T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SAKARYA BÜYÜKŞEHİR BELEDİYESİNE AİT EVRAK VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

Engin UÇAR

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

AĞUSTOS 2025

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SAKARYA BÜYÜKŞEHİR BELEDİYESİNE AİT EVRAK VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

Engin UÇAR

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Bilim Dalı

Tez Danışmanı: Prof. Dr. Nilüfer YURTAY

AĞUSTOS 2025

Engin Uçar tarafından hazırlanan "Sakarya Büyükşehir Belediyesine Ait Evrak Verilerinin Veri Madenciliği Yöntemleri İle İncelenmesi" adlı tez çalışması 19.08.2025 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı'nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	**Prof. Dr. Nilüfer YURTAY** (Danışman) Sakarya Üniversitesi	.....
Jüri Üyesi :	**Doç. Dr. Zeynep GARİP** Sakarya Uygulamalı Bilimler Üniversitesi	.....
Jüri Üyesi :	**Dr. Öğr. Üyesi Serap ÇAKAR KAMAN** Sakarya Üniversitesi	.....

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum "SAKARYA BÜYÜKŞEHİR BELEDİYESİNE AİT EVRAK VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE İNCELENMESİ" başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete'de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi'nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(19/08/2025).

Engin UÇAR

TEŞEKKÜR

Yüksek lisans eğitimim boyunca değerli bilgi ve deneyimlerinden yararlandığım, her konuda desteğini almaktan çekinmediğim değerli danışman hocam Prof. Dr. Nilüfer YURTAY'a ve tezimde kullanılan verileri edinme sürecinde yardımcı olan Sakarya Büyükşehir Belediyesi'ne teşekkürlerimi sunarım.

Engin UÇAR

İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ ..... **v**
TEŞEKKÜR ..... **vii**
İÇİNDEKİLER ..... **ix**
KISALTMALAR ..... **xi**
TABLO LİSTESİ ..... **xiii**
ŞEKİL LİSTESİ ..... **xv**
ÖZET ..... **xvii**
SUMMARY ..... **xix**
1. GİRİŞ **1**
1.1. Tezin Kapsamı 2
1.2. Tezin Amacı 2
1.3. Literatür Araştırması 2
1.4. Hipotez 5
2. VERİ MADENCİLİĞİ **7**
2.1. Veri Madenciliği Nedir? 7
2.2. Veri Madenciliği Uygulama Alanları 7
2.3. Veri Madenciliği Süreçleri 9
2.4. Veri Madenciliğinde Kullanılan Yöntemler 12
2.4.1. Sınıflandırma 12
2.4.2. Birliktelik analizi 12
2.4.3. Kümeleme 13
3. KÜMELEME **15**
3.1. Uzaklık Ölçüleri 16
3.2. Kümeleme Yöntemleri 18
3.3. K-Means Algoritması 18
4. UYGULAMA **21**
4.1. Veri Seti ve Özellikleri 21
4.2. Veri Ön İşleme ve Hazırlık Süreci 22
4.3. Analiz Sonuçları 29
4.4. K-Means Algoritmasının Uygulanması 38
4.4.1. K küme sayısının belirlenmesi 38
4.4.2. Kümeleme sonuçları 39
5. SONUÇ VE ÖNERİLER **45**
KAYNAKLAR **47**
EKLER **49**
ÖZGEÇMİŞ **51**

KISALTMALAR

CRISP-DM : Cross Industry Standard Process for Data Mining
EBYS : Elektronik Belge Yönetim Sistemi
SPSS : Statistical Package for the Social Sciences
TÜİK : Türkiye İstatistik Kurumu

TABLO LİSTESİ

Sayfa

Tablo 3.1. Örnek gözlem noktası değerleri. 19
Tablo 3.2. Gözlem noktalarının küme merkezine olan uzaklık değerleri. 20
Tablo 4.1. Daire başkanlıkları. 22
Tablo 4.2. Şube müdürlükleri. 24
Tablo 4.3. Ay dağılımları. 27
Tablo 4.4. Süre aralıkları. 28
Tablo 4.5. Dağıtım durumu. 28
Tablo 4.6. İlgi durumu. 28
Tablo 4.7. Ek durumu. 29
Tablo 4.8. db_id evrak dağılımı (ilk 6 ay). 30
Tablo 4.9. db_id evrak dağılımı (son 6 ay). 31
Tablo 4.10. Küme sayısı ve toplam hata (SSE) değişimi. 38
Tablo 4.11. Kümelerin eleman sayıları. 40

ŞEKİL LİSTESİ

Sayfa

Şekil 3.1. Örnek kümeleme görüntüsü. 15
Şekil 4.1. Kurum giden evrak dağılımı. 30
Şekil 4.2. Daire başkanlığı evrak dağılımı. 33
Şekil 4.3. Aylara göre evrak dağılımı (db_id). 33
Şekil 4.4. Aylara göre evrak dağılımı (mdr_id). 34
Şekil 4.5. Evrak süre aralığı. 35
Şekil 4.6. Evrak süre aralığı (db_id). 35
Şekil 4.7. Evrak karmaşıklığı. 36
Şekil 4.8. Değişkenler arası korelasyon matrisi. 37
Şekil 4.9. Elbow yöntemi 39
Şekil 4.10. db_id ve sure kümeleme grafiği. 41
Şekil 4.11. mdr_id ve sure kümeleme grafiği. 42
Şekil 4.12. ay_id ve sure kümeleme grafiği. 43

SAKARYA BÜYÜKŞEHİR BELEDİYESİNE AİT EVRAK VERİLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE İNCELENMESİ

ÖZET

Bu tez çalışmasında, kamu kurumlarında belge yönetim süreçlerinin analizi yoluyla hizmet kalitesinin artırılmasına katkı sağlanması amaçlanmaktadır. Bu kapsamda, Sakarya Büyükşehir Belediyesi'ne ait 2016 yılına ilişkin kurum giden evrak verileri kullanılarak, veri madenciliği yöntemlerinden biri olan kümeleme algoritması ile kapsamlı bir analiz gerçekleştirilmiştir. Belge yönetimi, kamu kurumlarında şeffaflık, izlenebilirlik ve verimlilik ilkeleri doğrultusunda kritik bir role sahiptir. Bu bağlamda, evrakların oluşturulma süresi, dağıtım durumu, ilgili belge referansları ve ek içerikleri gibi nitelikler aracılığıyla süreçlerin iyileştirilmesine yönelik veri temelli çıkarımlar yapılması hedeflenmiştir.

Çalışmanın yöntem kısmında, veri madenciliği sürecinde yaygın olarak kullanılan CRISP-DM (Cross-Industry Standard Process for Data Mining) modeli temel alınmıştır. Bu çerçevede, öncelikle iş probleminin tanımı yapılmış, ardından veri toplama, veri anlama, veri ön işleme, modelleme, değerlendirme ve uygulama aşamaları sistematik bir şekilde yürütülmüştür. Veri setinde yer alan eksik değerler temizlenmiş ve kategorik veriler sayısallaştırılma işlemleri uygulanarak analiz için uygun hale getirilmiştir. Modelleme sürecinde ise, veri seti üzerinde k-Ortalamalar (k-means) algoritması kullanılarak kümeler oluşturulmuştur. En uygun küme sayısının belirlenmesinde dirsek (elbow) yöntemi kullanılmıştır. Böylece, evrak yoğunluğunun belirli dönemsel örüntüler gösterip göstermediği ve bu örüntülerin personel izin planlamasına nasıl entegre edilebileceği ortaya konmuştur.

Uygulama aşamasında, Sakarya Büyükşehir Belediyesi'ne ait evrak yönetim sisteminden elde edilen 2016 yılına ait giden evrak verileri analiz edilmiştir. Yapılan kümeleme analizinde, evrakların çıkış süreleri, ait oldukları aylar ve belgeye eklenmiş ilgi veya ek içerik gibi nitelikler dikkate alınmıştır. K-means algoritması sonucunda, evrakların belirli dönemsel yoğunluklar sergilediği ve bu dönemlerde belge akışının daha fazla olduğu tespit edilmiştir. Özellikle yılın bazı aylarında, evrak çıkışlarının diğer dönemlere kıyasla anlamlı derecede arttığı gözlemlenmiş, bu durumun yıllık izin planlamaları açısından kritik bir gösterge olduğu değerlendirilmiştir. Bu tür dönemlerde personele ait izinlerin dikkatli planlanmaması durumunda, hizmet akışında aksamalar meydana gelebileceği anlaşılmıştır. Bu bağlamda, kümeleme sonucu elde edilen zaman dilimleri, yöneticilere hangi zaman aralıklarında yoğun belge işlemleri olduğunu göstererek karar destek sistemlerine katkı sağlamaktadır.

Analiz sürecinde ayrıca, evraklara ilişkin dağıtım, ilgi ve ek bilgisi gibi niteliklerin belge işleme süresi üzerindeki etkileri de incelenmiştir. Korelasyon analizleri neticesinde, özellikle ilgi ve dağıtım bilgisi içeren evrakların işlem sürelerinin daha uzun olduğu ortaya konmuştur. Bu durum, bu tür evrakların daha fazla kontrol ve içerik değerlendirmesi gerektirdiğini göstermektedir. Ayrıca korelasyon matrisleri kullanılarak değişkenler arası doğrusal ilişkiler değerlendirilmiş ve istatistiksel açıdan anlamlı bazı ilişkiler tespit edilmiştir. Elde edilen bulgular, evrakların işlenme

sürelerini etkileyen yapısal faktörlerin anlaşılmasına katkı sağlamış ve belge yönetim süreçlerinin iyileştirilmesi adına önemli iç görüler sunmuştur. Bu sayede, veri temelli yaklaşımların kamu kurumlarında sadece teknik değil, stratejik planlamalar için de rehberlik edebileceği vurgulanmıştır.

Sonuç olarak, bu tez çalışması, kamu kurumlarında yürütülen belge yönetim süreçlerinin yalnızca arşivleme, kayıt altına alma ya da yasal yükümlülükleri yerine getirme işleviyle sınırlı kalmadığını ortaya koymaktadır. Bilgi teknolojilerindeki gelişmelerin etkisiyle, kamu kurumları bünyesinde dijitalleşen belge yönetim sistemleri, kurumsal verinin hem takibi hem de analizi açısından önemli bir potansiyel sunmaktadır. Bu bağlamda, belge hareketlerinin sistematik biçimde incelenmesi; işlem süreleri, belge nitelikleri ve dönemsel yoğunluklar gibi parametrelerin analiziyle hizmet süreçlerinin iyileştirilmesine yönelik anlamlı örüntülerin keşfedilmesini mümkün kılmaktadır. Tez kapsamında gerçekleştirilen uygulama, belediye evraklarının barındırdığı bilginin veri madenciliği teknikleriyle analiz edilebileceğini ve bu analizler sonucunda kurumsal karar süreçlerine ışık tutacak veriye dayalı çıktılar üretilebileceğini göstermiştir.

K-means kümeleme algoritması ile elde edilen gruplamalar, özellikle insan kaynakları planlaması ve iş gücü yönetimi açısından değerli bilgiler sunmaktadır. Evrak çıkış süreleri ve dönemsel belge yoğunluğu verileri, kurumun yıl içerisindeki iş yükü dağılımını ortaya koymuş; bu da yıllık izinlerin hangi dönemlerde verilmesinin hizmet sürekliliği açısından daha uygun olacağını göstermiştir. Örneğin, belge hareketliliğinin yüksek olduğu ayların belirlenmesi, personelin bu dönemlerde izin kullanmasının hizmet performansını olumsuz etkileyebileceğini işaret etmektedir. Bu tür veri temelli çıkarımlar, klasik yöntemlerle planlanan izin süreçlerine kıyasla çok daha rasyonel, nesnel ve kurumun gerçekliğine uygun çözümler sunmaktadır. Böylece, kamu hizmetlerinde süreklilik ve kalite korunarak, çalışan memnuniyeti ile birlikte vatandaş memnuniyetinin de artırılması hedeflenebilmektedir.

Bu çalışmanın ortaya koyduğu bulgular ve uygulama sonuçları, yalnızca akademik katkı sunmakla kalmamakta; aynı zamanda yerel yönetimlerde stratejik planlama, süreç yönetimi ve kaynak tahsisi gibi temel yönetim alanlarına da pratik öneriler geliştirmektedir. Tezde önerilen model, kurumların sahip oldukları veri altyapısını daha etkin kullanmalarına olanak tanırken, veri madenciliği yöntemlerinin karar destek mekanizmalarına entegre edilmesiyle yönetimsel öngörülerin geliştirilmesine yardımcı olmaktadır. Özellikle belediye gibi vatandaşla doğrudan temas hâlindeki kurumlarda, hizmetlerin sürdürülebilirliğini sağlamak ve kaynakları verimli kullanmak açısından bu tür analitik yaklaşımlar kritik bir rol üstlenmektedir. Bu yönüyle tez, hem veri bilimi hem de kamu yönetimi literatürüne uygulamalı bir katkı sunmakta; benzer çalışmalar için örnek teşkil etmektedir.

ANALYSIS OF DOCUMENT DATAS OF SAKARYA METROPOLITAN MUNICIPALITY USING DATA MINING METHODS

SUMMARY

This thesis aims to contribute to the improvement of service quality by analyzing document management processes in public institutions. In this context, a comprehensive analysis was carried out with the clustering algorithm, one of the data mining methods, using the outgoing document data of Sakarya Metropolitan Municipality for the year 2016. Document management has a critical role in public institutions in line with the principles of transparency, traceability and efficiency. In this context, it is aimed to make data-based inferences for the improvement of processes through attributes such as the creation time of documents, distribution status, related document references and additional content.

The methodology of the study is based on the CRISP-DM (Cross-Industry Standard Process for Data Mining) model, which is widely used in the data mining process. In this framework, firstly, the business problem was defined and then the data collection, data understanding, data preprocessing, modeling, evaluation and application stages were carried out systematically. Missing values in the data set were cleaned, categorical data were digitized and normalization processes were applied to make them suitable for analysis. In the modeling process, clusters were formed using the K-Means algorithm on the data set. Elbow method was used to determine the optimal number of clusters. Thus, it was revealed whether document density shows certain periodic patterns and how these patterns can be integrated into personnel leave planning.

As the first step of the implementation phase, outgoing document data of Sakarya Metropolitan Municipality for the year 2016 were obtained and the data mining process was initiated on these data. The data set consists of a total of 16535 datas and 9 variables. These variables cover areas that are considered critical in document management systems such as record number, document duration, month information, directorate and department to which it belongs, distribution, attachment and interest information. The data set was first analyzed for missing data and 10 datas were identified as missing and these records were removed from the system. Then, categorical data were converted into numerical form and made suitable for the analysis process. In addition, data transformation processes were carried out to analyze the variables in a more meaningful way. At this stage, all data were digitized and a structure suitable for the k-means clustering algorithm was obtained.

In the second stage, the modeling step was started in line with the CRISP-DM process and the k-means algorithm was applied on the data set. At this stage, Elbow method was first used to determine the optimal number of clusters and as a result of graphical analysis, k=4 was selected as the optimal number of clusters. Then, clusters were formed using RapidMiner software and the common characteristics of the documents in the same clusters were analyzed. The results of the analysis revealed that the intensity of document output increased significantly in some periods compared to other periods. It was also found that during these intensity periods, the rate of inclusion of

interest or additional content to the document also increased. This increases not only the volume of document processing, but also document complexity and processing time.

In the last stage of the implementation, inferences were made in terms of internal processes in line with the clusters obtained. It was observed that document processing times are prolonged, especially during periods of high work intensity, and if the number of personnel using annual leave increases during these processes, disruptions in service flow may occur. In this context, it is recommended that personnel leave planning be made in a data-driven manner, taking into account document mobility and content density. In this way, service continuity can be maintained and staff satisfaction can be increased. This application is an important example to show that data mining methods can be integrated into decision support mechanisms in large public institutions such as municipalities.

In the analysis process, the effects of document attributes such as distribution, interest and attachment information on document processing time were also analyzed. Correlation analyses revealed that processing times were longer for documents that contained interest and distribution information. This indicates that such documents require more control and content evaluation. In addition, linear relationships between variables were evaluated using correlation matrices and some statistically significant relationships were identified. The findings contributed to the understanding of the structural factors affecting document processing times and provided important insights for improving document management processes. In this way, it is emphasized that data-driven approaches can guide not only technical but also strategic planning in public institutions.

As a result, this thesis reveals that document management processes carried out in public institutions are not limited to archiving, recording or fulfilling legal obligations. With the impact of developments in information technologies, digitalized document management systems within public institutions offer a significant potential for both tracking and analyzing institutional data. In this context, systematic examination of document movements makes it possible to discover meaningful patterns for improving service processes by analyzing parameters such as processing times, document qualities and periodic intensities. The application carried out within the scope of this thesis has shown that the operational information contained in municipal documents can be analyzed with data mining techniques and that data-based outputs that will shed light on corporate decision-making processes can be produced as a result of these analyzes.

