FHE algoritmalarına ise Gentry (2009), Dijk

vd. (2010), Brakerski ve Vaikuntanathan (2011) tarafından önerilen şemalar örnek olarak verilebilir.

HE yöntemleri arasında Paillier en çok kullanılan yaklaşımdır. Paillier özellikle şifreli veritabanı çözümlerinde benimsenmiştir (Cominetti & Simplicio, 2020). Temel olarak Paillier homomorfik şifreleme şeması i) anahtar oluşturma, ii) şifreleme ve iii) şifre çözme aşamalarından oluşmaktadır (Şenyürek ve Yakut 2013).

i) Anahtarları oluşturma;

- $EBOB\ (pq, p-1, q-1) = 1$ koşulunu sağlayan birbirinden farklı iki büyük asal sayı p ve q rasgele seçilir.
- $n = p\,.\,q$ ve Carmichel fonksiyonu ile en küçük ortak kat $\lambda = EKOK\ (p-1, q-1)$ ile hesaplanır.
- Üreteç için Z_n tamsayılar kümesinden rasgele α ve β seçilir.
- $g = (\propto n + 1)\ \beta^n\ mod\ n^2$ işlemi yapılır. Burada g sayısı ile n^2 aralarında asal olma koşulu sağlanmalıdır.
- L fonksiyonu $L(u) = (u-1)/n$ şeklinde tanımlanmak üzere $\mu = (L(g^\lambda\ mod\ n^2))^{-1}\ mod\ n$ ile hesaplanır.
- Şifreleme için kullanılacak (n, g) açık anahtar ve şifre çözmek için kullanılacak (λ, μ) anahtarı elde edilir.

ii) Şifreleme işlemi;

- Şifrelenecek metin Z_n tamsayılar kümesinden m değeri olsun
- Z_n tamsayılar kümesinden n ile aralarında asal rasgele bir r sayısı seçilerek $c = g^m r^n\ mod\ n^2$ şeklinde şifreleme yapılır.

iii) Şifre çözme işlemi;

Şifreli metin $m = L\left(c^{\lambda}\ mod\ n^2\right)\mu\ mod\ n$ işlemi ile çözülür.

Paillier'ın (1999) yaklaşımında iki homomorfik özellik bulunmaktadır. Düz metinlerin (m_1 ve m_2) homomorfik toplanması olarak bilinen ilk özelliği

$$D(E(m_1, G_k) \cdot E(m_2, G_k)(mod\ n^2)\ ) = \ m_1 + \ m_2(mod\ n) \tag{2.7}$$

şeklinde hesaplanmaktadır. Denklem 2.7, iki şifreli metnin çarpımının deşifresinin, düz metinlerin toplamına eşit olacağını göstermektedir. Paillier'in ikinci özelliği ise

$$D\left(E(m_1, G_k)^{m_2}\ (mod\ n^2)\right) = \ m_1 \cdot \ m_2(mod\ n) \tag{2.8}$$

olarak ifade edilmektedir. Denklem 2.8'de görüldüğü üzere, şifrelenen m_1 metninin, düz metin m_2. kuvvetinin deşifresi, iki düz metnin çarpımına eşit olacağını belirtmektedir.

2.4.2 Güvenli Çok Taraflı Hesaplama

Bilgisayar alanındaki gelişmelerle birlikte ölçüm yapma ve veri toplama oldukça yaygınlaşmıştır. İnternetin sunduğu olanakların artması ile benzer alanlarda faaliyet gösteren kuruluşların sahip oldukları veriler üzerinde ortak işlemler gerçekleştirmesi mümkün olmuştur. Güvenli Çok taraflı Hesaplama (Secure Multiparty Computation - SMC), kriptografinin alt dallarından biridir. Yao (1982) tarafından tanıtılan SMC, güvenilmeyen tarafların (iki veya daha fazla) birbirine özel girdilerini ifşa etmeden ortak bir fonksiyonun nasıl hesaplanacağı olarak tanımlanmaktadır. İki taraf olduğunda buna iki taraflı hesaplama denilmektedir. SMC protokolleri gizlilik ve doğruluk özelliklerini sağlamalıdır. SMC protokolleri, ya bozuk devre (Yao 1982) ya da gizli paylaşım şemalarına (Shamir, 1979) dayanmaktadır. Bozuk devre, güvensiz iki tarafın, güvenilir bir üçüncü taraf olmadan kendi ara yüzleri üzerinden bir işlevi birlikte değerlendirebildiği, iki taraflı güvenli hesaplama sağlayan bir şifreleme protokolüdür.

Üç temel güvenlik modeli vardır; güvenilir üçüncü taraf, yarı dürüst taraf ve kötü niyetli taraf. Güvenilir üçüncü taraf modelinde, tüm tarafların verilerini güvenilir üçüncü taraflara göndermesi gerekmektedir. Yarı dürüst model, tüm tarafların protokolün belirtilen adımlarını harfi harfine uygulayan, ancak diğer tarafların özel verileri hakkında bazı ek bilgiler elde etmek için tüm ara işlemleri kaydedebilen ve analizler yapabilen bir tür pasif saldırgan olabileceğini varsaymaktadır. Bu saldırgan sınıfı dürüst ancak aynı zamanda meraklı düşman olarak da bilinmektedir (Goldreich, 2004). Gerçek dünya uygulamasıyla oldukça ilgili olduğundan dolayı birçok çok taraflı hesaplama protokolü ve tasarımı, yarı dürüst modeli dikkate almaktadır (Barni vd., 2009). Son model olan kötü niyetli tarafta ise bazı tarafların, işlevin hesaplanmasını bozmak için yanlış veri aktarabileceğini kabul etmektedir.

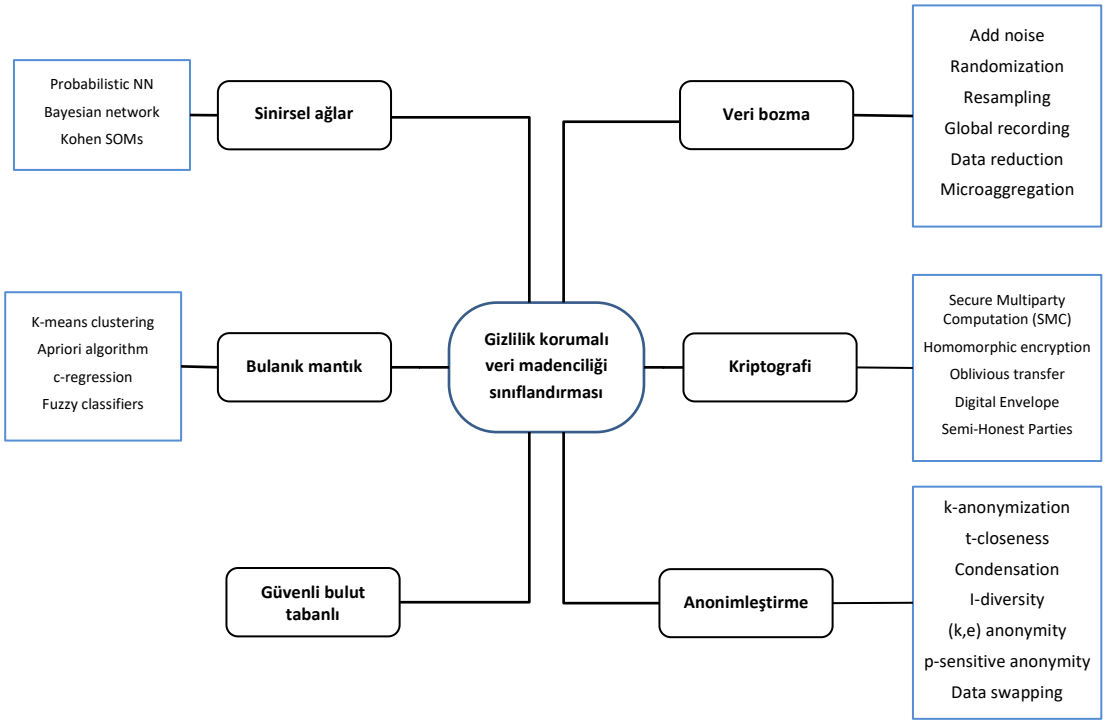
Ben-Or vd., (2019) ve Chaum vd. (1988) gizli veri paylaşıma dayalı çözümler önerdiler. Yao (1982), iki partili SMC için devre tasarımı önerirken, Micali vd., (1987) iki-taraflı yapıyı çok-taraflı yapıya geliştirmiştir. İki taraflı hesaplama ve çok taraflı hesaplama için önerilen protokoller arasında büyük farklılıklar vardır. Bogetoft vd. (2006), SMC tabanlı protokollerin ilk pratik uygulaması olarak güvenli bir müzayede modeli sunmaktadır. SMC protokolleri açık artırma, veri madenciliği ve e-oylama gibi birçok farklı alanda uygulanmaktadır.

2.4.3 Gizliliği Koruyan Veri Madenciliği

Gizlilik, birçok kişi ve kurum için önemli bir sorun haline gelmiştir. Günümüz internet toplumunda, kişiye özel verilerin toplanması, paylaşılması ve birçok alanda kullanılmasına yönelik büyük talepler gelmektedir. Söz konusu talepler kabul edilmeden bazı hizmetlerin kullanılmasına izin verilmemektedir. Cranor vd. (2000) tarafından yapılan ankete göre katılımcıların %17'sinin gizlilik konusunda oldukça katı olduğu görülmektedir. Öte yandan, yanıt verenlerin %56'sının veri kullanımı konusunda endişeli ve kalan %27'sinin ise çok az endişeli olduğu ifade edilmektedir. Durnell vd.'nin (2020) yaptıkları ankete göre katılımcıların en çok finansal verilerinin (%18,50) ardından kişisel verilerin gizliliğinden (%17,60) endişe ettiği belirtilmektedir. Uygulamaların gizlilik politikaları, verilerin yasal olarak toplanması, ifşa edilmesi ve

yönetilme biçimlerinin tanımlanması için kullanılmaktadır. Ancak kullanıcıların %74'ü kişisel verilerin gizliliği ile ilgili politikaları okumamaktadır (Obar ve Oeldorf-Hirsch 2018).

Veri madenciliğinde, verileri ve mahremiyeti koruma üzerine çalışmaları olan Agrawal ve Srikant (2000), gizliliği koruyan veri madenciliği (GKVM) terimini ilk defa ortaya atmaktadır. Agrawal ve Srikant, Lindell ve Pinkas (2000), Kantarcioglu ve Clifton'ın (2004) çalışmalarından sonra GKVM oldukça popüler olmuştur. GKVM'nin amacı, veri madenciliği süreci kullanılarak analiz edilen veri setinde bulunan hassas verilerin açığa çıkarılmasını önlemektir. GKVM'nin önemi, birçok ülkede kişisel verilerin korunması kanunlarının yürürlüğe girmesiyle daha fazla kabul görmektedir. GKVM'de kullanılan yöntemleri veri anonimleştirme, veri bozulması, krpitografi, nöral ağlar ve bulanık mantık algoritmaları başlıkları altında gruplayabiliriz. Veri madenciliği uygulamalarında yer alan dört farklı kullanıcı türü, veri sağlayıcı, veri toplayıcı, veri madenciliği ve karar verici olarak tanımlanmaktadır (Lei Xu vd. 2014). GKVM'de kullanılan yöntem ve tekniklerin sınıflandırılması şekil 2.10'da gösterilmektedir (Kreso vd. 2021).



Şekil 2.10 GKVM yöntem ve tekniklerin sınıflandırılması (Kreso vd. 2021)

Müşterilerin mahremiyetini korurken, onların ilgileri doğrultusunda ürünler önermek giderek daha fazla ilgi görmektedir. Kantarcioglu ve Clifton (2004) çalışmalarında, yatay olarak bölümlenmiş veriler üzerinde birliktelik kurallarının güvenli madenciliğini ele almaktadır. Paylaşılan bilgileri en aza indirmek için önerdikleri yöntemler kriptografik teknikler içermektedir. Polat ve Du (2005), müşterilere yönelik önerileri tahmin ederken müşterilerin özel verilerini maskelemek için rastgele dağıtılmış karıştırma tekniklerini kullanmaktadır. Kaleli ve Polat (2010), eşler arası ağlarda ikili derecelendirmeler hakkında özel tahminlerin nasıl sağlanacağını tartışmaktadır. İlaveten, ikili veriler üzerinde öneriler üretilirken gizliliği sağlamak adına rastgele karıştırma teknikleri gibi yöntemler kullanılmaktadır (Polat & Du, 2006). Zhang vd. (2006) ise müşterilerin her bir öğe için tercihlerini sunucunun rehberliğine göre maskeledikleri tahminleri kestirebilmek için iki yönlü bir iletişim gizliliğini koruma şeması önermektedir.

2.4.4 Güvenli Matris Çarpma Protokolü

Güvenli Matris Çarpma Protokolü (SMMP-Secure Matrix Multiplication Protocol) birçok araştırmacı tarafından önerilmektedir (Atallah vd. 2002; Lei vd. 2014). İlgili protokolde, Nassar vd. (2013) tarafından sunulan matris evirme protokolünde olduğu gibi aynı mimariye sahip şema tercih edilmektedir. Matrisler (A ve B), $A = A_1 + A_2$ ve $B = B_1 + B_2$'yi karşılayacak şekilde iki parçaya ayrılır ve her bir parça birbiriyle bağlantılı olmayan iki sunucuda saklanır. Protokol denklem 2.9'da ifade edilen temellere dayanmaktadır.

$$AB = (A_1 + A_2)(B_1 + B_2) = A_1B_1 + A_1B_2 + A_2B_1 + A_2B_2 \quad (2.9)$$

Burada $Sunucu_1$'in A_1 ve B_1 ve $Sunucu_2$'nin A_2 ve B_2 matrislerini depoladığı varsayılmaktadır. Bu durumda A_1B_1, $Sunucu_1$ tarafından $Sunucu_2$'ye ihtiyaç duyulmaksızın ve A_2B_2 de $Sunucu_2$ tarafından tek başına hesaplanabilir. A_1B_2 ve A_2B_1 terimleri ise ancak diğer bulut sunucularının yardımıyla hesaplanabilmektedir. Güvenli Matris Çarpma Protokolü, bulut sunucuları arasında en iyi durumda 4, en kötü durumda ise 8 matris aktarımı gerektirir.

2.4.5 Güvenli Matris Tersini Hesaplama Protokolü

Matrisin tersinin hesaplanması mühendislik ve bilimsel denklemlerde çok yaygın olarak kullanılan bir işlemdir. Bilgi güvenliğinin bir alt parçası olan ve bölüm 2.4.2'de bahsedilen Güvenli Çok Taraflı Hesaplama, özel girişleri ifşa etmeden iki veya daha fazla tarafın verilerini kullanarak bir girdinin çıktısını hesaplamayı sağlar. Nassar vd. (2013) Güvenli Matris Tersini Hesaplama Protokolünü (SMIP-Secure Matrix Inversion Protocol) önermektedir. Araştırmacılar çözümlerinde, A matrisini, A_1 + A_2 olan iki ayrı matris şeklinde parçaya ayırmaktadır. Her bir parça, birbiriyle bağlantılı olmayan iki bulut sunucuda depolanır. Böylelikle verinin tamamı tek bir sunucuda bulunmamaktadır. Denklem 2.10, Güvenli Matris Tersini Hesaplama işlemini vermektedir:

$$A^{-1} = A_1^{-1}(A_1^{-1} + A_2^{-1})^{-1}A_2^{-1} \tag{2.10}$$

(2.10) eşitliğinin sağ tarafı ters çevrilirse,

$$\begin{aligned}(A_1^{-1}(A_1^{-1} + A_2^{-1})^{-1}A_2^{-1})^{-1} &= A_2(A_1^{-1} + A_2^{-1})A_1 \\ &= A_2 A_1^{-1}A_1 + A_2 A_2^{-1}A_1 = A_2 I + I A_1 = A_2 + A_1 = A\end{aligned} \tag{2.11}$$

bulunmaktadır (Denklem 2.11). Güvenli Matris Tersini Hesaplama Protokolü'nde, ilgili hesaplamayı gerçekleştirmek için birbiriyle çakışmayan üç sunucuya ihtiyaç duyulmaktadır. A_1^{-1}, A_2^{-1} ve $A_1^{-1} + A_2^{-1}$ sırasıyla üç farklı sunucuda saklanır. A matrisinin tersi, Nassar vd.'nin (2013) çalışmalarında açıklandığı gibi güvenli 3-matris çarpım protokolü kullanılarak hesaplanmaktadır.

