DOKUZ EYLÜL ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

DENİZ TRAFİĞİ ANALİZİNDE ANOMALİ BELİRLEME YÖNTEMLERİNİN KULLANILMASI

Sibel ALKAN İZMİRLİ

Şubat, 2023

İZMİR

DENİZ TRAFİĞİ ANALİZİNDE ANOMALİ BELİRLEME YÖNTEMLERİNİN KULLANILMASI

Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü

Yüksek Lisans Tezi

İstatistik Anabilim Dalı, Veri Bilimi Yüksek Lisans Programı

Sibel ALKAN İZMİRLİ

Şubat, 2023

İZMİR

YÜKSEK LİSANS TEZİ SINAV SONUÇ FORMU

SİBEL ALKAN İZMİRLİ tarafından **DR. ÖĞR. ÜYESİ ENGİN YILDIZTEPE** yönetiminde hazırlanan "**DENİZ TRAFİĞİ ANALİZİNDE ANOMALİ BELİRLEME YÖNTEMLERİNİN KULLANILMASI**" başlıklı tez tarafımızdan okunmuş, kapsamı ve niteliği açısından bir Yüksek Lisans tezi olarak kabul edilmiştir.

.....

Dr. Öğr. Üyesi Engin YILDIZTEPE

Yönetici

.....	.....
Prof.Dr. Abdullah Fırat ÖZDEMİR	Prof.Dr. Femin YALÇIN
Jüri Üyesi	Jüri Üyesi

Prof. Dr. Okan FISTIKOĞLU
Müdür
Fen Bilimleri Enstitüsü

TEŞEKKÜR

Tez çalışması sürecimde, yönlendiriciliği, katkıları, anlayışlı tutumları ile destek sağlayan yardımlarını esirgemeyen ve sabırla çalışmama destek veren danışmanım sayın Dr. Engin YILDIZTEPE'ye, tez konumun belirlenmesinde ve deniz trafiği alanında sağlamış olduğu katkılar için sayın Prof. Dr. Selçuk NAS'a, AIS verileri hakkındaki bilgilendirmeleri ve açıklayıcı paylaşımları için Arş. Gör. Burak KUNDAKÇI'ya, desteği, katkıları ve anlayışı için sevgili eşim Mehmet İZMİRLİ'ye teşekkür ederim.

Sibel ALKAN İZMİRLİ

DENİZ TRAFİĞİ ANALİZİNDE ANOMALİ BELİRLEME YÖNTEMLERİNİN KULLANILMASI

ÖZ

Deniz trafiği analizinde, güvenliği sağlamak ve trafik akışını kontrol etmek için pek çok farklı yöntem kullanılmaktadır. Bu yöntemlerden biri de anomali tespitidir. Deniz trafiğinde anomali tespiti, deniz taşıtlarının olağan dışı davranışlarını veri madenciliği yöntemleri ile belirlemeyi amaçlayan bir çalışma alanıdır. Otomatik Tanımlama Sistemi (Automatic Identification System – AIS) verileri deniz trafiği analizindeki en önemli veri kaynağıdır. Bu tezde, AIS verileri ile deniz taşıtlarının anomali davranışlarının zamansal-mekansal kümeleme algoritmalarından ST-DBSCAN algoritması ile tespiti araştırılmıştır. Çalışmada, AIS verilerinin R programlama dili ile görselleştirilmesine de detaylı olarak yer verilmiştir. Deniz trafiğinde anomali gemi davranışı tespit süreci görsel içeriklerle desteklenmiştir. Uygulamada, Kuzey Amerika'da bulunan iki liman bölgesine ait AIS verileri kullanılmıştır. Elde edilen sonuçlara göre ST-DBSCAN algoritması ile gemilerin konumlarında meydana gelen anomali davranışlarının tespit edilebildiği görülmüştür. Algoritma olağan davranışların gerçekleştiği alanları kümeleyebilirken olağan davranıştan uzak noktaları ayırabilmiştir.

Anahtar kelimeler: AIS, anomali, deniz trafiği, ST-DBSCAN, zamansal-mekansal kümeleme

ANOMALY DETECTION METHODS IN MARINE TRAFFIC ANALYSIS

ABSTRACT

In marine traffic analysis, various methods are employed to ensure safety and manage the flow of traffic. One such method is anomaly detection, which is a research area that aims to identify unusual behaviors of marine vessels using data mining techniques. Automatic Identification System (AIS) data is a crucial source of data in marine traffic analysis. This thesis investigates the detection of the anomalous behavior of marine vessels using the ST-DBSCAN algorithm, a spatio-temporal clustering algorithm, with AIS data. The thesis provides a detailed explanation of visualizing AIS data using the R programming language. The application in this study utilizes AIS data from two harbors in North America. The results indicate that the ST-DBSCAN algorithm can effectively detect anomalous behaviors of marine vessels in their positions. The algorithm is able to cluster areas where typical behavior occurs while also identifying points that deviate from typical behavior.

Keywords: AIS, anomaly, marine traffic, ST-DBSCAN, spatio-temporal clustering

İÇİNDEKİLER

Sayfa

SINAV SONUÇ FORMU ii
TEŞEKKÜR iii
ÖZ v
ABSTRACT iv
ŞEKİLLER LİSTESİ ix
TABLOLAR LİSTESİ x

BÖLÜM BİR – GİRİŞ 1

BÖLÜM İKİ – ANOMALİ KAVRAMI 5

2.1 Anomali Türleri 6
2.2 Anomali Belirleme Yöntemleri 7

BÖLÜM ÜÇ – DENİZ TRAFİĞİNDE ANOMALİ KAVRAMI 10

3.1 Deniz Trafiğinde Anomali Gemi Davranışları 11

BÖLÜM DÖRT – VERİ VE YÖNTEM 16

4.1 AIS - Otomatik Tanımlama Sistemi 16
4.2 ST-DBSCAN 20
4.3 AIS Verilerini Kullanarak ST-DBSCAN Algoritması ile Anomali Belirleme 26

BÖLÜM BEŞ – UYGULAMA 30

5.1 Veri 30
5.2 Veri Ön İşleme 31
5.3 AIS Veri Görselleştirme 33
5.4 Charleston Limanı 33

5.4.1 Rota Görselleştirme 34
5.4.2 Kritik Bölge Seçimi 36
5.4.3 ST-DBSCAN ile Anomali Belirleme 42
5.4.4 Parametre Belirleme 45
5.5 Savannah Limanı 48

BÖLÜM ALTI – SONUÇLAR 55

KAYNAKLAR 57

ŞEKİLLER LİSTESİ

Sayfa

Şekil 2.1 Kolektif anomali örneği 7

Şekil 2.2 Yerel ve küresel anomali türleri 8

Şekil 4.1 ST-DBSCAN algoritma yapısı 23

Şekil 4.2 Eps1 parametre seçimi K-NN grafiği 25

Şekil 4.3 ST-DBSCAN ile anomali belirleme süreci 26

Şekil 5.1 Charleston Limanı iki haftalık trafik akışı 35

Şekil 5.2 Charleston Limanı kritik bölge seçimi 37

Şekil 5.3 Charleston Limanı kritik bölge 38

Şekil 5.4 Charleston Limanı kritik bölge gemilerin hız davranışı 38

Şekil 5.5 Charleston Limanı belirlenen kritik bölge 39

Şekil 5.6 Gemilerin COG ve heading açılarının dağılımı 40

Şekil 5.7 Charleston Limanı gemilerin hareket yönleri 41

Şekil 5.8 Charleston Limanı kritik bölgede liman doğrultusunda hareket eden gemilerin kanal içindeki dağılımları 42

Şekil 5.9 Oluşturulan anomali davranışlar 43

Şekil 5.10 Charleston Limanı normal ve anomali noktalar 44

Şekil 5.11 Eps1 parametre seçimi K-NN grafiği 45

Şekil 5.12 Charleston Limanı kümeleme sonuçları 47

Şekil 5.13 Savannah Limanı trafik akışı 48

Şekil 5.14 Savannah Limanı gemilerin hız davranışı 49

Şekil 5.15 Savannah Limanı kritik bölge 50

Şekil 5.16 Savannah Limanı trafik yön ayrımı için kesit alınan bölge 51

Şekil 5.17 Savannah Limanı anomali senaryosu 52
Şekil 5.18 Savannah Limanı kümeleme sonuçları 53

TABLOLAR LİSTESİ

Sayfa

Tablo 4.1 AIS veri iletim sıklığı 17
Tablo 4.2 AIS kinematik değişkenler 18
Tablo 4.3 AIS kinematik olmayan değişkenler 19
Tablo 4.4 AIS gemi tipi numaraları 20
Tablo 4.5 Yanılgı matrisi 28
Tablo 5.1 Ham veri ön izlemesi (ilk beş satır) 31
Tablo 5.2 Verilerin gemi tipine göre dağılımı 33
Tablo 5.3 Kritik bölgede yer alan gemilerin COG ve heading değerleri 40
Tablo 5.4 ST-DBSCAN algoritmasının farklı parametre değerleri için performansının değerlendirilmesi 46
Tablo 5.5 Charleston Limanı yanılgı matrisi 47
Tablo 5.6 ST-DBSCAN parametre seçimi 52
Tablo 5.7 ST-DBSCAN parametre seçimi-2 53
Tablo 5.8 Savannah Limanı yanılgı matrisi 54

BÖLÜM BİR
GİRİŞ

Küresel ticarette ürünlerin büyük bir kısmı deniz yoluyla taşınmaktadır. Genellikle büyük hacme sahip, birim fiyatı düşük, zaman duyarlılığı az yüklerin taşınmasında deniz yolu taşımacılığı tercih edilir. Ticaretin küreselleşmesinde de deniz taşımacılığı öncü ve önemli rol oynamaktadır. Dolayısıyla, deniz taşımacılığının uluslararası ticaretin bel kemiğidir (Grech, Horberry, ve Koester, 2019).

Deniz trafiğinin karmaşıklığı ve her geçen gün artan hacmi, deniz taşımacılığını güvenlik açısından hassas ve yüksek riskli bir hale getirmektedir. Deniz trafiğinde, güvenliği sağlamak, çevreyi korumak, trafik akışlarını optimize etmek, yasa dışı faaliyetlerden kaçınmak ve korunmak için farklı deniz gözetim teknikleri ve trafik analizleri gerekmektedir. Deniz taşıtlarının hareketleri hakkında bilgi edinmemizi sağlayan cihazların gelişmesi, artan veri depolama kapasiteleri deniz trafigi ile ilgili büyük boyutlarda veriye erişimi mümkün kılmaktadır (Izmirli ve Yıldıztepe, 2021).

Analistlerin bu alanda ki en önemli veri kaynağı Otomatik Tanımlama Sistemi (Automatic Identification System – AIS)'dir. AIS kullanan deniz taşıtları, konum, rota, hız, varış limanı ve gemi özellikleri gibi dinamik ve statik bilgileri düzenli aralıklarla iletirler. Bu veriler bir çok deniz gözetim sistemine ve deniz trafiği analiz modeline girdi olarak kullanılmaktadır. Varlamis, Tserpes, ve Sardianos (2018) çalışmalarında AIS verilerini kullanarak denizdeki arama kurtarma görevlerini otomatik olarak tespit eden bir araç sunmuşlardır. Pereira, Abreu, Polisciuc, ve Machado (2019) tarafından AIS verilerinin işlenmesine ve görselleştirilmesine dayalı gemi yörüngelerindeki kesişimleri analiz eden, çarpışmadan kaçınmak için bir yaklaşım geliştirilmiştir. Jiacai, Qingshan, Jinxing, ve Zheping (2012) deniz trafik durumunu değerlendirmek için AIS verilerine dayalı bir görselleştirme modeli önermişlerdir. AIS sisteminin, denizde güvenliği sağlama, çarpışmadan kaçınma, rota tahmini, trafik akışlarını analiz etme ve deniz trafiğinde anomali belirleme gibi yöntemler için önemli bir veri kaynağı olduğu bilinmektedir (Tu, Zhang, Rachmawati, Rajabally, ve Huang, 2017). Bu yöntemlerin ortak amacı denizcilik

alanında emniyeti ve güvenliği sağlamak, trafik akışlarını optimize etmek aynı zamanda yasa dışı faaliyetlerden korunmak olarak açıklanabilir. Tüm bu amaçlara hizmet eden yaklaşımlardan biri de anomali tespitidir.

Deniz trafiğinde anomali belirleme çalışmaları özellikle 2000'li yıllardan sonra artış göstermiştir. Deniz trafiğinde anomali kavramı güncel ve önemli bir çalışma alanıdır (Riveiro, Pallotta, ve Vespe, 2018). Deniz gözetim operatörleri yoluyla verilerin araştırılması ve izlenmesi zorlu bir süreçtir. Verilerin karmaşık ve heterojen doğası nedeniyle anlaşılması zordur. Aynı zamanda belirsizlik, yorgunluk veya diğer zaman kısıtlamaları gibi faktörler nedeniyle operatörler için zorlu bir faaliyet haline gelebilir. Gözetim operatörleri genellikle, tehlike arz eden durumları önlemek için, (örneğin çarpışma riski) denizdeki pek çok farklı gemiden gelen verileri araştırır ve durumlarını tahmin etmeye çalışır. Anormal ve tehlike arz eden durumların erken tespiti, potansiyel problemler ortaya çıkmadan önce uygun şekilde önlem almak için kritik bir zaman sağlayacaktır. Bu durumlarda operatörlere destek sağlamak için, anomali tespit özelliklerine sahip yöntemler ve sistemler yaygın bir şekilde kullanılmaktadır (Riveiro vd., 2018).