The groupings obtained with the K-means clustering algorithm provide valuable information especially for human resources planning and workforce management. The document exit times and periodic document density data revealed the workload distribution of the organization throughout the year, which in turn showed which periods of annual leave would be more appropriate in terms of service continuity. For example, identifying months with high document mobility indicates that staff taking leave during these periods may have a negative impact on service performance. Such data-driven inferences provide solutions that are much more rational, objective and in line with the operational reality of the organization compared to the leave processes planned by conventional methods. Thus, by maintaining continuity and quality in public services, it can be aimed to increase citizen satisfaction as well as employee satisfaction.

The findings and application results of this study not only contribute academically, but also provide practical recommendations for key management areas such as strategic planning, process management and resource allocation in local governments. While the model proposed in the thesis enables organizations to use their data infrastructure more effectively, it helps to develop managerial insights by integrating data mining methods into decision support mechanisms. Especially in organizations such as municipalities that are in direct contact with citizens, such analytical approaches play a critical role in ensuring the sustainability of services and using resources efficiently. In this respect, the thesis makes an applied contribution to both data science and public administration literature and sets an example for similar studies.

1. GİRİŞ

Kamu kurum ve kuruluşlarında bilgi yönetiminin en temel bileşenlerinden biri olan evrak yönetimi, idari süreçlerin etkin, şeffaf ve izlenebilir bir biçimde yürütülmesini sağlamaktadır. Bu bağlamda, özellikle yerel yönetim birimleri olan belediyelerde, evrakların doğru biçimde kaydedilmesi, sınıflandırılması, saklanması ve gerektiğinde erişilebilirliğinin sağlanması, kurumsal hafızanın oluşturulması ve hizmet kalitesinin artırılması açısından kritik öneme sahiptir. Elektronik belge yönetim sistemlerinin yaygınlaştırılmasının, donanım ve yazılım altyapısının sürekli güncellenmesinin ve kamu görevlilerine düzenli eğitim verilmesinin, kamu hizmetlerinin sunumunda niteliği artıracağını göstermektedir (Tamtürk, 2017). Evrak yönetim sistemleri, bu süreçleri dijital ortamda standartlara uygun biçimde gerçekleştirmeye yönelik olarak geliştirilmiş bilgi sistemleridir.

Kamu kurumlarında hizmet sürekliliği ve vatandaş memnuniyetinin sağlanması, etkili bir insan kaynakları yönetimini zorunlu kılmaktadır. Bu bağlamda, personelin yıllık izinlerinin planlı ve dengeli bir şekilde yönetilmesi, kurum içi iş akışının aksamadan sürdürülmesine katkı sunmaktadır. Özellikle büyük ölçekli kamu kuruluşlarında, personel sayısının fazlalığı ve hizmet çeşitliliği, izin planlamasını karmaşık ve zaman alıcı bir süreç hâline getirmektedir. Bu nedenle, izin planlamasının veri odaklı ve sistematik yaklaşımlarla desteklenmesi önemli bir ihtiyaç olarak ortaya çıkmaktadır.

Sakarya Büyükşehir Belediyesi, geniş hizmet yelpazesi ve kalabalık personel yapısıyla bu gereksinimi yakından hisseden kurumlardan biridir. Belediyede kullanılan evrak yönetim sistemleri; evrak kayıtları, personel işlemleri ve izin talepleri gibi birçok idari sürecin dijital ortamda yürütülmesine olanak tanımaktadır. Bu sistemlerde biriken büyük hacimli veriler, yalnızca idari kayıtlar olarak değil, aynı zamanda karar destek sistemleri için değerli birer kaynak niteliğindedir.

Bu tez çalışmasında, Sakarya Büyükşehir Belediyesi evrak yönetim sistemi üzerinden elde edilen giden evrak verileri kullanılarak, personelin yıllık izin planlamasına yönelik bir analiz gerçekleştirilmesi amaçlanmaktadır. Çalışma kapsamında, veri madenciliği yöntemlerinden biri olan k-means kümeleme algoritması kullanılacaktır.

Bu yöntemle, benzer nitelikteki dönemsel evrak yoğunlukları tespit edilerek, izin planlamasında dikkate alınabilecek dönemsel iş yükleri belirlenecektir. Böylece, izin taleplerinin hizmet kalitesini düşürmeyecek şekilde yönlendirilmesi mümkün olacaktır.

Yapılan çalışmada, Sakarya Büyükşehir Belediyesi'nin 2016 yılına ait kurum giden evrak verileri kullanılmıştır. Bu veriler kümeleme yöntemlerinden k-means algoritması kullanılarak analiz edilmiş ve sonuçları incelenmiştir.

1.1. Tezin Kapsamı

Çalışma kapsamında, veri madenciliği yöntemlerinden biri olan k-means kümeleme algoritması kullanılacaktır. Bu yöntemle, benzer nitelikteki dönemsel evrak yoğunlukları tespit edilecek ve evrak işlem süresini etkileyen faktörler incelenerek belge yönetim süreçlerinin iyileştirilmesine yönelik öneriler sunulacaktır. Böylece, yıllık izin taleplerinin, hizmet kalitesini düşürmeyecek ve iş süreçlerinde darboğazlar oluşturmayacak şekilde planlanması sağlanmış olacaktır.

1.2. Tezin Amacı

Bu çalışmada, Sakarya Büyükşehir Belediyesi'nin 2016 yılına ait kurum giden evrak verileri kullanılmıştır. Bu veriler kümeleme yöntemlerinden k-means algoritması kullanılarak analiz edilmiş ve sonuçları incelenmiştir.

Bu çalışmanın amacı veri madenciliği tekniklerinin kamu yönetiminde karar destek süreçlerine entegre edilmesine yönelik uygulamalı bir örnek sunmaktır. Aynı zamanda, belediyelerde dijital dönüşümün sadece işlem kolaylığı değil, aynı zamanda stratejik planlama için de fırsatlar sunduğunu göstermeyi hedeflemektedir.

1.3. Literatür Araştırması

Gelen ve giden evrak işlemleri, kurum dışı ve kurum içi iletişimin temel yapı taşlarını oluşturmaktadır. Belediyelere ulaşan her türlü belge, resmî yazılar, dilekçeler, bilgi talepleri ve benzeri içerikler, sistematik biçimde kaydedilerek iş akışına dahil edilmekte; aynı şekilde, kurum dışına gönderilen belgeler de belirli protokoller çerçevesinde düzenlenerek kayıt altına alınmaktadır. Bu süreçlerin sağlıklı bir şekilde yönetilebilmesi, yalnızca hukuki ve idari sorumlulukların yerine getirilmesi açısından

değil, aynı zamanda kurumsal performansın ve vatandaş memnuniyetinin artırılması bakımından da önem arz etmektedir.

Geleneksel kâğıt tabanlı evrak yönetim anlayışının yerini dijital tabanlı sistemlere bırakmasıyla birlikte, belediyelerde Elektronik Belge Yönetim Sistemleri (EBYS) gibi entegre çözümler ön plana çıkmıştır. Bu sistemler sayesinde belge akışının hızlanması, belgeye erişimin kolaylaşması, izlenebilirliğin artması ve arşivleme süreçlerinin daha verimli hâle getirilmesi mümkün olmuştur. Ayrıca, 5070 sayılı Elektronik İmza Kanunu ve ilgili diğer yasal düzenlemeler çerçevesinde belgelerin hukuki geçerliliğinin dijital ortamda da güvence altına alınması, elektronik evrak yönetimini kamu kurumları için kaçınılmaz bir gereklilik hâline getirmiştir.

Literatürde kümeleme yöntemleri kullanarak farklı yapılan çalışmalar mevcuttur.

Ahmed, Seraj ve Islam (2020) tarafından kaleme alınan bu derleme çalışması, k-means algoritmasının veri madenciliği alanındaki güçlü yönlerini ve sınırlılıklarını kapsamlı bir biçimde incelemektedir. Çalışmada, algoritmanın rastgele başlangıç merkezlerine duyarlılığı, küme sayısının önceden belirlenmesi zorunluluğu ve karma veri türlerini işleyememesi gibi temel problemler detaylandırılmıştır. Bu eksiklikleri gidermeyi amaçlayan çeşitli algoritma varyantları (örneğin: constrained k-means, x-means, k-prototype, kernel k-means) hem teorik hem de deneysel olarak değerlendirilmiştir. Altı farklı veri kümesi üzerinde yapılan karşılaştırmalı deneysel analizlerde, hiçbir algoritmanın tüm veri setleri üzerinde tutarlı üstünlük göstermediği, her bir varyantın uygulama veya veri türüne özgü avantajlar sunduğu sonucuna ulaşılmıştır.

Gül ve Taşyürek (2025) tarafından gerçekleştirilen bu çalışmada, çevre kirliliğini azaltmak amacıyla geri dönüşüm konteynerlerinin yerlerinin belirlenmesine yönelik çok boyutlu bir k-ortalamalar (k-means) kümeleme modeli önerilmiştir. Çalışmanın amacı, geri dönüşüm davranışlarını etkileyen sosyo-demografik ve çevresel faktörleri dikkate alarak uygun geri dönüşüm noktalarının optimizasyonunu sağlamaktır. Yöntem olarak, İstanbul iline ait mahalle düzeyinde nüfus, konut tipi, gelir düzeyi gibi çeşitli parametreler kullanılmış; bu çok boyutlu veri, k-means algoritmasıyla analiz edilmiştir. Yapılan analiz sonucunda, geri dönüşüm konteynerlerinin rastgele değil, veri temelli bir yaklaşımla konumlandırılmasının atık toplama verimliliğini ve çevresel sürdürülebilirliği artırabileceği ortaya konmuştur. Çalışma, kamu politikalarının veri odaklı planlanmasına katkı sağlayabilecek nitelikte önemli bulgular içermektedir.

Avcı ve diğerleri (2020) tarafından gerçekleştirilen bu çalışmanın amacı, büyükşehir belediyesi çağrı merkezi verilerinin veri madenciliği yöntemlerinden kümeleme analizi ile incelenerek anlamlı müşteri gruplarının ortaya çıkarılmasıdır. Araştırmada Türkiye'de nüfus yoğunluğu yüksek bir büyükşehre ait beş ilçeden toplanan çağrı verileri değerlendirilmiş ve k-ortalamalar algoritması kullanılmıştır. Analiz kapsamında ilişkili olduğu birim, başvuru tipi, başvuru ilçesi, eğitim durumu, cinsiyet, yaş ve anlık çözüm değişkenleri dikkate alınmıştır. SPSS Clementine ve WEKA programları aracılığıyla gerçekleştirilen kümeleme sonucunda, en uygun küme sayısı altı olarak belirlenmiş ve özellikle ulaşım birimine yönelik başvuruların yoğunlukta olduğu görülmüştür. Çalışma, çağrı merkezi performansının artırılması ve müşteri memnuniyetine yönelik stratejilerin geliştirilmesinde veri madenciliği tekniklerinin etkin bir araç olduğunu ortaya koymaktadır.

Mihevc ve Podržaj (2025) tarafından gerçekleştirilen bu çalışmada, elektronik iletişim verileri üzerinde kümelenme analizinde kullanılan Elbow yönteminin optimal küme sayısını belirlemedeki etkinliği incelenmiştir. Çalışmanın amacı, Elbow yönteminin nicel bir analiz aracı olarak ne ölçüde güvenilir sonuçlar sunduğunu, nitel değerlendirmelerle karşılaştırmalı olarak ortaya koymaktır. Araştırma kapsamında, Avrupa'daki 32 ülkeye ait FTTP, 5G kapsama oranı, yüksek bant genişliği erişimi ve nüfus yoğunluğu gibi dört parametre kullanılarak K-means algoritmasıyla kümelenme yapılmış; Python ortamında Elbow yöntemi ile elde edilen sonuçlar, RapidMiner yazılımı aracılığıyla nitel olarak değerlendirilmiştir. Sonuçlar, üç kümenin en uygun çözüm olduğunu göstermiş; bu yapı, ülkeleri iletişim altyapısı gelişmişlik düzeyine göre anlamlı biçimde sınıflandırmaktadır. Bu bulgular, Elbow yönteminin küçük ve homojen veri kümelerinde etkili bir küme sayısı tahmini sunduğunu ortaya koymuştur.

Zhao (2025) tarafından gerçekleştirilen bu çalışmada, nesnelerin interneti (IoT) cihazlarının veri taleplerini karşılamak üzere sis (fog) bilişim ve blokzincir teknolojilerinden yararlanılarak uç noktada önbellekleme performansının artırılması amaçlanmıştır. Bu kapsamda geliştirilen CUPE algoritması, kullanıcı tercihlerini K-means algoritması ile gruplayarak içerik önbelleklemesini optimize etmekte; derin sinir ağları ile içerik eğilimlerini tahmin etmektedir.

Baydan ve diğerleri (2025) tarafından yapılan çalışmada, Ankara ili mahalleleri kır–kent özellikleri açısından incelenmiş ve K-means algoritmasıyla sınıflandırılmıştır. Analiz sonucunda mahalleler, yoğun kır, yoğun kent, kırsal etkileşimli kentsel ve

kentsel etkileşimli kırsal olmak üzere dört grupta toplanmıştır. Elde edilen bulgular, K-means yönteminin yerleşim birimlerinin niteliğini belirlemede ve bölgesel planlama süreçlerine katkı sunmada kullanılabileceğini göstermektedir.

Kılıç ve diğerleri (2020) tarafından yapılan çalışmada, Türkiye'deki 81 il belediyesi, çevre hizmetleri açısından incelenmiştir. 2001–2016 yıllarına ait TÜİK verileri kullanılarak altı kriter üzerinden değerlendirme yapılmıştır. Kriter ağırlıkları Analitik Hiyerarşi Süreci (AHP) yöntemiyle belirlenmiş ve elde edilen değerler tek boyutlu kümeleme analizi ile sınıflandırılmıştır. Analiz sonucunda illerin yıllar içindeki belediye hizmet performanslarındaki değişim ortaya konmuştur.

Özari ve diğerleri (2018) tarafından yapılan çalışmanın amacı, TÜİK tarafından 2016 yılında yayımlanan "İllerde Yaşam Endeksi"ni temel alarak Türkiye'deki illerin yaşam kalitesini farklı bir bakış açısıyla değerlendirmektir. Söz konusu endeks; eğitim, sağlık, güvenlik, çevre, altyapı gibi 11 boyut ve toplam 41 alt göstergeden oluşmaktadır. Çalışmada öncelikle çok boyutlu ölçekleme yöntemi kullanılarak iller iki boyutlu düzlemde görselleştirilmiş ve aralarındaki benzerlikler noktasal dağılım üzerinden incelenmiştir. Sonraki aşamada ise K-ortalamalar algoritması ile iller kümelere ayrılmış ve elde edilen gruplar TÜİK'in yayımladığı sıralama sonuçlarıyla karşılaştırılmıştır. Analizler, bazı önemli farklılıklar olduğunu göstermiş ve endeks sonuçlarının kullanılan yönteme göre değişkenlik gösterebileceğini ortaya koymuştur. Bu durum, yaşam endeksinin yalnızca tek yıllık verilerle değerlendirilmesinin sınırlı olacağını; uzun dönemli verilerin ve farklı sıralama tekniklerinin kullanılması gerektiğini göstermektedir. Çalışma, yaşam kalitesinin ölçümünde çok kriterli yaklaşımların ve veri madenciliği tekniklerinin önemini vurgulamaktadır.

1.4. Hipotez

Bu çalışmanın temel amacı, Sakarya Büyükşehir Belediyesi'ne ait 2016 yılı kurum giden evrak verileri üzerinde k-means algoritması kullanılarak gerçekleştirilecek kümeleme analizi ile evrakların ait olduğu birim, evrağın kurumdan çıkış süresi ve zamanı, dayandığı belge veya yazışma (ilgi), yazıya eklenen belge, tablo, çizim, fotoğraf gibi dokümanlar (ek) ve yazının gönderildiği kişi veya kurumlar (dağıtım) gibi nitelikler doğrultusunda belirli örüntüler ortaya koymak ve bu örüntüler aracılığıyla kurum içi belge yönetim süreçlerinin iyileştirilmesine katkı sağlayacak anlamlı bilgi kümeleri elde etmektir.

2. VERİ MADENCİLİĞİ

2.1. Veri Madenciliği Nedir?

Veri madenciliği (data mining), büyük ve karmaşık veri yığınları içerisinden anlamlı, yeni, yararlı ve daha önce bilinmeyen örüntülerin, ilişkilerin ve eğilimlerin keşfedilmesi sürecidir (Han, Kamber & Pei, 2011). Veri madenciliği, istatistik, yapay zeka, makine öğrenmesi ve veritabanı sistemlerinin bir araya gelmesiyle oluşmuş disiplinler arası bir çalışma alanıdır.

Günümüzde kamu kurumları, özel sektör ve akademik çevreler veri madenciliğini karar destek sistemlerinde, tahminleme çalışmalarında ve süreç optimizasyonlarında aktif olarak kullanmaktadır. Özellikle kamu yönetiminde kaynakların etkin kullanımı, hizmet kalitesinin artırılması ve süreçlerin iyileştirilmesi gibi amaçlar doğrultusunda veri madenciliği önemli katkılar sağlamaktadır.