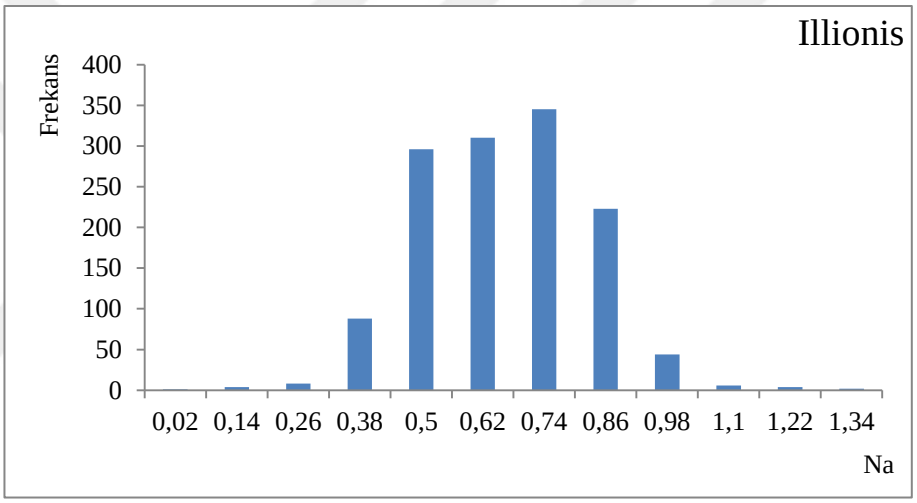
2.5 Veri Seti

Bu çalışmada, Birleşik Devletler Ulusal Jeokimyasal Araştırma Veritabanı'nda (U.S. National Geochemical Survey Database) yer alan üç gerçek veri seti kullanılmaktadır. Veri setlerinde Colorado, Illinois ve Wisconsin eyaletlerinde toprakta bulunan sodyum (Na) miktarı yer almaktadır. Söz konusu veri setinde, eyaletlere göre sırasıyla 1.150, 1.331 ve 912 örnek bulunmaktadır. Kullanılan veri setleri, sodyum içeriği için normal

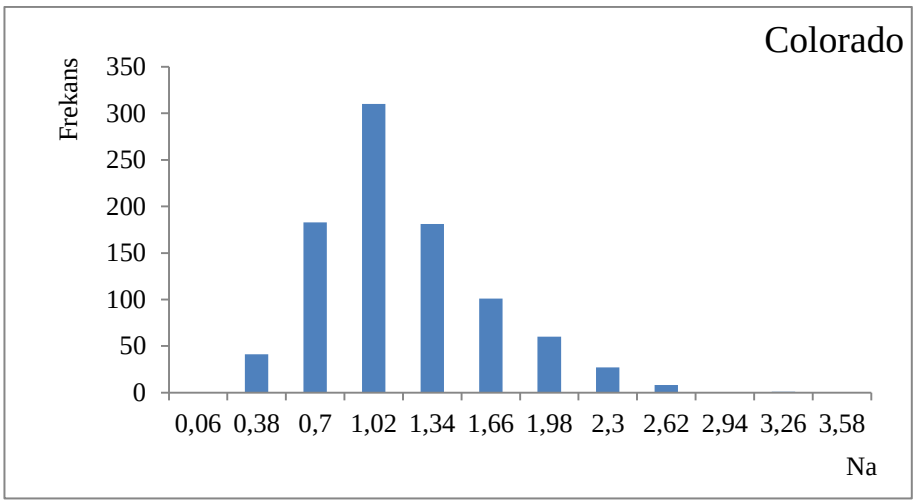
dağılım göstermektedir. İlgili durum şekil 2.11, şekil 2.12 ve şekil 2.13'de gösterilmektedir. Verilerin normal dağılması bulguların güvenirliği açısından önemli bir etkendir. Her üç veri setinin tanımlayıcı istatistik değerleri çizelge 2.1'de verilmektedir.

Çizelge 2.1 Illinois, Colorado ve Wisconsin veri setlerinin tanımlayıcı istatistikleri

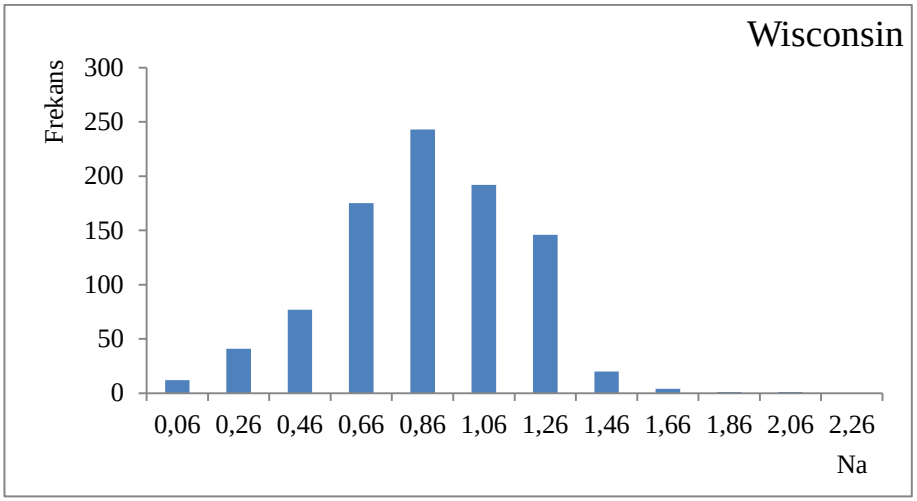
	Colorado	Illinois	Wisconsin
Ortalama	0.999	0.602	0.773
Medyan	0.912	0.597	0.771
Minimum	0.063	0.018	0.007
Maksimum	3.230	1.273	2.043
Standart sapma	0.460	0.162	0.299
Çarpıklık	0.912	0.111	-0.177
Basıklık	4.034	3.156	3.131



Şekil 2.11 Illinois veri setinin histogramı



Şekil 2.12 Colorado veri setinin histogramı



Şekil 2.13 Wisconsin veri setinin histogramı

2.6 Değerlendirme Metrikleri

Enterpolasyon yöntemlerini değerlendirmek için çok sayıda metrik vardır. Genelde, modellerin doğruluğunu karşılaştırmada, Ortalama Mutlak Hata (OMH) ve Kök Ortalama Kare Hatası (KOKH) kullanılmaktadır. OMH, tahmin edilen değerlerin ölçülen değerlerden ne kadar farklı olduğunu göstermektedir. Yerbilim çalışmalarında en çok tercih edilen model değerlendirme ölçütü olan KOKH ise büyük hataları daha fazla dikkate almaktadır (Chai ve Draxler 2014). OMH ve KOKH değerlerinin küçük olması, tahmin modelinin daha iyi sonuçlar ürettiğini göstermektedir. OMH ve KOKH, denklem 2.12 ve 2.13'de gösterildiği gibi hesaplanmaktadır.

$$OMH = \left[\frac{1}{n}\sum_{i=1}^{n}|\hat{z}(x_i, y_i) - z(x_i, y_i)|\right] \tag{2.12}$$

$$KOKH = \left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{z}(x_i, y_i) - z(x_i, y_i)\right)^2\right]^{1/2} \tag{2.13}$$

Burada n veri kümesindeki toplam gözlemlenen değer sayısıdır. $z(x_i, y_i)$, (x_i, y_i) noktasında gözlenen değerdir, $\hat{z}(x_i, y_i)$ ise aynı noktadaki tahmin değeridir.

Değerlendirme metrikleri kullanılarak yapılan doğruluk analizinin amacı, işbirliğinin etkilerini göstermektir. İki veri sahibinin verilerinin entegrasyonunun daha doğru tahmin modelleri üreteceği belirgindir. Söz konusu hipotezi kanıtlamak adına yapılan deneyler ile paylaşılan veriler üzerine bir tahmin modeli inşa edildiğinde, doğruluğun ne kadar arttığı görülmektedir.

3. İLGİLİ ÇALIŞMALAR

Konumsal veri analizi son zamanlarda kullanıcıların cep telefonlarında yaygın olarak kullandıkları birçok uygulama aracıyla büyük boyutta verinin toplamasıyla daha çok müşteri bulmak veya iyi hizmet vermek amacıyla kurumlar tarafından tercih edilmektedir. Böylece bir firma yeni bir mağaza açmak istediğinde önce bölgedeki potansiyel müşteri gruplarını belirlemekte ve rakip firmaların var olan mağazalarına göre kendi mağazasının konumuna karar verebilmektedir. Benzer şekilde ambulans ve itfaiye merkezleri gibi acil durum müdahale merkezlerinin konumu insan ve trafik yoğunluğu durumuna göre araçların en kısa zamanda ulaşabilmesi için ey uygun yer belirlemesini yine konuma bağlı analizler ile yapılabilmektedir (Y. Liu vd., 2021; Yin & Mu, 2012)

Konum verisine bağlı diğer çalışmalar bilinen ölçüm değerlerini kullanarak değeri bilinmeyen noktalar için bir tahmini değer üreten konumsal enterpolasyon yöntemleridir. Bu alanda ilk çalışma Krige tarafından kömür madenlerinde kömür rezervlerin konumu ve miktarını belirlemek amacıyla gerçekleştirilmiştir. Bu yöntem daha sonraları geliştirilmesiyle Kriging adıyla kullanılmaya başlanmıştır. Benzer bir şekilde Shepard (1968) tarafından ortaya atılan TMA yöntemi nesnelerin birbirlerine benzerliğinin mesafeyle ters orantılı bir şekilde hesaplanmasına dayanmaktadır. Sonraları her iki yönteme bağlı ve yeni teknikler içeren birçok çalışma ortaya atılmıştır. Li ve Heap (2014), yaygın olarak uygulanan yöntemleri özelliklerine göre sınıflandırmıştır. Tüm konumsal enterpolasyon yöntemleri üç farklı sette sınıflandırmıştır: jeositatistik olmayan, jeoistatistik ve hibrit yöntemler. Meng vd. (2013), yedi farklı konumsal enterpolasyon yöntemini üç gerçek veri seti kullanarak elde ettikleri sonuçları karşılaştırmışlardır. İlgili araştırmacılar deterministik yöntemlerin hesaplanmasının basit ve anlaşılmasının kolay olduğu sonucuna varırlar.

YEA, yaygın olarak uygulanan konumsal enterpolasyon yöntemlerindendir. YEA, belirli bir bölgedeki örneklenmiş veri noktalarından geçen en uygun polinom yüzeyini bulmayı amaçlar. Burada YEA, örneklenen verilerin genel eğilimlerini tespit etmek için kullanılır. Ayrıca yerel ve global olmak üzere iki farklı seçeneği vardır. Yerel

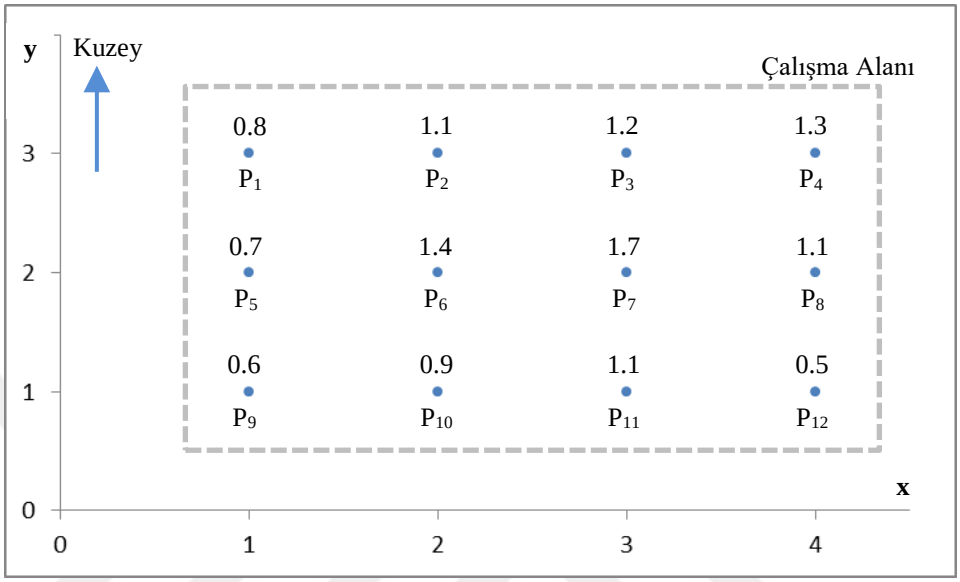
yöntemler, ölçülmeyen noktaların çevresinde sadece belirli sayıda örneklenmiş veri kullanarak en iyi yüzeyi belirler. Global yöntemler ise ilgili bölge için en iyi yüzeyi bulmada örneklenmiş verilerin tümünü kullanmaktadır. İlk kez YEA'yı Oldham ve Sutherland (1955) önermişlerdir. Agterberg (1984), YEA'nın temellerini tanımlayıp, çevre ile ilgili çalışmalarda uygulama olanağı bulmuşlardır. Grohmann (2005) morfometrik parametreler olan izobaz ve hidrolik gradyandan gelen veriler üzerinde YEA'yı gerçekleştirmiştir. Burada ikinci ve altıncı dereceden yüzey ifadelerinin en iyi doğruluğu sağladığı sonucuna varmıştır. Luo vd. (2008) İngiltere ve Galler'den elde edilen rüzgâr verilerinin oluşturduğu yüzeyleri oluşturmak için YEA ve diğer altı konumsal enterpolasyon yöntemini uyguladılar. Yao vd. (2014), Kuzey Çin'deki yeraltı suyu seviyesini belirlemek için YEA ve diğer birçok konumsal enterpolasyon yöntemini kullanmaktadır. Yuan vd. (2015), Çin genelinde günlük iklim değişkenlerinin enterpolasyon sonuçlarını değerlendirirler. Çalışmalarında YEA'nın düşük irtifa ve homojen arazi koşulları için en iyi sonuçlar ürettiğini öne sürmektedirler. Gao ve Prasad (2016), binalar ve fabrikalar gibi kapalı alanlarda pozisyon bulmak için geliştirdikleri mobil uygulamada YEA'yı konumsal bir enterpolasyon aracı olarak kullandılar. Tisseuil vd. (2016), biyolojik yayılım hızındaki konumsal değişimi ölçmek için YEA ve diğer konumsal enterpolasyon yöntemlerini kullanırlar. Sonuç olarak YEA'nın farklı örnekleme koşulları ve yayılım modelleri altında en güvenilir yöntemlerden biri olduğunu öne sürmektedirler.

Araştırmacılar, aynı veya komşu bölgeler için topladıkları veri analizi çözümlerini üç farklı veri dağıtım mimarisi üzerine kurmaktadırlar. Bunlar yatay, dikey ve hibrittir. Abdel Wahab vd. (2014) yatay olarak bölümlenmiş veriler üzerinde gizliliği koruyan bir birliktelik kuralı madenciliği önerdi. Çalışmalarında dağıtılmış veri tabanı mimarisinde özel verileri koruyan birliktelik kuralı sorguları geliştirdiler. Sun vd. (2015), gizliliği koruyan bir destek vektör makinesi çözümü sunmuştur. Çözümleri, sınıflandırma amacıyla dikey olarak bölümlenmiş veriler üzerinde tasarlanmıştır. Yang vd. (2015), bulut sunucular üzerinde dikey olarak bölümlenmiş tıbbi verilerin paylaşımına yönelik bir prototip önermişlerdir.