Anomali tespiti ise, en basit anlamıyla verideki beklenmedik durumların veya desenlerin bulunmasını amaçlayan bir süreçtir (Chandola, Banerjee, ve Kumar, 2009). Bu bağlamda deniz trafiğinde anomali tespiti, beklenen veya bilinen davranışından sapan deniz taşıtlarının durumlarını tespit etmeyi amaçlayan yöntemler bütünü olarak tanımlanabilir. Denizcilikte anomali tespiti, iki aşamada açıklanabilir. İlk olarak geçmiş verilerden olağan gemi hareket modellerini çıkarmaktır. İkinci aşama ise olağan hareket modellerinden önemli ölçüde sapan şüpheli davranışları tespit etmektir. Örneğin anomali gemi davranışı, beklenmedik şekilde ani bir hızlanma ya da yavaşlama veya konum değişikliği, standart deniz yolundan sapma, gemiler arası yakınlaşma ve gemilerin izinsiz bir bölgeye girişi olarak örneklendirebilir (Lane, Nevell, Hayward, ve Beaney, 2010). Deniz trafiğinde anomali gemi davranışları, gemilerin manevralarında, konumlarında, seyahat hız ve sürelerinde ayrıca yörünge ve rotalarında meydana gelebilir. Handayani, Sediono, ve Shah (2013) çalışmalarında iki farklı senaryoda anomali gemi davranışı örnekleri sunmuşlardır. Bunlar rotada

beklenmedik durma ve rotada u dönüşü davranışlarıdır. Sidibé ve Shu (2017) çalışmalarında ne tür AIS verilerini kullanarak hangi anomali gemi davranışlarının tespit edilebileceğinin incelemesini sunmuşlardır.

Anomali tespitinde önemli olan bir diğer konu da yöntem seçimidir ve veri yapısına göre belirlenmelidir. Deniz trafiğinde gemi davranışlarını tanımlamak ve incelemek için en önemli veri kaynağı AIS verilerdir. AIS verilerinin zamansal ve mekansal boyutunu göz önüne alan az sayıda çalışma vardır. Zamansal-mekansal veriler, hem mekan hem de zaman bilgisi barındırmaktadır. Literatürde pek çok çalışma AIS verilerini anomali gemi davranışı tespiti için geleneksel kümeleme algoritmaları ile kullanmıştır. Zhao ve Shi (2019) çalışmalarında anormal gemi davranışlarını tespit etmek için, DBSCAN algoritması ile LSTM sinir ağı algoritmasını kullanan bir yaklaşım önermişlerdir. Kümeleme sonuçlarından gemi yörüngelerini yapay sinir ağı algoritması LSTM için eğitim seti olarak kullanmışlardır. Doğan, Kart, Kundakçi, ve Nas (2021) çalışmalarında AIS verileri ile Kmeans++ ve geliştirilmiş DBSCAN kümeleme algoritmalarını kullanarak yasa dışı transit gemi izlerinin keşfini ele almaktadır. Geleneksel kümeleme algoritmaları zamansal-mekansal verileri kümelemek üzerine kurgulanmamıştır. Kümeleme algoritmalarının zamansal-mekansal kümelemeye odaklanan Birant ve Kut (2007) tarafından öne sürülen bir versiyonu vardır. Bu algoritma Spatial-Temporal Density-Based Spatial Clustering of Applications with Noise (ST-DBSCAN) algoritmasıdır. Bu algoritmanın önerilen çalışma alanlarından biri de AIS verileridir. Wang, Yao, Meng, ve Claramunt (2020) gemilerin, COVİD-19'un yayılımındaki etkisini araştıran çalışmasında ST-DBSCAN algoritmasını, gemilerin uğradıkları ülkelerin ve limanların tespiti için kullanmışlardır. Ancak literatürde ST-DBSCAN algoritmasının AIS verileri ile kullanımı oldukça kısıtlıdır.

Bu çalışmada, deniz trafiği analizinde anomali gemi davranışlarını tespit etmek için ST-DBSCAN algoritmasının kullanımını araştırıyoruz. Bugüne kadar ST-DBSCAN algoritmasının deniz trafiği analizinde AIS verileri ile anomali tespiti için kullanımı sunulmamıştır. ST-DBSCAN algoritmasının bu alanda en önemli avantajı zamansal-mekansal niteliklere odaklanarak kümeleme yapıyor olmasıdır.

Kümelerden elde edilecek rutinler anomali belirlemede kullanılabilir. Oluşacak kümelerden yola çıkılarak normal davranışa uymayan gemi hareketleri tespit edilmeye çalışılacaktır.

BÖLÜM İKİ
ANOMALİ KAVRAMI

Anomali terimi, birçok alanda kullanılmaktadır. Kullanılan alana göre anomali kavramı farklılık gösterebilir. En yaygın tanımı, bir verinin beklenen davranışına uymayan durumlar veya desenler anomalidir. Anomaliler, nadir meydana gelen ve normal verilerden önemli ölçüde farklı gözlemlerdir. Bu durumlar literatürde aykırı değer, istisnai durum, sapan değer olarak da adlandırılır.

Anomali tespiti, en basit anlamıyla verideki beklenmedik durumların veya desenlerin bulunmasını sağlayan bir süreçtir. Bu beklenmedik durumlar veya desenler aslında bir verinin beklenen davranışına uymayan desenler veya kalıplardır (Chandola vd., 2009). Anomali tespitinin yaygın ve özelleşmis kullanım alanları vardır. Örneğin, sahtekarlık tespiti alanında bir sistemin yanlış kullanımlarını, sahtekarlığı belirten şüpheli olayları, kötüye kullanımı veya çalınan kredi kartlarını tespit etmek amacıyla uygulamalar yapılabilir. Bilişim sistemlerindeki veri kaçaklarını tespit etmek, hassas bilgilerin erken aşamada kaybını önlemek gibi amaçlar için de kullanılabilir. Bir başka yaygın uygulama alanı da sağlıktır. Hayati tehlike arz eden durumları, anormal hücre veya tümörleri saptamak için anomali tespiti kullanılabilir. Uygulama alanlarına bakıldığında anomali tespitinin, kredi kartı dolandırıcılığı, sağlık hizmetleri, siber güvenlik için izinsiz giriş tespiti, güvenlik ağlarından kritik sistemlerde arıza tespiti, adli tıp, dolandırıcılık tespiti gibi çok çesitli alanlarda kullanıldığı görülmektedir.

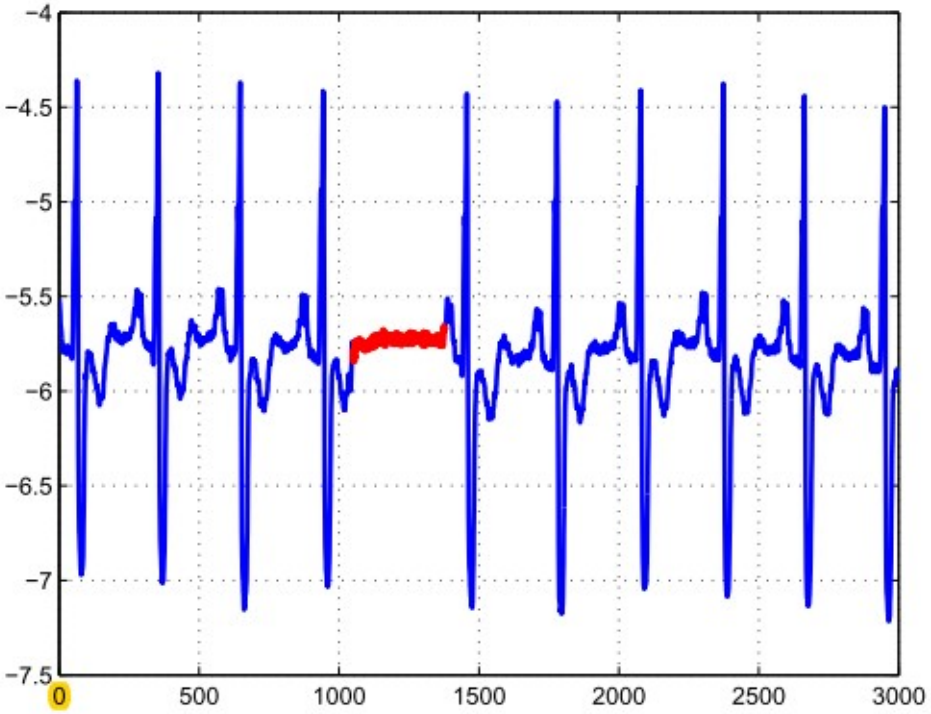
Tüm bu uygulama alanlarının ortak noktası, normal davranış tanımlamasına ihtiyaç duyulmasıdır. Normal davranış tanımlaması yapılırken, her olası normal davranışı tanımlayan bir bölge oluşturmak, normal ve anormal davranış arasındaki sınırı belirlemek dikkat edilmesi gereken noktalardır. Bir anomalinin tam tanımı farklı uygulama alanları için farklılık gösterebilir. Örneğin, tıbbi alanda normalden küçük bir sapma (vücut sıcaklığındaki dalgalanmalar) bir anormallik olabilirken, borsa alanındaki benzer dğerdeki sapma (bir stok değerindeki dalgalanmalar) normal olarak kabul edilebilir. Dolayısıyla bir alanda gelistirilen bir tekniğin diğerine

uygulanması kolay değildir. Anomali tespit teknikleri tarafından kullanılan modellerin eğitimi-doğrulanması için normal ve anormal olarak etiketlenmiş verilere ihtiyaç duyulur. Kullanılacak anomali tespit yöntemi, verilerin niteliği, etiketli verilerin mevcudiyeti, tespit edilecek anomalilerin tipi vb. gibi faktörler göz önünde bulundurularak belirlenmelidir. Bu konuda çalışan araştırmacılar, istatistik, makine öğrenmesi, veri madenciliği ve bilgi teorisi gibi çesitli disiplinlerden kavramları benimsemiş ve bunları bir arada kullanmışlardır (Chandola vd., 2009).

Kısacası, anomali tespiti yapılmak istenen uygulama alanında, anomali davranış tanımının alana özgü yapılması gerekebilir. Normal davranış olarak kabul edilecek durum titizlikle belirlenmelidir. Ayrıca anomali tespiti yapılacak verinin doğası yöntem seçimini etkileyecektir. Yöntem seçiminde tespit edilmek istenen anomalinin türüne ve veri yapısı da dikkat edilmelidir.

2.1 Anomali Türleri

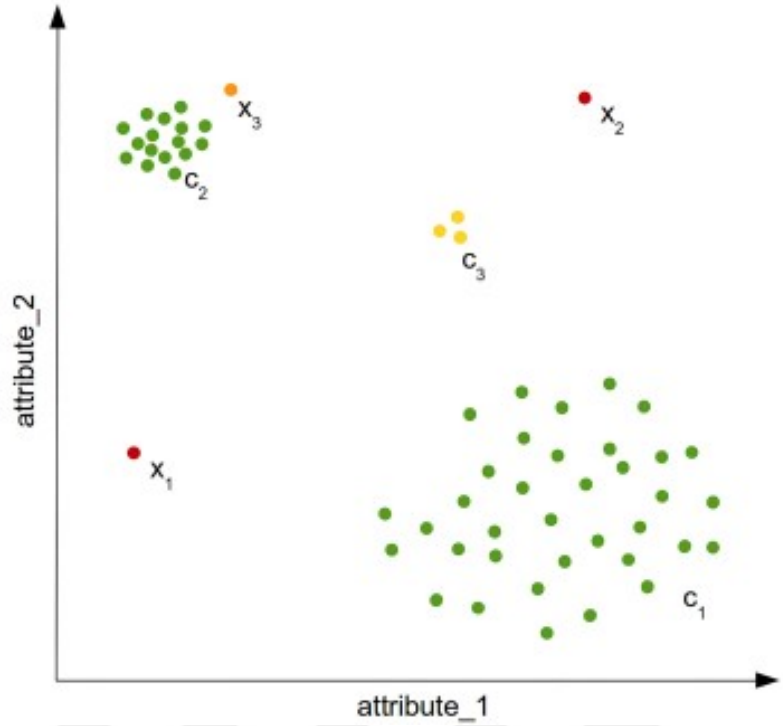
Anomaliler üç kategoride sınıflandırılır. Bunlar nokta anomaliler, ba˘glamsal anomaliler ve kolektif anomalilerdir. Bireysel bir veri örne˘gi verinin geri kalanına göre farklılık gösteriyor ise bu veri örne˘gi nokta anomali olarak adlandırılır. Nokta anomali, sapan de˘ger veya aykırı de˘ger olarak da isimlendirilebilir. Bir bireyin, normal kredi kartı harcama davranı¸sına göre çok yüksek bir harcama yapmı¸s olması, nokta anomaliye örnek verilebilir. Bu bireyin, daha önceki kredi kartı harcamalarına göre farklı bir davranı¸s söz konusudur. Tek ba¸sına bakıldı˘gında anomali olarak de˘gerlendirilmeyen fakat özel bir ba˘glamda bakıldı˘gında anomali olarak nitelendirilebilen durumlar ba˘glamsal anomali olarak isimlendirilir. Yaz ayları için ortalama hava sıcaklı˘gının 30 derece olması anormal bir durum olarak algılanmazken, sonbahar ve kı¸s aylarında ortalama hava sıcaklı˘gının 30 derece civarında olması, mevsimsel açıdan bakıldı˘gında ba˘glamsal bir anomali örne˘gidir. Bireyselde anomali olarak de˘gerlendirilmeyen, ancak bir araya gelerek olu¸sturdukları desen ile verinin geri kalanına göre farklılık gösteren gözlemler ise, kolektif anomali olarak adlandırılır (Chandola vd., 2009).



Şekil 2.1 Kolektif anomali örneği (Chandola vd., 2009)

Şekil 2.1 ’de bir bireyin elektrokardiyografi (EKG) grafiği yer almaktadır. Kırmızı bölgedeki değerler tek başlarına incelendiklerinde anomali olmayan, tek tek bakıldığında oluşabilecek değerlerdir. Ancak art arda bir araya gelerek anormal şekilde bir süre devam etmişlerdir. Bu durum bir kolektif anomali örneğidir.

Anomaliler veri setindeki konumlarına göre yerel (local) ve küresel (global) anomaliler olmak üzere iki sınıfa ayrılır. Şekil 2.2’de X1 ve X2 gözle bakıldığında kolayca tespit edilebilen, yoğun alanlardan belirgin şekilde uzakta olan gözlemlerdir. Bu gözlemler küresel anomali olarak adlandırılır. X3 ise veri kümesinde ilk bakışta belirgin şekilde uzakta olmayan, normal bir örnek gibi gözükmekteyken, C2 kümesine odaklanıldığında anomali olduğu anlaşılan bir gözlemdir. Bu gözlemler yerel anomali olarak adlandırılır.