Veri madenciliği konusunda bir çok farklı tanım bulunmaktadır. Bir tanım yapmak gerekirse, Büyük veri kümeleri içerisinden daha önce bilinmeyen, geçerli, kullanılabilir ve anlamlı bilgilerin otomatik ya da yarı otomatik yöntemlerle keşfedilmesi süreci veri madenciliği olarak tanımlanabilmektedir.

Veri madenciliğinin temel amacı, büyük ve karmaşık veri yığınları içerisinden anlamlı bilgi ve içgörü elde etmektir. Bu sayede, mevcut verilerden stratejik çıkarımlar yaparak daha bilinçli kararlar alınabilmektedir. Elde edilen bilgiler, karar destek sistemlerine entegre edilerek yöneticilerin daha etkili ve hızlı karar vermelerine katkı sağlamaktadır. Ayrıca, geçmiş verilere dayalı örüntüler ve eğilimler analiz edilerek geleceğe yönelik tahmin ve modellemeler yapılabilmektedir. Böylece belirsizliklerin azaltılması ve kaynakların daha verimli kullanılması mümkün hâle gelmektedir.

2.2. Veri Madenciliği Uygulama Alanları

Veri madenciliği, günümüzde birçok sektörde yaygın olarak kullanılan bir analiz yöntemidir. Veri madenciliği, sektör bağımsız olarak geniş bir uygulama alanına sahiptir ve her sektörde farklı ihtiyaçlara yönelik olarak özgün çözümler üretmektedir.

Finans, sağlık, eğitim, pazarlama ve e-ticaret gibi alanların yanı sıra, kamu yönetimi ve belediyecilik hizmetlerinde etkin biçimde uygulanmaktadır.

Belediyeler; ruhsat işlemleri, şikâyet yönetimi, sosyal yardımlar, imar faaliyetleri ve yazışma süreçleri gibi birçok alanda yoğun veri üretmektedir. Bu bağlamda, veri madenciliği teknikleri kullanılarak vatandaş taleplerinin analizi, hizmet yoğunluğu haritalarının çıkarılması ve kaynakların daha etkin planlanması mümkün hâle gelmektedir. Özellikle evrak yönetim sistemleri üzerinden elde edilen kullanılarak kurum içi iş akış süreçlerinde sıkışmaların tespiti, belge dolaşımının optimizasyonu ve personel görev dağılımlarının veriye dayalı şekilde yeniden yapılandırılması sağlanmaktadır. Böylece hem hizmet kalitesi artırılmakta hem de zaman ve iş gücü kaybı en aza indirilmiş olmaktadır.

Finans sektöründe, dolandırıcılık tespiti (fraud detection), kredi risk analizleri ve müşteri segmentasyonu gibi konular ön plandadır. Özellikle bankacılık işlemleri üzerinden elde edilen büyük veri setleri kullanılarak şüpheli işlem örüntüleri tespit edilebilmekte ve riskli kredi başvuruları önceden öngörülebilmektedir.

Sağlık alanında ise hasta verileri üzerinden teşhis destek sistemleri geliştirilebilmekte, hastalık sınıflandırmaları yapılabilmekte ve bireyselleştirilmiş tedavi yaklaşımları sunulabilmektedir. Örneğin, hasta geçmişi ve laboratuvar verilerine dayalı olarak kanser türlerinin erken evrede tanısı mümkün hâle gelebilmektedir.

Eğitim alanında veri madenciliği, öğrencilerin başarı durumlarının analizi, öğrenme sürecine etkide bulunan faktörlerin belirlenmesi ve erken uyarı sistemlerinin geliştirilmesinde kullanılmaktadır. Öğrenme yönetim sistemlerinden elde edilen veriler ile öğrencilerin davranışsal eğilimleri sınıflandırılabilir, böylece akademik başarısızlık riski taşıyan bireyler önceden tespit edilebilmektedir.

Pazarlama ve e-ticaret sektörlerinde ise müşteri davranışlarını analiz etme, çapraz satış fırsatlarını belirleme, kişiselleştirilmiş öneri sistemleri oluşturma ve sadakat programları tasarlama gibi uygulamalar ön plana çıkmaktadır. Özellikle müşteri alışveriş geçmişi, tıklama verileri ve sosyal medya etkileşimleri gibi çoklu veri kaynakları bir arada değerlendirilerek daha etkili pazarlama stratejileri geliştirilebilmektedir.

Veri madenciliğinin bu denli geniş bir uygulama alanına sahip olmasının temel nedenlerinden biri, yapay zekâ ile benzer algoritmaları ve yöntemleri paylaşmasıdır

(Silahtaroğlu, 2016). Bu iki disiplin, birbirlerini karşılıklı olarak besleyen ve geliştiren nitelikte çalışmakta; aralarındaki teknik yakınlık, çeşitli sektörlerdeki yaygın kullanımını da desteklemektedir. Ayrıca veri madenciliği, yalnızca mevcut verileri anlamlandırmakla kalmayıp sektörel düzeyde rekabet avantajı elde etmeye de olanak tanımaktadır.

2.3. Veri Madenciliği Süreçleri

Veri madenciliği giderek yaygınlaşan bir alan haline gelmesi ile farklı disiplinlerin farklı uygulama yaklaşımları oluşmuştur ve ortak bir metodoloji üzerinde çalışma gereksinimi ortaya çıkmıştır. Bu gereksinimin karşılanması için Avrupa Birliği destekli bir araştırma konsorsiyumu olan CRISP-DM (Cross-Industry Standard Process for Data Mining), 1996 yılında kurumsal ve akademik iş birlikleriyle geliştirilmiştir. 2000'li yıllardan beri hem sektör hem akademi tarafından benimsenmiş ve kullanılmaktadır. Bu süreç altı aşamadan oluşmaktadır. Bu adımlardan sırasıyla aşağıda bahsedilecektir.

İşi anlama: Proje hedeflerinin ve başarı kriterlerinin iş açısından tanımlandığı aşamadır. Bu aşama, projenin sadece teknik yönleriyle değil, iş hedefleriyle doğrudan ilişkili şekilde ele alınmasını sağlamaktadır. Bu adımda:

- Projenin amacı ve başarım ölçütleri tanımlanmaktadır.
- Mevcut iş süreçleri analiz edilmektedir.
- Veri madenciliği hedefleri iş hedefleriyle hizalanmaktadır.
- Proje planı oluşturulmaktadır.

Kamu kurumlarında bu aşama, hizmet kalitesini artırmak, süreçleri iyileştirmek ya da kaynak yönetimini optimize etmek gibi hedefleri kapsayabilmektedir.

Veriyi anlama: Mevcut verinin toplanması, incelenmesi ve iş problemleriyle olan ilişkilerinin analiz edildiği bölümdür. Bu aşama, eldeki verilerin keşfi ve analizine odaklanmaktadır. Amaç, verilerin yapısını, kalitesini ve potansiyel problemlerin belirlenmesini sağlamaktır. Bu adımda:

- Veri kaynakları tanımlanmaktadır.
- İlk örneklem veriler bu aşamada toplanmaktadır.
- Verilerle ilgili temel istatistikler, eksik veriler veya uç değerler tespit edilmektedir.

Belediyelerde evrak kayıt sistemlerinden, izin taleplerinden ya da hizmet süreçlerinden gelen veriler bu aşamada kontrol edilmektedir.

Veri hazırlığı: Analize uygun veri kümesinin oluşturulması için veri temizleme, dönüştürme, seçme gibi işlemler gerçekleştirilmektedir. Veri madenciliği için gerekli veri kümesinin oluşturulduğu bu aşama, zamanın büyük kısmını kapsar. Bu adımda:

- Eksik Verilerin İşlenmesi: Verilerde bazı gözlem değerleri eksik olabilmektedir. Bu eksik verilerin giderilmesi için bazı yaklaşımlar mevcuttur.
    - Eksik verilerin sayısı veri setinde göz ardı edilecek boyutta ise eksik kayıtların silinmesi
    - Eksik verilerin diğer verilerden yararlanılarak ortalama, medyan, mod gibi istatistiklerle doldurulması.
    - Eksik verilerin diğer verilerden yararlanılarak knn, regresyon tahmini, çoklu atama gbi yöntemler ile doldurulması.
    - Veritabanına ulaşım kolya sağlanılabiliyorsa eksik olan veriler in el yordamı ile tamamlanması.
- Gürültülü Verilerin Temizlenmesi: Ölçüm hataları, veri girişindeki yazım yanlışları gibi sorunlar gürültü olarak adlandırılmaktadır. Gürültülü verilerin temizlenmesinde çeşitli yöntemlerden yararlanılmaktadır. Bu yöntemler arasında; kutulama tekniği ile veri değerlerinin belirli aralıklara bölünerek sadeleştirilmesi, sınır tabanlı düzeltme yöntemleriyle verilerin belirli eşiklere göre düzenlenmesi ve kümeleme algoritmaları yardımıyla aykırı değerlerin tespit edilerek elimine edilmesi gibi yaklaşımlar yer almaktadır.
- Yinelenen Verilerin Kaldırılması: Aynı kayıtların birden fazla kez veri setinde yer alması analiz sonuçlarını bozabilmektedir. Bu nedenle bu kayıtların tespit edilip tekararlayan verinin veritabanından çıkarılması gerekmektedir.
- Veri Dönüştürme: Verilerin belirli bir formata dönüştürülmesi gerekebilmektedir.
    - Kategorik verilerin sayısallaştırılması: Veri madenciliği ve makine öğrenmesi algoritmalarının büyük çoğunluğu yalnızca sayısal verilerle çalışabilmektedir. Ancak gerçek dünyadaki birçok veri kümesinde kategorik değişkenler de bulunmaktadır. Bu nedenle, analizden önce bu tür verilerin uygun biçimde sayısal forma dönüştürülmesi gerekmektedir.

- Görsel etiketleme: Bu yöntemde, veri setindeki sayısal değişkenler için histogram grafiği oluşturularak dağılım incelenmektedir. Histogram üzerinde verilerin yoğunluk gösterdiği ve ani geçişlerin yaşandığı noktalar analiz edilerek, bu ani değişimlerin görüldüğü bölgeler veri aralıklarının doğal sınırları olarak kabul edilerek bu sınırlara göre kategorik etiketler tanımlanmaktadır.
- Normalizasyon: Veri madenciliği uygulamalarında, farklı ölçekteki değişkenlerin birlikte analiz edilmesi durumunda modelin güvenilirliğini ve doğruluğunu artırmak amacıyla normalizasyon işlemi yaygın biçimde kullanılmaktadır. Normalizasyon, özellikle değişkenler arasında birim ve değer aralığı farkı olduğunda, bu farkların analizi yanıltıcı şekilde etkilemesini önlemek için uygulanmaktadır. Bu işlem, her bir özelliğin ortak bir ölçeğe genellikle [0,1] aralığına indirgenmesini sağlar. Özellikle mesafe tabanlı algoritmalar (örneğin k-nearest neighbors, k-means) için verilerin benzer ölçeklerde olması, doğru benzerlik ölçümlerinin yapılabilmesi açısından kritik öneme sahiptir. Bu bağlamda en sık başvurulan yöntemlerden biri min-max normalizasyonu olup, her bir veri noktasının minimum ve maksimum değerler arasındaki konumuna göre yeniden ölçeklenmesini sağlamaktadır. Böylece, tüm değişkenlerin analize eşit ağırlıkla katkı sunması mümkün hâle gelmekte ve modelin önyargılı sonuç üretmesi engellenmiş olmaktadır.

Modelleme: Veri madenciliği algoritmalarının uygulandığı aşamadır. Bu adımda:

- Uygun algoritmanın seçilmesi (örneğin: kümeleme için K-means),
- Model parametrelerinin ayarlanması (örneğin: küme sayısı),
- Modelin uygulanması,
- Gerekirse alternatif algoritmaları denenmesi

Bu aşamada teknik bilgi ön plana çıkmaktadır. Model performansını etkileyen faktörler yapılan çalışmaların sonucunda belirlenmektedir.

Değerlendirme: Modelleme sonucunda elde edilen bulguların yalnızca teknik değil, iş hedefleri açısından da uygunluğu değerlendirilmektedir. Bu adımda;

- Modelin doğruluğu, istatistiksel ölçütlerle test edilmektedir.
- İş hedefleri ile model sonuçları karşılaştırılmaktadır.
- Modelin anlamlılığı ve uygulanabilirliği sorgulanmaktadır.

Elde edilen kümelerin pratikte işe yarayıp yaramadığı (örneğin hangi personelin hangi dönemde izin kullandığına göre yapılan gruplamanın yöneticilere katkı sağlayıp sağlamadığı) bu aşamada belirlenmektedir.

Uygulama: Sürecin son adımıdır. Modelin gerçek hayatta uygulanması ve sonuçlarının karar destek süreçlerine entegre edilmesi aşamasıdır. Bu adımda sonuçların görselleştirilmesi ve raporlanması yapılmaktadır. Gerekirse yapılan uygulama karar destek sistemlerine entegre edilmektedir.

2.4. Veri Madenciliğinde Kullanılan Yöntemler

Veri madenciliğinde birden fazla yöntem ve algoritma bulunmaktadır.kullanılan yöntemler genellikle üç ana başlık altında incelenmektedir. Bunlar sınıflandırma, birliktelik analizi ve kümelemedir. Bu üç ana başlık için aşağıda bilgiler verilecektir.

2.4.1. Sınıflandırma

Sınıflandırma, önceden tanımlanmış sınıf etiketlerine sahip verilerden öğrenerek yeni verilerin hangi sınıfa ait olduğunu tahmin etmeye yarayan denetimli bir öğrenme yaklaşımıdır. Bu yaklaşımda veri setinin bir kısmı eğitim bir kismı test amacıyla ayrılmaktadır. Veri seti eğitildikten sonra sınıflandırma kuralları oluşturulmaktadır. Kurallar ortaya çıktıktan sonra yeni bir durum oluştuğunda nasıl davranılacağına oluşan bu kurallar sayesinde karar verilmektedir (Özkan, 2016). Örneğin, bir belediyeye gelen başvuruların “onaylandı”, “beklemede” veya “reddedildi” şeklinde etiketlenmesi için sınıflandırma teknikleri kullanılabilir. Sık kullanılan sınıflandırma algoritmaları arasında Karar Ağaçları (Decision Trees), Destek Vektör Makineleri (SVM), Yapay Sinir Ağları (ANN) ve Naive Bayes yer almaktadır.

2.4.2. Birliktelik analizi

Veritabanlarında yer alan kayıtlar arasındaki ilişkilerin analiz edilmesi ve olayların birlikte ya da eş zamanlı gerçekleşme olasılıklarının belirlenmesi amacıyla kullanılan yöntem birliktelik analizi olarak adlandırılmaktadır. Bu yöntem, geçmişte özellikle süpermarket veri setlerinin incelenmesi kapsamında yoğun olarak kullanıldığından, literatürde sıklıkla market sepet analizi veya pazar sepeti analizi olarak da

anılmaktadır. Birliktelik analizi, müşterilerin bir ürünü satın aldıklarında hangi diğer ürünleri satın alma eğiliminde olduklarını ortaya koyar. Bu ilişkiler, mağaza raflarının müşteri alışkanlıklarına göre yeniden düzenlenmesi ve alışveriş deneyiminin kolaylaştırılması amacıyla kullanılmaktadır.

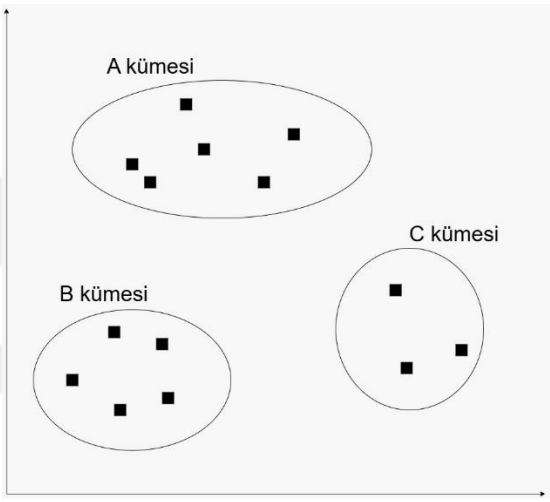
Birliktelik kuralları genellikle "X ürününü alan müşteriler, aynı zamanda Y ürününü de alma eğilimindedir" biçiminde ifade edilir. Benzer bir yaklaşım, kamu yönetimi bağlamında da kullanılabilir. Örneğin, belirli türdeki evrakların çoğunlukla hangi ek belgeler veya ilgi yazılarıyla birlikte gönderildiğinin tespit edilmesi mümkündür. Bu tür analizlerde sıklıkla Apriori ve FP-Growth algoritmaları tercih edilmektedir.