Konumsal enterpolasyon yöntemleri ile ilgili önceki çalışmalar veri mahremiyetini dikkate almamıştır. Konumsal veriye sahip taraflarının gizliliğini koruyan enterpolasyon yöntemleri literatürde ilk defa Tuğrul ve Polat tarafından çalışılmıştır. Yazarlar hem TMA hem de Kriging yöntemlerinin 3 farklı veri dağıtım mimarisi altında inceleyerek çözümler önermişlerdir. İlk makale (Tugrul & Polat, 2013), tek parti mimarisine dayalı güvenli bir Kriging modelinin nasıl oluşturulacağını ele almaktadır. Bu şemada iki taraf vardır; bir istemci ve sunucu. Sunucu, bir Kriging modeli oluşturmak için gerekli tüm örneklenmiş verileri tutar. Model ve örneklenen veriler, sunucunun gizli verileri olarak kabul edilir. İstemci, veri sahibinden belirli bir koordinat için bir tahmin değeri ister. Tahmin değeri ve koordinat verisi istemci için gizli veri olarak kabul edilmiştir. Diğer iki makale ise iki ve çok partili mimari yapısında çözümler önermektedir (Tugrul & Polat, 2014a, 2014c). Bu çalışmalara ek olarak, tek parti mimarisi için mahremiyete dayalı TMA üzerine bir çalışma da yayınlamışlardır (Tugrul & Polat, 2014b).

4. YÖNTEM

Bu bölümde, i) Merkezi ii) Dağıtılmış verilerde ve iii) Bulut ortamları için geliştirilen üç yeni gizlilik-korumalı tabanlı YEA yaklaşımları anlatılmaktadır.



Şekil 4.1 Örnek çözüm için koordinatlar ve ölçüm değerleri

İkinci dereceden bir modelin katsayılarının çözümü aşağıdaki gibidir. Öncelikle (x, y) koordinatlarındaki z değerleri ile çizelge 4.1'de gösterildiği gibi matris hesaplamasında kullanılacak ortalamalar hesaplanır.

Çizelge 4.1 Matris hesaplamasında kullanılacak ortalamalar

	x	y	z	x^2	y^2	xy	yz	zx	x^3	y^3	xy^2	x^2y	x^2y^2	x^3y	xy^3	x^4	y^4	x^2z	y^2z	xyz
1	1	3	0,8	1	9	3	2,4	0,8	1	27	9	3	9	3	27	1	81	0,8	7,2	2,4
2	2	3	1,1	4	9	6	3,3	2,2	8	27	18	12	36	24	54	16	81	4,4	9,9	6,6
3	3	3	1,2	9	9	9	3,6	3,6	27	27	27	27	81	81	81	81	81	10,8	10,8	10,8
4	4	3	1,3	16	9	12	3,9	5,2	64	27	36	48	144	192	108	256	81	20,8	11,7	15,6
5	1	2	0,7	1	4	2	1,4	0,7	1	8	4	2	4	2	8	1	16	0,7	2,8	1,4
6	2	2	1,4	4	4	4	2,8	2,8	8	8	8	8	16	16	16	16	16	5,6	5,6	5,6
7	3	2	1,7	9	4	6	3,4	5,1	27	8	12	18	36	54	24	81	16	15,3	6,8	10,2
8	4	2	1,1	16	4	8	2,2	4,4	64	8	16	32	64	128	32	256	16	17,6	4,4	8,8
9	1	1	0,6	1	1	1	0,6	0,6	1	1	1	1	1	1	1	1	1	0,6	0,6	0,6
10	2	1	0,9	4	1	2	0,9	1,8	8	1	2	4	4	8	2	16	1	3,6	0,9	1,8
11	3	1	1,1	9	1	3	1,1	3,3	27	1	3	9	9	27	3	81	1	9,9	1,1	3,3
12	4	1	0,5	16	1	4	0,5	2	64	1	4	16	16	64	4	256	1	8	0,5	2
Ort.	**2,5**	**2**	**1,03**	**7,5**	**4,67**	**5**	**2,18**	**2,71**	**25**	**12**	**11,67**	**15**	**35**	**50**	**30**	**88,5**	**32,67**	**8,18**	**5,19**	**5,76**

Hesaplanan ortalama değerleri denklem 2.4'teki gibi matrise yerleştirilir. Daha sonra gerekli hesaplamalar yapılarak $(a_0, a_1, a_2, a_3, a_4, a_5)$ katsayı değerleri bulunur (Şekil 4.2).

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 1 & 2{,}5 & 2 & 7{,}5 & 5 & 4{,}67 \\ 2{,}5 & 7{,}5 & 5 & 25 & 15 & 11{,}67 \\ 2 & 5 & 4{,}67 & 15 & 11{,}67 & 12 \\ 7{,}5 & 25 & 15 & 88{,}5 & 50 & 35 \\ 5 & 15 & 11{,}67 & 50 & 35 & 30 \\ 4{,}67 & 11{,}67 & 12 & 35 & 30 & 32{,}67 \end{bmatrix}^{-1} \begin{bmatrix} 1{,}03 \\ 2{,}71 \\ 2{,}18 \\ 8{,}18 \\ 5{,}76 \\ 5{,}19 \end{bmatrix}$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} -1{,}4579 \\ 1{,}0858 \\ 1{,}3204 \\ -0{,}22 \\ 0{,}0604 \\ -0{,}3254 \end{bmatrix}$$

Şekil 4.2 YEA modelindeki katsayıların hesaplanması

Böylece, şekil 4.1'de temsil edilen alanda tahmin üretmek için geliştirilen ikinci dereceden modelimiz $z = -1{,}4579 + 1{,}0858x + 1{,}3204y - 0{,}22x^2 + 0{,}0604y^2 - 0{,}3254xy$ şeklinde olacaktır. Daha üst dereceden denklemler de benzer şekilde çözülmektedir.

4.1 Gizlilik-Korumalı Yüzey Eğilimi Analizi (GKYEA)

Veri gizliliğini önemsemeden YEA enterpolasyon tekniğini kullanarak belirli konumlar için bilinmeyen değerleri tahmin etmek kolay bir iştir. Temel olarak istemci, tahmin değeri almak istediği konumun bilgilerini veri sahibine gönderir. Veri sahibi, denklem 2.2'te verilen YEA enterpolasyonu ile sahip olduğu ölçümleri kullanarak tahmin değerini hesaplar ve sonucu istemciye gönderir. Geleneksel YEA'da konum ve tahmin değerlerinin açık bir şekilde istemci-veri sahibi arasında iletilmesi gizlilik endişesini öne çıkarmaktadır.

4.1.1 Yöntem

Geleneksel YEA'da gizlilik endişesinin olmaması önemli bir problemdir. Hâlbuki istemciler ve sunucular korunmasız bir durumda olabilir. Ayrıca kuruluşların geleceği

sahip oldukları verilere bağlıdır. Bu nedenle istemciler ve veri sahipleri, özel verilerini açıklamadan bilimsel hesaplamaları yapmak zorundadır. Geliştirdiğimiz Gizlilik-Korumalı Yüzey Eğilim Analizi (GKYEA) yöntemi, bölüm 2.4.1'de bahsedilen Paillier şifreleme ile hem istemciler hem de sunucular için gizlilik sağlamaktadır. Ortaya konulan çözümün adımları;

i. İstemci, tahmine ihtiyaç duyduğu hedef koordinatlara (x_t, y_t) karar verir.

ii. Veri sahibi, örneklem verileri kullanarak en uygun yüzey eğilim tahmin modelini oluşturur.

iii. Veri sahibi, tahmin modeline göre bir giriş dizisi oluşturur. Model ikinci dereceden olduğu varsayıldığında denklem 4.1'teki gibi olacaktır.

$$z = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy \tag{4.1}$$

Dolayısıyla giriş dizisi $[1, x, y, x^2, y^2, xy]$ şeklindedir. Söz konusu model, daha yüksek dereceli denklemlere de uygulanabilmektedir. Veri sahibi tahmin modelinin katsayılarını $[a_1, a_2, a_3, a_4, a_5]$ açık veya okunabilir bir biçimde göndermemelidir; aksi takdirde istemci modeli öğrenecektir.

iv. İstemci, her bir giriş dizisinin karşılık gelen değerini hesaplar $[1, x_t, y_t, x_t^2, y_t^2, x_t y_t]$

Ardından istemci, Paillier homomorfik şemasını kullanarak giriş dizisini özel anahtarı ile şifreler ve elde ettiği değerleri $[\varepsilon_c(1), \varepsilon_c(x_t), \varepsilon_c(y_t), \varepsilon_c(x_t^2), \varepsilon_c(y_t^2), \varepsilon_c(x_t y_t)]$ veri sahibine gönderir. Veri sahibi, istemcinin anahtarını bilmediği sürece, gönderilen değerlerin şifresini çözemez ve hedef koordinatı öğrenemez.

v. Veri sahibi şifrelenmiş giriş dizisini tahmin modelinin katsayıları ile üstel işleme tabi tutar.

$[\varepsilon_c(1)^{a_1}, \varepsilon_c(x_t)^{a_2}, \varepsilon_c(y_t)^{a_3}, \varepsilon_c(x_t^2)^{a_4}, \varepsilon_c(y_t^2)^{a_5}, \varepsilon_c(x_t y_t)^{a_6}]$.

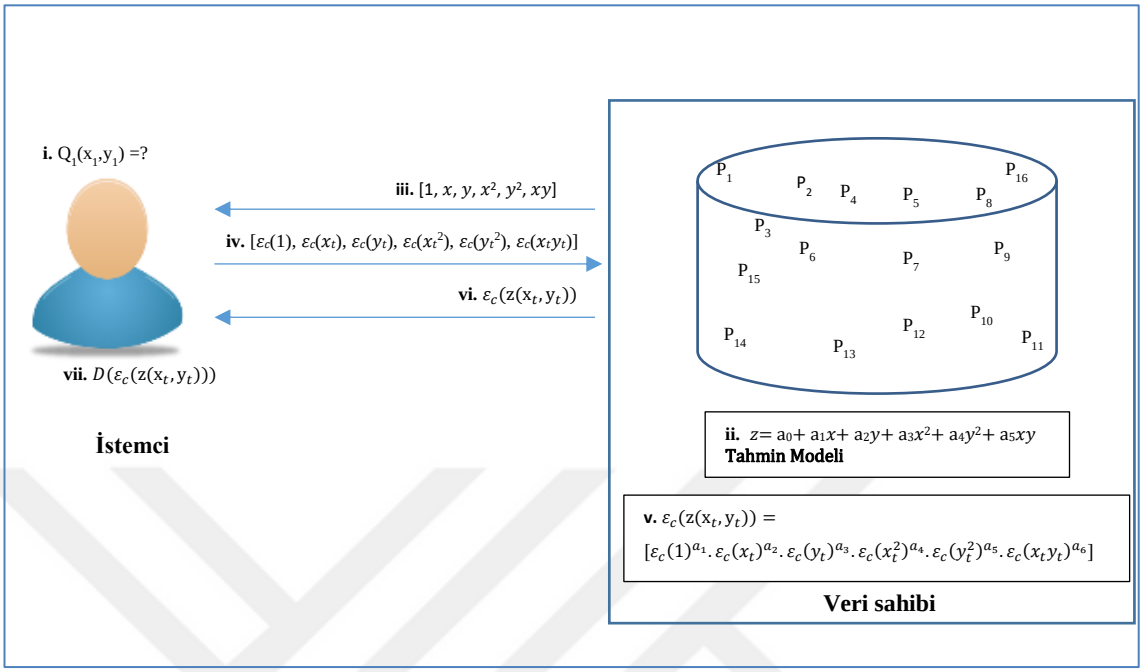
vi. Daha sonra veri sahibi, şifrelenmiş tüm değerleri çarpar.

$\varepsilon_c(\mathrm{z}(\mathrm{x}_t, \mathrm{y}_t)) = [\varepsilon_c(1)^{a_1}.\varepsilon_c(x_t)^{a_2}.\varepsilon_c(y_t)^{a_3}.\varepsilon_c(x_t^2)^{a_4}.\varepsilon_c(y_t^2)^{a_5}.\varepsilon_c(x_t y_t)^{a_6}]$

Sonuç z'nin şifrelenmiş değerine eşittir.

vii. Son olarak, istemci tahmin değerini şifrelenmiş biçimde $\varepsilon_c(\mathrm{z}(\mathrm{x}_t, \mathrm{y}_t))$ alır ve şifresini yalnızca istemci tarafından bilinen anahtarla çözer.

GKYEA işlem adımları şekil 4.3'de gösterilmektedir.



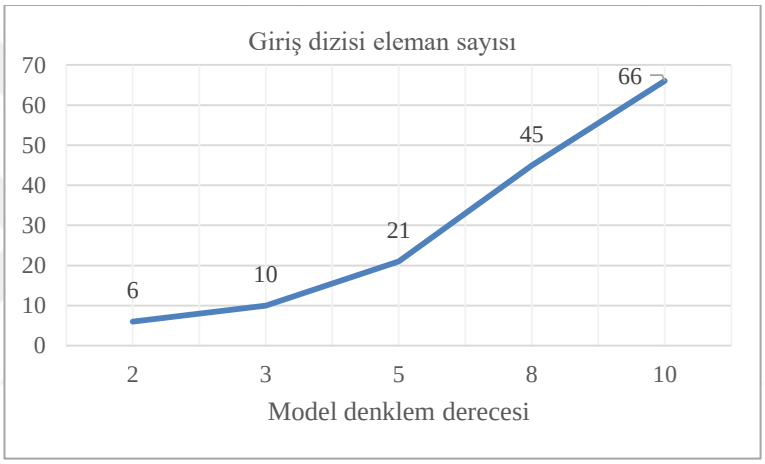
Şekil 4.3 Gizlilik-Korumalı YEA modeli ve işlem adımları

4.1.2 Maliyet analizi

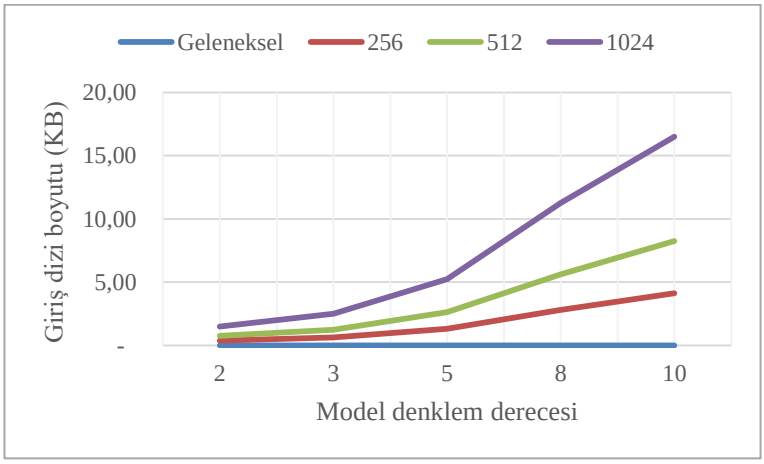
Bu kısımda, GKYEA'nın ek depolama, iletişim ve hesaplama maliyeti analizleri yapılmıştır. Geleneksel YEA'da istemci sadece hedef koordinatı ve tahmin değerini saklamaktadır. Önerilen GKYEA'da ise istemcinin, sunucu tarafından gönderilen modelin giriş dizi sırasını ve karşılık gelen değerlerini kaydetmesi gerekmektedir. Giriş dizisinin uzunluğu da sunucu tarafından oluşturulan modele bağlıdır. Örnek olarak, model ikinci dereceden ise giriş dizisi $[1, x, y, x^2, y^2, xy]$ ve karşılık gelen değerleri $[1, x_t, y_t, x_t^2, y_t^2, x_ty_t]$ şeklinde olacaktır.