Şekil 2.2 Yerel ve küresel anomali türleri (Goldstein ve Uchida, 2016)

2.2 Anomali Belirleme Yöntemleri

Anomali tespiti yapılmak istenen verinin etiketli olup olmadığı yöntem seçimini etkileyecektir. Verinin etiketli olması, tüm veri örneklerinin normal ya da anomali olarak sınıflandırılması anlamına gelir. Bu bağlamda anomali belirleme yaklaşımları denetimli, denetimsiz ve yarı denetimli olmak üzere üç grup altında incelenebilir. Denetimli anomali belirleme yaklaşımları, eğitim ve test verilerinin tamamının etiketli olduğu durumlarda kullanılır. Bu yöntemde tüm verilerin doğru etiketlendiği varsayılır. Bu pratikte her zaman mümkün olmayabilir. Pek çok uygulama için anomaliler önceden bilinmemektedir. Yarı denetimli anomali belirleme yaklaşımları, eğitim ve test verilerinden, eğitim verilerinin normal örneklerden oluştuğu durumlarda kullanılır. Denetimsiz anomali belirleme yaklaşımında, eğitim kümesine ihtiyaç duyulmaz, gözlemler önceden normal-anomali olarak etiketlenmez. Denetimsiz anomali belirleme yöntemleri diğer yöntemlere kıyasla hızlı, esnek ve gerçek hayat problemlerine daha uygundur. Verileri etiketlemek her zaman mümkün olmayan zor ve maliyetli bir iş olabilir.

Bir anomali tespit algoritmasının çıktısı iki şekilde olabilir. Birincisi, bir veri noktasının normal ya da anomali olduğunu gösteren bir etiket olabilir. İkincisi ise, anormallik derecesini gösteren bir puan (skor) değeridir. Bu tür durumlarda bir eşik değer belirlenerek veri noktaları belirlenen eşik değere göre normal ya da anomali olarak etiketlenebilir. Genellikle denetimli anomali belirleme algoritmaları çıktı olarak etiket kullanırlar. Diğer yandan yarı-denetimli ve denetimsiz anomali belirleme algoritmaları çıktı olarak anomali puanı üretirler (Goldstein ve Uchida, 2016).

BÖLÜM ÜÇ
DENİZ TRAFİĞİNDE ANOMALİ KAVRAMI

Deniz ulaşımı, büyük miktarlarda yükü uzak mesafelerden taşımanın en ekonomik ve en etkin yolunu sağlar. Uluslararası Denizcilik Örgütü'nün (International Maritime Organization-IMO) yaptığı istatistiksel tahminler, dünya ticaretinin yaklaşık % 90'ının deniz yolu ile gerçekleştiğini ve ticaret hacminin giderek büyüdüğünü vurgulamaktadır (Kaluza, Koelzsch, Gastner, ve Blasius, 2010). Dünya ticaretindeki büyüme, daha fazla kargo kapasitesine ve daha yüksek seyahat hızına sahip gemi talebine yol açmaktadır. Bu nedenle deniz trafiğinde güvenliği ve emniyeti ilişkin kaygılar daha fazla ön plana çıkmaktadır. Deniz trafiğinde ya da gemilerde meydana gelebilecek sorunlar oldukça büyük ekonomik problemleri ve çevresel tehditleri meydana getirebilir.

İstatistiksel çalışmalar, deniz trafiğinde meydana gelen kazaların %75 ila %96'sının insan hatalarından meydana geldiğini göstermektedir (Harati Mokhtari, Wall, Brooks, ve Wang, 2007). Deneyimli ve ilgili personel kaza oranını azaltmaktadır ancak tek başına yeterli değildir. Aynı zamanda deniz trafiğini analiz eden, olası kazaları önleyebilecek yardımcı sistemlere de ihtiyaç duyulmaktadır. Liman görevlileri ve gemilerde bulunan navigasyon gibi akıllı sistemler, deniz trafiğini izleyerek şüpheli durumları veya olası çarpışmaları belirleyip uyarılar verirler. Deniz trafiğindeki artış ve deniz trafiğinin karmaşıklığı, gözetleme sistemlerinin de gelişmesine neden olmuştur. Bu alanda, deniz trafiği emniyetini sağlamak için geliştirilen algoritmalar ve veri madenciliği yöntemleri oldukça önemlidir. Veri madenciliği yöntemleri ve algoritmaları deniz gözetleme sistemleri haricinde çeşitli denizcilik faaliyetlerinde de kullanılmaktadır. Örneğin; sefer optimizasyonu, gemi bakım onarım tahminleri, taşımacılığın sürdürülebilirliği, enerji verimliliği ve nakliye sorunları gibi. Deniz güvenliğinin iyileştirilmesi için kullanılan veri madenciliği yöntemlerinden birisi de anomali tespitidir. Deniz trafiğinde anomali tespiti, olası güvenlik ve seyrüsefer emniyeti tehlikelerini belirleyebilir bu nedenle gemideki akıllı navigasyon sistemi ve liman yetkilileri için oldukça değerlidir.

Deniz trafiğinde anomali tespiti, deniz taşıtlarının olağan dışı davranışlarını veri madenciliği yöntemleriyle belirlemeyi amaçlayan bir çalışma alanıdır. Burada ki temel mantık, olağan gemi hareket davranışlarını geçmiş verilerden çıkarmayı ve olağan hareket davranışlarından önemli ölçüde sapan şüpheli gemi davranışlarını belirlemeye çalışmaktır. Olağan gemi hareketlerinin genellikle bir deseni takip etmesi beklenebilir. Belirli bir alanda gemilerin belirli hız sınırları içinde seyahat etmesi, belirli iki liman arasında deniz taşıtlarının genellikle aynı yolu izleyerek seyahat etmesi gibi. Olağan dışı hareket yani anomali davranışın doğru tanımlanabilmeside önemlidir. Bu doğrultuda alan bilgisine sahip kişilerin görüşlerine başvurulabilir ya da ilgili literatür incelenebilir.

3.1 Deniz Trafiğinde Anomali Gemi Davranışları

Deniz trafiği anomalileri, gemilerin manevralarında, konumlarında, seyahat hız ve sürelerinde ayrıca yörünge ve rotalarında oluşan olağan dışı davranışlarla meydana gelebilir. Deniz trafiğinde anomali gemi davranışlarını sınıflandırarak inceleyen bazı çalışmalar vardır. Tu vd. (2017) çalışmalarında anomali gemi davranışlarını konum, hız ve zaman olarak üç genel başlık altında incelemişken, Horn, Eisler, Dobias, ve Collins (2016) çalışmasında anomali gemi davranışlarını manevra, yer ve etkileşim anomalileri olarak incelemiştir.

Tu vd. (2017) konum, hız ve zaman anomalilerini şöyle tanımlar: Konum anomalisi, bir geminin yasak bir bölgede veya beklenmedik bir konumda olmasıyla oluşabilir. Bir geminin hızı, bölgeye göre normal hızın önemli ölçüde üstünde veya altında ise hız anomalisi yaratabilir. Zaman anomalisi ise bir geminin limanda kalma süresinin beklenen ziyaret süresine göre uzun veya kısa olması ile oluşabilir.

Horn vd. (2016) manevra, yer ve etkileşim anomalilerini şöyle tanımlar: Manevra anomalisi, hız içeren anomaliler, beklenmedik bir yön değişikligi, durma, dolaşma veya olağandışı bir hızda seyahat etme gibi bir davranıştır. Tutarsız yön değişimi, gözlemlenen zaman içinde hız değişimi gibi manevra hareketlerini içerir. Yer anomalisi, bir geminin bulunduğu konum ile ilgilidir. İzinsiz bölgelere veya

kısıtlanmış alanlara girmek, gemi trafik ayrım düzeni dışında seyahat etmek, normal standart rotaların dışına çıkmak örnek verilebilir. Etkileşim anomalisi, gemiler arasındaki yakınlaşma olarak açıklanabilir. Etkileşim anomalisine hem açık denizde, hem de kıyı sularında gemi buluşması örnek verilebilir. Gemilerin denizde buluşmaları iyimser bir neden ile olabileceği gibi, kaçakçılık gibi yasa dışı bir faaliyet içinde olabilir.

İlgili literatürdeki çalışmaların çoğunluğunun anomali gemi davranışlarını sınıflamak yerine, belirli olayları incelemeye odaklandığı da görülmüştür. Yaygın olarak belirlenmeye çalışılan anomali gemi davranışları vardır. En genel beş anomali gemi davranışı standart rotadan sapma (Deviation from standard route), izinsiz bölgeye giriş (Zone Entry), beklenmeyen AIS aktivitesi(Unexpected AIS activity), beklenmedik limana varış (Unexpected port arrival) ve çok yaklaşmadır (Close approach) (Lane vd., 2010).

Gemilerin iki nokta arasındaki en ekonomik yoldan seyahat etmesi avantajlıdır. Bir geminin okyanus yüzeyindeki hareketini kısıtlayan, kara parçaları, su derinliği, hava durumu ve yasak bölgeler gibi engeller olabilir. Bu engeller gemi yollarının belirli kalıpları izlemesine neden olur (Lane vd., 2010). Böylelikle standart rota olarak adlandırabileceğimiz gemi yolları oluşur. Gemilerin standart rotalarından sapmaları incelenmesi gereken bir durumdur. Standart rotasından çıkan bir gemi karaya oturma, çarpışma, dümen kitlenmesi vb. tehditler ile karşı karşıya kalabilir. Bomberger, Rhodes, Seibert, ve Waxman (2006), gemilerin mevcut hareketlerinden davranışlarını öğrenerek bir sonraki hareketlerini tahmin etmeye dayalı bir yaklaşım izlemişlerdir. Amaçları normal davranışı öğrenerek normallikten sapmaları tahmin edebilmektir. Gemilerin farklı durumlar ve koşullar için normal davranışlarını keşfetmişlerdir. Tahmin edilen bir sonraki hareket ile gerçekleşen hareketlerini karşılaştırmışlardır. Beklenen davranış gerçekleşmedi ise bu anomali bir durum olabilir. Burada hareket, bir sonraki konum olarak düşünülmüştür.

Çevre alanlarını, askeri tesisleri, uluslar arası suları veya kötü hava alanlarına giren gemileri korumak için yasaklı bölgeler tanımlanabilir. Gemilerin yasaklı ya da izinsiz

bölgelere girişleri bir anomali gemi davranışı olarak incelenebilir. Burada ilgilenilen olay, bir geminin izni olmayan bir bölgeye girip girmediğidir. Bu duruma ek olarak, bir geminin gelecekte belirli bir zaman diliminde bir bölgeye girme olasılığını belirlemek de izinsiz bölgeye giriş tespiti için faydalı olabilir. Bu durum yetkililere, duruma müdahale etmek için zaman tanıyacaktır.

Denizde Can Güvenliği Uluslararası Sözleşmesi gereğince, gemilerin yerlerini AIS kullanarak düzenli aralıklarla iletmeleri gerekmektedir. AIS sinyallerinde meydana gelebilecek beklenmedik durumlar anomali gemi davranışı olarak incelenebilir. Örneğin, geminin rotasına göre beklenmedik bir alandan gelen AIS sinyali veya geminin AIS alıcı menzilinin dışına çıkması, anomali gemi davranışını işaret ediyor olabilir. Bununla birlikte, bir rotada uzun bir AIS sessizliği varsa, bu geminin AIS vericisinin kapalı olduğunu gösterebilir. Bir şekilde yasa dışı faaliyetlerde bulunmak için bilinçli olarak kapatıldığı düşünülebilir. Geminin gerçek AIS iletim sinyalleri yanlış yönlendirilmiş ya da bu AIS sinyali bilinçli olarak bozulmuş olabilir. Geminin yasa dışı faaliyetlere karıştığının bir göstergesi olabilir. Kowalska ve Peel (2012) çalışmasında AIS veri tabanından gemilere ait gerçek bir veri seti almış ve bazı gemileri yeni bir konuma taşıyarak anomali davranışı yapay olarak oluşturulmuştur. Ancak gemilerin hızı ve yönü etkilenmeden bırakılmış yanlızca konumları değiştirilmiştir. Yapay olarak oluşturulan konumlar tespit edilmeye çalışılmıştır. Gemilerin AIS sinyallerinin beklenmedik bir alandan gelmesi konusuna örnek bir çalışmadır.

Beklenmedik limana varış, bu tür bir anomali gemi davranışı için iki durum olabilir. Bunlardan ilki gemi ve liman tesislerine dayanmaktadır. Belirli tipteki bir gemi, gemiyi idare edecek tesisleri olmayan bir limana ulaşırsa, bu durum anomali kabul edilebilir. Tesisler ve gemi tipleri arasında eşleşme yapıldıktan sonra bu bilgiye dayalı bir tespit yapılabilir. İkinci durumda ise gemilerin liman ziyaret modellerine bakılır. Zamanla, gemiler diğer limanları değil, belirli limanları ziyaret etme eğilimindedir ve limanlar belirli bir sırada ziyaret edilir. Bu sıralamadan herhangi bir sapma, beklenmedik bir liman ziyareti anomali kabul edilebilir.

Çok yaklaşma veya çarpışma riski, iki geminin denizde olağan dışı şekilde birbirine yakın olması ve belirli bir hızın altında seyahat etmesi olarak tanımlanır. Bu davranışı sergileyen gemiler, malların veya insanların yasa dışı transferini gerçekleştiriyor olabilir. Bir gemiden alınan her ölçüm için, konumu, hızı ve istikameti belirli bir zaman dilimi içinde geçmişe ve geleceğe doğru bir yörünge oluşturmak için kullanılır. İzlenen tüm gemiler için yörüngeler üretilir ve bunlar, zamanın bir fonksiyonu olarak herhangi iki gemi arasındaki mesafeyi tahmin etmek için kullanılır. İki geminin tahmini yerleri ve zamanları en yakın yaklaşma noktası için kaydedilir. Böylelikle iki gemi tahmin edilenden daha yakın seyahat ediyorsa bu durum incelenebilir.