2.4.3. Kümeleme

Kümeleme, veri setindeki benzer özelliklere sahip gözlemleri bir araya getirerek etiketsiz verilerden anlamlı gruplar oluşturmayı hedefleyen denetimsiz bir öğrenme tekniğidir. Kümeleme algoritmaları, gözlemler arasındaki mesafeleri veya benzerlik ölçülerini kullanarak kümeler ortaya çıkarmaktadır. Örneğin, bir belediyede yıllık evrak işleme sürelerinin benzer özelliklere göre gruplandırılması veya vatandaş şikâyetlerinin türlerine göre kümelenmesi mümkündür. Yaygın kullanılan kümeleme yöntemleri arasında hiyeraşik kümeleme yöntemlari arasında yer alan en yakın komşu algoritması ve en uzak komşu algoritması; hiyerarşik olmayan kümeleme yöntemine ise k-means algoritması örnek olarak verilebilmektedir (Özkan, 2016). Kümeleme yöntemleri, uzaklık hesaplamaları ve k-means algoritması detaylı bir şekilde bölüm 3' te incelenecektir.

3. KÜMELEME

Kümeleme, veri madenciliğinde kullanılan ve denetimsiz öğrenme kategorisine giren bir yöntemdir. Bu yöntem, benzer özellikler taşıyan verileri kendi içinde homojen, birbirleri arasında ise heterojen olacak şekilde gruplandırmayı amaçlamaktadır. Kümeleme yötemlerinin çoğu veriler arasındaki uzaklıkları hesaplayarak veriyi benzer alt gruplara ayırma işlemini gerçekleştirmektedir (Özkan, 2016).



Şekil 3.1. Örnek kümeleme görüntüsü.

Kümeleme, etiketsiz verilerde gizli örüntülerin keşfedilmesi amacıyla birçok sektörde yaygın biçimde kullanılan bir veri madenciliği tekniğidir. Pazarlama alanında, müşteri davranışları, alışveriş alışkanlıkları ve demografik veriler analiz edilerek benzer özelliklere sahip müşteri segmentleri belirlenmekte ve buna uygun kişiselleştirilmiş kampanyalar geliştirilebilmektedir. Sağlık sektöründe, hasta verileri ve test sonuçları kümelenerek benzer hastalık tiplerinin tespit edilmesi ve tedavi planlarının bireyselleştirilmesi sağlanmaktadır. Finans alanında, yatırımcıların risk profilleri veya dolandırıcılık faaliyetlerinin tespiti amacıyla kümeleme tekniklerinden yararlanılmaktadır. Eğitim ve e-öğrenme sistemlerinde, öğrenci performansları ve öğrenme eğilimleri gruplandırılarak akademik başarıya yönelik öngörüler elde edilmektedir. Kamu yönetiminde ve belediyelerde ise vatandaş talepleri, şikâyet kayıtları veya evrak işlem süreleri kümelenerek hizmet yoğunluğu analiz edilmekte ve personel planlaması yapılabilmektedir. Ayrıca, biyoinformatik ve genetik

çalışmalarında gen ifadeleri ve DNA dizileri kümelenerek benzer genetik yapıların sınıflandırılması sağlanmaktadır. Bu çeşitlilik, kümelemenin hem akademik hem de endüstriyel uygulamalarda stratejik bir araç olduğunu göstermektedir.

3.1. Uzaklık Ölçüleri

Veri madenciliği ve makine öğrenmesi uygulamalarında, özellikle kümeleme ve sınıflandırma gibi algoritmalarda veri noktaları arasındaki benzerlik veya farklılıkların hesaplanması kritik öneme sahiptir. Bu amaçla kullanılan yöntemler uzaklık ölçüleri olarak adlandırılmaktadır. Uzaklık ölçüleri, veri uzayındaki iki nokta arasındaki mesafeyi sayısal olarak ifade etmektedir.

En yaygın kullanılan ölçütlerden biri öklid uzaklığı olup, iki nokta arasındaki en kısa mesafeyi hesaplamaktadır ve sürekli değişkenler için uygundur. Manhattan uzaklığı, noktalar arasındaki mesafeyi yalnızca dikey ve yatay eksenler boyunca hesaplar ve özellikle şehir blokları benzetmesiyle ifade edilir. Minkowski uzaklığı, öklid ve manhattan uzaklıklarının genel bir formu olup, farklı parametre değerleriyle her iki ölçütü kapsayabilektedir. Cosine benzerliği ise özellikle metin madenciliği veya vektör tabanlı veri setlerinde iki vektör arasındaki açıya dayalı bir benzerlik ölçüsüdür. Bunun yanında, Jaccard Benzerliği kümeler arasındaki ortak öğelerin oranını baz alır ve genellikle kategorik veya ikili (binary) verilerde kullanılmaktadır.

Uzaklık ölçüsünün seçimi, veri türüne (sayısal, kategorik, metin) ve analiz amacına bağlı olarak değişmektedir. Yanlış seçilen bir uzaklık metriği, algoritmanın yanlış kümeler oluşturmasına veya hatalı sınıflandırma sonuçları üretmesine yol açabilmektedir. Bu nedenle, veri ön işleme sürecinde uygun uzaklık ölçüsünün belirlenmesi büyük önem taşımaktadır.

Öklid uzaklığı: Uygulamalarda en sık kullanılan uzaklık bağıntısıdır. Öklid uzaklığının genel denklemi aşağıdaki gibidir.

$$d(i,j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \tag{3.1}$$

$A(x_1, y_1)$ ve $B(x_2, y_2)$ noktaları arasında uzaklık öklid uzaklığı kullanılarak aşağıdaki verilen denklemdeki gibi hesaplanmaktadır.

$$d(A,B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{3.2}$$

Manhattan uzaklığı: Bu uzaklık ölçüsü iki nokta arasındaki mutlak uzaklıkların toplamı alınarak hesaplanmaktadır (Özkan, 2016). Manhattan uzaklığının genel denklemi aşağıdaki gibidir.

$$d(i,j) = \sum_{k=1}^{p} (|x_{ik} - x_{jk}|) \tag{3.3}$$

$A(x_1, y_1)$ ve $B(x_2, y_2)$ noktaları arasında uzaklık manhattan uzaklığı kullanılarak aşağıdaki verilen denklemdeki gibi hesaplanmaktadır.

$$d(A,B) = |x_1 - x_2| + |y_1 - y_2| \tag{3.4}$$

Minkowski uzaklığı: Bu uzaklık ölçüsü ölkid ve manhattan uzaklık ölçülerinin genel bir formudur.

$$d(i,j) = \left[\sum_{k=1}^{p} (|x_{ik} - x_{jk}|^m)\right]^{\frac{1}{m}} \tag{3.5}$$

Örnek: A(2,3), B(4,1) şeklinde iki nokta arasındaki uzaklıkların öklid, manhattan, minkowski uzaklık ölçüleri hesaplanarak bulunması.

Öklid ile hesaplanması;

$$d(A,B) = \sqrt{(2-4)^2 + (3-1)^2} = 2{,}83$$

Manhattan ile hesaplanması;

$$d(A,B) = |2 - 4| + |3 - 1| = 4$$

Minkowski ile hesaplanması;

3.5'te verilen bağıntıya göre m =2 olarak alınacaktır.

$$d(A,B) = [|2 - 4|^2 + |3 - 1|^2]^{\frac{1}{2}} = 2{,}83$$

3.2. Kümeleme Yöntemleri

Kümeleme, veri madenciliğinde en yaygın kullanılan veri analiz yöntemlerinden biridir ve gözlemler arasındaki benzerliklere dayalı olarak verileri anlamlı gruplara ayırmayı amaçlamaktadır. Gözetimsiz öğrenme yöntemleri kapsamında değerlendirilen bu teknik, verilerdeki yapıları keşfetmek ve örüntüler oluşturmak için kullanılmaktadır. Kümeleme yöntemleri genel olarak hiyerarşik kümeleme yöntemleri ve hiyerarşik olmayan kümeleme yöntemleri olmak üzere iki ana gruba ayrılmaktadır.

Hiyerarşik kümeleme yöntemleri, gözlemler arasında hiyerarşik (ağaç benzeri) bir yapı oluşturarak kümeleme işlemini gerçekleştirmektedir.

Hiyerarşik kümeleme yöntemleri kendi içinde çeşitli bağlantı stratejilerine ayrılmaktadır. Bunlar arasında tekli bağlantı, tam bağlantı ve ortalama bağlantı yöntemleri öne çıkmaktadır.

Tekli bağlantı yöntemi, literatürde zaman zaman yakın komşu yöntemi olarak da adlandırılmaktadır. Bu yaklaşımda iki küme arasındaki uzaklık, her iki kümedeki herhangi birer gözlem çifti arasındaki en kısa mesafe üzerinden hesaplanmaktadır. Yani A ve B kümeleri arasındaki mesafe, A kümesindeki bir eleman ile B kümesindeki bir eleman arasındaki en küçük mesafedir. Bu yöntem, kümeleri birbirine zincirleme bağlayabilmekte ve uzun, ince kümeler oluşmasına neden olabilmektedir.

Tam bağlantı yöntemi, uzak komşu yaklaşımı olarak da bilinmektedir. Bu yöntemde iki küme arasındaki mesafe, aralarındaki tüm gözlem çiftleri içerisinden en büyük mesafe dikkate alınarak hesaplanmaktadır. İki küme arasındaki mesafe, kümeler içindeki tüm gözlemler arasından en uzun mesafe olarak tanımlanmaktadır. Bu yöntem, daha kompakt ve dengeli kümeler oluşturma eğilimindedir.

Ortalama bağlantı yöntemi ise iki küme arasındaki uzaklığı, aralarındaki tüm gözlem çiftlerinin ortalama mesafesi üzerinden tanımlamaktadır.

Hiyerarşik olmayan kümeleme yöntemlerine ise en bilinen örnek olarak k-ortalamalar (k-means) algoritması verilebilmektedir. K-means algoritması detaylı şekilde bir alt bölümde incelenecektir.

3.3. K-Means Algoritması

Hiyerarşik olmayan kümeleme yöntemlerinin en bilinen ve en sık kullanılan örneği K-Means algoritmasıdır. K-Means (K-Ortalamalar) algoritması, verileri k sayıda küme

olacak şekilde gruplandıran ve her bir kümenin merkezini (centroid) belirleyerek iteratif bir süreçle kümeleri güncelleyen bir algoritmadır (MacQueen, 1967). Algoritmanın temel amacı, her gözlemi en yakın merkeze atayarak kümeler içindeki varyansı (intra-cluster variance) minimize etmektir. K-Means algoritmasının temel adımları şu şekildedir:

a) Başlangıç: Küme sayısı (k) belirlenir ve rastgele k adet merkez seçilir.
b) Atama Adımı: Her veri noktası, kendisine en yakın merkeze atanır.
c) Merkez Güncelleme: Her küme için, ona ait verilerin ortalaması alınarak yeni merkez hesaplanır.
d) Yineleme: Atama ve merkez güncelleme adımları, kümeler değişmeyene kadar tekrarlanır.

Bu işlem sonucunda her bir veri noktası en yakın olduğu merkeze atanarak gruplandırılmış olmaktadır.

Örnek: K-means algoritması için aşağıda ufak bir örnek verilecektir.

Tablo 3.1. Örnek gözlem noktası değerleri.

Nokta	X	Y
A	1	2
B	1	4
C	1	0
D	10	2
E	10	4
F	10	0

1. Küme sayısı rastgele k=2 olarak belirlenmiştir.
2. Veri setinden rastgele iki küme merkezi seçilmiştir.

   1 nolu kümenin merkezi A (1,2) noktası, 2 nolu kümenin merkezi ise D (10,2) noktası olarak seçilmiştir.
3. Her nokta en yakın merkeze atanır.

Tablo 3.2. Gözlem noktalarının küme merkezine olan uzaklık değerleri.

Nokta	(1,2)	(10,2)	Küme
A(1,2)	0	9	1
B(1,4)	2	9.2	1
C(1,0)	2	9.2	1
D(10,2)	9	0	2
E(10,4)	9.2	2	2
F(10,0)	9.2	2	2

Tablo 3.2'de noktalar arasındaki uzaklıklar öklid uzaklık ölçüsüne göre hesaplanmıştır. Noktalar küme merkezine olan uzaklık hangisinde daha küçük ise o kümeye atanmıştır. Yani küme merkezine en yakın olana atanmıştır.

4. Yeni küme merkezlerini hesaplama;

Birinci küme için küme merkezi: A(1,2), B(1,4), C(1,0) noktalarının bulunduğu kümenin merkezi x ve y değerlerinin ayrı ayrı toplanıp aritmetik ortalamasının hesaplanması ile bulunmaktadır. Hesap yapıldıktan sonra küme merkezi değişmeyip aynı kalmıştır.

İkinci küme için küme merkezi: D(10,2), E(10,4), F(10,0) noktalarının bulunduğu kümenin merkezi x ve y değerlerinin ayrı toplanıp aritmetik ortalamasının hesaplanması ile bulunmaktadır. Hesap yapıldıktan sonar küme merkezi değişmeyip aynı kalmıştır.

5. Oluşan kümeler:

1 nolu küme A, B, C noktalarından, 2 nolu küme D, E ve F noktalarından oluşmaktadır.

4. UYGULAMA

4.1. Veri Seti ve Özellikleri

Bu çalışmada kullanılan veri seti, Sakarya Büyükşehir Belediyesi'ne ait 2016 yılı kurum giden evraklarına ilişkin kayıtları içermektedir. Veri seti, toplamda 9 değişkenden oluşmakta olup, belge yönetim süreçlerinin analizine olanak tanıyacak nitelikte bilgiler içermektedir. Bu değişkenlerin bir kısmı gösterge panelleri (dashboard) için, bir kısmı da kümeleme analizi için kullanılacaktır.

Veri setinde yer alan değişkenleri şu şekilde açıklayabiliriz:

- kayit_no: Her bir evrağa sistem tarafından verilen benzersiz kayıt numarasını ifade eder. Belediyede yer alan birimlerden kurum dışına gönderilen tüm evraklara, Yazı İşleri ve Arşiv Şube Müdürlüğü tarafından verilen numaradır.
- mdr_id: Evrağı hazırlayan şube müdürlüğüne karşılık gelen sayısal kimlik bilgisidir.
- db_id: Evrağı hazırlayan şube müdürlüğünün bağlı bulunduğu daire başkanlığına karşılık gelen sayısal kimlik bilgisidir.
- sure: Evrağın ilgili birim tarafından oluşturulma tarihi ile Yazı İşleri ve Arşiv Şube Müdürlüğü'nden çıkış tarihi arasında geçen süreyi gün cinsinden ifade eder.
- sure_aralik: sure değişkenine göre oluşturulmuş kategorik bir değişkendir. Süreler beş aralıkta kategorize edilmiştir. 0-1 gün için 1, 2-3 gün için 2, 4-7 gün için 3, 8-14 gün için 4, 15 gün ve üzeri için 5 olarak kabul edilmiştir.
- ay_id: Evrağın Yazı İşleri ve Arşiv Şube Müdürlüğünden kayıt numarası alarak çıktığı ayı temsil eden sayısal değerdir. Tüm aylara 1'den 12'ye kadar ay_id'si atanmıştır.
- dagitim: Evrağın birden fazla kişi yada kurumla paylaşılıp paylaşılmadığını belirten ikili değişkendir. Dağıtımsız evraklar için 0, dağıtımlı evraklar için 1 değeri kullanılmıştır.

- ilgi: Evrağın başka bir belgeye atıf (ilgi) içerip içermediğini gösteren ikili değişkendir. İlgi olmayan evraklar için 0, ilgi olan evraklar için 1 değeri kullanılmıştır.
- ek: Evrakla birlikte gönderilen ek belge olup olmadığını belirten ikili değişkendir. Ek olmayan evraklar için 0, ek olan evraklar için 1 değeri kullanılmıştır.

Veri setindeki tüm değişkenler sayısal biçimde düzenlenmiş olup, analiz sürecinde doğrudan kullanılabilir durumdadır. Bu yapı, gösterge panellerinin oluşturulması ve kümeleme analizinin uygulanması süreçlerini kolaylaştırmaktadır.

4.2. Veri Ön İşleme ve Hazırlık Süreci

Veri madenciliği süreçlerinde güvenilir ve anlamlı sonuçlara ulaşabilmek için, ham verilerin analiz öncesinde uygun biçimde işlenmesi gerekmektedir. Bu çalışmada, Sakarya Büyükşehir Belediyesi 2016 yılına ait kurum giden evrak verileri üzerinde çeşitli ön işleme adımları gerçekleştirilmiştir. Söz konusu veri seti toplam 16.535 kayıt ve 9 değişkenden oluşmaktadır. Veriler sayısal biçimde düzenlenmiştir.

Veri ön işleme ve hazırlık sürecinde, Sakarya Büyükşehir Belediyesi 2016 yılına ait kurum giden evrak verileri Bilgi İşlem Daire Başkanlığı'ndan alınmış ve veri setinin 16.545 kayıttan oluştuğu görülmüştür. Tüm evrak verileri detaylı olarak incelendiğinde 10 evrak verisinin boş olduğu tespit edilmiştir. Veri temizleme adımında bu 10 kayıt, veri setinden silinmiştir. Böylece her bir değişkenin tüm kayıtlar için dolu olması, doğrudan analiz yapılabilmesine olanak sağlamıştır.