Öte yandan, geleneksel YEA'da sunucu örneklenmiş verileri, modeli ve tahmin değerini saklar. GKYEA'da ise sunucunun, modelin giriş sırasına karşılık gelen açık değerlerini $[1,\ x,\ y,\ x^2,\ y^2,\ xy]$ ve istemciden gelen şifreli $[\varepsilon_c(1),\ \varepsilon_c(x_t),\ \varepsilon_c(y_t),\ \varepsilon_c(x_t^2),\ \varepsilon_c(y_t^2),\ \varepsilon_c(x_ty_t)]$ değerlerini depolaması gerekmektedir.

Geleneksel YEA'da, istemci sadece hedef koordinatı (x, y) sunucuya gönderir ve sunucu da hesaplanan tahmin değerini istemciye gönderir. İstemci ve sunucu arasındaki toplam iletişim sayısı ikidir. Önerilen çözümde ise açık/şifresiz $[1, x, y, x^2, y^2, xy]$ ve şifreli $[\varepsilon_c(1), \varepsilon_c(x_t), \varepsilon_c(y_t), \varepsilon_c(x_t^2), \varepsilon_c(y_t^2), \varepsilon_c(x_t y_t)]$ giriş dizileri sunucu ve istemci arasında değiştirilmelidir. Bu nedenle oluşan iletişim sayısı artmaktadır. Tahmin modelinin denklem derecesine bağlı olarak sunucu ve istemci arasında transfer edilen giriş dizisi eleman sayısının değişimi Şekil 4.4'de gösterilmektedir. İstemcinin tahmin modelinin denkleme derecesi ve şifreleme anahtarının boyutuna bağlı olarak değişen giriş dizisi boyutu ise şekil 4.5'te gösterilmektedir.



Şekil 4.4 Model denklem derecesine göre model giriş dizisi eleman sayısı

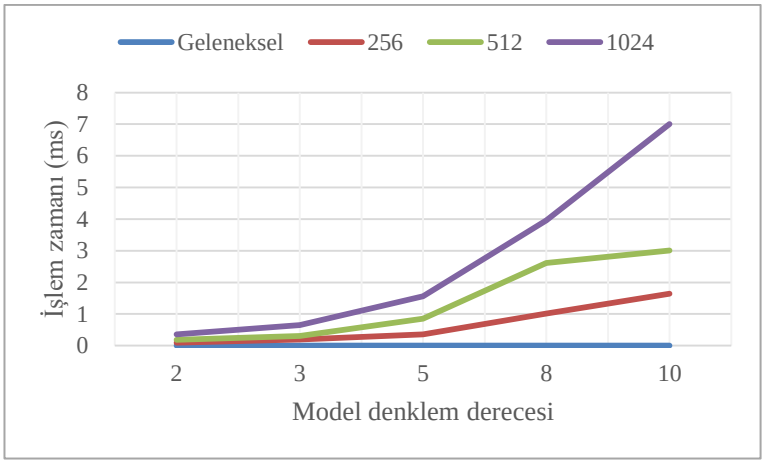


Şekil 4.5 Model denklem derecesine göre model giriş dizisi boyutu

İstemci, geleneksel YEA'da herhangi bir hesaplama yapmamaktadır. Gerekli tüm hesaplamalar sunucu tarafından yapılır. Ancak GKYEA'da, istemcinin giriş sırasını şifrelemesi ve sunucu tarafından hesaplanıp şifreli biçimde gönderilen tahmin değerini kendi anahtarı ile çözmesi gerekmektedir. Geleneksel YEA'da, sunucu hedef koordinatı istemciden alır ve modeli önceden oluşturur. Bu nedenle, tahmin değerini üretmek için gerekli tüm verilere sahiptir. Öte yandan, GKYEA'da tahmin değerini hesaplamak için sunucunun $[\varepsilon_c(1)^{a_1},\ \varepsilon_c(x_t)^{a_2},\ \varepsilon_c(y_t)^{a_3},\ \varepsilon_c(x_t^2)^{a_4},\ \varepsilon_c(y_t^2)^{a_5},\ \varepsilon_c(x_t y_t)^{a_6}]$ ve $[\varepsilon_c(1)^{a_1}.\varepsilon_c(x_t)^{a_2}.\varepsilon_c(y_t)^{a_3}.\varepsilon_c(x_t^2)^{a_4}.\varepsilon_c(y_t^2)^{a_5}.\varepsilon_c(x_t y_t)^{a_6}]$ yi hesaplaması gerekmektedir. Şifrelemede kullanılan anahtar boyutu arttıkça şifrelenmiş sayının boyutu da artmaktadır. Yine aynı şekilde tahmin modelinin denklem derecesi arttıkça modelde kullanılan değişken sayısı da artmaktadır. Örneğin ikinci derece model 6 değişken içerirken onuncu derece modelde 66 değişken yer almaktadır. Tahmin modelinin denklem derecesi ve şifreleme anahtarının uzunluğuna bağlı olarak işlem süresi artmaktadır. Çizelge 4.2 ve şekil 4.6'da tahmin modeli denklem derecesi ve anahtar uzunluğuna bağlı olarak hesaplama süreleri gösterilmektedir. Geliştirilen tüm modellere ait ölçümler, Intel i7 1.7 GHz CPU ve 4 GB bellekli bilgisayarda Java Sanal Makinesi üzerinde çalıştırılarak elde edilmiştir. İstemci ve sunucular, performans karşılaştırma amacıyla aynı makinede sırayla çalıştırılmıştır. Ölçümlerde sunucu-istemci arasında ağ üzerinden veri aktarım maliyetleri göz ardı edilmiştir.

Çizelge 4.2 Model denklem derecesi ve anahtar boyutuna bağlı işlem zamanı (milisaniye)

Denklem derecesi	**Geleneksel**	**Anahtar boyutu**		
		256 bit	**512 bit**	**1024 bit**
2	0,00089	0,09	0,18	0,36
3	0,00089	0,20	0,30	0,65
5	0,00089	0,35	0,85	1,56
8	0,00133	1,02	2,62	3,96
10	0,00178	1,64	3,01	7,00



Şekil 4.6 Model denklem derecesi ve anahtar boyutuna göre sunucu işlem zamanı

4.1.3 Doğruluk analizi

Gizlilik koruma yaklaşımları, kullandıkları yöntemler nedeniyle tahmin modelinin doğruluğunu kötüleştirebilir. Verilere rasgele değer ekleyerek gizlilik koruma sağlayan tekniklerde durum böyledir (Chen vd., 2011). Önerdiğimiz yöntemde homomorfik şifreleme algoritması olan Paillier şifreleme sistemi kullanılmaktadır. Paillier şifreleme sistemi, düz metinlerde Homomorfik toplama ve çarpma işlemi olmak üzere iki özelliği yerine getirmektedir. İlgili sistem, aynı zamanda deterministik olduğundan, şifre metinleri üzerinde hesaplama yapmak sonucu etkilememektedir. Bu nedenle, istemci gizlilik önlemi yokmuş gibi kesin bir tahmin değeri alır. Sonuç olarak, GKYEA geleneksel şemada olduğu gibi aynı tahmin değerini üretmektedir. Böylece, hem istemcinin hem de sunucunun verileri korunmuş olmaktadır.

4.1.4 Gizlilik analizi

Tahmin değerinin hesaplanması sırasında hem istemcinin hem de sunucunun verilerinin birbirlerine ve üçüncü taraflara açıklanmamasını sağlamak adına önerilen çözümümüz doğrulanmalıdır. Hedef koordinat ve tahmin değeri istemcinin özel verileri olarak kabul edilmektedir. Ayrıca, sunucu tahmin modelinin ve örneklenmiş verilerin katsayıları da istemciden gizli tutulmalıdır. Geliştirilen yöntemde, sunucunun tahmin modelinin derecesinin öğrenilmesi, hassas bilgileri ortaya çıkarmamaktadır. Hedef koordinatta

tahmin değeri üretmek için gerekli tüm veri transferleri ve hesaplamaları şifreli olarak yapılır. Paillier yöntemi (1999), istemcinin özel anahtarını diğer kullanıcılardan gizlediği sürece Homomorfik şifrelemenin güvenilir olduğunu açıkça belirtmektedir.

4.2 Dağıtılmış Verilerde Gizlilik-Korumalı Yüzey Eğilimi Analizi (D-GKYEA)

Konumsal enterpolasyon yöntemlerinin güvenilirliği ve doğruluğu, çalışma bölgesinde toplanan veri miktarına bağlıdır. Elde edilecek veri miktarını artırmak da çok büyük miktarda zaman ve bütçe gerektirmektedir. Geçmiş zamanlara ait verileri toplama ise mümkün olmamaktadır. Bazı durumlarda farklı kuruluşlar, hatta rakip kuruluşlar bile aynı bölgeden veri toplayabilmektedirler. Söz konusu kuruluşlar tarafından toplanan veriler birleştirilebilirse daha doğru tahmin modellerinin oluşturulması mümkün olacaktır. Ancak gizlilik endişesi nedeniyle kuruluşlar verilerini paylaşmayabilirler. Birçok ülkede kişisel, sağlık ve mali kayıtlar gibi özel verilerin açık bir şekilde paylaşılması yasalarla engellenmiştir. Bu nedenle, tarafların veri gizliliğini sağlayan bir çözüm, tarafları işbirliği yapmaya teşvik edebilir.

Bu bölümde, hem istemcilerin hem de veri sahiplerinin gizliliğini sağlayan, dağıtılmış verilerde gizlilik-korumalı yüzey eğilim analizi yöntemi (D-GKYEA) önerilmektedir. Çözümümüz, gizliliklerini ihlal etmeden iki veri sahibinin verileri birleştirerek daha doğru ve güvenilir bir tahmin modeli sağlamaktadır. Geliştirilen D-GKYEA, maliyet, doğruluk ve gizlilik açısından analiz edilmiştir. Gizliliği koruyan çözüm ekstra depolama, hesaplama ve iletişim maliyetleri getirse de, enterpolasyon sürecinin genel performansını etkilememektedir.

4.2.1 Yöntem

D-GKYEA yönteminde veriler iki sunucuda bulunduğundan, istemci-sunucu ve sunucu-sunucu şeklinde iki tür iletişim yapılmaktadır. İstemci-veri sahibi ve veri sahibi-veri sahibi arasındaki iletilen verilerin gizliliğinin korunması önemlidir ve önerilen GKYEA'da gerekli önlem Homomorfik şifreleme ile yapılmaktadır. Ayrıca sunucu-

sunucu arasındaki verilerin gizliliğin temini ile birleştirilmesi için bölüm 2.5.1'deki SMC yaklaşımı kullanılmaktadır.

Sunucularda verilerin yatay olarak bölümlendiğini varsayılmaktadır. Bu durumda veri sahipleri aşağıdaki gibi veriye sahiptir.

$$A = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_k & y_k & z_k \end{bmatrix} \qquad B = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_m & y_m & z_m \end{bmatrix}$$

Sunucu A (S_A) Verileri — Sunucu B (S_B) Verileri

İkinci derece polinom ifadesinin katsayılarını hesaplamak için denklem 2.4'e göre her iki sunucunun da şu değerlere sahip olması gerekir:

$$\left[\bar{x}, \bar{y}, \bar{z}, \overline{x^2}, \overline{xz}, \overline{xy}, \overline{yz}, \overline{y^2}, \overline{x^3}, \overline{x^2y}, \overline{x^2z}, \overline{xyz}, \overline{y^2z}, \overline{xy^2}, \overline{y^3}, \overline{x^4}, \overline{x^3y}, \overline{x^2y^2}, \overline{xy^3}, \overline{y^4}\right]$$

Çözümümüzde iki sunucu (S_A ve S_B) vardır ve ortak verileri üzerinde bir tahmin modeli oluşturmaktadırlar. Çözümümüz, A notasyonu ile ifade edilen S_A üzerinden açıklanmıştır. S_B için model oluşturulmak istendiğinde de aynı adımlar takip edilmektedir (Şekil 4.7).

i. S_A ilk önce şu hesaplamaları yapmalıdır.

$$[V_1^A=\Sigma_{i=1}^k x_i,\ V_2^A=\Sigma_{i=1}^k y_i,\ V_3^A=\Sigma_{i=1}^k z_i,\ V_4^A=\Sigma_{i=1}^k x_i^2,\ V_5^A=\Sigma_{i=1}^k x_i z_i,\ V_6^A=\Sigma_{i=1}^k x_i y_i,\ V_7^A=\Sigma_{i=1}^k y_i z_i,$$
$$V_8^A=\Sigma_{i=1}^k y_i^2,\ V_9^A=\Sigma_{i=1}^k x_i^3,\ V_{10}^A=\Sigma_{i=1}^k x_i^2 y_i,\ V_{11}^A=\Sigma_{i=1}^k x_i^2 z_i,\ V_{12}^A=\Sigma_{i=1}^k x_i y_i z_i,\ V_{13}^A=\Sigma_{i=1}^k y_i^2 z_i,$$
$$V_{14}^A=\Sigma_{i=1}^k x_i y_i^2,\ V_{15}^A=\Sigma_{i=1}^k x_i^3,\ V_{16}^A=\Sigma_{i=1}^k x_i^4,\ V_{17}^A=\Sigma_{i=1}^k x_i^3 y_i,\ V_{18}^A=\Sigma_{i=1}^k x_i^2 y_i^2,\ V_{19}^A=\Sigma_{i=1}^k x_i x_i^3,$$
$$V_{20}^A=\Sigma_{i=1}^k x_i^4\ ]$$

ii. Aynı şekilde S_B de şu hesaplamaları yapmalıdır.

$$[V_1^B=\Sigma_{i=1}^m x_i,\ V_2^B=\Sigma_{i=1}^m y_i,\ V_3^B=\Sigma_{i=1}^m z_i,\ V_4^B=\Sigma_{i=1}^m x_i^2,\ V_5^B=\Sigma_{i=1}^m x_i z_i,\ V_6^B=\Sigma_{i=1}^m x_i y_i,\ V_7^B=\Sigma_{i=1}^m y_i z_i,$$
$$V_8^B=\Sigma_{i=1}^m y_i^2,\ V_9^B=\Sigma_{i=1}^m x_i^3,\ V_{10}^B=\Sigma_{i=1}^m x_i^2 y_i,\ V_{11}^B=\Sigma_{i=1}^m x_i^2 z_i,\ V_{12}^B=\Sigma_{i=1}^m x_i y_i z_i,\ V_{13}^B=\Sigma_{i=1}^m y_i^2 z_i,$$
$$V_{14}^B=\Sigma_{i=1}^m x_i y_i^2,\ V_{15}^B=\Sigma_{i=1}^m x_i^3,\ V_{16}^B=\Sigma_{i=1}^m x_i^4,\ V_{17}^B=\Sigma_{i=1}^m x_i^3 y_i,\ V_{18}^B=\Sigma_{i=1}^m x_i^2 y_i^2,\ V_{19}^B=\Sigma_{i=1}^m x_i x_i^3,$$
$$V_{20}^B=\Sigma_{i=1}^m x_i^4\ ]$$

iii. S_A, her bir *V* için olmak üzere $p_1, p_2, \dots, p_{20}$ şeklinde rasgele bir tamsayı üreterek saklar.

iv. S_A ürettiği tüm rasgele tamsayıları açık anahtarı ile şifreleyerek S_B'ye gönderir. $\xi_{K_A}(p_1), \xi_{K_A}(p_1), \dots, \xi_{K_A}(p_{20})$

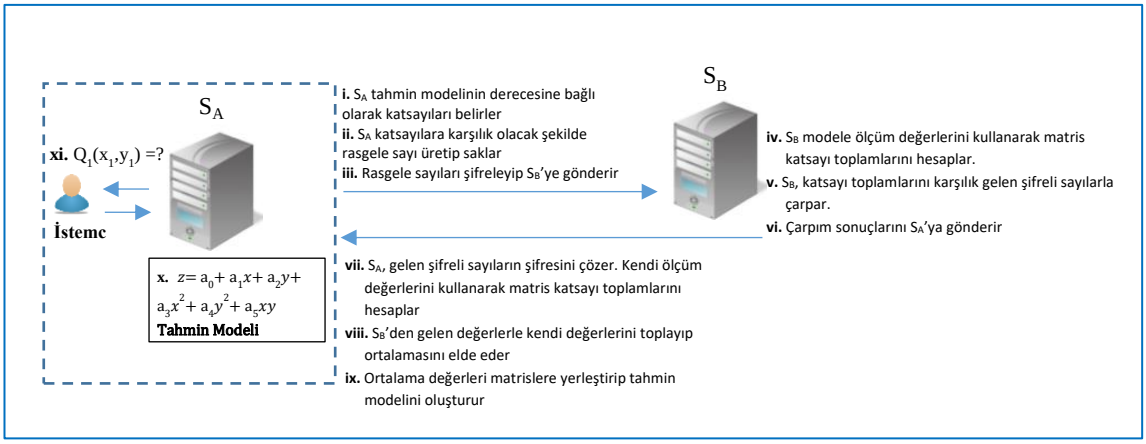
v. S_B, S_A'den gelen değerlerle $[\xi_{K_A}(p_1)^{V_1^B}, \xi_{K_A}(p_2)^{V_2^B}, \dots, \xi_{K_A}(p_{20})^{V_{20}^B}]$ hesaplar ve S_A'e geri gönderir.

vi. S_A, gelen şifreli değerleri gizli anahtarını kullanarak çözer ve $p_1V_1^B, p_2V_2^B, \dots, p_{20}V_{20}^B$ değerlerini elde eder.

vii. Son olarak S_A, $\bar{x} = \frac{1}{k+m}\left(V_1^A + \frac{p_1V_1^B}{p_1}\right)$, $\bar{y} = \frac{1}{k+m}\left(V_2^A + \frac{p_2V_2^B}{p_2}\right), \dots, \overline{y^4} = \frac{1}{k+m}\left(V_{20}^A + \frac{p_{20}V_{20}^B}{p_{20}}\right)$ hesaplamalarını yapar.