Yasa dışı davranış sergileyen gemiler, genellikle standart rotalardan sapma, beklenmeyen AIS faaliyeti, beklenmeyen limana varış, yakın yaklaşma ve izinsiz bölgeye giriş gibi davranışlar sergilemektedir. Gerçekleştirmek istedikleri faaliyete bağlı olarak niyetlerini gizlemeye çalışacaklardır (d'Afflisio, Braca, Millefiori, ve Willett, 2018).

Bu beş genel davranışın dışında deniz trafiğinde anomali olarak incelenen başka davranışlar da vardır. Handayani vd. (2013) yakıt ihmalinden ya da dümen kilitlenmesinden kaynaklanabilecek yanal kayma ve beklenmedik şekilde yavaşlama davranışı üzerine bir çalışma yapmışlardır. Bu davranışı, bir damar davranış anomalisi olarak ele almışlar ve destek vektör makinesi(support vector machine (SVM)) yöntemini kullanmışlardır. Pallotta, Vespe, ve Bryan (2013) çalışmalarında geminin yörüngesi üzerinde meydana gelebilecek bazı anomali davranışları incelemişlerdir. Traffic Route Extraction for Anomaly Detection (TREAD) adını verdikleri bir yöntem sunmaktadırlar. Bu yöntem, gemilerin iki liman arasında seyahat davranışlarından yörüngeler çıkarmakta ve yörüngeye göre normal davranış sınıflaması yaratmaktadır. Bir geminin, iki liman arasında beklenen yörüngeden sapma davranışını anomali olarak belirlemektedir. Ayrıca yörüngeye göre geminin hızının yüksek ya da düşük olması ve gemi burnunun dönmesi pruva (baş) anomali olarak incelenmektedir. Rong, Teixeira, ve Soares (2020) 'de bir rota boyunca gemilerin seyir davranışlarını incelemişlerdir. Belirli tipte gemilerin benzer hız ve

manevra hareketlerini sergiledikleri görülmüştür. Bu davranış kalıbından yola çıkarak beklenen rota davranışından çıkan gemiler anomali olarak incelenmiştir. Yörüngelerinde, beklenen davranış kalıplarını sergilemeyen gemilerde teknik bir problem olduğu ya da başka bir gemi tarafından tahrik edildiği düşünülebilir. Wang, Hu, ve Fan (2018) çalışmalarında iki liman arası rotalarda çalışmışlardır. Belirli yaklaşımlarla keşfettikleri iki liman arası en çok kullanılan rotalarda hız anomalisi araştırmışlardır. Bu rotalarda gemilerin hız profillerini çıkarmışlar ve beklenen hız davranışına uymayan gemileri anomali olarak incelemişlerdir. Doğan vd. (2021) Marmara Denizi transit gemi rotalarını ele almışlardır. Marmara Denizi, dünyanın en yoğun transit gemi trafiğine sahip bölgelerinden birisidir. Transit gemi rotaları, Uluslararası Denizcilik Örgütü tarafından belirlenir; kaptanlar ise hava, deniz ve diğer şartlar nedeniyle anlık olarak güzergah değiştirebilirler. Gemilerin bu hareketleri anomali gemi davranışı olarak algılanır. Veri madenciliği algoritmaları ile bu davranışı dikkate alan özgün bir yaklaşım uygulanmıştır.

BÖLÜM DÖRT
VERİ VE YÖNTEM

Bu çalışmada denetimsiz öğrenme yöntemlerinden biri olan kümeleme algoritmalarından ST-DBSCAN algoritması AIS verileri ile anomali belirlemede kullanılmıştır. Takip eden bölüm 4.1'de AIS sistemi tanıtılmış, bölüm 4.2'de ST-DBSCAN algoritması anlatılmıştır. Bölüm 4.3'de uygulamada kullanılan süreç açıklanmıştır.

4.1 AIS - Otomatik Tanımlama Sistemi

AIS, yakın mesafeler içinde deniz araçlarının takip edilmesini sağlayan bir araçtır. AIS Sistemi VHF deniz telsiz frekanslarını kullanır. Gemiden gemiye, gemiden deniz gözetim merkezlerine, gemiden AIS uydularına bilgi transferi gerçekleşir. Sistemi kullanan gemiler, bu sistem üzerinden çevrelerindeki gemiler hakkında oldukça detaylı bilgileri, kendi AIS ekranları üzerinde görebilirler. Çevrelerindeki deniz trafiği hakkında detaylı bilgiye sahip olurlar. Özellikle hassas ve yüksek riskli bölgelerde deniz trafiğinin deniz gözetim operatörleri tarafından izlenebilmesi ve yönlendirilmesi oldukça önemlidir. AIS cihazlarının kullanımı sayesinde deniz trafiğinin daha güvenli şekilde yönetilmesi sağlanır.

Uluslararası denizcilik işletmelerinde en büyük zorluklardan biri, özellikle gemi çarpışmalarından kaçınmak, gemilerin güvenliğini sağlamak olmuştur. Gemilerin boyut, tasarım ve hızlarındaki değişiklikler göz önüne alındığında, rotalarını değiştirmeleri ve çarpışmalardan kaynaklanan kazalardan kaçınmaları giderek daha önemli bir konu haline gelmiştir. Bu sebeplerle, gemilerin güvenliğini sağlamak amacıyla Uluslararası Denizcilik Örgütü (International Maritime Organization-IMO) himayesinde 1974 yılında birçok devlet tarafından Denizde Can Güvenliği (Safety Of Life At Sea-SOLAS) adlı uluslararası denizcilik anlaşması imzalanmıştır (Jha, 2020).

AIS, SOLAS tarafından, 300 gros tonun üzerindeki gemiler ve kargo, yolcu gemisi, tanker vb. gibi çoğu ticari gemi için zorunlu hale getirilmiştir. Bu zorunluluk 2002 yılında gerçekleşmiştir. Böylelikle denizcilik alanındaki mevcut veri miktarında

ciddi artış olmuştur. AIS ilk olarak denizde çarpışmadan kaçınmak için kullanılmıştır. Ancak zaman içerisinde deniz gözetimi ve anormallik tespiti için bir veri kaynağı haline gelmiştir (Obradović, Miličević, ve Žubrinić, 2014). AIS verileri çok uzun bir mesafeye iletilebilir ve alınabilir. AIS, deniz koşulları ve hava koşulları gibi dış faktörlerden daha az etkilenir.

Ayrıca arama kurtarma çalışmalarına destek sağlaması, açık denizde avcılık yapan balıkçı teknelerinin kontrolü, avlanmaya kapalı bölgelere giriş çıkışın ve denizlerdeki kaçakçılık ve yasa dışı göç gibi konuların kontrol altına alınabilmesi gibi avantajları vardır.

AIS verileri iki türe ayrılabilir. Bunlar, kinematik ve kinematik olmayan verilerdir. Kinematik, hareketli cisimlerdeki noktaların uzaydaki konumlarını, bu noktaların birim zamanda aldıkları yolu ve hızlarında birim zamanda ortaya çıkan değişmeleri (ivme) açıklamayı amaçlar. Bu nedenle kinematik tipteki AIS verileri gemilerin hareketleri ve varış noktaları ile ilgilidir. Kinematik AIS verileri ve ölçüm birimleri tablo 4.2'de yer almaktadır. Kinematik verilere gemi konumu, hız ve rota örnek verilebilir. Kinematik olmayan verilere ise gemi adı ve gemi MMSI bilgisi gibi statik bilgilerdir. Kinematik olmayan AIS verileri ve ölçüm birimleri tablo 4.3'de yer almaktadır.

Tablo 4.1 AIS veri iletim sıklığı

Gemi Hız Durumu	**Güncelleme Sıklığı**
Demirli Gemi	3 dakika
Gemi hızı 0-14 knot	12 saniye
Gemi hızı 0-14 knot ve rota değiştiriliyor	4 saniye
Gemi hızı 14-23 knot	6 saniye
Gemi hızı 14-23 knot ve rota değiştiriliyor	2 saniye
Gemi hızı >23 knot	3 saniye
Gemi hızı >23 knot ve rota değiştiriliyor	2 saniye

AIS yayını geminin o anki hızına göre farklı aralıklarla yapılır. Geminin kinematik bilgileri (hız, rota, zaman vb.) sabit duran bir gemide 3 dakikada bir yayınlanmasına karşın, hareket halindeki gemiler hızlarıyla doğru orantılı 2-10 saniye arasında sıklıklarla sinyal gönderirler. Kinematik olmayan bilgiler (çağrı kodu, gemi boyutları,

tahmini varış zamanı gibi.) 6 dakikada bir yayınlanır (Tablo 4.1) (Tu vd., 2017).

Tablo 4.2 AIS kinematik değişkenler

Kinematik Değişkenler	**Birim**
TIME	Zaman (y/m/d - h/m/s)
LATITUDE	Ondalık Derece
LONGITUDE	Ondalık Derece
COURSE OVER GROUND (COG)	Açı (0-359)
SPEED OVER GROUND (SOG)	Knot
HEADING	Açı (0-359)
NAVSTAT	Tam Sayı
DESTINATION	Karakter
ETA	Karakter

Time değişkeninde AIS'ten alınan sinyallerin tarih ve saat bilgileri tutulmaktadır. Latitude ve longitude değişkenleri enlem ve boylam bilgilerini barındırır. COG ve heading değişkenleri kuzey noktası referans alınarak saat yönünde ölçülen açısal değerlerdir. Heading değeri geminin gerçekte baktığı yönü gösterir. Heading değeri 90 derece ise gemi yönü doğu yönündedir. COG değeri ise geminin rüzgar, gelgit ve akıntıların etkilerine bağlı olarak gerçekte gittiği yönü göstermektedir. Gemiye kuşbakışı bakıldığında, geminin çevresel şartlar sonucu gerçekte gittiği yöndür. SOG değişkeni geminin hız bilgisini barındırır. Navstat değişkeni seyir durumunu ifade eden tam sayı değerleri almaktadır. Bu tam sayı değerleri gemi demirli, motor kullanılıyor gibi durumları ifade eder. Destination değişkeni ise AIS sinyallerine göre bir sonraki varış noktası bilgisini barındırır. ETA değişkeni de tahmini varış zamanıdır.

Tablo 4.3 AIS kinematik olmayan değişkenler

Kinematik Olmayan Değişkenler	Birim
MMSI	Tam Sayı
IMO	Karakter - Tam Sayı
VESSEL NAME	Karakter
CALLSIGN	Karakter - Tam Sayı
DIMENSIONS	Metre
A	Metre
B	Metre
C	Metre
D	Metre
VESSEL TYPE	Tam Sayı

Maritime Mobile Services Identity (MMSI), mobil hizmet kimliği olarak söylenebilir. Kendine özgü dokuz haneli bir sayıdır. Bu numara telefon numarası işlevi görür. Gemi bilgilerine bu numara ile ulaşılabilir. IMO ve ardından SOLAS tarafından verilen yedi haneli benzersiz rakamlardan oluşmaktadır. Name değişkeni, gemi adı bilgisini tutar. Callsign değişkeni ise sayı ve harflerden oluşan bir lisans numarasıdır. Otomobil ruhsat sicil no gibi düşünülebilir. Bu kod, geminin tescil edildiği ülkedenin ilgili makamları tarafından verilir. Dimensions, A, B, C ve D değişkenleri geminin boyutları ile ilgilidir. Vessel type değişkeni 0-99 arasında bir tam sayı değeri alır. AIS özelinde her sayının karşılık geldiği bir gemi tipi bilgisi saklar. Tablo 4.4'de gemi tipi listesi yer almaktadır.

Tablo 4.4 AIS gemi tipi numaraları

AIS KODU	VESSEL TYPE
20-29	Deniz Uçağı (Wing in ground)
30	Balıkçı Teknesi (Fishing)
31-32	Acil Durum Çekici Gemisi (Towing)
33-34	Dalış Destek Gemisi (Underwater/Diving ops)
35	Askeri Operasyon Gemisi (Military ops)
36	Yelken (Salling)
37-49	Tekne/ Feribot (Craft/High speed craft)
50	Klavuz Gemi (Pilot vessel)
51	Arama Kurtarma Gemisi (Search and Rescue vessel)
52	Römorkör (Tug)
53	Destek Gemisi (Ship's tender)
54	Kirlilik Önleyici Ekipman (Anti-pollution equipment)
55	Sahil Güvenlik (Law Enforcement)
56-57	Yerel Gemi (Local Vessel)
58	Medikal Taşıma (Medical Transport)
59	Savaş Dışı Gemi (Noncombatant ship)
60-69	Yolcu Gemisi (Passenger)
70-79	Kargo Gemisi (Cargo)
80-89	Tanker Gemisi (Tanker)
90-99	Diğer
0	Varsayılan Değer (NA)
1-19	İleride Kullanmak için Saklanmış

4.2 ST-DBSCAN

Kümeleme analizi, nesnelerin ya da noktaların benzer özelliklerine göre kümeler halinde gruplandırıldığı bir yöntemdir. Aynı kümedeki nesneler birbirine daha çok benzerken farklı kümelerdeki nesneler birbirine daha az benzemektedir (Kassambara, 2017). Veri setindeki bir dizi noktanın bir küme oluşturup oluşturmayacağına karar verilirken, bir uzaklık ölçüsü metriği (Öklit, Manhattan ve Minkowski gibi) aracılığıyla hesaplanan uzaklık değerleri kullanılır. Veri setindeki benzer gözlemleri gruplandırmak için kullanılan bir çok kümeleme algoritması vardır. Kümeleme algoritmaları, küme oluşturma yöntemlerine göre; bölünmeye dayalı, hiyerarşik, yoğunluğa ve modele dayalı olarak sınıflandırılabilir. Kümeleme algoritmaları, sapan değer tespiti, müşteri segmentasyonu, örüntü tanıma ve görüntü analizi dahil olmak üzere birçok alanda kullanılmaktadır. Kümeleme algoritmaları genellikle

zamansal-mekansal olmayan verileri kümelemek üzerine kurgulanmamıştır. Örneğin yoğunluk tabanlı DBSCAN algoritması, mesafe ölçümü için yalnızca eps adında bir uzaklık parametresi kullanır. Ancak, günümüzde çeşitli teknolojiler tarafından giderek daha fazla zamansal-mekansal veri üretilmektedir. Örneğin, GPS ve mobil ağlar gibi. Bu nedenle zamansal-mekansal kümeleme giderek daha popüler hale gelen bir veri madenciliği alanıdır (Ansari, Ahmad, Khan, Bhushan, vd., 2020). Zamansal-mekansal veriler, mekansal verinin zamansal dilimler halinde saklanan halidir. Zamansal-mekansal veriler yapı olarak iki boyutlu olduğundan dolayı benzerlikleri ölçmek için eps1 ve eps2 olmak üzere iki adet uzaklık parametresi gereklidir. Zamansal-mekansal kümeleme, verilerin hem zamansal hem mekansal komşularını dikkate almaktadır. Bu kümeleme algoritmaları coğrafi bilgi sistemleri (CBS), hava durumu öngörüleri, medikal görüntüleme gibi birçok alanda kullanılmaktadır (Birant ve Kut, 2007). AIS verileri zamana bağlı konum verileri içermektedir. Bu nedenle geleneksel kümeleme algoritmaları yerine zamansal-mekansal kümeleme algoritmalarının kullanımı uygun olacaktır.