2016 yılında Sakarya Büyükşehir Belediyesi'nde 15 daire başkanlığının mevcuttur. Herhangi bir daire başkanlığına bağlı olmayan ve bu dönemde kurum dışına evrak gönderen Özel Kalem, Hukuk Müşavirliği ve Halkla İlişkiler Şube Müdürlüğü'ne ait evrak sayılarının tespit edilebilmesi için hem db_id hem de mdr_id atamaları yapılmıştır. Tablo 4.1.'de daire başkanlıklarına ait db_id bilgileri verilmiştir.

Tablo 4.1. Daire başkanlıkları.

Daire Başkanlığı	db_id
Bilgi İşlem	1

Tablo 4.1. (Devamı) Daire başkanlıkları.

Daire Başkanlığı	db_id
Destek Hizmetleri	2
Fen İşleri	3
Halkla İlişkiler	4
Hukuk Müşavirliği	5
Kültür ve Sosyal İşler	6
Mali Hizmetler	7
Sağlık İşleri	8
Sosyal Hizmetler	9
Tarımsal Hizmetler ve Muhtarlık İşleri	10
Ulaşım	11
Yol Bakım ve Altyapı Koordinasyon	12
Zabıta	13
Çevre Koruma ve Kontrol	14
Özel Kalem	15
İmar ve Şehircilik	16
İnsan Kaynakları ve Eğitim	17
İtfaiye	18

Sakarya Büyükşehir Belediyesi'nin organizasyon şeması incelendiğinde şube müdürlüklerinin genellikle bir daire başkanlığına bağlı olduğu görülmektedir. Tablo

4.2.'de bu dönemde kurum dışına evrak gönderen şube müdürlüklerine ait mdr_id bilgileri verilmiştir.

Tablo 4.2. Şube müdürlükleri.

Şube Müdürlüğü	mdr_id
Araştırma ve Kaynak Geliştirme	1
Atık Yönetimi	2
Aykome	3
Ail eve Çocuk Hizmetleri	4
Bütçe ve Raporlama	5
Bilişim Sistemleri ve Donanım	6
Bitkisel Üretim Destekleme	7
Coğrafi Bilgi Sistemleri ve Yazılım	8
Destek Hizmetleri	9
Engelli Hizmetleri	10
Gelirler	11
Gençlik ve Spor	12
GSM Ruhsat Denetim	13
Hal	14
Halkla İlişkiler	15
Harita, Emlak ve İstimlak	16
Hukuk Müşavirliği	17

Tablo 4.2. (Devamı) Şube müdürlükleri.

Şube Müdürlüğü	mdr_id
Kararlar	18
Kentsel Dönüşüm	19
Keşif ve Metraj	20
Kültür ve Sanat	21
Maaş Tahakkuk	22
Makina İkmal	23
Mezarlıklar	24
Muhasebe ve Finansman	25
Muhtarlık İşleri	26
Müdahale	27
Park ve Bahçeler	28
Satın Alma	29
Sağlık Hizmetleri	30
Sosyal Hizmetler	31
Stratejik Yönetim	32
Tarımsal Hizmetler	33
Taşınır Konsolide	34
Terminal	35
Toplu Taşıma	36

Tablo 4.2. (Devamı) Şube müdürlükleri.

Şube Müdürlüğü	mdr_id
Trafik	37
UKOME	38
Veteriner Hizmetleri	39
Yapı İşleri	40
Yaygın Eğitim	41
Yazı İşleri ve Arşiv	42
Yol Bakım ve Asfalt	43
Zabıta	44
Çevre Koruma ve Geliştirme	45
Ön Mali Kontrol	46
Önleme ve Eğitim	47
Özel Kalem	48
İdari İşler	49
İhale İşleri	50
İmar ve Şehircilik	51
İnsan Kaynakları ve Eğitim	52
İş Sağlığı ve Güvenliği	53
İşletme ve İştirakler	54
Şehir Planlama	55
Şehir Tiyatro	56

Kümeleme analizinin yapılabilmesi için evrağın kurumdan çıkış tarihine ait ay bilgisi sayısallaştırılarak Tablo 4.3'te verilmiştir.

Tablo 4.3. Ay dağılımları.

Ay adı	ay_id
Ocak	1
Şubat	2
Mart	3
Nisan	4
Mayıs	5
Haziran	6
Temmuz	7
Ağustos	8
Eylül	9
Ekim	10
Kasım	11
Aralık	12

Süre değişkeni, evrağın oluşturulma tarihi ile kurumdan çıkış tarihi arasındaki farkı gün cinsinden temsil etmektedir. Bu sürekli değişken, analiz kolaylığı ve kümeler arası karşılaştırma yapılabilmesi adına sure_aralik adlı kategorik bir değişkene dönüştürülmüştür. Tablo 4.4.'te evrak sürelerine göre belirlenen sure_aralik bilgileri verilmiştir.

Tablo 4.4. Süre aralıkları.

sure_aralik	Açıklama
1	0-1 gün (çok kısa)
2	2-3 gün (kısa)
3	4-7 gün (orta)
4	8-14 gün (uzun)
5	15+ gün (çok uzun)

Evraklarda ek, ilgi ve dağıtım bilgilerine ilişkin alanlar, bu özelliklerin varlığı (1) veya yokluğu (0) esasına göre ikili olarak kodlanmıştır. Bu sayede söz konusu nitelikler, sayısal yöntemlere uygun hale getirilmiştir. Evraklarda bulunan ek, ilgi ve dağıtım bilgileri evrak karmaşıklığı hakkında bilgi vermektedir.

Tablo 4.5.'te evrakların dağıtım bilgileri verilmiştir.

Tablo 4.5. Dağıtım durumu.

Dağıtım	Açıklama
0	Dağıtımsız
1	Dağıtımlı

Tablo 4.6.'da evrakların ilgi bilgileri verilmiştir.

Tablo 4.6. İlgi durumu.

İlgi	Açıklama
0	Evrakta ilgi yok
1	Evrakta ilgi var

Tablo 4.7.'de evrakların ek bilgileri verilmiştir.

Tablo 4.7. Ek durumu.

Ek	Açıklama
0	Evrakta ek yok
1	Evrakta ek var

Bu ön işleme adımları sayesinde, veri seti analiz ve modelleme aşamaları için hazır hale getirilmiş; özellikle k-means kümeleme algoritması gibi sayısal girişlere duyarlı tekniklerin sağlıklı çalışması garanti altına alınmıştır.

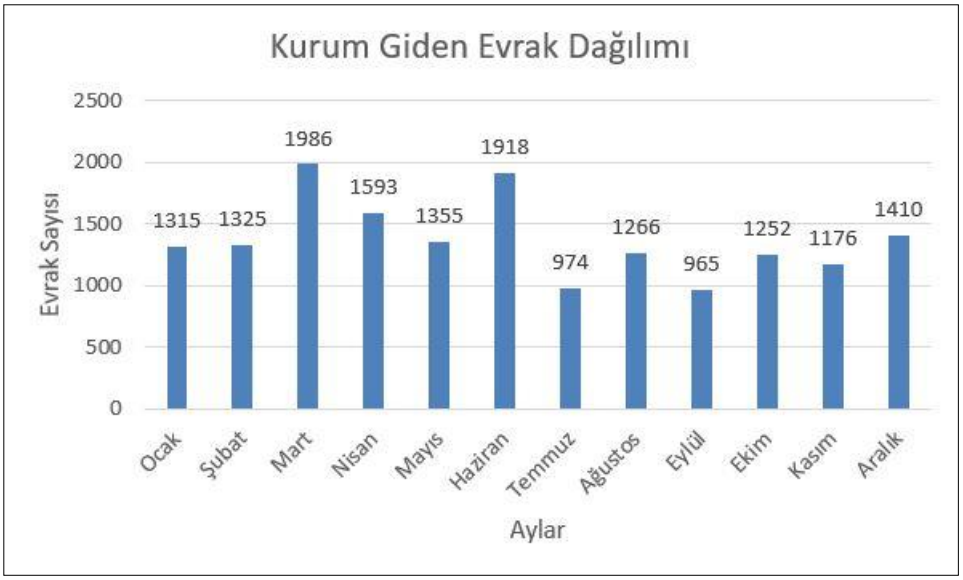
4.3. Analiz Sonuçları

Veriler düzenlendikten sonra veri seti üzerinde bazı analizler yapılmıştır. Bu başlık altında yapılan analiz sonuçları gösterge panelleri (dashboard) kullanılarak yorumlanacaktır.

Veri madenciliğinde gösterge panelleri (dashboard), büyük ve karmaşık veri setlerinden elde edilen bilgilerin görsel ve anlaşılır bir şekilde sunulmasını sağlayan araçlardır. Dashboard'lar; grafikler, tablolar, göstergeler ve diğer görselleştirme bileşenleri aracılığıyla karar vericilerin, veriye dayalı bir görüş elde etmesine olanak tanır.

Yapılan analizler sonucunda aşağıda Sakarya Büyükşehir Belediyesi'nin 2016 yılında kurum dışına gönderdiği evrak verilerine ait grafikler sunulmuştur. Toplam 16.535 evrak verisi bulunmaktadır.

Kurum dışına gönderilen evrakların tümü Yazı İşleri ve Arşiv Şube Müdürlüğü'nden bir kayıt numarası almaktadır. Aylara göre evrak dağılım grafiği Şekil 4.1'de verilmiştir. En çok evrak çıkışı Mart ayında (1986 evrak), en az ise Eylül ayında (965 evrak) gerçekleşmiştir. Aylık bazlı evrak yoğunluğuna bakarak, evrak çıkışının az olduğu Temmuz ve Eylül aylarında Yazı İşleri ve Arşiv Şube Müdürlüğü personeline yıllık izin verilmesinin daha uygun olacağı öngörülmektedir. Evrak çıkışının çok olduğu Mart ve Haziran aylarında personele izin verilmesi ise iş akışını olumsuz etkileyebilir.



Şekil 4.1. Kurum giden evrak dağılımı.

Daire başkanlığı bazında (db_id) ilk 6 aya ait evrak dağılımı Tablo 4.8.'de verilmiştir.

Tablo 4.8. db_id evrak dağılımı (ilk 6 ay)

db_id	Ocak	Şubat	Mart	Nisan	Mayıs	Haziran
1	5	3	4	11	6	5
2	28	19	37	23	28	23
3	83	82	105	155	158	118
4	3	5	15	4	5	3
5	11	12	10	7	8	6
6	9	10	50	16	19	14
7	321	282	300	249	224	305
8	36	62	21	40	48	12
9	12	16	16	9	18	11
10	10	19	30	15	11	9

Tablo 4.8. (Devamı) db_id evrak dağılımı (ilk 6 ay)

db_id	Ocak	Şubat	Mart	Nisan	Mayıs	Haziran
11	157	168	245	139	191	192
12	0	0	0	0	0	5
13	46	63	77	84	60	82
14	57	43	63	65	62	70
15	2	2	3	4	3	2
16	285	281	728	534	267	823
17	170	155	152	164	155	124
18	80	103	130	74	92	114

Daire başkanlığı bazında (db_id) son 6 aya ait evrak dağılımı ise Tablo 4.9.'da verilmiştir.

Tablo 4.9. db_id evrak dağılımı (son 6 ay)

db_id	Temmuz	Ağustos	Eylül	Ekim	Kasım	Aralık
1	4	5	3	11	1	6
2	19	36	32	18	32	57
3	48	80	68	85	51	74
4	2	4	0	3	1	3
5	10	11	23	13	18	36
6	3	2	14	11	14	6
7	167	179	258	247	266	186
8	10	61	52	71	57	37

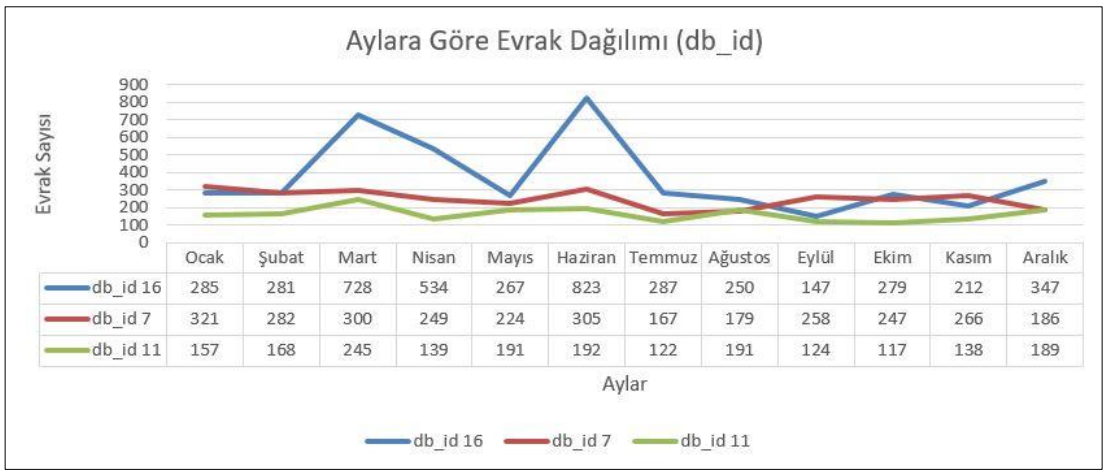
Tablo 4.9. (Devamı) db_id evrak dağılımı (son 6 ay)

db_id	Temmuz	Ağustos	Eylül	Ekim	Kasım	Aralık
9	7	17	11	16	12	20
10	15	9	9	12	11	9
11	122	191	124	117	138	189
12	23	24	24	36	44	27
13	56	53	26	60	52	51
14	49	78	32	77	81	66
15	2	2	2	3	2	4
16	287	250	147	279	212	347
17	99	182	88	113	98	167
18	51	82	52	80	86	125

Daire başkanlıklarına ait evrak dağılım grafiği Şekil 4.2.'de verilmiştir. En çok evrak çıkışı yapan üç daire başkanlığı sırasıyla; İmar ve Şehircilik Dairesi Başkanlığı (4440 evrak), Mali Hizmetler Dairesi Başkanlığı (2984 evrak) ve Ulaşım Dairesi Başkanlığı'dır (1973 evrak).

Şekil 4.2. Daire başkanlığı evrak dağılımı.

En çok evrak çıkışı yapan üç daire başkanlığına ait aylara göre evrak dağılımı grafiği Şekil 4.3'te verilmiştir. İmar ve Şehircilik Dairesi Başkanlığı'nın (db_id 16) Mart ve Haziran aylarında evrak çıkışının çok olduğu görülmektedir. Mali Hizmetler Dairesi Başkanlığı (db_id 7) ve Ulaşım Dairesi Başkanlığın'da (db_id 11) evrak sayılarının aylara göre homojen olarak dağıldığı görülmektedir. Aylık bazlı evrak yoğunluğuna bakarak, bu birimlerde evrak işleriyle ilgilenen personellerin yıllık izin planlaması yapılabilir.

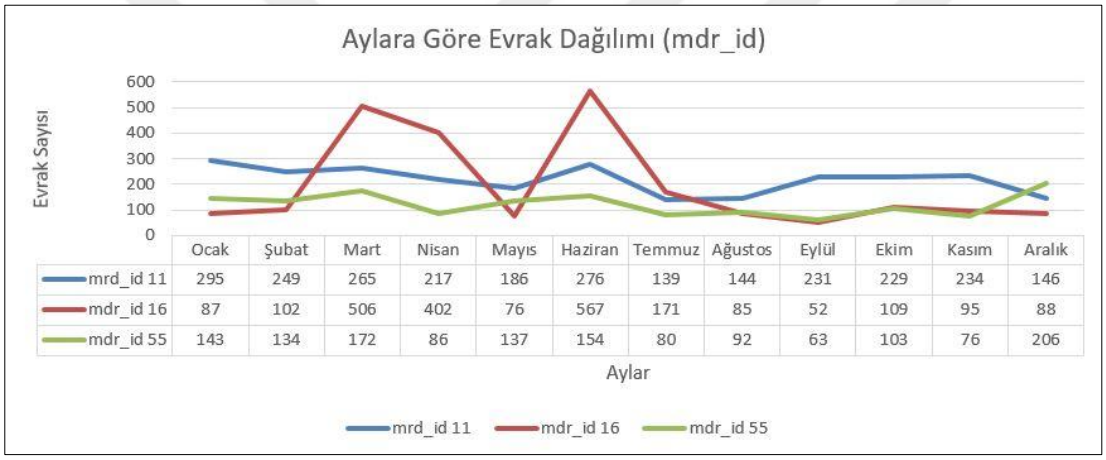


Şekil 4.3. Aylara göre evrak dağılımı (db_id).

Sakarya Büyükşehir Belediyesi'nin organizasyon şeması incelendiğinde, şube müdürlüklerinin genellikle bir daire başkanlığına bağlı olduğu Ek A'da görülmektedir. En çok evrak çıkışı yapan üç şube müdürlüğü sırasıyla; Gelirler Şube Müdürlüğü

(2611 evrak), Harita, Emlak ve İstimlak Şube Müdürlüğü (2340 evrak) ve Şehir Planlama Şube Müdürlüğü'dür (1446 evrak). Gelirler Şube Müdürlüğü, Mali Hizmetler Dairesi Başkanlığı'na, Şehir Planlama Şube Müdürlüğü ile Harita, Emlak ve İstimlak Şube Müdürlüğü ise İmar ve Şehircilik Dairesi Başkanlığı'na bağlıdır.