Böylece, S_A ikinci derece YEA tahmin modeli oluşturabilmek için gerekli tüm değerlere sahip olmaktadır. Her iki taraf da gizliliklerinden ödün vermeden birleştirilmiş verileri üzerinde daha doğru ve güvenilir bir tahmin modeli üretmek için şunları yapmalıdır:

- Taraflar denklem 2.4'de belirtildiği gibi gerekli toplamları hesaplar.
- Sunucular gizli anahtarlarıyla şifrelediği rasgele 20 tam sayı üretirler.
- Daha sonra sunuculardan biri *v*. adımda ifade edilen hesaplamaları yapar ve karşı tarafa gönderir.
- Her iki taraf da alınan verileri şifreler ve sahip oldukları parçaları toplar.
- Son olarak, her iki taraf artık birleştirilen verileri kullanarak bir YEA tahmin modeli oluşturabilir.



Şekil 4.7 Dağıtılmış Verilerde Gizlilik-Korumalı YEA Modeli ve işlem adımları

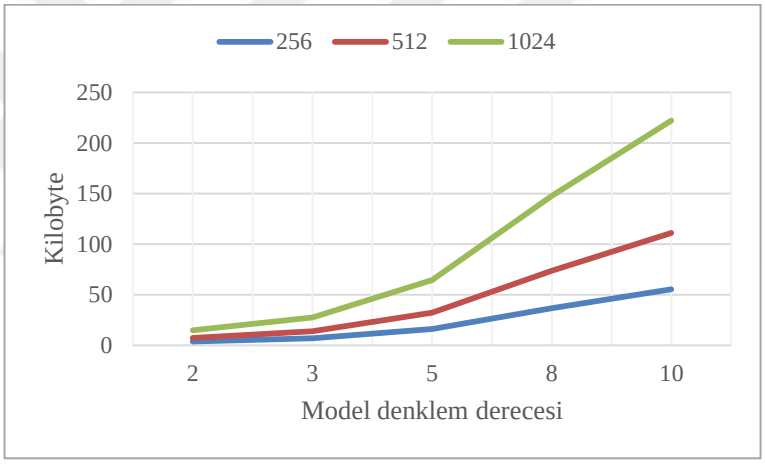
4.2.2 Maliyet analizi

Çözümümüzde gizlilik sağlamak adına SMC yöntemi kullanıldığından, geleneksel YEA'ya göre depolama, iletişim ve hesaplama gereksinimleri artmaktadır. Bu nedenle, çözümümüz geleneksel YEA ile ek depolama, iletişim ve hesaplama maliyetleri açısından karşılaştırılmaktadır. Geleneksel YEA'da, bir tahmin modeli oluşturmak için gereken tüm veriler tek bir sunucuda bulunmaktadır. Veri sahibi, tahmin modeli için gerekli olan tüm matris ve polinom denklemindeki değerleri hesaplamalı ve saklamalıdır. D-GKYEA'da, iki veri sahibinin (S_A ve S_B) ortak verileri üzerinde bir tahmin modeli oluşturulmaktadır. Yöntemde gösterildiği gibi S_A tahmin modeline göre polinom ifadesinin katsayılarını hesaplamak için kullanacağı değişkenler kadar rastgele sayı şifreleyerek S_B'ye gönderir. İkinci dereceden bir polinom denklemine sahip tahmin modeli için S_A, 20 rasgele tamsayı değeri oluşturur ve saklar. Ardından, ilgili sayıları şifreler ve S_B'ye gönderir. S_B ise kendine ait ölçüm değerlerinin toplam değerleri ile S_A'dan gelen şifreleri sayılarla üssel işlem yaparak onları S_A'ya gönderir. S_A bunların şifresini çözer ve geri kalan işlemleri yapar. Bu nedenle S_A'nın 20 tam sayı ve 20 reel sayı saklaması gerekmektedir. Ek olarak, 40 şifreli metin değerini de depolar. Ayrıca, S_B'nin fazladan 40 şifreli metin saklaması gerekir. Hesaplama maliyetleri açısından, S_A'nın ek olarak 20 rasgele sayı, 20 şifreleme, 20 şifre çözme ve 20 bölme işlemi oluşturması gerekir. Ayrıca, S_B'nin de 20 üslü işlem yapması gerekir. Geleneksel YEA'da iletişim maliyeti yoktur. Ancak D-GKYEA'da 20+20 şifreli metin S_A ve S_B

arasında aktarılması gerekmektedir. İki sunucunun ortak verilerinden üretilecek Tahmin modelinin polinom denklem derecesine bağlı olarak S_A ve S_B sunucularının kullandıkları bellek miktarı çizelge 4.3 ve çizelge 4.4'de gösterilmektedir.

Çizelge 4.3 S_A tarafından kullanılan bellek miktarı

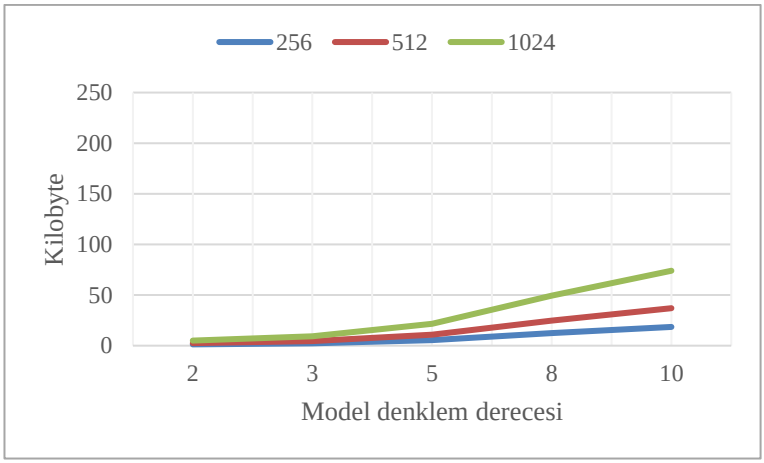
Denklem derecesi	**Anahtar boyutu**		
	256 bit	**512 bit**	**1024 bit**
2	3,75 KB	7,50 KB	15,00 KB
3	6,93 KB	13,87 KB	27,75 KB
5	16,12 KB	32,25 KB	64,50 KB
8	36,93 KB	73,87 KB	147,75 KB
10	55,50 KB	111,00 KB	222,00 KB



Şekil 4.8 S_A tarafından kullanılan bellek miktarı

Çizelge 4.4 S_B tarafından kullanılan bellek miktarı

Denklem derecesi	**Anahtar boyutu**		
	256 bit	**512 bit**	**1024 bit**
2	1,25 KB	2,50 KB	5,00 KB
3	2,31 KB	4,63 KB	9,25 KB
5	5,38 KB	10,75 KB	21,50 KB
8	12,31 KB	24,63 KB	49,25 KB
10	18,50 KB	37,00 KB	74,00 KB

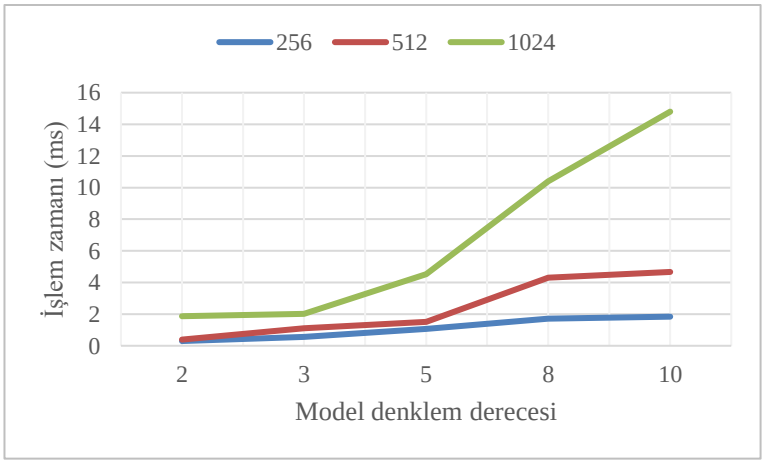


Şekil 4.9 S_B tarafından kullanılan bellek miktarı

S_B'den S_A'ya transfer edilen veriler yeni bir tahmin modeli oluşturmak için kullanılmaktadır. S_A, tahmin modelini oluşturduktan sonra S_B'nin verilerine tekrar ihtiyaç duymamaktadır. S_B'nin sahip olduğu ölçüm değerlerinin değişmesi durumunda yeni değerlere göre yeni bir tahmin modeli üretmek için işlem adımları baştan işletilmelidir.

Çizelge 4.5 S_B sunucu işlem zamanı (milisaniye)

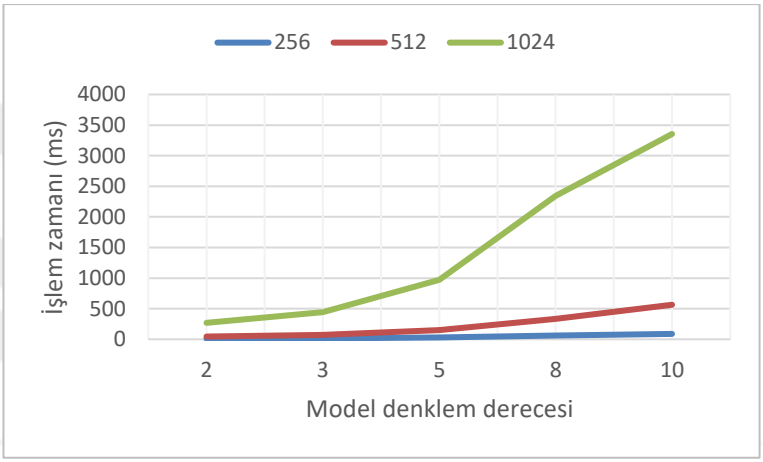
Denklem derecesi	**Anahtar boyutu**		
	256 bit	**512 bit**	**1024 bit**
2	0,30	0,39	1,87
3	0,56	1,10	2,01
5	1,07	1,52	4,53
8	1,72	4,30	10,39
10	1,84	4,67	14,80



Şekil 4.10 S_B sunucu işlem zamanı

Çizelge 4.6 S_A sunucu işlem zamanı (milisaniye)

Denklem derecesi	Anahtar boyutu		
	256 bit	512 bit	1024 bit
2	11	45	269
3	16	71	442
5	31	149	973
8	61	331	2.340
10	88	564	3.354



Şekil 4.11 S_A sunucu işlem zamanı

Uygulamada, geleneksel YEA gerçek-zamanlı zor bir sürece sahip değildir. Araştırmacılar, gizlilik endişesinden kaynaklanan ek maliyetleri beklemek ve hesaplamak için yeterli zaman ve kaynak bulabilirler. Çizelge 4.5 ve çizelge 4.6 gösterildiği gibi S_A ve S_B sunucularının ortak verileri üzerinde geliştirilen D-GKYEA her ne kadar ek maliyet getirse de genel performansı etkilenmemektedir. Sunucu ve istemci arasındaki işlemleri gerçekleştirmek için gereken ek maliyetler GKYEA'da olduğu gibidir.

4.2.3 Doğruluk analizi

İşbirliğinin tahmin modelinin doğruluğunu nasıl geliştirdiğini değerlendirmek için gerçek veri kümeleri üzerinde deneyler yapılmıştır. Bir veri sahibinin bir model oluşturmada yeterli veriye sahip olmaması durumunda D-GKYEA'da işbirliğinin

tahmin modelinin doğruluğunu nasıl geliştirdiğini değerlendirmek için gerçek veri kümeleri üzerinde deneyler yapılmıştır. Veri sahibinin bir model oluşturmak için yeterince veriye sahip olmaması durumunda tahmin modelinin doğruluğu tatmin edici olmamaktadır. Veri sahiplerinin işbirliği yapması sonuçları iyileştirmektedir.

Doğruluk analizinin amacı, işbirliğinin etkilerini göstermektir. Önerildiği gibi iki veri sahibinin verilerinin birleştirilmesi daha doğru tahmin modelleri üretmektedir. Bu hipotezi kanıtlamak adına yapılan deneyler ile paylaşılan veriler üzerine bir tahmin modeli inşa edildiğinde, doğruluğun ne kadar arttığı görülmektedir.

Gözlenen tüm verilerin iki veri sahibi tarafından paylaşıldığı varsayılmıştır. Ancak veriler taraflar arasında eşit olarak paylaşılamayabilir. Bu nedenle, eşit olmayan dağılım senaryoları da dikkate alınmıştır. β, bir tarafın, $(1-\beta)$ ise diğer tarafın elinde tuttuğu veri oranlarını tanımlamaktadır. Yani, β 0,25 seçilirse, bir taraf tüm verinin dörtte birini, diğer taraf ise dörtte üçünü elinde tutmaktadır. OMH ve KOKH sonuçları sırasıyla çizelge 4.7 ve çizelge 4.8'te gösterilmektedir.