ST-DBSCAN, Birant ve Kut (2007) tarafından önerilmiş yoğunluk tabanlı bir kümeleme algoritmasıdır. ST-DBSCAN, Density-based spatial clustering of applications with noise (DBSCAN), kümeleme algoritmasının zamansal-mekansal veriler için geliştirilmiş versiyonudur. DBSCAN, algoritması yoğunluk tabanlı popüler bir mekansal kümeleme algoritmasıdır (Ester, Kriegel, Sander, Xu, vd., 1996). ST-DBSCAN algoritması zamansal-mekansal değişkenlerin benzerliklerini tanımlamak için DBSCAN algoritmasından farklı olarak, eps1 ve eps2 adında 2 adet uzaklık parametresi kullanmaktadır. ST-DBSCAN algoritmasının parametleri eps1, eps2, $\Delta_{\in}$ ve MinPts'dir. Eps1, mekansal değişkenler (enlem ve boylam) için mesafe parametresidir. Eps2, mekansal olmayan değişkenler için mesafe parametresidir. Eps1 ve eps2 için Öklid, Manhattan veya Minkowski uzaklığı gibi bir uzaklık ölçüsü kullanılabilir. $\Delta_{\in}$ birbirine komşu konumlardaki mekansal olmayan gözlemlerin, birbirleri ile uzaklık farkının çok düşük olmasından kaynaklanan iç içe geçmiş kümeleri ayırt etmek için kullanılmaktadır. MinPts parametresi ise bir noktanın eps1 ve eps2 mesafesi içindeki minimum nokta sayısıdır (Birant ve Kut, 2007).

ST-DBSCAN algoritmasının yapısını incelemeden önce aşağıda yer alan tanımlara değinmek faydalı olacaktır.

Tanım 1 Komşuluk (Neighborhood): P ve q noktaları için komuşuluk, bir uzaklık metriği tarafından belirlenen dist(p,q) ile gösterilir.

Tanım 2 Eps-Komşuluk (Eps-Neighborhood): Bir p noktasının eps komşuluğu şu şekilde tanımlanır.

$$N_Eps(p) = (q \in D | dist_(p,q) <= Eps) \tag{4.1}$$

Tanım 3 Çekirdek Gözlem (Core Object): Çekirdek gözlem, etrafında eps değeri şartına uyan MinPts sayısı kadar komşu nokta bulunan gözlemdir.

Tanım 4 Sınır Gözlemi (Border Object): Bir p gözlemi, bir çekirdek gözlem değilse, ancak başka bir çekirdek gözlemden yoğunluğa erişilebilir ise bir sınır gözlemdir.

Tanım 5 Doğrudan yoğunluğa erişilebilirlik (Directly Density-Reachable) : Aşağıda yer alan koşulların sağlanması halinde p gözlemi q gözlemi için doğrudan yoğunluğa erişilebilirdir.

$$p \in N_Eps(q) \quad ve \quad N_Eps(q)| >= MinPts \tag{4.2}$$

Tanım 6 Yoğunluğa erişilebilirlik (Density-Reachable) : Eps ve MinPts parametre değerleri göz önünde bulundurularak, p_1, ... , p_n gözlemler kümesi varsa, p_1 = q ve p_n = q ise $p_(i+1)$, p_i den doğrudan yoğunluğa erişebilirdir. Böylece, p noktası q noktası üzerinden yoğunluğa erişilebilirdir.

Tanım 7 Gürültü (Noise) : Eps ve MinPts değerleri göz önünde bulundurularak C_1, ... ,C_k veri setindeki kümeler olsun. Herhangi bir gözlem bir C_i, kümesinde yer almıyorsa, bu gözlem gürültü olarak adlandırılır.

Algorithm 1 ST-DBSCAN (D, Eps1, Eps2, MinPts, Δ_ϵ)

```
// Inputs :
    // D=(o1, o2, . . . , on) Set of objects
    // Eps1 : Maximum geographical coordinate (spatial) distance value.
    // Eps2 : Maximum non-spatial distance value.
    // MinPts : Minimum number of points within Eps1 and Eps2 distance.
    // Δε : Threshold value to be included in a cluster
// Output :
    // C = (C1, C2, . . . , Ck) set of clusters

Cluster_Label = 0

for i=1 to n do
    if oi is not in a cluster then
        X = Retrieve_Neighbors (oi, Eps1, Eps2 )

        if |X| < MinPts then
            Mark oi as noise
        else                                          ▹ construct a new cluster
            Cluster_Label = Cluster_Label + 1

            for j=1 to |X| do
                Mark all objects in X with current Cluster_Label
            end for

            Push (all objects in X)

            while not IsEmpty( ) do
                CurrentObj = Pop( )
                Y = Retrieve_Neighbors (CurrentObj, Eps1, Eps2)

                if |Y| >= MinPts then
                    for all Objects o in Y do
                        if (o is not marked as noise or it is not in a cluster)
                            |ClusterAvg( )-o.Value| <= Δε then
                            Mark o with current Cluster_Label
                            Push (o)
                        end if
                    end for
                end if
            end while
        end if
    end if
end for
```

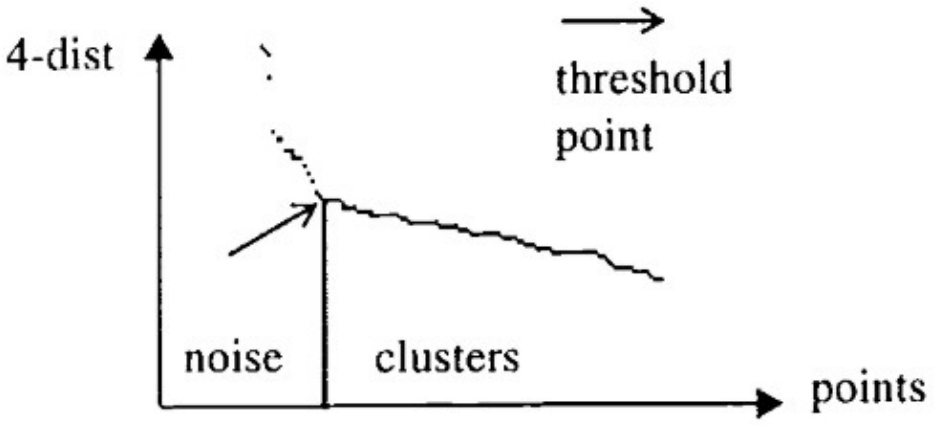
Şekil 4.1 ST-DBSCAN algoritma yapısı

Algoritma, veri tabanındaki ilk gözlem noktası ile başlar. Eps1 ve eps2 mesafelerine göre gözlem noktasından yoğunlukla erişilebilen tüm noktaları alır. Bu nokta bir çekirdek gözlem ise bir küme oluşur. Eğer bu gözlem noktası bir sınır gözlem ise buradan ulaşılabilen hiçbir gözlem noktası yoktur. Sonuç olarak algoritma veri tabanının bir sonraki gözlem noktasını ziyaret eder. İşlem tüm gözlem noktaları işlenene kadar devam eder.

Algoritmayı detaylı incelersek, şekil 4.1'de göründüğü gibi süreç D veri setindeki ilk gözlem $0_(i)$ ile başlar. Bu gözlemde tüm aşamalar gerçekleştirilir ve ardından sonraki gözleme geçilir. Sonraki gözlem herhangi bir kümeye dahil olamıyorsa, Retrieve_Neighbors fonksiyonu çağrılır. Retrieve_Neighbors (Object, Eps1, Eps2) fonksiyonu, seçili gözleme eps1 ve eps2 mesafelerinden daha kısa bir mesafeye sahip gözlemleri döndürür. Başka bir ifade ile, seçilen gözlemden eps1, eps2 ve MinPts değerlerine göre yoğunluğa erişilebilen tüm gözlemleri alır. Alınan gözlemler Eps-neighborhood nesnesini oluşturur. Retrieve_Neighbors (Object, Eps1, Eps2), Retrieve_Neighbors (Object, Eps1) ve Retrieve_Neighbors (Object, Eps2) kesişimine eşittir. Eps-neighborhood içindeki gözlem sayısı MinPts değeri ile karşılaştırılır. Eğer MinPts değerinden daha küçükse, seçilen gözlem gürültü olarak tanımlanır. Seçilen gözleminin kümelenmek için yeterli komşusu yoktur. Gürültü olarak tanımlanan gözlemler, doğrudan yoğunluğa erişilebilir değil ise, ancak veri setindeki başka bir gözlemden yoğunluğa erişilebilir ise daha sonra başka bir kümeye dahil olabilir. Bu genellikle kümelerin sınır gözlemleri için gerçekleşir. Ardından gelen gözlem eps1 ve eps2 değerleri içerisinde komşu gözlemlere sahipse ve seçilen gözlem bir çekirdek gözlem ise, yeni bir küme oluşturulur. Daha sonra bu çekirdek gözlemin doğrudan yoğunluğa erişilebilir komşuları da bu kümeye dahil edilir. Devamında algoritma, çekirdek gözlem ile yoğunluğa erişilebilir olan gözlemleri yinelemeli olarak toplar. Toplanan yığın, yoğunluğa erişilebilir gözlemleri, doğrudan yoğunluğa erişilebilir gözlemlerden bulmak için gereklidir. Eğer gözlem gürültü olarak işaretlenmediyse ve bir kümeye ait değilse ayrıca küme ortalaması ile yeni gelecek olan gözlem arasındaki fark Δ_ϵ eşik değerinden daha küçükse, gözlem mevcut kümeye yerleştirilir. Seçilen gözlemin işlem süreci tamamlandıktan sonra algoritma bir sonraki gözlemi işleme sokar ve tüm gözlemler bu sürece dahil edilene kadar algoritma sürdürülür (Birant ve Kut, 2007).

ST-DBSCAN algoritması eps1, eps2 ve MinPts olmak üzere üç parametre ile çalışmaktadır. Parametre değerlerinin belirlenmesinde bir çok durumda başarılı olan basit bir sezgisel yaklaşım önerilmektedir (Ester vd., 1996). MinPts değerinin belirlemek için ln(n) kullanılabilir. Burada n veri tabanının boyutudur. Eps

parametresi bir noktanın komşularını seçeceği en uzak mesafedir. Bu nedenle, sezgisel olarak, bir noktanın kaç tane komşu keşfedeceğine karar verecektir. Eps seçiminde her nesnenin k en yakın komşularına olan uzaklığı hesaplanır. K-nn için k parametresi MinPts'e eşittir. Daha sonra bu k en yakın komşu değerleri, azalan sırada sıralanmalıdır. Ardından bu sıralanan değerlerin grafiğinden eps değeri belirlenmelidir. Bu grafiği k-mesafe grafiği olarak adlandırabiliriz. K-mesafe grafiğinin ilk kırılım noktası eps değeri olarak alınabilir. Kırılımdan daha yüksek k-mesafe değerine sahip tüm noktalar (eşiğin solunda) gürültü olarak kabul edilir, diğer tüm noktalar (eşiğin sağında) bir kümeye atanır (Ester vd., 1996). Buradaki anlatım Şekil 4.2'de yer almaktadır.



Şekil 4.2 Eps1 parametre seçimi K-NN grafiği (Ester vd., 1996)

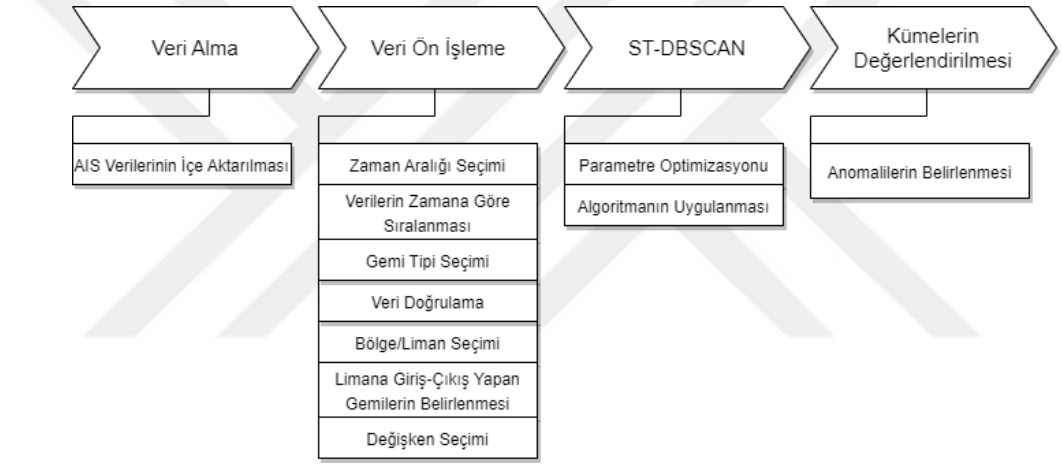
Eps parametresi belirlenirken k en yakın komşu grafiğinde bir dirsek noktası aranmaktadır. Bu nokta ve yakını parametre değerleri alınarak algoritma çalıştırılabilir. Dirsek noktasından -pozitif yönde- büyük seçilen değerlerde küme sayılarının artması beklenmektedir. Optimum küme sayısı için farklı eps1 değerleri seçilerek denemeler yapılabilir.