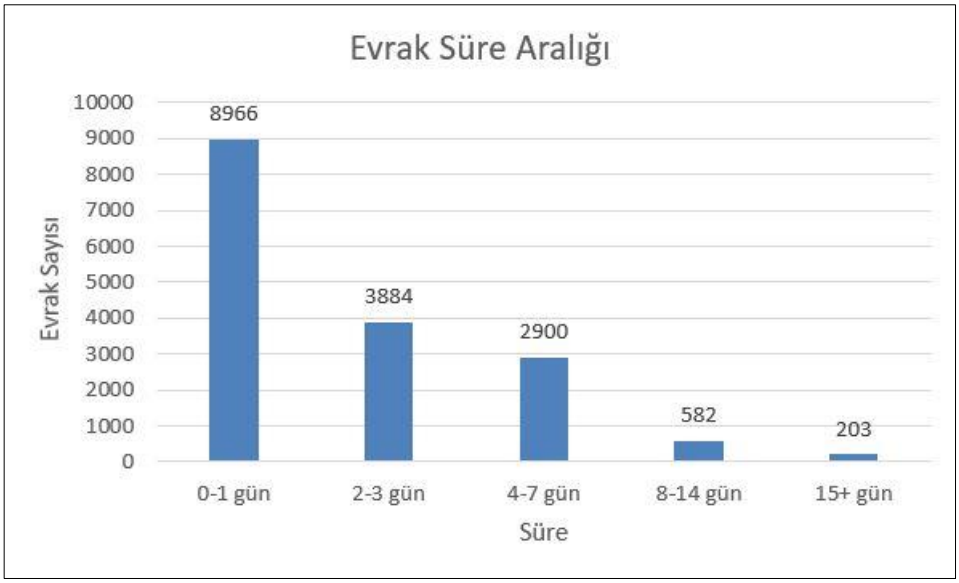
En çok evrak çıkışı yapan üç şube müdürlüğüne ait aylara göre evrak dağılımı grafiği Şekil 4.4'te verilmiştir. Harita, Emlak ve İstimlak Şube Müdürlüğü'nün (mdr_id 16) Mart, Nisan ve Haziran aylarında evrak çıkışının çok olduğu görülmektedir. Gelirler Şube Müdürlüğü (mdr_id 11) ve Şehir Planlama Şube Müdürlüğü'nde (mdr_id 55) evrak sayılarının aylara göre homojen olarak dağıldığı görülmektedir. Aylık bazlı evrak yoğunluğuna bakarak, bu müdürlüklerde evrak işleriyle ilgilenen personellerin yıllık izin planlaması yapılabilir.



	Ocak	Şubat	Mart	Nisan	Mayıs	Haziran	Temmuz	Ağustos	Eylül	Ekim	Kasım	Aralık
mrd_id 11	295	249	265	217	186	276	139	144	231	229	234	146
mdr_id 16	87	102	506	402	76	567	171	85	52	109	95	88
mdr_id 55	143	134	172	86	137	154	80	92	63	103	76	206

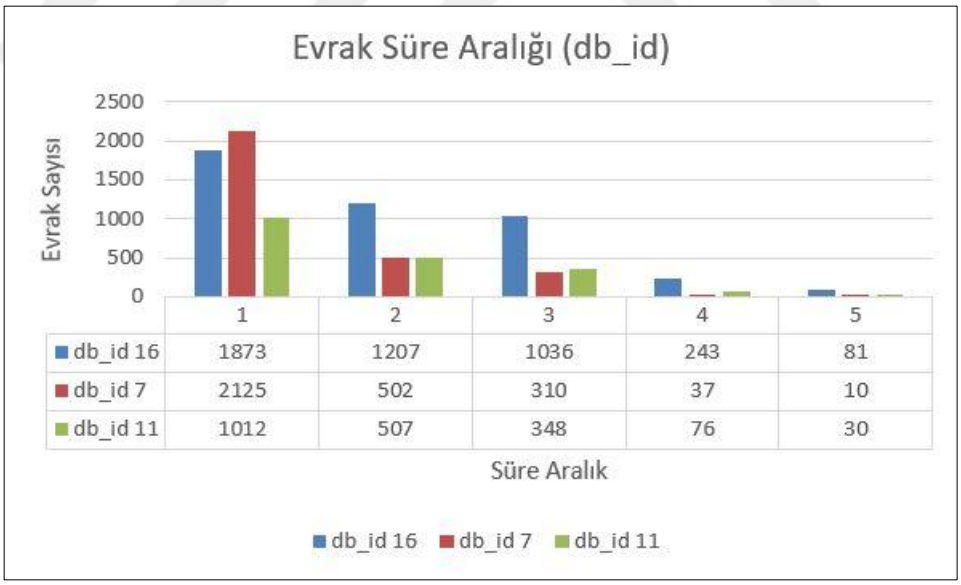
Şekil 4.4. Aylara göre evrak dağılımı (mdr_id).

Kurum dışına gönderilen bir evrağın yazılmaya başlandığı gün ile Yazı İşleri ve Arşiv Şube Müdürlüğü'nden çıkışının yapıldığı gün arasındaki fark, evrak süresini temsil etmektedir. Evrak süre aralığı grafiği Şekil 4.5'te verilmiştir. Buna göre evrak yazışmalarıyla ilgili süreçlerin %54.22'si 0-1 günde, %1.23'ü ise 15+ günde sonuçlanmaktadır. En uzun yazışma süreci 198 gün ile İmar ve Şehircilik Dairesi Başkanlığı'na ait bir evraktır. Yazışma süreci 100 günün üzerinde süren 7 evrak vardır. Bu evrakların 3'ü İmar ve Şehircilik Dairesi Başkanlığı'na aittir.



Şekil 4.5. Evrak süre aralığı.

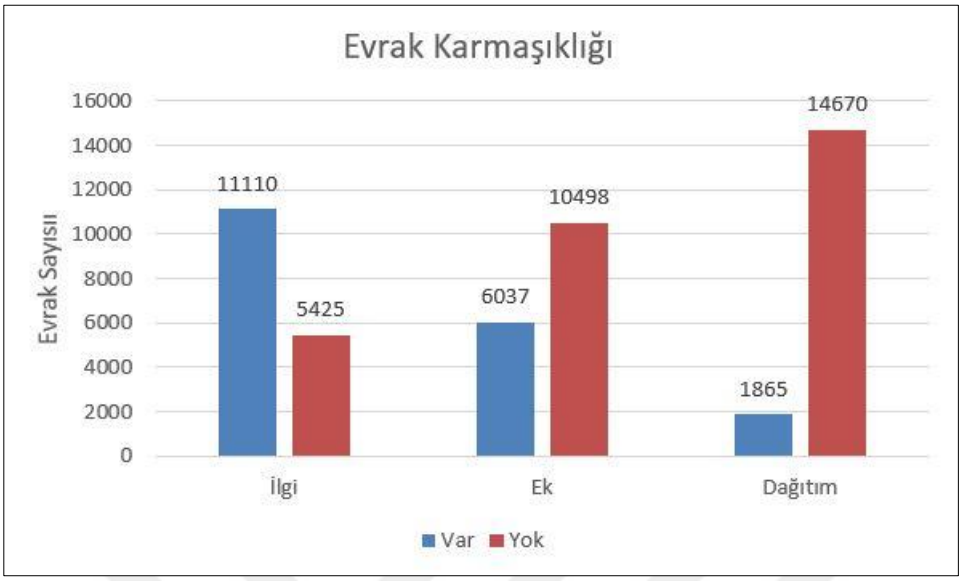
En çok evrak çıkışı yapan üç daire başkanlığına ait evrak süre aralığı grafiği Şekil 4.6.'da verilmiştir. İmar ve Şehircilik Dairesi Başkanlığı evrak süreçlerinin %92.7'sini, Mali Hizmetler Dairesi Başkanlığı %98.42'sini ve Ulaşım Dairesi Başkanlığı ise %94.63'ünü ilk 7 gün içersinde sonuçlandırmaktadır.



Şekil 4.6. Evrak süre aralığı (db_id).

Evraklarda yer alan ilgi, ek ve dağıtım bilgileri evrak karmaşıklığı hakkında bilgi vermektedir. Kurumdan çıkış yapan evraklara ait evrak karmaşıklığı grafiği Şekil 4.7.'de verilmiştir. Buna göre, evrakların %67.19'unda ilgi yazısının olduğu ve %36.51'inde ek dosyaların bulunduğu görülmüştür. Evrakların %88.72'sinin tek bir

kuruma yada kişiye gönderildiği (dağıtımsız), %11.28'inin ise birden fazla kurum veya kişiye gönderildiği (dağıtımlı) görülmüştür.

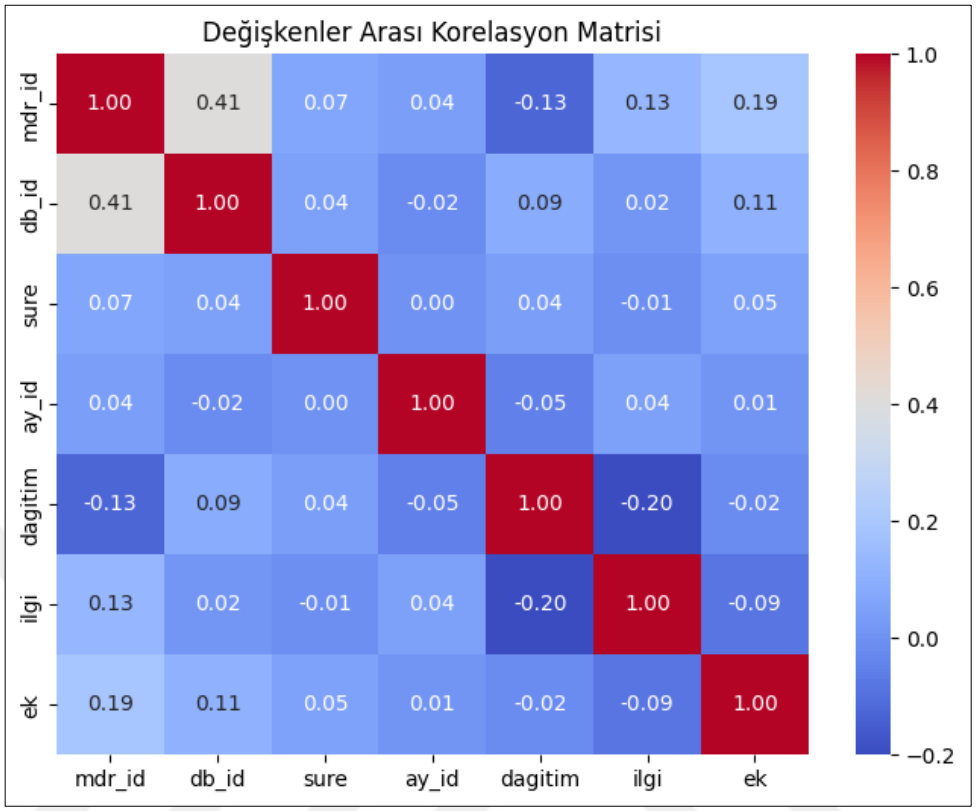


Şekil 4.7. Evrak karmaşıklığı**.**

Şekilde 4.8.'de veriye ait korelasyon matrisi verilmiştir. Sakarya Büyükşehir Belediyesi'ne ait kurum giden evrak verilerindeki değişkenler arasındaki doğrusal ilişkileri analiz etmek amacıyla oluşturulmuştur. Matris incelendiğinde, en yüksek pozitif korelasyonun mdr_id ile db_id arasında olduğu (%41) görülmektedir. Bu durum, aynı daire başkanlığına bağlı müdürlüklerin evrak düzenlemelerinde benzerlik gösterdiğine işaret etmektedir. Diğer değişkenler arasındaki korelasyonlar ise oldukça düşüktür ve genellikle ±0.20 sınırları içerisinde kalmaktadır; bu da değişkenler arasında güçlü doğrusal ilişkilerin olmadığını göstermektedir. Örneğin, süre değişkeni ile diğer tüm değişkenler arasında oldukça zayıf ilişkiler gözlemlenmiştir; bu da evrakların işlem süresinin, belgeye ait ek, ilgi ya da dağıtım gibi diğer unsurlardan bağımsız değiştiğini düşündürmektedir.

Negatif korelasyonlar açısından dagitim değişkeni ile ilgi (-0.20) ve ek (-0.12) değişkenleri arasında zayıf düzeyde negatif ilişkiler tespit edilmiştir. Bu durum, bir evrakın dağıtım bilgisinin olması durumunda genellikle ilgi ya da ek bilgilerle desteklenmediğini gösterebilir. Genel olarak korelasyon katsayılarının düşük çıkması, veri setindeki değişkenlerin birbirinden büyük ölçüde bağımsız olduğunu ortaya koymakta ve bu bağımsızlık, kümeleme gibi denetimsiz öğrenme algoritmalarının uygulanabilirliğini güçlendirmektedir. Bu bağlamda, korelasyon analizi sonuçları,

değişkenler arası çoklu doğrusal ilişki bulunmadığını doğrulamakta ve yapılacak veri madenciliği uygulamaları için uygun bir temel sağlamaktadır.



Şekil 4.8. Değişkenler arası korelasyon matrisi.

Bu çalışmada kullanılan Sakarya Büyükşehir Belediyesi'ne ait kurum giden evrak verileri, belirli dönemlerde evrak hareketliliği, birimlerin iş yükü ve evrak süreleri gibi önemli göstergeler sunmaktadır. Dashboard kullanımı, bu verilerden elde edilen bilgilerin görsel ve dinamik bir şekilde sunulmasına imkân tanıyarak hem analiz sürecini hızlandırır hem de karar alma süreçlerine somut katkılar sağlar. Ham veriler tablolar halinde incelendiğinde karmaşık ve anlaşılması zor bir yapı sunarken, dashboard sayesinde bu veriler grafikler, göstergeler ve karşılaştırmalı görselleştirmeler aracılığıyla kolayca yorumlanabilmektedir.

Kurum giden evrak verilerinin dashboard üzerinden incelenmesiyle, belirli bir dönem için evrak yoğunluğunun hangi birimlerde arttığı, hangi tarihlerde evrak sayısının yoğunlaştığı veya hangi evrakların süreçlerde gecikmeye neden olduğu net bir şekilde görülebilir. Bu durum, yöneticilere birim bazlı iş yükünü dengeleme, süreçlerdeki darboğazları tespit etme ve verimlilik artırıcı önlemler alma imkânı tanır. Ayrıca, evrak çıkış sürelerinin analiz edilmesiyle süreçlerin ne kadar etkin yürütüldüğü değerlendirilebilir ve gerekli iyileştirmeler için veri destekli öneriler geliştirilebilir.

4.4. K-Means Algoritmasının Uygulanması

4.4.1. K küme sayısının belirlenmesi

Bu çalışmada k-means kümeleme analizinde en uygun k sayısını belirlemek için Elbow (dirsek) yöntemi kullanılmıştır. Elbow yöntemi, kümeleme analizinde yaygın olarak kullanılan bir yöntemdir.