Çizelge 4.7 Değişen β ve derece değerleri (OMH) İşbirliğinin Etkileri (β (sütunlar) bölme oranıdır ve Derece (satırlar) YEA için polinomun derecesini göstermektedir)

Derece	**Colorado**				**Illinois**				**Wisconsin**			
	β											
	25	**50**	**75**	**100**	**25**	**50**	**75**	**100**	**25**	**50**	**75**	**100**
2	0.320	0.320	0.319	0.318	0.106	0.105	0.103	0.100	0.199	0.199	0.198	0.197
3	0.288	0.287	0.285	0.285	0.105	0.103	0.101	0.099	0.178	0.174	0.173	0.171
5	0.273	0.264	0.260	0.260	0.103	0.101	0.099	0.096	0.166	0.158	0.157	0.154
8	0.286	0.255	0.248	0.246	0.111	0.102	0.098	**0.096**	0.194	0.157	0.152	**0.147**
10	0.312	0.256	0.247	**0.243**	0.131	0.107	0.101	0.098	0.291	0.169	0.158	0.149

Çizelge 4.8 Değişen β ve derece değerleriyle (KOKH) İşbirliğinin Etkileri (β (sütunlar) bölme oranıdır ve Derece (satırlar) YEA için polinomun derecesini göstermektedir)

Derece	**Colorado**				**Illinois**				**Wisconsin**			
	β											
	25	**50**	**75**	**100**	**25**	**50**	**75**	**100**	**25**	**50**	**75**	**100**
2	0.417	0.417	0.415	0.414	0.140	0.140	0.140	0.140	0.254	0.254	0.253	0.251
3	0.392	0.389	0.385	0.384	0.138	0.137	0.136	0.136	0.230	0.224	0.223	0.222
5	0.376	0.364	0.360	0.359	0.137	0.135	0.133	0.133	0.223	0.210	0.209	0.207
8	0.435	0.353	0.342	0.340	0.159	0.137	0.131	**0.130**	0.350	0.215	0.206	**0.201**
10	0.527	0.355	0.339	**0.334**	0.260	0.144	0.135	0.132	0.863	0.268	0.238	0.207

Çizelge 4.7 ve çizelge 4.8 incelendiğinde, sonuçların hipotezimizi doğruladığı açıktır. Tüm tahmin modeli derecelerinde yeterli değere sahip olmayan veri sahipleri daha yüksek OMH ve KOKH değerleri almaktadır. Aksine, yeterli veriye sahip olanlar için daha küçük OMH ve KOKH değerleri gözlemlenmektedir. İlaveten, Colorado kümesi için 10. dereceden polinom denklemi kullanıldığında minimum OMH ve KOKH değerleri çıkmaktadır. Ancak, Illinois ve Wisconsin kümeleri için en iyi sonuçları 8. dereceden polinom denklemleri vermektedir. Verilerin yarısına sahip iki sunucunun entegre verileri üzerine kurulan bir tahmin modelinin, doğruluğu ortalama olarak %6 artmaktadır. Veri sahipleri tahmin modellerini entegre veriler üzerine kurarlarsa istemcilere daha güvenilir hizmetler sunabileceklerdir.

4.2.4 Gizlilik analizi

Önerilen D-GKYEA, veri sahiplerinin birbirlerinin özel verileri hakkında bilgi çıkarmasını önlemelidir. Ayrıca kötü niyetli üçüncü şahıslar, veri sahiplerine ait sunucular arasındaki iletişimlerden herhangi bir faydalı veri elde etmemelidir. Her sunucuya ait veriler de birbirlerinden gizli olmalıdır. Ayrıca tahmin modellerinin oluşturulmasında veri sahiplerinin yarı dürüst olduğu varsayılmaktadır. Bundan dolayı her sunucunun algoritmayı izlemesi gerektiği anlamına gelmektedir. Ancak veri sahipleri diğer tarafın verilerini merak ediyor olabilirler. Yukarıda bahsi geçen durumda, gizliliği sağlamak adına homomorfik şifreleme yöntemi kullanılmaktadır. S_A, genel anahtarıyla tüm rasgele değerleri şifrelemektedir. Bu nedenle, yalnızca S_A, metinlerin şifresini çözebilir ve verileri okuyabilir. S_B, polinom denkleminin hesaplanması sırasında x_i, y_i ve z_i'nin birleştirilmiş değerlerini kullanır. S_A, S_B ile ölçülen belirli bir koordinat hakkında hiçbir bilgi öğrenemez.

D-GKYEA'da şifreleme ve birleştirilmiş verilerin kullanılması, veri sahipleri arasındaki gizliliği garanti etmektedir. Kötü niyetli üçüncü şahısların sunucular arasında aktarılan verileri dinlemesi mümkündür. Ancak D-GKYEA'da sunucular arasındaki tüm iletişim, şifrelenmiş veriler kullanılarak gerçekleştirilmektedir. Yöntemimizde kullanılan HE'nin Paillier, güvenli olduğunu ve kötü niyetli kişilerin şifrelenmiş değerlerden herhangi bir bilgi elde etmesinin mümkün olmadığını beyan etmiştir (Paillier, 1999). D-GKYEA

yukarıda açıklanan tüm güvenlik gereksinimlerini karşılamaktadır. Bu nedenle, veri sahiplerinin kişisel verilerini koruyarak işbirliği yapmaları mümkündür.

4.3 Bulut Sunucular Üzerinde Gizlilik-Korumalı Yüzey Eğilimi Analizi

Geleneksel YEA'nın, en çok uygulanan ve güvenilir konumsal enterpolasyon yöntemlerinden olduğu daha önceki bölümlerde bahsedilmiştir. YEA, bilindiği üzere, verilen veri seti için en uygun polinom ifadesini aramaktadır. Bu tür işlemler çok fazla zaman, hesaplama ve depolama kapasitesi gerektirmektedir Son yıllarda, bu tür yükleri bilgi işlem ve depolama hizmetleri vermeye adanmış bulut sunucularına aktarmak adına yeni bir eğilim vardır. Ancak kurumların büyük maliyetler sonucu elde ettikleri veriler en değerli varlıklarından biridir. Dolayısıyla her bir tarafın mahremiyeti oldukça önemlidir. Bu bölümde, veri sahiplerinin, verilerini hem istemcilerden hem de bulut sunucularından gizlemeye çalışan bir yöntem geliştirilmiştir (B-GKYEA). Çalışmamız bu çerçevede güvenli bir çözüm önermektedir. Ayrıca, çözümümüz maliyet, doğruluk ve güvenlik açısından incelenmiştir.

4.3.1 Yöntem

B-GKYEA yönteminde veriler bulut sunucuda bulunduğundan, veri sahibi-bulut sunucu, bulut sunucu-istemci şeklinde iki tür iletişim vardır. Burada bulut sunucu (BS), bulut ortamında hizmet veren kuruluşları ve istemci ise müşterileri temsil etmektedir. Öncelikle B-GKYEA'da güvenli bir dış kaynak kullanım protokolünde bulunması gereken doğruluk, gizlilik ve verimlilik unsurları tanımlanmıştır

- Doğruluk: tüm taraflar protokolde yer alan adımları dürüstçe takip ettiğinde sonuç geleneksel şemadakiyle aynı olacaktır.
- Gizlilik: Tüm tarafların özel verileri birbirlerinden korunmalıdır.
- Verimlilik: Gizliliği sağlamak için kullanılan protokollerden kaynaklanan bellek, iletişim ve hesaplama maliyetleri kabul edilebilir düzeyde olmalıdır.

B-GKYEA aşağıdaki adımları takip eder;

i. Veri sahibi, tüm değerleri (x_i, y_i, z_i) kendi gizli anahtarıyla şifreler ve bulut sunucular arasındaki iletişimi yöneten bulut sunuculardan birine (BS1) gönderir. Veri sahibinin verileri kendi anahtarıyla şifrelemesi, yalnızca onun şifresini çözebileceği ve değerleri şifreli metin biçiminde okuyabileceği anlamına gelir. Bu nedenle bulut sunucu herhangi bir girişi öğrenemez.

Çizelge 4.9 Veri sahibinin sahip olduğu konumsal veri tablosu

ID	x	y	z
1	x_1	y_1	z_1
2	x_2	y_2	z_2
3	x_3	y_3	z_3
...	...	...	...
n	x_n	y_n	z_n

Çizelge 4.10 Bulut sunucu tarafından depolanan şifreli konumsal veri tablosu

ID	x	y	z
1	$\xi c(x_1)$	$\xi c(y_1)$	$\xi c(z_1)$
2	$\xi c(x_2)$	$\xi c(y_2)$	$\xi c(z_2)$
3	$\xi c(x_3)$	$\xi c(y_3)$	$\xi c(z_3)$
...	...	...	...
n	$\xi c(x_n)$	$\xi c(y_n)$	$\xi c(z_n)$

ii. Bulut sunucu (BS1), polinom ifadesinin katsayılarını hesaplamak için gerekli matrisleri oluştururken, homomorfik toplama ve çarpma özelliklerini uygular ve bunları veri sahibine geri gönderir.

iii. Veri sahibi, matrislerdeki değerlerin şifresini çözmek için gerekli anahtara sahiptir. Veri sahibi, A ve B matrislerini bölüm 2.4.5'de anlatılan SMIP protokolünde açıklandığı gibi iki yarıya (A = A1 + A2 ve B = B1 + B2) böler ve her bir yarıyı çakışmayan iki bulut sunucusuna gönderir. Bulut sunucuları, sahip oldukları yarılarla tahmin modeli oluşturamaz. Bulut sunucular, YEA'ya dayalı bir tahmin modeli oluşturmak için matrislerin diğer parçalarına ihtiyaç duymaktadır.

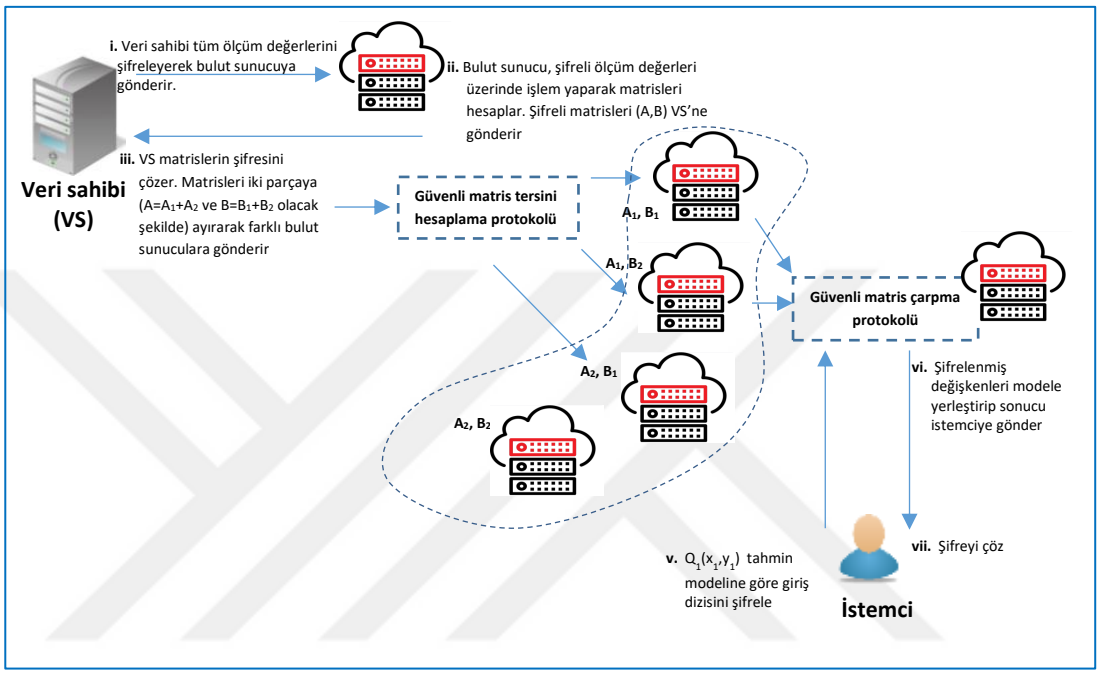
iv. (A1 + A2) (B1 + B2) ile bölüm 2.4.4'de ifade edilen SMMP çarpıldıktan sonra polinom ifadesinin katsayıları hesaplanabilir.

v. İstemci, bir tahmine ihtiyaç duyduğu koordinat değerlerini (x_p, y_p) kendi özel anahtarı $[\xi c(1), \xi c(x_p), \xi c(y_p), \xi c(x_p^2), \xi c(y_p^2), \xi c(x_p y_p)]$ ile şifreler ve bulut sunucuya gönderir.

vi. Bulut sunucu (BS1), düz metinde "$a_0, a_1 x_p, a_2\, y_p, a_3\, x_p^2, a_4\, y_p^2, a_5\, x_p y_p$" ye eşit olan Paillier şifreleme sistemi ile

"$a_0^{\xi c(1)} . a_1^{\xi c(x_p)} . a_2^{\xi c(y_p)} . a_3^{\xi c(x_p^2)} . a_4^{\xi c(y_p^2)} . a_5^{\xi c(x_p y_p)}$" i hesaplar ve şifreli metni istemciye gönderir.

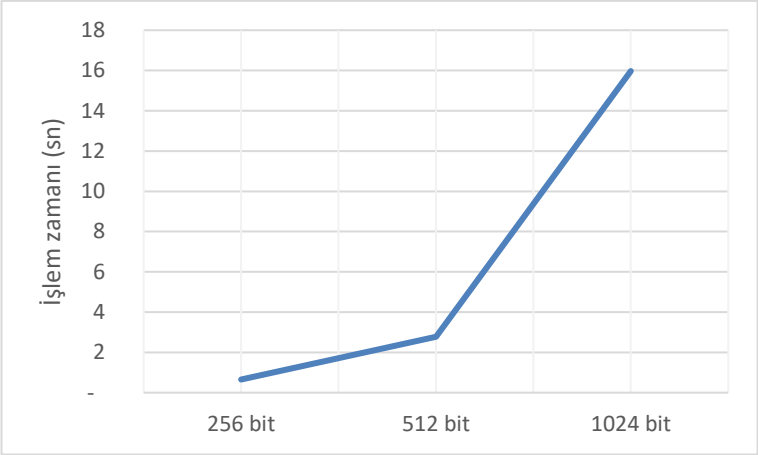
vii. İstemci, sahip olduğu gizli anahtar ile koordinat (x_p, y_p) için kendisine gönderilen şifreli tahmin değerini açar.



Şekil 4.12 Bulut sunucular üzerinde Gizlilik-Korumalı YEA Modeli ve işlem adımları

4.3.2 Maliyet analizi

Geleneksel YEA'da veri sahibi, bir istemci tarafından talep edilen konuma tahmin değeri üretmek için verileri toplar, depolar ve hesaplar. Geleneksel yöntemler taraflar için gizlilik sağlamaz. Bu nedenle mahremiyeti sağlamak için gizlilik koruma yöntemleri uygulanabilir. Gizliliği sağlamak adına veri sahibi kendi özel anahtarıyla tüm koordinat ve ölçüm değerlerini şifreleyerek bulut sunucuya göndermektedir. Şifreleme işlemi veri miktarına ve şifreleme anahtarının büyüklüğüne göre değişmektedir. Şekil 4.13'de veri sahibi tarafından 1000 ölçüm değerinin farklı boyutlardaki anahtarla şifreleme süreleri gösterilmektedir.