Eps2 parametresini belirlenirken önceki çalışmalar gözden geçirilmiştir. Eps2 mekansal olmayan parametre yani zaman değişkeni için uzaklık parametresidir. İncelenen çalışmalara göre olayın gerçekleştiği zaman aralığı parametre değeri olarak seçilebilir.

4.3 AIS Verilerini Kullanarak ST-DBSCAN Algoritması ile Anomali Belirleme

Bu çalışmada, zamansal mekansal verilerde kümeleme amacıyla önerilen ST-DBSCAN (Birant ve Kut, 2007) algoritması deniz trafiğinde AIS verileri ile anomali belirleme amacıyla kullanılmıştır. Bildiğimiz kadarıyla, ST-DBSCAN algoritması AIS verileri ile anomali belirleme amacıyla daha önce kullanılmamıştır.

ST-DBSCAN ile konum verilerinden elde edilen rutinler, yani normal davranış kalıpları anomali belirlemede kullanılabilir. Uygulamada, tüm kümelerden daha uzak mesafede bulunan nokta veya noktalar (normal davranış kalıpları dışında kalanlar) bir anomali davranışı olarak kabul edilmiştir.



Şekil 4.3 ST-DBSCAN ile anomali belirleme süreci

Şekil 4.3'de ST-DBSCAN ile anomali belirleme sürecinin adımları yer almaktadır. Süreç AIS verilerinin elde edilmesi ile başlamaktadır. Elde edilen veriler üzerinde veri ön işleme adımları uygulanmıştır. Bu adımlar, veri modellenmeden önce veri seti üzerinde yapılan bir takım düzeltme, eksik veriyi tamamlama, tekrarlanan verileri kaldırma, dönüştürme, temizleme, normalleştirme, boyut indirgeme vb. işlemlerdir. Veriler çalışmanın amacına uygun olarak işlenir ve veri seti üzerinde bilgi keşfi yapılmış olur. Veri seti üzerinde uygulanan veri ön işleme adımları ilerleyen paragraflarda anlatılmaktadır.

Zaman Aralığı Seçimi: Veriler çeşitli zaman aralıkları ile günlük, haftalık, aylık ya da yıllık olarak elde edilmiş olabilir. Veri seti çalışılmak istenen zaman aralığına

uygun olarak elde edilmeli ya da hazırlanmalıdır.

Verilerin Zamana Göre Sıralanması: Veriler tarih ve saate göre sıralanmalı ve tarih saat formatına dikkat edilmelidir. Analizlerin ilerleyen adımlarında tarih ve saat formatı seçimi önemli olabilir.

Gemi Tipi Seçimi: Deniz taşıtları farklı boy, ağırlık ve manevra kabiliyetlerine sahip olabilirler. Bir balıkçı teknesi ile bir kargo gemisinin seyahat davranışları farklı olabilir. Benzer özelliklere sahip deniz taşıtlarının benzer davranışları sergileyecekleri düşünülürse fiziki özellikleri benzer gemiler bir arada değerlendirilmelidir.

Veri doğrulama: Değişkenlerin değerleri birimlerine göre mantıksal olarak sınanmalı ve gerekli görülen düzeltmeler yapılmalıdır. (Örneğin negatif bir açı değeri gibi.) Boş gözlem değerleri ya da değişkenler için gerekli işlemler yapılmalıdır. Gerekli ise veri standartlaştırma gibi işlemler de bu adımda yapılabilir.

Bölge Liman Seçimi: Büyük deniz alanlarına odaklanmak yerel tehlikelerin göz ardı edilmesine sebep olabilir. Bu nedenle belirli bir liman bölgesi seçilmeli ve bu bölge için kritik trafik akışlarının olabileceği alanlara odaklanılmalıdır.

Limana Giriş-Çıkış Yapan Gemilerin Belirlenmesi: Gemilerin limana giriş ve limandan ayrılış hareketleri farklılık gösterebileceğinden veriler tek yönde filtrelenmelidir.

Değişken Seçimi: Algoritmaya uygun olan girdi değişkenler bu adımda seçilmelidir.

Şekil 4.3’ de yer alan veri alma ve veri ön işleme adımlarından sonra veri, belirlenen liman için kritik trafik akışlarının olabileceği bölgenin verilerinden oluşacaktır. Bu veriler yanlızca limana giriş yapan gemilerin bilgilerini içerecektir. Veriler gerçek hayatta var olan durumu yansıtmaktadır. Bu haliyle tüm veri örnekleri normal kabul edilecektir. Algoritmanın performansını ölçebilmek için veriler üzerinde anomali senaryoları yaratılacaktır. Bu durumlar ise anomali kabul edilecektir. Yaratılan anomali durumlarını algoritmanın tespit etmesi beklenmektedir.

Daha sonraki adım da ST-DBSCAN algoritmasının parametreleri ayarlanır. ST-DBSCAN algoritması eps1, eps2 ve MinPts olmak üzere üç parametre ile çalışmaktadır. Parametre değerlerinin nasıl belirlenebileceği bölüm 4.2'de açıklanmıştır. Optimum parametre seçimi için farklı parametre değerleri ile algoritma çalıştırılacaktır. Farklı parametrelerin anomali belirleme açısından performanslarına göre optimum parametre seçimi yapılacaktır. Her bir farklı parametre için yanılgı matrisi oluşturulacaktır (Tablo 4.5). Yanılgı matrisi veri setinde var olan durum ile ST-DBSCAN algoritmasının tespit ettiği durumları karşılaştırmak için kullanılacaktır. Algoritmanın uygulanmasından sonra oluşacak kümeler ve gürültü değerler anomali davranışı açısından değerlendirilecektir.

Tablo 4.5 Yanılgı matrisi

		GERÇEK		
		Anomali	**Normal**	**Toplam**
TAHMİN	**Anomali**	*Doğru Pozitif (DP)*	*Yanlış Negatif (YN)*	*DP+YN*
	Normal	*Yanlış Pozitif (YP)*	*Doğru Negatif (DN)*	YP+DN
	Toplam	*DP+YP*	*YN+DN*	

- Doğru Pozitif : Modelin, anomali olarak etiketlenmiş gözlemlerden kaçını anomali olarak belirlediğinin sayısıdır.
- Doğru Negatif : Modelin, normal olarak etiketlenmiş gözlemlerden kaçını normal olarak belirlediğinin sayısıdır.
- Yanlış Pozitif : Modelin, normal olarak etiketlenmiş gözlemlerden kaçını anomali olarak belirlediğinin sayısıdır.
- Yanlış Negatif : Modelin, anomali olarak etiketlenmiş gözlemlerden kaçını normal olarak belirlediğinin sayısıdır.

Doğru Pozitif Oranı/Duyarlılık/Hassasiyet (True Positive Rate – Recall) : Doğru olarak tahmin edilen anomali sayısının gerçekte anomali gözlem sayısına oranıdır. Modelin anomalileri belirlemek konusundaki etkinliği denilebilir.

$$DP0 = \frac{DP}{(DP + YN)} \tag{4.3}$$

Yanlış Pozitif Oranı (False Positive Rate): Gerçekte normal olarak etiketlenmiş ancak anomali diye tahmin edilen gözlemlerin gerçekte normal olan gözlemlere oranıdır.

$$YPO = \frac{YP}{(YP + DN)} \tag{4.4}$$

Doğru Negatif/Seçicilik/Özgüllük (True Negative Rate – Specifity) : Normal olarak etiketlenmiş gözlemlerin normal olarak tahmin edilmesinin oranıdır.

$$DNO = \frac{DN}{(DN + YP)} \tag{4.5}$$

F1 Puanı : Bu değer kesinlik (precision) ve hassasiyet (recall) değerlerinin harmonik ortalamasını göstermektedir.

Hassasiyet (DPO-Recall): Doğru olarak tahmin edilen anomali sayısının gerçekte anomali gözlem sayısına oranıdır.
Kesinlik (Precision) : Gerçekte anomali olan gözlemlerin anomali olarak tahmin edilmesinin oranıdır.

Algoritmanın performans ölçüm kriteri olarak F1 puanı kullanılacaktır.

$$Hassasiyet = DP0 = \frac{DP}{(DP + YN)} \tag{4.6}$$

$$Kesinlik = \frac{DP}{(DP + YP)} \tag{4.7}$$

$$F1 = 2 * \frac{Kesinlik * Hassasiyet}{(Kesinlik + Hassasiyet)} \tag{4.8}$$

BÖLÜM BEŞ
UYGULAMA

Uygulamada R (4.0.4) istatistiksel programlama dili ve dbscan, leaflet, dplyr, readxl, lubridate, chron, caret ve factoextra kütüphaneleri kullanılmıştır. R istatistiksel programlama dili ile AIS verileri görselleştirilmiş ve algoritmanın uygulaması yapılmıştır. Dinamik ve statik zengin içerikli grafikler oluşturulmuştur. ST-DBSCAN algoritması için https://github.com/CKerouanton/ST-DBSCAN sayfasında paylaşılan kodlardan yararlanılmıştır. Uygulamada kullanılan veriler, veri ön işleme adımları, AIS veri görselleştirme kodları ve ST-DBSCAN ile anomali belirleme R kodları https://github.com/sblalkn/SaDeuTez ve https://github.com/eyildiztepe/AIS_Project adreslerinde paylaşılmıştır.

5.1 Veri

Uygulamada kullanılan veri, https://marinecadastre.gov/ais/ adresinden elde edilen halka açık, ABD limanlarına ait AIS verilerdir. İndirilen veriler yirmi dört saatlik dilimlerde tutulan 2020/01/01 ve 2020/01/15 tarihlerine ait günlük verilerdir. Tablo 5.1'de verinin ilk beş satırı yer almaktadır. Veriye ait değişkenlerin açıklamaları bölüm 4.1'de verilmiştir. Ham veri, kinematik ve kinematik olmayan 17 değişken içermektedir. Veri içerisinde farklı kategoriden deniz taşıtı verileri bulunmaktadır. Bu deniz taşıtları, kargo, tanker, yolcu gemisi, balıkçı gemileri gibi farklı türlerde deniz taşıtlarıdır. Uygulamada sadece kargo ve tanker gemilerine ait veriler kullanılmıştır. Ham veri örneklem büyüklüğü, 17 değişkene ait yaklaşık 18 milyon satırdır. Elde edilen büyük veri, veri ön işleme adımlarından geçirilerek uygulamanın amacına uygun hale geitirilmiştir. Uygulama, ABD Güney Carolina eyaletinde bulunan Charleston Limanı ve Savannah Limanı verileri ile sınırlandırılmıştır. Seçilen limanlar, bölge için önemli ve büyük konteyner limanlarıdır.

Tablo 5.1 Ham veri ön izlemesi (ilk beş satır)

MMSI	BaseDateTime	LAT	LON	SOG	COG
477628100	2020-01-01 00:00:01	36.80096	-75.22302	12.8	231.3
538002845	2020-01-01 00:00:10	26.08420	79.48273-	13.2	189.0
477214900	2020-01-01 00:00:06	48.13155	-123.43617	0.1	157.5
373746000	2020-01-01 00:00:39	29.87873	-89.94732	0.0	132.5
538004188	2020-01-01 00:00:40	37.75965	-122.34710	0.0	43.0

Heading	VesselName	IMO	CallSign	VesselType
283	NINGO SEAL	IMO9579066	VRJD3	70
187	DARDANEL	IMO9339985	V7ME9	80
68	PTI HERCULES	IMO9358307	VRBY7	80
121	COMMANDER	IMO9610652	3EWU5	70
153	NAVE POLARIS	IMO9457749	V7VQ6	80

Status	Length	Width	Draft	Transciever Class
0	225	32	14.3	B
0	245	42	15.0	B
0	183	32	13.2	B
1	186	27	11.0	A
1	170	26	10.1	A

5.2 Veri Ön İşleme

Veri ön işleme adımında şekil 4.3'de açıklanan adımlar uygulanmıştır. Veri, kaynağından günlük olarak elde edilmiş ancak iki haftalık trafik akışlarını kapsayacak şekilde birleştirilmiştir. Zaman aralığı olarak 2020-01-01 / 2020-01-15 aralığı seçilmiştir. Tüm veriler zamana göre artan sırada sıralanmıştır.

Veri içerisinde başlangıçta kargo, tanker, yolcu gemisi, balıkçı gemileri gibi çeşitli deniz taşıtlarına ait AIS verileri bulunmaktadır. Ancak çalışmamızda gemi tipi olarak tanker ve kargo gemilerine ait AIS verileri kullanılmıştır. Bu seçim yapılırken gemilerin boyutları, hızlanmaları, manevra kabiliyetleri gibi farklılıklar göz önünde tutulmuştur. Birbirine benzer özelliklerde ve boyutlarda deniz taşıtları seçilmiştir. Bu

filtreleme, "Vessel Type" değişkeninin değerlerine bakılarak yapılmıştır. Veri içerisinde, ABD limanlarına ait kargo ve tankerlerin AIS verileri bırakılmıştır.

Veri doğrulama adımında ise öncelikli olarak, değişkenler birimlerine göre mantıksal olarak sınanmıştır. Hız ve açı gibi değerlerin özet istatistiklerine bakılmıştır. COG değişkeninin negatif açı değerleri aldığı görülmüş ve bu değişken üzerinde düzeltme yapılmıştır. Tüm negatif açı değerleri 360 derece ile toplanmıştır. Daha sonra heading ve COG değerlerinin tutarlılığı kontrol edilmiştir. Bölüm 4.1'de değişkenler ve birimleri hakkında detaylı açıklama yapılmıştır. Tüm değişkenlerin birimleri (saat ve tarih formatı gibi) kontrol edilmiştir. Boş değerli gözlemler veriden kaldırılmıştır.