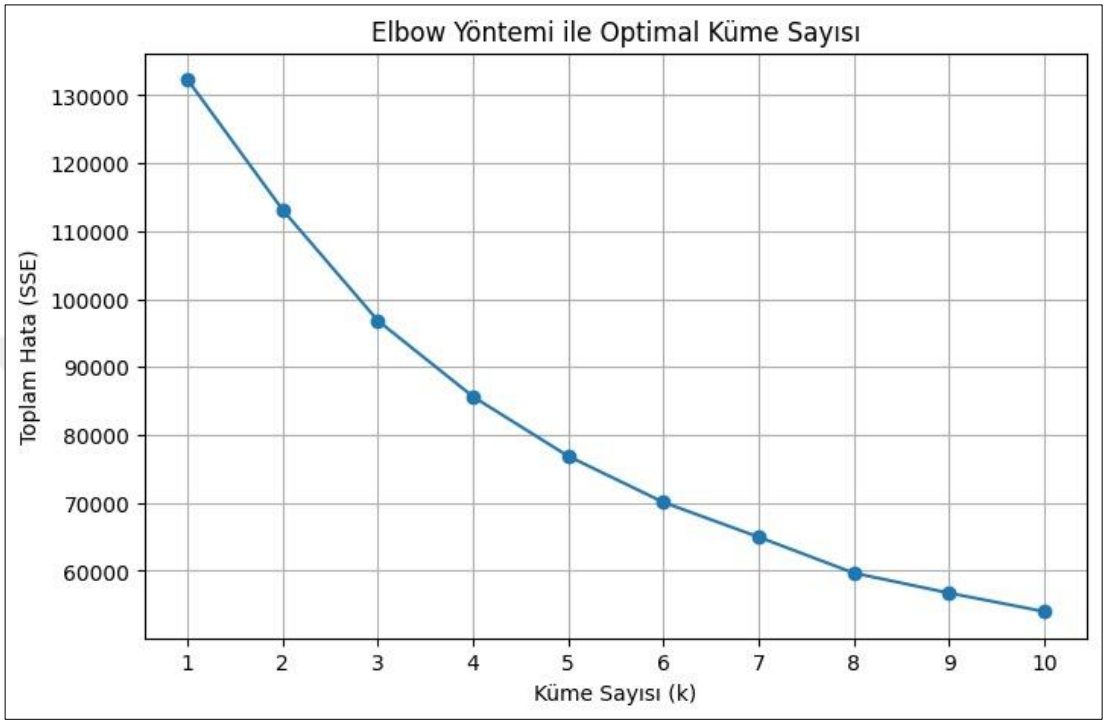
Elbow yönteminin çalışma mantığı şu şekildedir (Ketchen ve Shook, 1996):

1. Küme sayısı belirlenmeden önce analiz yapılır ve k-means algoritmasının çalışabilmesi için kullanıcıdan bir k (küme sayısı) değeri istenmektedir.
2. Belirli bir aralıkta k değerleri denenir. Genellikle k=2'den başlayarak mantıklı bir üst sınıra kadar farklı k değerleri için k-means algoritması uygulanır.
3. Her bir k değeri için toplam hata (SSE) hesaplanır. Her kümeleme sonucunda, veri noktalarının kendi küme merkezlerine olan uzaklıklarının kareleri toplanarak SSE (Sum of Squared Errors) değeri bulunur.
4. k – SSE tablosu oluşturulur. Tablo 4.10'da küme sayısı (k) ve hata sayısı (SSE) tablosu verilmiştir.

Tablo 4.10. Küme sayısı ve toplam hata (SSE) değişimi.

Küme Sayısı (k)	Toplam Hata (SSE)	Azalma (Bir önceki k'ye göre)
1	132280.00	-
2	113108.69	19171.31
3	96816.13	16392.56
4	85636.07	11180.06
5	76880.27	8755.80
6	70095.55	6784.72
7	64985.59	5109.96
8	59691.35	5294.24
9	56746.57	2944.78
10	54022.34	2724.23

5. Grafikteki dirsek noktası analiz edilir. SSE değeri, k arttıkça azalır; ancak belirli bir k değerinden sonra bu azalma yavaşlar. Bu noktada grafik dirsek görünümünü alır.
6. Dirsek noktası en uygun k değeri olarak seçilir. Grafikteki bu kırılma noktası, kümeler arasındaki ayrımın anlamlı olduğu optimal değeri verir.



Şekil 4.9. Elbow yöntemi.

Elbow yöntemi ile optimal küme sayısının belirlendiği grafik yukarıda Şekil 4.9'da verilmektedir. Grafikte net bir kırılım gözlenmese de, toplam hatadaki (SSE) azalmaların önemli ölçüde yavaşlamaya başladığı noktanın k=4 olduğu görülmektedir.

4.4.2. Kümeleme sonuçları

Bu çalışmada, Sakarya Büyükşehir Belediyesi'nin 2016 yılına ait kurum giden evrak verileri üzerinde db_id, mdr_id, sure, ay_id, dagitim, ilgi ve ek değişkenlerine göre kümeleme yapılmıştır. Çalışma gerçekleştirilirken RapidMiner Studio yazılımı kullanılmıştır. Kümeleme yöntemi olarak, hiyerarşik olmayan kümeleme yöntemlerinden biri olan k-means algoritması seçilmiş ve uzaklık fonksiyonlarından öklid seçilerek 4 ayrı küme (cluster) oluşturulmuştur. Küme sayısının 4 olacağına dirsek (elbow) yöntemiyle karar verilmiştir. Tablo 4.10'da her bir kümede yer alan evrak sayıları verilmiştir. Aynı kümede yer alan evraklar benzer özellik göstermektedir.

Tablo 4.11. Kümelerin eleman sayıları.

Küme	Eleman sayısı
Cluster 0	5004
Cluster 1	7808
Cluster 2	954
Cluster 3	2769

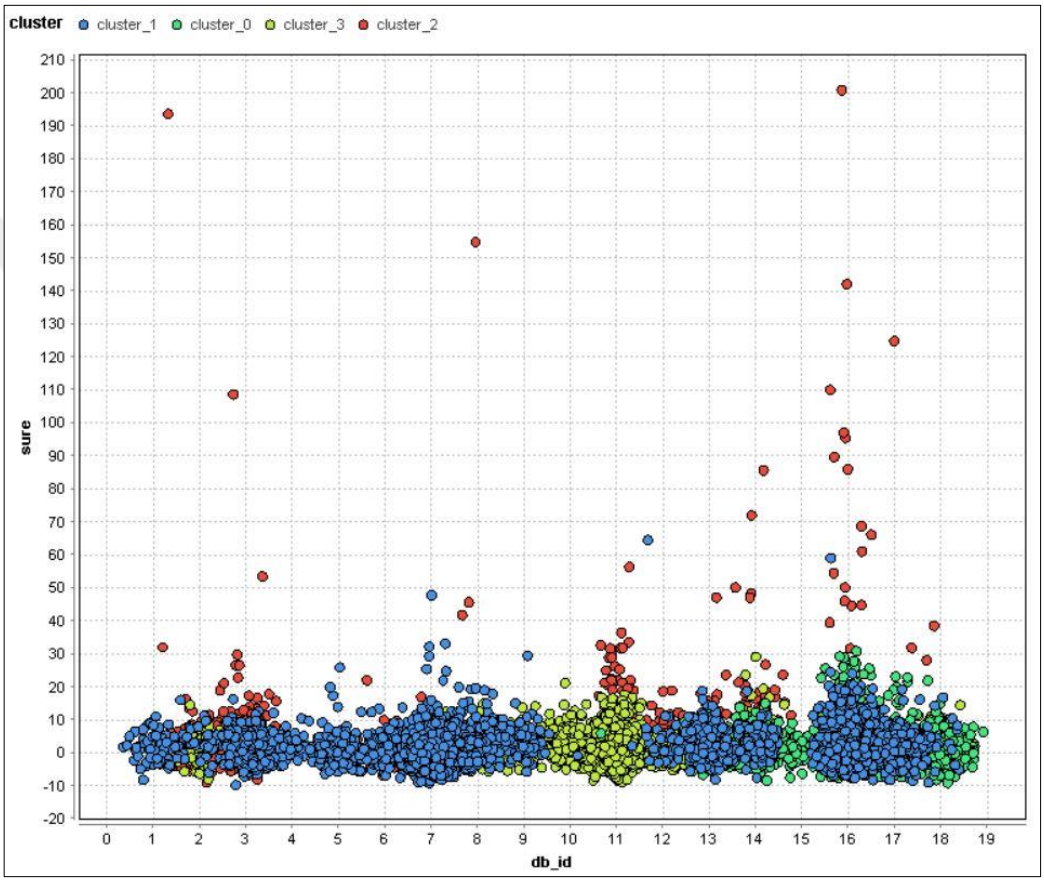
Cluster 0'da, 5004 evrak yer almaktadır. 2.44 gün ortalaması ile evrak işlemleri genellikle hızlıdır. İlgi (%80.4) ve ek (%46.8) oranları yüksektir, bu da evrakların büyük kısmının önceki belgelerle bağlantılı olduğunu göstermektedir. Yazıların daha çok dağırımsız evraklardan oluştuğu görülmektedir. Şube müdürlüğü çeşitliliği (11 farklı müdürlük) düşüktür ve evraklar belirli birimlerde yoğunlaşmıştır. Bu kümenin ortak özellikleri incelendiğinde, evrakların ağırlıklı olarak İmar ve Şehircilik Dairesi Başkanlığı'na bağlı Şehir Planlama Şube Müdürlüğü'ne, İnsan Kaynakları ve Eğitim Dairesi Başkanlığı'na bağlı İnsan Kaynakları ve Eğitim Şube Müdürlüğü'ne ve İtfaiye Dairesi Başkanlığı'na bağlı Önleme ve Eğitim Şube Müdürlüğü'ne ait olduğu görülmektedir.

Cluster 1, 7808 evrak ile en büyük kümedir. 2.09 gün ortalaması ile evrak işlemleri hızlıdır. Dağıtım oranı %15 ile diğer kümelere göre yüksektir. Bu durum daha fazla paydaşla iletişim kurulduğunu göstermektedir. Yoğun belge üretiminin olduğu ve çok sayıda müdürlüğün (25 farklı müdürlük) aktif rol aldığı bir kümedir. Bu kümenin ortak özellikleri incelendiğinde, evrakların ağırlıklı olarak Mali Hizmetler Dairesi Başkanlığı'na bağlı Gelirler Şube Müdürlüğü'ne ve İmar ve Şehircilik Dairesi Başkanlığı'na bağlı Harita, Emlak ve İstimlak Şube Müdürlüğü'ne ait olduğu görülmektedir. Evrakların en fazla çıkış yapıldığı ayların Mart ve Haziran olduğu görülmektedir.

Cluster 2, 954 evrak ile en küçük kümedir. 7.22 gün ortalaması ile evrak süresinin en uzun olduğu kümedir. Dağıtım (%5.9) ve ilgi (%37.7) oranları düşük, ek oranı (%35.4) ortalama düzeydedir. Bu kümede 23 farklı müdürlüğe ait evraklar yer almaktadır. Bu kümenin ortak özellikleri incelendiğinde, evrakların ağırlıklı olarak Fen İşleri Dairesi

Başkanlığı'na bağlı Yapı İşleri Şube Müdürlüğü'ne ait olduğu görülmektedir. Evrakların en fazla çıkış yapıldığı ayın Aralık olduğu görülmektedir.

Cluster 3'de, 2769 evrak yer almaktadır. Evrak işlem süreleri ortalama 2.44 gündür. Evrakların %78.9 ilgi oranı ile oldukça yüksektir. Ek oranı ise %40.3 ile dikkat çekici seviyededir. Bu kümede 16 farklı müdürlüğe ait evraklar yer almaktadır. Bu kümenin ortak özellikleri incelendiğinde, evrakların ağırlıklı olarak Ulaşım Dairesi Başkanlığı'na bağlı Toplu Taşıma Şube Müdürlüğü'ne ait olduğu görülmektedir.

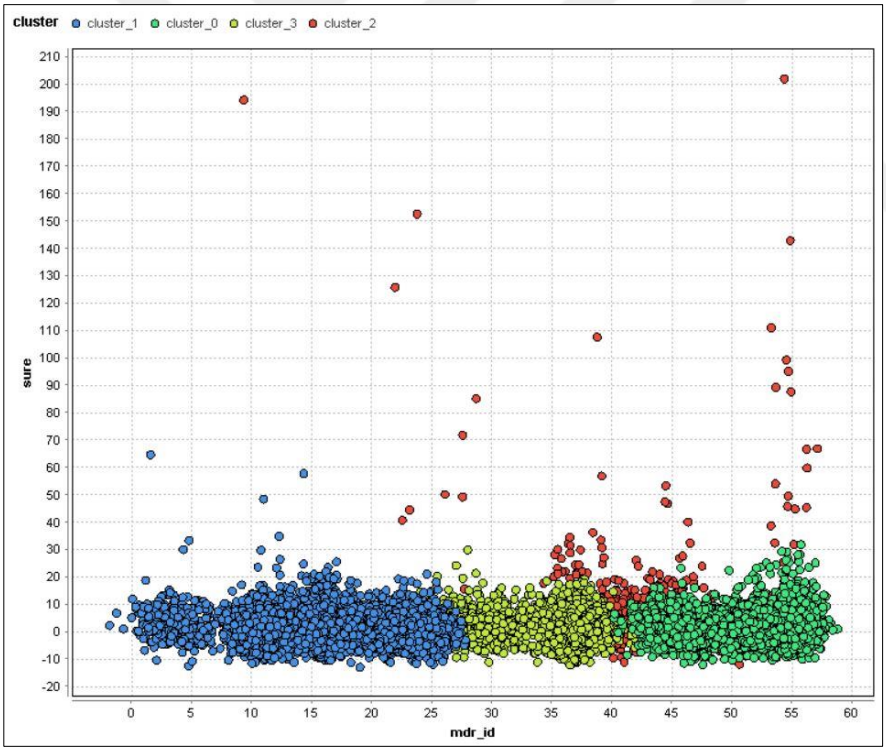


Şekil 4.10. db_id ve sure kümeleme grafiği.

Yukarıdaki Şekil 4.10.'da yer alan kümeleme grafiği incelendiğinde, belgelerin büyük çoğunluğunun işlem sürelerinin 0 ile 20 gün arasında yoğunlaştığı ve mavi renkli küme (cluster_1) içerisinde yer aldığı görülmektedir. Bu durum, evrakların büyük kısmının standart sürelerde işlem gördüğünü ve sistematik biçimde yönetildiğini göstermektedir. Mavi kümeye ait belgeler neredeyse tüm db_id aralığına yayılmış olup, kurum genelinde belge işleme süreçlerinin çoğunlukla tutarlı yürütüldüğü anlaşılmaktadır.

Öte yandan, kırmızı renkli küme (cluster_2) dikkat çekici şekilde yüksek süre değerlerine sahiptir. Bu kümeye ait belgelerin büyük bölümü 30 gün ve üzeri işlem sürelerine sahip olup, bazı evrakların 150–200 güne kadar uzadığı gözlemlenmektedir. Bu durum, belirli daire başkanlıklarında belge işleme sürecinde ciddi gecikmeler yaşandığını veya belgelerin içeriği bakımından daha karmaşık onay süreçlerine tabi olduğunu düşündürmektedir. Bu veriler, ilgili birimlerde süreç iyileştirme ihtiyacının değerlendirilmesi gerektiğini ortaya koymaktadır.

Yeşil (cluster_0) ve sarı (cluster_3) kümeler, genellikle 0 ile 20 gün arasındaki sürelerde, ancak mavi kümeden farklı bir işlem yapısına sahip oldukları anlaşılan belgeleri içermektedir. Bu belgeler orta yoğunlukta işlem sürelerine sahip olmakla birlikte, kendi içlerinde belirli daire başkanlıklarına yoğunlaşmaktadır. Bu kümeler, kurum içinde belge tipine veya işlem türüne göre farklı prosedürlerin varlığına işaret etmektedir.



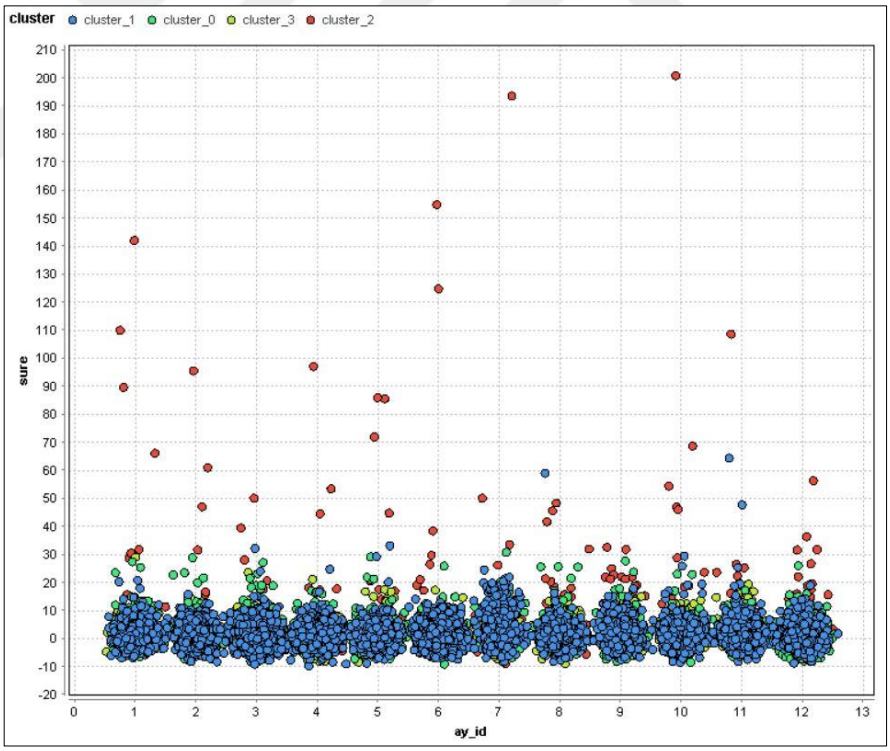
Şekil 4.11. mdr_id ve sure kümeleme grafiği.

Yukarıdaki Şekil 4.11.'de yer alan kümeleme grafiği incelendiğinde, evrakların büyük bir çoğunluğunun mavi renkli küme (cluster_1) içerisinde yoğunlaştığı ve genellikle 0–20 gün işlem süresi aralığında toplandığıdır. Bu durum, belediyeye bağlı pek çok müdürlüğün belge işlemlerini kısa sürede tamamladığını ve belge akışında

standartlaşmış bir süreç işlediğini göstermektedir. Özellikle mdr_id 0–30 aralığında kalan müdürlükler bu kümeye yoğunlukla katkı sağlamaktadır.