Şekil 4.13 Ölçüm değerlerinin şifrelenmesi

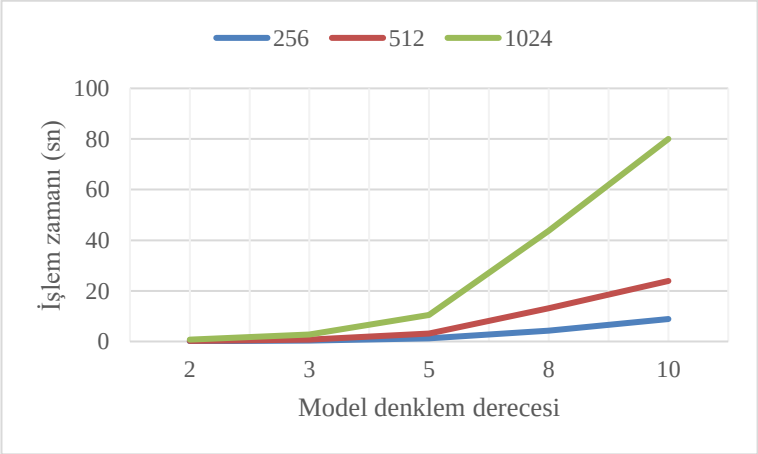
Geleneksel sistemde gerekli tüm hesaplamalar veri sahibi tarafından yapılmaktadır. Verimlilik, ölçeklenebilirlik ve ekonomik nedenlerden dolayı bulut bilişim tercih ediliyorsa, bulut sunucuların gerekli tüm hesaplamaları yapması da beklenmektedir. Bulut sunucular, veri sahiplerinden bir tahmin modeli oluşturmanın yükünü üstlenir, ancak gizlilik konusu önemli bir endişe olarak devam etmektedir. Önerilen B-GKYEA, veri sahipleri, bulut sunucular ve istemciler için gizliliği koruyan bir çözüm sunmaktadır. B-GKYEA'da, bölüm 4.2'de de olduğu gibi SMC protokolleri ve homomorfik şifreleme yöntemleri kullanılmaktadır. Bu yöntemler ise taraflara güvenli bir dış kaynak tahmin hizmeti sunarken hesaplama sayısını da artırmaktadır. Konumsal enterpolasyon hizmetleri gerçek zamanlı çalışan sistemler değildir. Bu nedenle, tahmin üretirken küçük gecikmeler tarafları etkilememektedir. Veri sahibi, tüm konumsal veri tablosunu bulut sunucuya göndermek için şifrelemelidir. Bulut sunucu, şifreli değerler kullanarak matrislerin toplu değerlerini bulmak adına toplama işlemi yerine çarpma işlemini kullanmalıdır. Çarpma işlemleri, toplama işlemlerinden daha fazla işlem gücü gerektirmektedir. Veri sahibi, YEA denkleminin parçası olan iki matrisin şifresini çözmeli ve bunları iki parçaya bölmek için gizli paylaşım şemaları uygulamalıdır. Bulut sunucular, denklem 2.3'de $(z = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy)$ açıklanan katsayıları bulmak için SMIP ve SMMP protokollerini uygulamalıdır. İstemci, hedef koordinatı (x_p, y_p) anahtarıyla şifreler ve $[\xi c(1), \xi c(x_p), \xi c(y_p), \xi c(x_p^2), \xi c(y_p^2), \xi c(x_p y_p)]$ alır. Bu nedenle şifreleme işlevini altı kez çağırmalıdır. İstemcinin şifreleyerek giriş dizisini oluşturma maliyetli ilk modelde

gösterildiği gibidir. Bulut sunucu, şifreli metindeki son tahmin değerine karşılık gelen $a_0^{\xi c(1)}.a_1^{\xi c(x_p)}.a_2^{\xi c(y_p)}.a_3^{\xi c(x_p^2)}.a_4^{\xi c(y_p^2)}.a_5^{\xi c(x_p y_p)}$ işlemini hesaplar. İstemci, tahmin değerini elde etmek için yalnızca kendisi tarafından bilinen anahtarı kullanarak şifre çözme işlevini çağırmalıdır.

Tüm verilerin bulut sunuculara aktarılmasından dolayı iletişim maliyetleri artmaktadır. Şifrelenmiş konumsal veri tablosu, bulut sunuculardan birine gönderilmelidir. İlgili bulut sunucusu, YEA'da kullanılan matrisleri şifreli biçimde hesaplar ve veri sahibine geri gönderir. Bulut sunucunun şifrelenmiş 1000 örnekli test verisini hesaplama süreleri Çizelge 4.11 ve şekil 4.14'de gösterilmektedir. İşlem zamanı, kullanılan anahtar boyutuna ve veri miktarına göre artan bir eğilim göstermektedir. Bunun yanı sıra veri miktarına bağlı olarak artan işlemler sunucuda bir defa yapılacağı için süreler veri sahipleri için kabul edilebilir büyüklüktedir.

Çizelge 4.11 Bulut sunucu matris hesaplama süresi (milisaniye)

Denklem derecesi	**Anahtar boyutu**		
	256 bit	**512 bit**	**1024 bit**
2	107	231	735
3	366	795	2.698
5	1.269	3.158	10.417
8	4.288	13.129	43.770
10	8.872	23.916	80.027



Şekil 4.14 Bulut sunucu matris hesaplama zamanı

Veri sahibi tarafından gizli paylaşım ilkelerine göre iki matris (A1+A2 ve B1+B2) oluşturulur ve bulut sunuculara gönderilir. Nassar vd.'nin (2013) çalışmalarında belirttiği gibi bu matrisler SMMP ve SMIP'yi çalıştırmak için diğer bulut sunucularına aktarılması gerekmektedir. Bu nedenle B-GKYEA, bulut sunucular ve sunucu arasında 1 x (konumsal veri tablosunun boyutu) + 12 x (matrislerin boyutu) (en iyi durum) ek iletişim oluşturur (matrisler veri tablosundan çok daha küçüktür).

Gizlilik ve verimlilik kaygısı yoksa tüm konumsal veriler ve matrisler veri sahiplerinin sunucularında saklanır. Ancak hem gizlilik hem de verimlilik sunan B-GKYEA daha fazla depolama kapasitesi kullanmaktadır. Ek olarak, konumsal veri tablosu şifrelenerek bulut sunuculardan birinde saklanmalıdır. Geleneksel YEA, polinom ifadesinin katsayılarını hesaplamak için iki matris gerektirir, ancak çözümümüz için üç kat daha fazla kapasite (SMMP ve SMIP protokolleri nedeniyle) gerekmektedir. Bu nedenle B-GKYEA, 1 x (konumsal veri tablosunun boyutu) + 4 x (matrislerin boyutu) (en iyi durum) ek depolama kapasitesi gerektirir (matrisler veri tablosundan çok daha küçüktür).

4.3.3 Doğruluk analizi

Özel verileri gizlemek için kullanılan bazı gizlilik-korumalı veri madenciliği yöntemleri, tahmin modellerinin doğruluğunu azaltabilir. Agrawal ve Srikant (2000), rasgele değerleri kullanıcıların özel verilerine ekleyerek gizlilik sağlamayı önermiştir. Bu yöntemde, bir dağılıma göre üretilen rasgele bir gürültü orijinal verilere eklenmektedir. Kargupta vd. (2005) ise ham verilerin, spektral filtre teknikleri kullanılarak bozulmuş verilerden tahmin edilebileceğini göstermişlerdir. Bu nedenle B-GKYEA'da, homomorfik şifreleme ve gizli paylaşım yöntemlerini kullanılmaktadır. Kullanılan bu yöntemler deterministiktir. Bu bakımdan geleneksel YEA'nın ve B-GKYEA'nın nihai tahmin değeri aynı olmaktadır. Ayrıca YEA'nın güvenli dış kaynak kullanımına dâhil olan taraflar da, aynı doğrulukta bir tahmin modeli oluşturacaktır. Ek olarak, tüm tarafların gizliliği önerdiğimiz B-GKYEA ile sağlanmıştır.

4.3.4 Gizlilik analizi

B-GKYEA, güvenli bir dış kaynak kullanma protokolü doğruluğu, gizliliği ve verimliliği sağlamalıdır. Çözümümüzün tüm tarafların gizliliğini koruduğunu göstermek için, kimsenin özel verilerinin yetkisiz taraflara ifşa edilmediğini kanıtlamalıyız.

- Veri sahibi konumsal veri kümesini bulut sunucuya göndermeden önce yalnızca kendisi tarafından bilinen anahtarla şifreler. Bu nedenle, sunucularda depolanan şifreli metin hiçbir bulut sunucu tarafından öğrenilemez. Veri sahibi-bulut sunucu arasındaki işlem hattından, kötü niyetli taraflar faydalı bilgiler elde edememektedirler.
- Homomorfik özelliklere sahip şifreleme yöntemi, bulut sunucunun YEA için gerekli olan gerekli matrisleri oluşturmasını ve hesaplamasını sağlamaktadır. Matris değerleri, konumsal verilerin toplamıdır. Bu nedenle, bulut sunucu veya kötü niyetli taraflar, matrislerden belirli bir konum için karşılık gelen ölçüm değerini tahmin edemez.
- Ayrıca, matrisler, gizli paylaşım şemaları kullanılarak ikiye bölünmektedir (esas olarak eşit parçalara değil). Bulut sunucular matrisleri öğrenmek isterse kötü niyetli davranmak zorundadır. Ancak bu davranış bulut hizmeti verenlerin itibarlarına ve güvenilirliklerine zarar verecektir. Vurgulamak gerekirse, B-GKYEA'da kötü niyetli davransalar bile, herhangi bir belirli konumsal veriyi değil, yalnızca matrislerin içindeki değerleri öğreneceklerdir.
- İstemci, koordinat değerini şifreli biçimde (istemci sadece kendi tarafından bilinen anahtar kullanır) bulut sunuculara gönderir. B-GKYEA, bulut sunucuların şifreli metin biçiminde nihai tahmin değerini hesaplamasına olanak tanımaktadır. Bu nedenle, veri sahibi, bulut sunucular ve diğer üçüncü taraflar, istemcinin tahmin değerine ihtiyaç duyduğu konumu öğrenemez.

Yukarıda açıklanan maddelerden, B-GKYEA'daki güvenli dış kaynak kullanım protokolünün, gizlilik gereksinimlerini karşıladığı sonucuna varılmaktadır.

5. SONUÇ

Eldeki ölçüm değerlerini kullanarak değeri bilinmeyen veya ölçüm yapılmayan noktalar için değer tahmini yapmak konumsal enterpolasyon olarak bilenmektedir. Bu amaca yönelik birçok yöntem literatürde tanımlanmıştır. YEA, temel ve yaygın olarak tercih edilen konumsal enterpolasyon yöntemlerinden biridir. Sunucular tarafından tutulan konum ve bu konumlara ait ölçüm değerleri, istemci tarafından tahmin değeri istenen noktaların koordinatları ve tahmin değerleri gizli veri olarak kabul edilir. Bu tür verilerin ifşa edilmesi taraflar için maddi ve manevi zararlara neden olabilmektedir. Ancak geleneksel YEA yöntemi hem istemcilerin hem de veri sahiplerinin gizliliğini garanti etmemektedir. Bu nedenle konumsal enterpolasyon işlemlerinde veri gizliliği sağlamak çok önemlidir.

Bu tezde sunucu ve istemcilere ait verilerin gizliliğini sağlamak adına merkezi, dağıtılmış ve bulut sunuculu veri dağıtım senaryolarında gizlilik korumalı YEA enterpolasyon yöntemleri önerilmiştir. Veri gizliliği homomorfik şifreleme algoritmaları ve güvenli çok-parti hesaplama çözümleri ile gerçekleştirilmiştir. Çözüm önerilen yöntemler için merkezi, dağıtık ve bulut hesaplama veri dağıtım senaryoları dikkate alınmıştır. Ayrıca her bir çözüme ait gizlilik ve ek maliyet analizleri gerçekleştirilmiş ve geleneksel yönteme ait değerler ile karşılaştırmalar yapılmıştır. Önerilen çözümlerin yeterli düzeyde doğru tahmin değerleri ürettiklerini göstermek için gerçek veriler kullanılarak çeşitli deneyler yapılmış ve elde edilen sonuçlar tartışılmıştır.

Kurumlar aynı veya komşu bölgelerden konumsal veri toplayabilmektedir. Tahmin modellerinin güvenilirliği ve doğruluğu iki farklı sunucuya ait ortak verilerin kullanılması durumunda artış göstermektedir. İlgili artışın sebebi YEA algoritmasının kullanmış olduğu veri miktarının artmasındandır. Ancak veri sahipleri, verilerini hem yasalar gereğince hem de ekonomik nedenlerden korumak istemektedirler. Verilerini açıktan paylaşmadan analiz sonuçları üretecek çözümler veri sahiplerini iş birliğine ikna etmeye yardımcı olmaktadır. Bu tez çalışmasında iki tarafın ortak verilerini birbirlerine veya üçünü taraflara ifşa etmeden YEA analizi gerçekleştiren bir çözüm önerilmiştir. Her iki tarafın gizliliklerinden ödün verilmeden birleştirilmiş veriler üzerinde bir tahmin

modeli oluşturulmuştur. Önerilen çözümde gizliliği sağlamak için güvenli matris çarpma işlemlerinin kullanılması ekstra iletişim ve depolama maliyetine yol açmaktadır. Bu maliyetler, analiz bölümünde ifade edildiği gibi ihmal edilebilir seviyededir.

Son olarak dış kaynak kullanımına yönelik güvenli YEA çözümleri önerilmiştir. Çözüm önerisinde gizliliği sağlamak için güvenli matris çarpma ve güvenli matris tersi hesaplama protokolleri kullanılmıştır. Güvenlik ve verimlilik gibi etmenler ise bu protokollerin seçiminde belirleyici olmaktadır. Ayrıca, veri sahipleri ve istemcilerin girdilerini aktarmak ve hesaplamak için Paillier homomorfik şifreleme sistemi kullanılmıştır. Diğer senaryolarda olduğu gibi önerilen yöntem ek maliyet, doğruluk ve gizlilik açısından analiz edilmiştir. Sonuç olarak ek maliyetin ihmal edilebilir olduğu ifade edilebilir. Hem geleneksel yöntemin hem de YEA'nın güvenli dış kaynak kullanımının doğruluğunun aynı olduğu hem de tüm tarafların gizliliğinin sağlandığı gerçekleştirilen deney sonuçlarıyla gösterilmiştir.

KAYNAKLAR

Abdel Wahab, O., Hachami, M. O., Zaffari, A., Vivas, M., and Dagher, G. G. 2014. DARM: A privacy-preserving approach for distributed association rules mining on horizontally-partitioned data. IDEAS'14: 18th International Database Engineering & Applications Symposium. 7-9 July 2014; Porto/Portugal.

Aggarwal, C. C., and Yu, P. S. 2008. A General Survey of Privacy-Preserving Data Mining Models and Algorithms, In: Privacy-preserving data mining, Springer, 11-52, Boston/USA.

Agrawal, R., and Srikant, R. 2000. Privacy-preserving data mining. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. 15-18 May 2000, 439-450, New York/USA

Agterberg, F. P. 1984. Trend Surface Analysis. Spatial Statistics and Models. Springer, 147-171, Dordrecht/Netherlands.

Anonymous. 2012. Web sitesi. http://www.gitta.info/SpatChangeAna/en/html/spatial_dist_an_TrendSurf.html. Erişim tarihi: 02.06.2022

Armstrong, M. 1998. Basic linear geostatistics. Springer-Verlag Berlin Heidelberg, 153, New York/USA.

Atallah, M. J., Pantazopoulos, K. N., Rice, J. R., and Spafford, E. E. 2002. Secure outsourcing of scientific computations. In: Advances in Computers (Vol. 54, pp. 215-272). Elsevier.

Aumond, P., Can, A., Mallet, V., De Coensel, B., Ribeiro, C., Botteldooren, D., and Lavandier, C. 2018. Kriging-based spatial interpolation from measurements for sound level mapping in urban areas. The Journal of the Acoustical Society of America, 143(5).

Barni, M., Failla, P., Kolesnikov, V., Lazzeretti, R., Sadeghi, A.-R., and Schneider, T. 2009. Secure evaluation of private linear branching programs with medical applications. In European symposium on research in computer security (pp. 424-439). Springer, Berlin, Heidelberg.

Ben-Or, M., Goldwasser, S., and Wigderson, A. 2019. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali (pp. 351-371).

Benaloh, J. 1994. Dense probabilistic encryption. In Proceedings of the workshop on selected areas of cryptography, 120-128.

Benaloh, J. C. 1987. Secret sharing homomorphisms: Keeping shares of a secret secret. In Conference on the theory and application of cryptographic techniques (pp. 251-260). Springer, Berlin, Heidelberg.

Bogetoft, P., Damgård, I., Jakobsen, T., Nielsen, K., Pagter, J., and Toft, T. 2006. A practical implementation of secure auctions based on multiparty integer computation. In International Conference on Financial Cryptography and Data Security (pp. 142-147). Springer, Berlin, Heidelberg.

Boneh, D., Goh, E. J., and Nissim, K. 2005. Evaluating 2-DNF formulas on ciphertexts. In Theory of cryptography conference (pp. 325-341). Springer, Berlin, Heidelberg.

Brakerski, Z., and Vaikuntanathan, V. 2011. Fully homomorphic encryption from ring-LWE and security for key dependent messages. In Annual cryptology conference (pp. 505-524). Springer, Berlin, Heidelberg.