Anomali gemi davranışlarını tespit etmek üzere belirli liman bölgeleri seçilmiştir. Belirli bir limana ait verileri seçmek için, liman enlem ve boylam değerleri dikkate alınmıştır. Belirlenen limanın trafik akışı görselleştirilmiş, gemilerin bu bölgede sergiledikleri seyir davranışları incelenmiştir. Liman trafik akışının görselleştirilmesi, aynı zamanda liman için kritik olabilecek alanlarıda işaret etmektedir. İlk aşamada görsel yöntemlere dayanarak kritik bir bölge belirlenmiştir. Anomali gemi davranışı bu kritik bölge için incelenecektir. Belirlenen kritik bölge liman için önemli ve gemilerin genellikle ortak davranışları sergilediği bir alan olmalıdır. Gemilerin limana giriş ve çıkış davranışları farklı olabilir. Bu nedenle yanlızca limana giriş yapan gemiler seçilmiştir. Bu seçim COG ve heading açıları referans alınarak gerçekleştirilmiştir. Gemilerin yön bilgisi, tek bir hat doğrultusunda hareket ettikleri ve doğrusal bir bölge seçilerek yapılmıştır. Bu bölge öyle bir bölge olmalıdır ki buradan geçen her gemi o limana uğramalıdır. Böylelikle gemilerin limana doğru ya da limandan ayrılış yönünde hareket ettikleri doğru olarak anlaşılabilir. Nihayetinde anomali tespiti yapılacak veri, belirlenen liman için kritik bölgede limana doğru hareket eden gemilerinin verilerinden oluşacaktır.

Amerika Güney Karolina eyaletinin en işlek limanlarından biri olan Charleston Limanı, ayrıca bölgenin önemli limanlarından Savannah Limanı üzerinde uygulama yapılmıştır. Ancak detaylı uygulama anlatımı ve görseller Charleston Limanı

üzerinden gösterilecektir. Bölüm 5.3'de görsellere ve detaylı anlatımlara yer verilmiştir.

Veri ön işleme, veri doğrulama adımı nihayetinde, veri içerisinde ABD limanlarına ait kargo ve tanker gemilerinin gözlem sayılarının dağılımı tablo 5.2'da belirtilmiştir.

Tablo 5.2 Verilerin gemi tipine göre dağılımı

Gemi Tipi	**Gemi Sayısı**	**Gözlem Sayısı**
Tanker	871	1423680
Kargo	1783	2888233
Toplam	2654	4311913

5.3 AIS Veri Görselleştirme

Deniz trafiği analizinde kullanılan yaklaşımlar incelendiğinde, bir çok yaklaşımda AIS verilerinin görselleştirilmesine ihtiyaç duyulduğu görülmüştür. AIS verileri, anomali gemi davranışı tespiti için farklı yöntemlerle birçok çalışmada kullanılmıştır. Ancak gemi yollarında gemiler arası kesişme gibi durumların tespitinde görselleştirme faaliyetlerinin daha da önem kazandığı görülmüştür. Pereira vd. (2019) tarafından veri işleme ve görselleştirmeye dayalı, deniz trafiğinde gemiler arası kesişimleri analiz eden bir yaklaşım önerilmektedir. Sonuçlar, birkaç kesişimin yalnızca önerilen görselleştirme yaklaşımları ile fark edilebilir olduğunu göstermektedir. R programlama dili ile pek çok adımda AIS verileri görselleştirilmiştir. Elde edilen grafikler ile deniz trafiğindeki olağan dışı davranışların nasıl belirlenebileceği tartışılacaktır. Veri görselleştirmede genellikle R leaflet paketinden yararlanılmıştır. Bu paket ile etkileşimli haritalar, çeşitli özelliklerin dahil edilebildiği görseller yaratılmıştır. Gemilerin rotaları görselleştirilirken her bir gemi farklı bir renk ile temsil edilebilir. Her bir geminin rotasını temsil edecek sıralı noktalar, farklı çaplarda çizdirilebilir. Örneğin, gemilerin anlık hızlarına göre noktaların çapları orantılı olarak çizdirilebilir. Görseller üzerinde pek çok uyarlama yapmak mümkündür.

5.4 Charleston Limanı

Charleston ve Savannah Limanları üzerinde uygulamalar yapılmıştır. Ancak Charleston Limanı üzerinden uygulama adımları detaylı olarak anlatılmıştır. Veri görselleştirme kodları grafiklerin altlarında yer almaktadır.

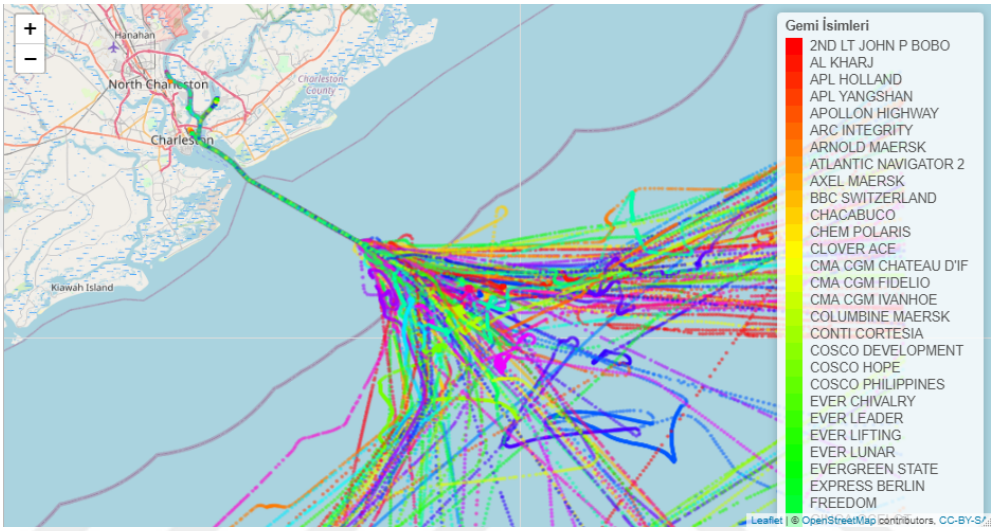
5.4.1 Rota Görselleştirme

Şekil 5.1'de Charleston Limanı iki haftalık kargo ve tanker gemilerine ait trafik akışı görülmektedir. Grafik elde edilmeden önce veriler Cahrleston Limanı enlem ve boylam (32.79487° / -79.914565°) değerleri dikkate alınarak filtrelenmiştir. Filtreleme işlemi LAT ve LON değişkenleri üzerinden bölgenin trafik akışını kapsayacak bir aralık oluşturularak yapılmıştır. Her renk farklı bir gemiyi temsil etmektedir. Şekil 5.1'de 140 adet gemiye ait veri bulunmaktadır. Liman trafik akışlarını görselleştirmek, gemilerin limana giriş-çıkışlarında sergiledikleri davranışlar hakkında fikir vermektedir. Bu davranışlardan yola çıkarak liman hakkında pek çok bilgi elde edilebilir. Limanın kritik bölgeleri, liman trafik yoğunluğu, limanın coğrafi özellikleri gibi. Bu aşamada beklenen, olağan deniz taşıtı davranışı öğrenilir. Bu davranışların modellenmesi sayesinde, bu davranışa uymayan durumlar anomali gemi davranışı olarak adlandırılabilir. Olağan gemi davranışlarının öğrenilmesi, liman için anomali gemi davranışının tanımlanması konusunda önemli bir adımdır.

Charleston Limanı trafik akışı incelendiğinde, kargo ve tanker gemilerinin belirli bir noktadan sonra aynı davranışı sergiledikleri görülmektedir. Gemiler belirgin bir yoldan limana giriş ve çıkış yapmaktadırlar. Liman incelendiğinde bu durumun su derinliği ile ilgili olduğu görülmektedir. Eğer gemiler belirgin yoldan çıkar ya da zamanından önce belirgin yola girmezler ise karaya oturma tehlikesi ile karşı karşıya kalacaklardır. Charleston Limanı için anomali gemi davranışı bu bağlamda tanımlanmıştır. İlerleyen kısımlarda bu durum detaylandırılmıştır.

Grafiklerin altına eklenen kodların çalıştırılması için yüklü olması ve çağırılması gereken R kütüphaneleri;

- library(leaflet)
- library(leaflet.extras)
- library(dplyr)



Şekil 5.1 Charleston Limanı iki haftalık trafik akışı (Kargo-Tanker)

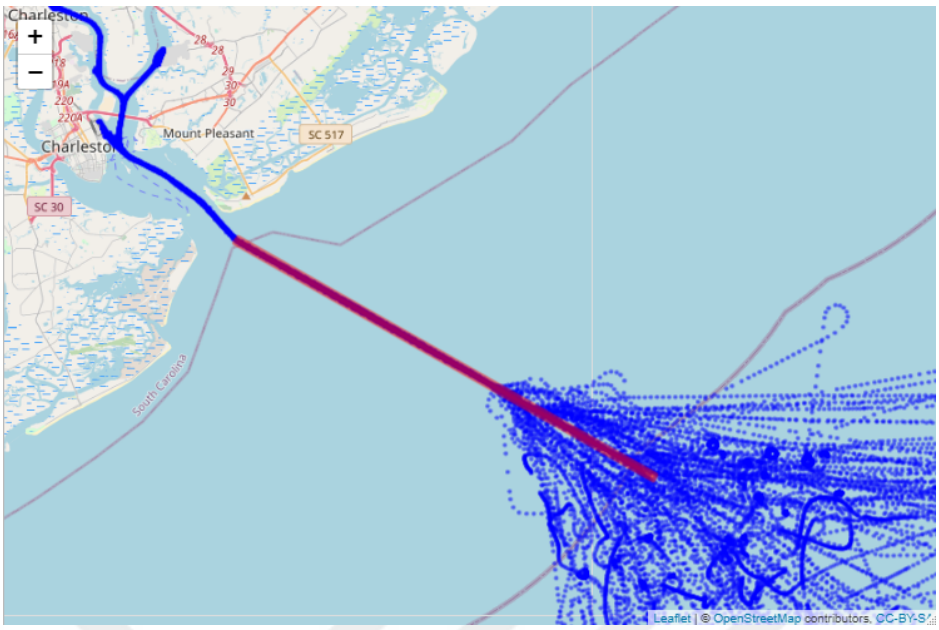
paldata <-colorFactor(palette = rainbow(1000), domain = charleston$ VesselName)
m<-leaflet(charleston) %>%
addTiles() %>%
addCircles(lat = LAT, lng = LON, weight = 2, radius = SOG,
label = paste(LAT, LON), color = paldata(VesselName)) %>%
addLegend("topright", pal = paldata, values = VesselName,
title = "Gemi İsimleri",
labFormat = labelFormat(suffix = " "), opacity = 2)
m

Leaflet kütüphanesi ile oluşturulan bu grafikler dinamik yapıdadırlar. Grafiklere yakınlaşıp uzaklaşılabilir. Kod satırında yer alan "label= paste(LAT,LON)" ifadesi ile noktaların üzerine geldiğinizde görmek istediğiniz bilgileri ekleyebilirsiniz. Bu

grafikte herhangi bir noktanın üzerine fare imleci ile geldiğinizde o noktanın enlem ve boylam değerlerine erişebilirsiniz. Veri içerisinden, enlem ve boylamdan farklı değişkenler de seçilebilir. Her bir grafiğe "addLegend" ifadesi ile istenen bilgilerin yazılabileceği bir katman eklenebilir. Aynı zamanda grafiklerdeki noktaların çaplarını belirli bir sayısal değişkene göre belirleyebilirsiniz. Yukarıdaki kodda yer alan "radius= SOG" ifadesi ile noktaların çapları, gemi hızlarına göre değişmektedir. Hızlı olan gemiler daha büyük noktalarla temsil edilirken, yavaş olan gemiler daha küçük noktalarla temsil edilmektedir. Şekil 5.1'de "pcolor= paldata(VesselName)" ifadesi sayesinde her gemi farklı bir renk ile temsil edilmiştir. Değişken seçimi elde edilmek istenen grafiğin görünümüne bağlı olarak değişebilmektedir.

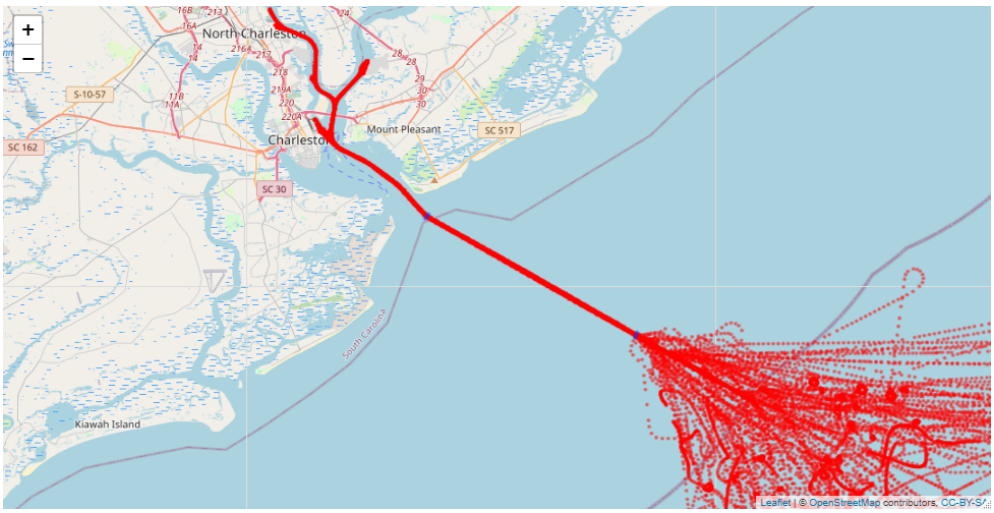
5.4.2 Kritik Bölge Seçimi

Şekil 5.2'de Charleston Limanı için kritik bölge kırmızı ile kanal oluşturularak görselleştirilmiştir. Kırmızı ile doldurulan alan belirlenirken www.navionics.com deniz haritalarından yararlanılmıştır. Şekil 5.2'de görünen kanalın uzunlupu 27.4 km genişliği ise 308 metredir. Suyun derinliği ve gemilerin hızları göz önüne alınarak Charleston Limanına girmek isteyen gemilerin bu kanalı takip etmeleri uygundur. Ancak trafik akışı görselleştirildiğinde görülmüştür ki, gemiler bu kanala sonradan girebilmektedir. Bunun nedeni gemi hızlarının yavaş olması, kaptanların bu bölgeyi iyi tanımaları, su derinliğinin kısmen daha fazla olduğu alanları biliyor olmaları olabilir. Şekil 5.4'de gemilerin kanala girerken sergiledikleri hız davranışı görülmektedir. Kanala vaktinde girmeyen gemilerin hızlarını düşürdükten sonra kanala girdikleri görülmektedir. Kritik bölgeye karar verilirken veride gerçekleşen durum referans alınmıştır. Gemilerin en son kanala girdikleri nokta sınır kabul edilmiştir. Şekil 5.3'de iki mavi nokta ile belirtilen alan uygulama için kritik bölge olarak seçilmiştir.