Öte yandan, kırmızı renkli küme (cluster_2), belgelerin işlem süresi bakımından istisnai değerlere sahip olduğunu göstermektedir. Bu kümeye ait belgeler genellikle 30 günden 200 güne kadar sürebilen işlem sürelerine sahiptir ve çoğunlukla mdr_id 30'dan büyük olan müdürlüklerde toplanmaktadır. Bu durum, bu müdürlüklerde belge işlem sürecinin olağandan uzun sürdüğünü ve sistematik bir iyileştirmeye ihtiyaç duyulduğunu işaret etmektedir.

Yeşil (cluster_0) ve sarı (cluster_3) kümelerde yer alan belgeler ise genellikle orta düzeyde işlem süresine sahip olup 0–30 gün arasında değişmektedir. Bu kümeler, diğer kümelere göre daha dengeli bir dağılım sergilemekte olup mdr_id 30–45 aralığındaki müdürlüklerde yoğunlaşmaktadır. Bu durum, bazı müdürlüklerde işlem süresinin kontrol altında tutulmakla birlikte, belirli karmaşıklıklar nedeniyle sürenin uzayabildiğini göstermektedir.



Şekil 4.12. ay_id ve sure kümeleme grafiği.

Yukarıdaki Şekil 4.12.'de yer alan kümeleme grafiği incelendiğinde, belgelerin büyük çoğunluğu mavi renkli küme (cluster_1) içinde yer almakta ve işlem süreleri 0 ile 20 gün arasında yoğunlaşmaktadır. Bu durum, yılın tüm aylarında belge işlemlerinin

büyük ölçüde zamanında ve standart bir süreçle yürütüldüğünü göstermektedir. Belgelerin işlem süresi bakımından bu düzenlilik, kurumun genel belge yönetim sisteminin etkinliğine işaret etmektedir.

Ancak, kırmızı renkli küme (cluster_2) yılın her ayında istisnai belgelerin varlığına dikkat çekmektedir. Bu kümeye ait belgelerin işlem süreleri 30 gün ile 198 gün arasında değişmekte olup, bazı aylarda (özellikle 2, 6, 7, 10 ve 11. aylar) yüksek işlem süresiyle dikkat çeken kümelenmeler oluşmuştur. Bu durum, yılın belirli dönemlerinde belge işleyişinde sistem dışı gecikmeler yaşandığını ya da bazı belgelerin daha karmaşık, çok aşamalı süreçlere tabi olduğunu göstermektedir. Söz konusu ayların yıllık izin dönemleri ya da dönemsel iş yükü artışıyla ilişkili olabileceği değerlendirilmelidir.

Yeşil (cluster_0) ve sarı (cluster_3) kümeler ise genel olarak orta seviye işlem sürelerine sahip belgeleri kapsamaktadır ve yıl boyunca dağınık biçimde gözlemlenmektedir. Bu kümeler, belirli belge türlerinin veya işleyiş biçimlerinin ay bazında değişkenlik gösterdiğini ve süreçlerin her bir belgede homojen olmadığını ortaya koymaktadır.

Sonuç olarak, bu grafik aylar bazında belge işlem süresinin genel olarak sabit bir düzende ilerlediğini, ancak yılın belirli dönemlerinde aşırı süreli evrakların ortaya çıktığını göstermektedir. K-means algoritmasının bu örüntüleri başarılı şekilde ayrıştırması, kurumun operasyonel süreçlerini dönemsel olarak değerlendirme imkânı sunmaktadır. Yüksek işlem süresine sahip belgelerin ait olduğu aylar özelinde süreç analizi yapılması, verimliliğin artırılması ve insan kaynağı planlamasında isabetli kararlar alınması açısından stratejik bir öneme sahiptir.

5. SONUÇ VE ÖNERİLER

Bu çalışmada temel amaç, Sakarya Büyükşehir Belediyesi'ne ait giden evrak verilerinin veri madenciliği yöntemleriyle analiz edilerek, belge yoğunluğu örüntülerinin ortaya çıkarılması ve bu örüntüler aracılığıyla yıllık izin planlamalarına katkı sağlayacak stratejik bilgiler sunmaktır. Bu doğrultuda uygulanan K-means kümeleme algoritması, benzer özellikler taşıyan evrakların gruplanmasına imkân tanımış ve belge yönetim süreçlerine ilişkin önemli çıkarımlar yapılmasını mümkün kılmıştır. Analiz sonucunda, iş yükünün belirli dönemlerde arttığı ve bu artışların insan kaynakları planlamasında dikkate alınması gerektiği anlaşılmıştır.

Kullanılan yöntem kapsamında, veri madenciliği süreci CRISP-DM çerçevesinde yapılandırılmış ve veri hazırlama aşamasında eksik veriler temizlenmiş, sayısal ve kategorik dönüşümler uygulanmıştır. Elbow yöntemi ile K-means algoritması için en uygun küme sayısı belirlenmiş, bu sayede analiz çıktılarının güvenilirliği artırılmıştır. Süre, ay, ek, ilgi ve dağıtım gibi nitelikler temel alınarak yapılan kümeleme sonucu, belge yönetimi açısından anlamlı gruplar ortaya çıkarmıştır.

Kümeleme analizinde elde edilen kümeler, evrak çıkış süreleri ve ait oldukları aylar bakımından belirgin farklılıklar göstermiştir. Özellikle yılın bazı aylarında evrak işlem hacminin arttığı tespit edilmiş ve bu dönemlerin yoğun hizmet talepleriyle ilişkili olabileceği değerlendirilmiştir. Bu durum, yıllık izinlerin bu dönemlerin dışında planlanması gerektiğini ortaya koymakta ve hizmet kalitesinin sürdürülebilirliğini sağlamaya yönelik somut bir veri temeli sunmaktadır.

Kümeleme sonuçları, özellikle ilgi ve ek içeriğe sahip evrakların işlem süresinin daha uzun olduğunu göstermiştir. Bu tür belgelerin işlenmesi sırasında daha fazla dikkat ve onay sürecine ihtiyaç duyulması, sürelerin uzamasına neden olmaktadır. Bu bulgu, belge niteliğine göre iş gücü planlamasının önemini vurgulamakta ve farklı belge türlerinin yönetiminde önceliklendirme gerekliliğine işaret etmektedir.

Korelasyon analizi, bazı değişkenler arasında anlamlı ilişkiler olduğunu ortaya koymuştur. Özellikle mdr_id ve db_id değişkenleri arasında gözlemlenen yüksek korelasyon, müdürlük ve daire başkanlıkları arasında belge üretiminde yapısal bir

ilişki olduğunu göstermektedir. Bu yapı, kurum içi belge trafiğinin örgütsel hiyerarşiyle doğrudan bağlantılı olduğunu desteklemektedir.

Analiz bulguları, belge çıkış süresi ile belgeye ait dağıtım ve ilgi bilgileri arasında zayıf ama anlamlı bir ilişki olduğunu ortaya koymuştur. Bu durum, daha karmaşık belgelerin işlem sürecinde zaman kaybına neden olduğunu ve bu belgelerin uygun personel yoğunluğuyla planlanması gerektiğini göstermektedir. Yani, sadece dönemsel değil, belge tipi bazında da insan kaynağı planlaması yapılmalıdır.

Yılın farklı dönemlerine ait kümeler arasında belirgin farklılıklar olması, belge trafiğinin homojen bir şekilde dağılmadığını ve operasyonel yükün belli dönemlerde yoğunlaştığını kanıtlamaktadır. Bu durum, belediyenin planlama ve yönetim süreçlerinde dönemsel analizlerin ne denli kritik olduğunu göstermektedir. Böylece, sadece geçmişe dönük değil, ileriye dönük izin ve iş gücü tahminlemesi de yapılabilir hâle gelmiştir.

K-means algoritmasının kullanımı, özellikle verilerin sayısal olması ve gözlemler arasında doğal grupların varlığı açısından uygun bir tercih olmuştur. Her bir küme, belge işlem süresi, içerik karmaşıklığı ve dönemsel yoğunluk gibi unsurlar çerçevesinde anlamlı farklılıklar sunmuştur. Bu yönüyle algoritma, evrakların içeriksel ve operasyonel benzerliklerine göre gruplanmasına olanak sağlayarak veri madenciliği süreçlerinde etkinlik sağlamıştır.

Elde edilen sonuçlar, yalnızca belge trafiği analizi ile sınırlı kalmamakta, aynı zamanda kurumsal kaynak planlamasında somut fayda sunmaktadır. Örneğin, benzer iş yüküne sahip müdürlüklerin aynı zaman dilimlerinde yoğun belge üretmesi, bu birimlerin eş zamanlı izin kullanmasının hizmette aksamalara neden olabileceğini göstermektedir. Bu bulgu, yöneticilerin çapraz planlama yapmalarını teşvik edecek niteliktedir.

Sonuç olarak, yapılan analizler, Sakarya Büyükşehir Belediyesi özelinde olmakla birlikte, tüm kamu kurumlarına yönelik uygulanabilir öneriler sunmaktadır. Veri madenciliği ve özellikle kümeleme temelli yaklaşımlar, belge yönetimi süreçlerinin yalnızca kayıt tutma işleviyle sınırlı olmadığını, aynı zamanda stratejik karar destek süreçlerine de entegre edilebileceğini göstermektedir. Bu tez, analitik kamu yönetimi anlayışının somut bir örneği olarak değerlendirilebilir

KAYNAKLAR

Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, *9*(8), 1295.

Avcı, S., Özkan, E., & Aladağ, Z. (2020). Bir Büyükşehir Belediyesi Çağrı Merkezi Verilerinin Kümeleme Analizi ile İncelenmesi. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi,* 35(3), 78-91.

Baydan, E., Yenigül, S. B., & Gürel, Z. A. (2025). Kır-kent ikileminde yeni arayışlar: K-means algoritması ile yerleşim birimlerinin kümelendirilmesi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 40(3), 1467-1478. https://doi.org/10.17341/gazimmfd.1417659

Gül, E., & Taşyürek, M. (2025). Çevre Kirliliği Artışının Azaltılması için Geri Dönüşüm Noktalarının Çok Boyutlu K-ortalamalar Algoritması ile Belirlenmesi. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, 41(1), 110-119.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, 17(6), 441-458.

Kılıç, G., Budak, İ., & Organ, A. (2020). K-ortalamalar Kümeleme Analizi İle Belediyelerin Çevre Hizmetlerinin Değerlendirilmesi. *Eskişehir Osmangazi Üniversitesi İktisadi Ve İdari Bilimler Dergisi*, 15(1), 209-230. https://doi.org/10.17153/oguiibf.545524

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. In L. M. Le Cam & J. Neyman (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics (pp. 281–297). University of California Press.

Mihevc, A., & Podržaj, P. (2025, March). Analyzing the elbow method in predicting the optimal number of clusters in Electronic Communications dataset. In *2025 24th International Symposium INFOTEH-JAHORINA (INFOTEH)* (pp. 1-7). IEEE.

Özari, Ç., & Eren, Ö. (2018). İllerin Yaşam Endeksi Göstergelerinin Çok Boyutlu Ölçekleme ve K-ortalamalar Kümeleme Yöntemi ile Analizi. Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi, 20(2), 303-313. https://doi.org/10.32709/akusosbil.427746

Özkan, Y. (2016). Veri Madenciliği Yöntemleri (Üçüncü baskı). Papatya Yayıncılık Eğitim.

Silahtaroğlu, G. (2016). *Veri madenciliği: Kavram ve algoritmalar* (3. baskı). Papatya Yayıncılık Eğitim.

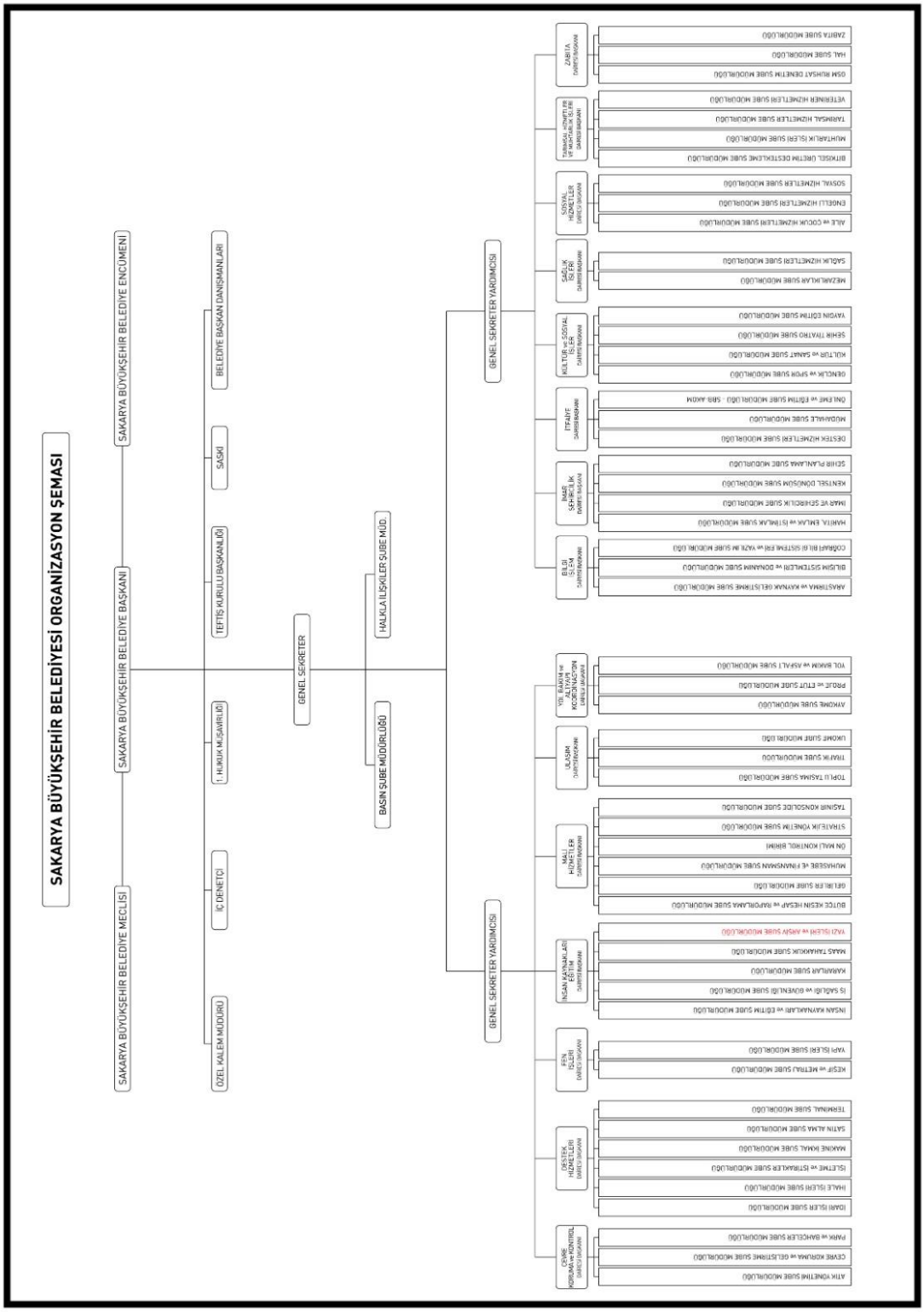
Tamtürk, E. (2017). Kamu Yönetiminde Elektronik Belge Yönetim Sistemi. *Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi*, 5(3), 851-863. https://doi.org/10.18506/anemon.288221

Zhao, F. (2025). Application and Performance Improvement of K-Means Algorithm in Collaborative. In *2025 International Conference on Intelligent Systems and Computational Networks (ICISCN)* (pp. 1-6). IEEE.

EKLER

EK A. Şekiller

EK A:



Şekil A.1. 2016 yılı Sakarya Büyükşehir Belediyesi organizasyon şeması

ÖZGEÇMİŞ

Ad-Soyad : Engin UÇAR

ÖĞRENİM DURUMU:

- **Lisans :** 2013, Sakarya Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2013-2016 yılları arasında özel bir firmada iş analisti olarak çalıştı.
- 2016-2018 yılları arasında Sakarya Üniversitesi Karasu Meslek Yüksek Okulu'nda öğretim görevlisi olarak göreve başladı.
- 2018-2020 yılları arasında Sakarya Uygulamalı Bilimler Üniversitesi'nde bilgi işlem dairesi başkanı olarak görev aldı.
- Şu anda Sakarya Uygulamalı Bilimler Üniversitesi Bilişim Teknolojileri Meslek Yüksekokulu'nda öğretim görevlisi olarak görevine devam etmektedir.

TEZDEN TÜRETİLEN ESERLER:

- Uçar E., Yurtay N. (2017,7-8, Aralık). Veri Madenciliği Yöntemleri İle Kamudaki Yıllık İzinlerin Planlanarak Hizmet Kalitesinin Arttırılması - Sakarya Büyükşehir Belediyesi Örneği. *International Conference on Quality in Higher Education*, Sakarya, Turkey.