Bringer, J., Chabanne, H., and Patey, A. 2013. Privacy-preserving biometric identification using secure multiparty computation: An overview and recent trends. IEEE Signal Processing Magazine, 30(2), 42-52.

Burrough, P. A., McDonnell, R., McDonnell, R. A., and Lloyd, C. D. 2015. Principles of geographical information systems. Oxford university press.

Chai, T., and Draxler, R. R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature. Geoscientific model development, 7(3), 1247-1250.

Chaum, D., Crépeau, C., and Damgard, I. 1988. Multiparty unconditionally secure protocols. In Proceedings of the twentieth annual ACM symposium on Theory of computing (pp. 11-19).

Chen, K., Liu, L., Chen, K., and Liu, L. 2011. Geometric data perturbation for privacy preserving outsourced data mining. Knowledge and information systems, 29(3), 657-695.

Cho, Y., Cao, Y., Zagayevskiy, Y., Wong, T., and Munoz, Y. 2020. Kriging-based monitoring of reservoir gas saturation distribution using time-lapse multicomponent borehole gravity measurements: Case study, Hastings Field. Journal of Petroleum Science and Engineering, 190.

Cominetti, E. L., and Simplicio, M. A. 2020. Fast additive partially homomorphic encryption from the approximate common divisor problem. IEEE Transactions on Information Forensics and Security, 15, 2988-2998.

Cranor, L. F., Reagle, J., and Ackerman, M. S. 2000. Beyond Concern: Understanding Net Users' Attitudes About Online Privacy. Retrieved April.

Curtis, A. J., Mills, J. W., and Leitner, M. 2006. Spatial confidentiality and GIS: re-engineering mortality locations from published maps about Hurricane Katrina. International Journal of Health Geographics, 5(1), 1-12.

Domadiya, N., and Rao, U. P. 2019. Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data. Procedia computer science, 148, 303-312

Durnell, E., Okabe-Miyamoto, K., Howell, R. T., and Zizi, M. 2020. Online privacy breaches, offline consequences: Construction and validation of the concerns with the protection of informational privacy scale. International Journal of Human-Computer Interaction, 36(19), 1834-1848.

El. Saad, A., and Dong, Z. 2015. An Extension of DIRECT Algorithm Using Kriging Metamodel for Global Optimization. In Proceedings of the The International Conference on Engineering & MIS 2015 (pp. 1-6).

ElGamal, T. 1985. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE transactions on information theory, 31(4), 469-472.

Evfimievski, A., Srikant, R., Agrawal, R., and Gehrke, J. 2004. Privacy preserving mining of association rules. Information Systems, 29(4), 343-364

Exeter, D. J., Rodgers, S., and Sabel, C. E. 2014. "Whose data is it anyway?" The implications of putting small area-level health and social data online. Health policy, 114(1), 88-96

Ferrer, J. D. i. 1996. A new privacy homomorphism and applications. Information Processing Letters, 60(5), 277-282.

Gambs, S., Killijian, M.-O., Moise, I., and del Prado Cortez, M. N. 2013. Mapreducing gepeto or towards conducting a privacy analysis on millions of mobility traces. In 2013 IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum (pp. 1937-1946). IEEE

Gao, S., and Prasad, S. 2016. Employing spatial analysis in indoor positioning and tracking using Wi-Fi access points. In Proceedings of the Eighth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness (pp. 27-34).

Gentry, C. 2009. Fully homomorphic encryption using ideal lattices. In Proceedings of the forty-first annual ACM symposium on Theory of computing (pp. 169-178).

Goldreich, O. 2004. Foundations of Cryptography: Basic Applications. Cambridge University Press, 449, New York/USA.

Goldwasser, S., and Micali, S. 1982. Probabilistic encryption & how to play mental poker keeping secret all partial information. In Proceedings of the Annual ACM Symposium on Theory of Computing, (pp. 173-201).

Grohmann, C. H. 2005. Trend-surface analysis of morphometric parameters: a case study in southeastern Brazil. Computers and Geosciences, 31(8), 1007-1014.

Hariss, K., Noura, H., and Samhat, A. E. 2017. Fully Enhanced Homomorphic Encryption algorithm of MORE approach for real world applications. Journal of Information Security and Applications, 34, 233-242.

Henecka, W., and Schneider, T. 2013. Faster secure two-party computation with less memory. In Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security, 8-10 May, (pp. 437-446).

Huang, C.-Y., Tong, H., He, J., and Maciejewski, R. 2019. Location Prediction for Tweets. Frontiers in Big Data, 2, 5.

Johnson, K., Hoef, J. M. Ver, Krivoruchko, K., and Lucas, N. 2001. Using ArcGIS geostatistical analyst (Vol. 380). Redlands: Esri.

Kaleli, C. 2012. Privacy-preserving concordance-based recommendations on vertically distributed data. In 2012 Tenth International Conference on ICT and Knowledge Engineering, 21-23 November, (pp. 19-24). IEEE.

Kaleli, C., and Polat, H. 2010. P2P collaborative filtering with privacy. Turkish Journal of Electrical Engineering and Computer Sciences, 18(1), 101-116.

Kantarcioglu, M., and Clifton, C. 2004. Privacy-preserving distributed mining of association rules on horizontally partitioned data. IEEE Transactions on Knowledge and Data Engineering, 16(9), 1026-1037.

Kargupta, H., Datta, S., Wang, Q., and Sivakumar, K. 2005. Random-data perturbation techniques and privacy-preserving data mining. Knowledge and Information Systems, 7(4), 387-414.

Kartheeban, K., and Murugan, A. D. 2017. Privacy preserving data storage technique in cloud computing. In 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1-6). IEEE.

Kounadi, O., and Leitner, M. 2014. Why does geoprivacy matter? The scientific publication of confidential data presented on maps. Journal of Empirical Research on Human Research Ethics, 9(4), 34-45.

Kreso, I., Kapo, A., and Turulja, L. 2021. Data mining privacy preserving: Research agenda. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(1), e1392.

Krige, D. G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy, 52(6), 119-139.

Kwan, M.-P., Casas, I., and Schmitz, B. 2004. Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks?. Cartographica: The International Journal for Geographic Information and Geovisualization, 39(2), 15-28.

Lei, X., Liao, X., Huang, T., and Heriniaina, F. 2014. Achieving security, robust cheating resistance, and high-efficiency for outsourcing large matrix multiplication computation to a malicious cloud. Information sciences, 280, 205-217.

Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren. 2014. Information Security in Big Data: Privacy and Data Mining. IEEE Access, 2, 1149-1176.

Leitner, M., and Curtis, A. 2006. A first step towards a framework for presenting the location of confidential point data on maps—results of an empirical perceptual study. International Journal of Geographical Information Science, 20(7), 813-822.

Li, J., and Heap, A. D. 2008. A Review of Spatial Interpolation Methods for Environmental Scientists. Geoscience Australia, Record 2008/23, 137 pp.

Li, J., and Heap, A. D. 2014. Spatial interpolation methods applied in the environmental sciences: A review. Environmental Modelling and Software, 53, 173-189.

Li, L., and Goodchild, M. F. 2013. s privacy still an issue in the era of big data? Location disclosure in spatial footprints. In 2013 21st International Conference on Geoinformatics, 20-22 June, (pp. 1-4). Henan, China.

Lindell, Y., and Pinkas, B. 2000. Privacy preserving data mining. In Annual International Cryptology Conference, 17-21 August, (pp. 36-54). Springer, Berlin, Heidelberg.

Liu, X., Deng, R. H., Choo, K.-K. R., Yang, Y., and Pang, H. 2020. Privacy-Preserving Outsourced Calculation Toolkit in the Cloud. IEEE Transactions on Dependable and Secure Computing, 17(5), 898-911.

Liu, Y., Yuan, Y., Shen, J., and Gao, W. 2021. Emergency response facility location in transportation networks: A literature review. Journal of Traffic and Transportation Engineering (English Edition), 8(2), 153-169.

Luo, W., Taylor, M. C., and Parker, S. R. 2008. A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. International Journal of Climatology: A Journal of the Royal Meteorological Society, 28(7), 947-959.

Mell, P., and Grance, T. 2011. The NIST definition of Cloud Computing. Gaithersburg, MD: NIST, Special Publication 800-145.

Meng, Q., Liu, Z., and Borders, B. E. 2013. Assessment of regression kriging for spatial interpolation-comparisons of seven GIS interpolation methods. Cartography and Geographic Information Science, 40(1), 28-39.

Micali, S., Goldreich, O., and Wigderson, A. 1987. How to play ANY mental game. In Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 25-27 May, New York, NY, USA (pp. 25-27).

Myers, D. E. 1994. Spatial interpolation: an overview. Geoderma. vol. 62, pp. 17-28

Nassar, M., Erradi, A., Sabri, F., and Malluhi, Q. M. 2013. Secure Outsourcing of Matrix Operations as a Service. 2013 IEEE Sixth International Conference on Cloud Computing, June 28-July 3, 918-925. California, USA

Obar, J. A., and Oeldorf-Hirsch, A. 2018. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. Information, Communication & Society, 23(1), 128-147.Oldham, C. H. G., and Sutherland, D. B. 1955. Orthogonal polynomials: Their use in estimating the regional effect. Geophysics, 20(2), 295-306.

Paar, C., and Pelzl, J. 2010. The advanced encryption standard (AES). In Understanding cryptography (pp. 87-121). Springer, Berlin, Heidelberg.

Paillier, P. 1999. Public-key cryptosystems based on composite degree residuosity classes. In International conference on the theory and applications of cryptographic techniques (pp. 223-238). 2-6 May, Springer, Berlin, Heidelberg.

Polat, H., and Du, W. 2005. Privacy-preserving collaborative filtering. International journal of electronic commerce, 9(4), 9-35.

Polat, H., and Du, W. 2006. Achieving private recommendations using randomized response techniques. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 637-646). 9-12 April, Springer, Berlin, Heidelberg.

Rivest, R. L., Shamir, A., and Adleman, L. 1983. A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM, 26(1), 96-99.

Rizwan, M., Wan, W., and Gwiazdzinski, L. 2020. Visualization, spatiotemporal patterns, and directional analysis of urban activities using geolocation data extracted from LBSN. ISPRS International Journal of Geo-Information, 9(2), 137.

Sathya, S. S., Vepakomma, P., Raskar, R., Ramachandra, R., and Bhattacharya, S. 2018. A review of homomorphic encryption libraries for secure computation.

Shamir, A. 1979. How to share a secret. Communications of the ACM, 22(11), 612-613.

Sheikhalishahi, M., and Martinelli, F. 2017. Privacy preserving clustering over horizontal and vertical partitioned data. In 2017 IEEE Symposium on Computers and Communications (ISCC), 3-6 July, (pp. 1237-1244).

Shepard, D. 1968. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM National Conference, 27-29 August, (pp. 517-524). New York.

Shrestha, R., and Kim, S. 2019. Integration of IoT with blockchain and homomorphic encryption: Challenging issues and opportunities. In Advances in Computers (Vol. 115, pp. 293-331). Elsevier.

Silva, A., and Bellala, G. 2017. Privacy-preserving multi-party clustering: an empirical study. In 2017 IEEE 10th International Conference on Cloud Computing (CLOUD), 25-30 June, (pp. 326-333). Hawaii, USA

Simmon, E. 2018. Evaluation of cloud computing services based on NIST SP 800-145. NIST Special Publication, 500, 322.

Sun, L., Mu, W.-S., Qi, B., and Zhou, Z.-J. 2015. A new privacy-preserving proximal support vector machine for classification of vertically partitioned data. International Journal of Machine Learning and Cybernetics, 6(1), 109-118.

Sun, Z., Strang, K. D., and Pambel, F. 2020. Privacy and security in the big data paradigm. Journal of Computer Information Systems, 60(2), 146-155.

Swanlund, D., Schuurman, N., Zandbergen, P., and Brussoni, M. 2020. Street masking: a network-based geographic mask for easily protecting geoprivacy. International journal of health geographics, 19(1), 1-11.

Şen, Z. 2009. Spatial Modeling Principles in Earth Sciences. Springer, 108-120, New York

Şenyürek, E., and Yakut, İ. 2013. A Brief Survey on Electronic Voting. Technological Applied Sciences, 8(3), 46-54.

Tisseuil, C., Gryspeirt, A., Lancelot, R., Pioz, M., Liebhold, A., and Gilbert, M. 2016. Evaluating methods to quantify spatial variation in the velocity of biological invasions. Ecography, 39(5), 409-418.

Tobler, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, 46, 234.

Tugrul, B., and Polat, H. 2013. Estimating kriging-based predictions with privacy. International Journal of Innovative Computing, Information and Control, 9(8), 3197-3209.

Tugrul, B., and Polat, H. 2014a. Privacy-preserving kriging interpolation on distributed data. In International Conference on Computational Science and Its Applications, 30 June-3 July, (pp. 695-708). Guimaraes, Portugal.

Tugrul, B., and Polat, H. 2014b. Privacy-Preserving Inverse Distance Weighted Interpolation. Arabian Journal for Science and Engineering, 39(4), 2773-2781.

Tugrul, B., and Polat, H. 2014c. Privacy-preserving kriging interpolation on partitioned data. Knowledge-Based Systems, 62, 38-46.

Vaidya, J., and Clifton, C. 2002. Privacy preserving association rule mining in vertically partitioned data. In Proceedings of The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 23-26 July, (pp. 639-644). Edmonton, Canada.

van Dijk, M., Gentry, C., Halevi, S., and Vaikuntanathan, V. 2010. Fully homomorphic encryption over the integers. In Annual International Conference on The Theory

and Applications of Cryptographic Techniques, 30 May-3 June, (pp. 24-43). Springer, Berlin, Heidelberg.

Wang, L., Li, L., Li, J., Li, J., Gupta, B. B., and Liu, X. 2019. Compressive Sensing of Medical Images With Confidentially Homomorphic Aggregations. IEEE Internet of Things Journal, 6(2), 1402-1409.

Webster, R., and Oliver, M. 2007. Geostatistics for environmental scientists. Chichester: Wiley

Will, M. A., and Ko, R. K. L. 2015. A guide to homomorphic encryption, In The Cloud Security Ecosystem, Syngress, (pp. 101-127), New York

Yang, J.-J., Li, J.-Q., and Niu, Y. 2015. A hybrid solution for privacy preserving medical data sharing in the cloud environment. Future Generation Computer Systems, 43-44, 74-86.

Yao, A. C. 1982. Protocols for secure computations. In 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982), 3-5 November, (pp. 160-164), Washington DC, USA.

Yao, L., Huo, Z., Feng, S., Mao, X., Kang, S., Chen, J., Xu, J., and Steenhuis, T. S. 2014. Evaluation of spatial interpolation methods for groundwater level in an arid inland oasis, northwest China. Environmental Earth Sciences, 71(4), 1911-1924.

Yin, P., and Mu, L. 2012. Modular capacitated maximal covering location problem for the optimal siting of emergency vehicles. Applied Geography, 34, 247-254.

Yiu, M. L., Assent, I., Jensen, C. S., and Kalnis, P. 2012. Outsourced Similarity Search on Metric Data Assets. IEEE Transactions on Knowledge and Data Engineering, 24(2), 338-352.

Young, C., Martin, D., and Skinner, C. 2009. Geographically intelligent disclosure control for flexible aggregation of census data. International Journal of Geographical Information Science, 23(4), 457-482.

Yuan, W., Xu, B., Chen, Z., Xia, J., Xu, W., Chen, Y., Wu, X., and Fu, Y. 2015. Validation of China-wide interpolated daily climate variables from 1960 to 2011. Theoretical and Applied Climatology, 119(3-4), 689-700.

Zhong, X., Kealy, A., and Duckham, M. 2016. Stream Kriging: Incremental and recursive ordinary Kriging over spatiotemporal data streams. Computers & Geosciences, 90, 134-143.

Zhou, L., Zhu, Y., and Castiglione, A. 2017. Efficient k-NN query over encrypted data in cloud with limited key-disclosure and offline data owner. Computers and Security, 69, 84-96.