Şekil 5.2 Charleston Limanı kritik bölge seçimi

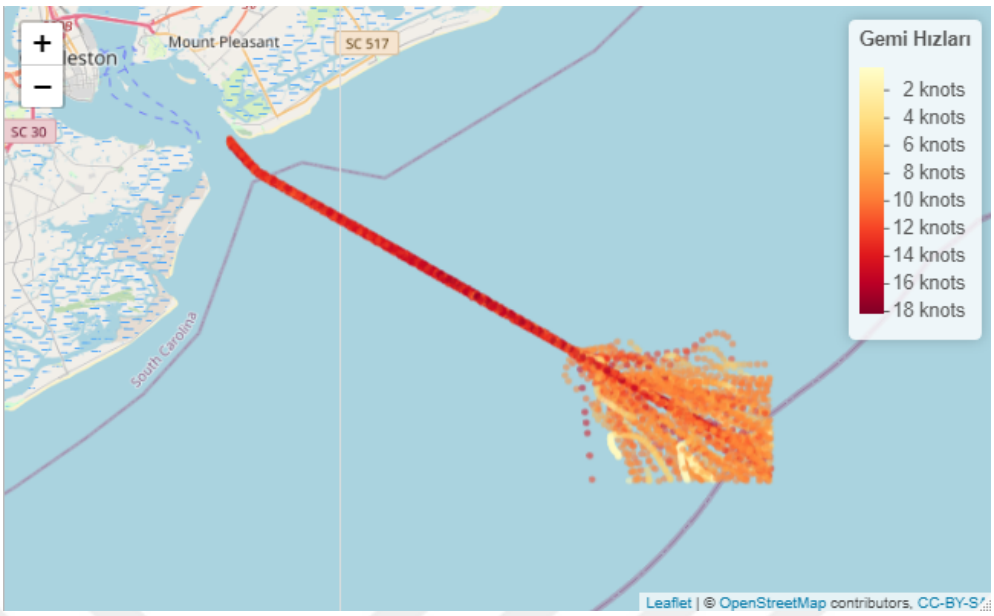
```
m<-leaflet(charleston) %>%
addTiles()%>%
addCircles(lat =  LAT, lng =  LON, weight = 2, radius =  SOG,
label =  paste(LAT, LON, NewCOG), color = "blue" )
m
m %>%
addPolygons(
lng = c(-79.591750, -79.590133, -79.844233, -79.846283),
lat = c(32.616750, 32.619167, 32.740250, 32.738050),
color = "red", opacity = 0.2)
```



Şekil 5.3 Charleston Limanı kritik bölge

```
m<-leaflet(charleston) %>%
addTiles() %>%
addCircles(lat =  LAT, lng =  LON, weight = 2, radius =  SOG,
label =  paste(LAT, LON, NewCOG), color = "red" )
m
m %>%
addPolylines(c(-79.844333, -79.846233), c(32.740300, 32.738067),
fillColor = "transparent") %>%
addPolylines(c(-79.68797, -79.6892), c(32.66623, 32.66287),
fillColor = "transparent")

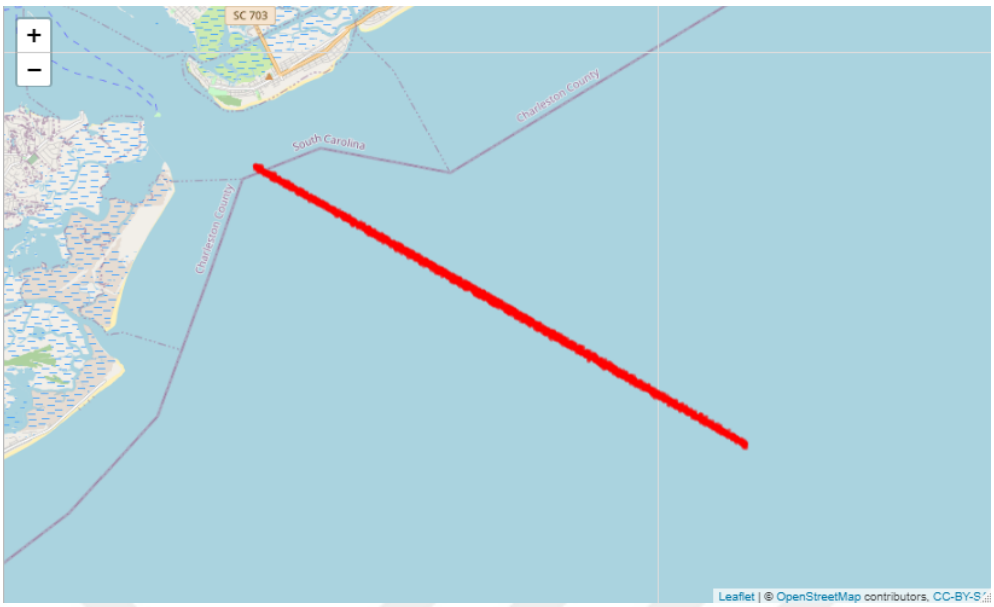
pal<- colorNumeric(palette = "YlOrRd", domain = kanal$ SOG)
m<-leaflet(kanal) %>%
addTiles() %>%
addCircles(lat =  LAT, lng =  LON, weight = 3, radius =  SOG,
label =  paste(SOG,VesselName), color =  pal(SOG) ) %>%
addLegend("topright", pal = pal, values =  SOG,
title = "Gemi Hızları",
labFormat = labelFormat(suffix = " knots" ), opacity = 1 )
m
```



Şekil 5.4 Charleston Limanı kritik bölge gemilerin hız davranışı

Charleston Limanı için belirlenen ve üzerinde uygulama yapılacak kritik bölge Şekil 5.5'de yer almaktadır. Bu alanda iki haftalık süre içerisinde limana giriş çıkış yapmış kargo ve tanker gemilerinin verileri bulunmaktadır. Gemilerin limana doğru ya da limandan ayrılırken sergiledikleri davranışlar farklılık gösterebileceğinden, veri limana doğru hareket edenler ve limandan ayrılanlar olarak ayrılmıştır. Bu iki durum için "Limana Giriş" ve "Limandan Ayrılış" etiketleri atanmıştır. Etiketleme COG ve heading açılarına bakılarak yapılmıştır. Kritik bölge doğrusala yakın yapısı nedeniyle aynı zamanda yön ayrımı içinde kullanılmıştır. Eğer bölge doğrusal değil eğrisel bir yapıda olsaydı COG ve heading açı değerleri istenen bilgiyi sağlayamazdı.

Bu nedenle yön ayrımı yapılabilmesi için seçilen bölgenin doğrusal yapıda olması önemli bir noktadır.



Şekil 5.5 Charleston Limanı belirlenen kritik bölge

kesit<-charleston[charleston$ LAT >= 32.66287 & charleston$ LAT <= 32.740300,]

kesit<-kesit[kesit$ LON >= -79.844333 & kesit$ LON <= -79.6892,] m<-leaflet(kesit)

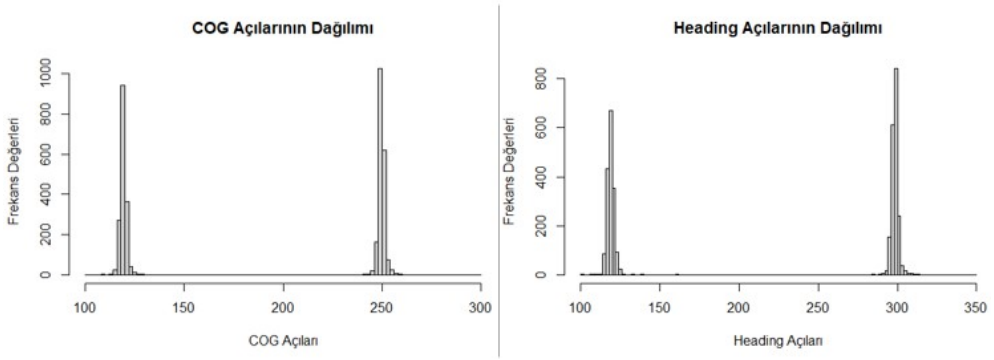
%>%

addTiles() %>%

addCircles(lat = LAT, lng = LON, weight = 2, radius = SOG,

label = paste(LAT, LON, NewCOG), color = "red")

m



Şekil 5.6 Gemilerin COG ve Heading açılarının dağılımı

hist(kesit$ NewCOG,breaks = seq(from=100, to=300, by=2),

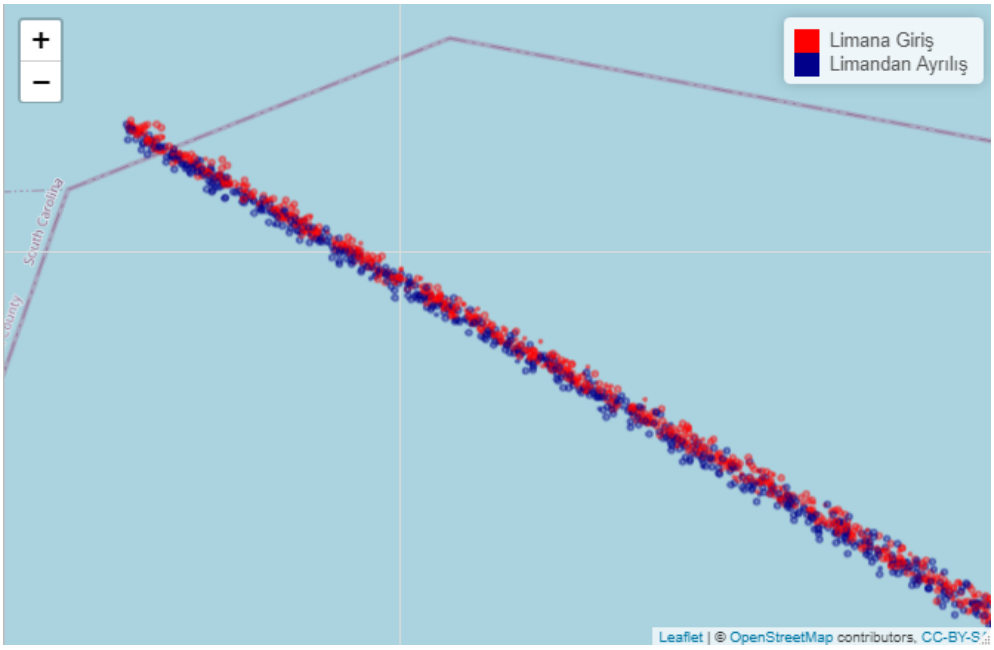
xlab = "COG Açıları", ylab = "Frekans Değerleri",

main = "COG Açılarının Dağılım Histogramı")

Tablo 5.3 Kritik bölgede yer alan gemilerin COG ve heading değerleri

YÖN				
	Min(COG)	**Max(COG)**	**Min(Heading)**	**Max(Heading)**
Limana Giriş	239	260	285	313
Limandan Ayrılış	105	145	102	162

Şekil 5.6 incelendiğinde COG ve heading değerleri için iki tepeli bir dağılım gözlenmektedir. Bu durum bize incelenen alanda iki yönlü trafik akışı olduğunu göstermektedir. Tablo 5.3'de COG ve heading açılarının gemilerin hareket yönlerine göre maksimum ve minimum değerleri yer almaktadır. Verilerin "Limana Giriş"-"Limandan Ayrılış" olarak etiketlenebilmesi için bir eşik değer belirlenmiştir. COG açılarının dağılımına bakılarak orta nokta eşik değer olarak alınmıştır. COG açısı 190 dereceden büyük olan açı değerleri limana giriş, diğerleri ise limandan ayrılış olarak etiketlenmiştir.



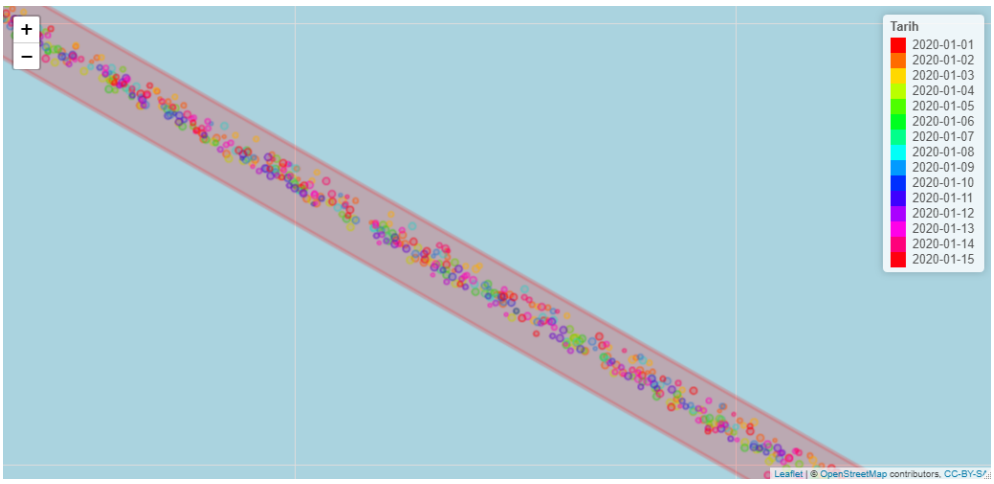
Şekil 5.7 Charleston Limanı gemilerin hareket yönleri

```
paldata <- colorFactor(palette = c("red", "darkblue"), domain = kesit$ Direction)
m<-leaflet(kesit) %>%
addTiles() %>%
addCircles(lat = LAT, lng = LON, weight = 2, radius = SOG,
label = paste(LAT, LON ,NewCOG), color= paldata(Direction) ) %>%
addLegend("topright", pal = paldata, values = Direction,
title = "Tarih",
labFormat = labelFormat(suffix = " " ),
opacity = 2)
m
```

Şekil 5.7'de gemiler limana doğru gidiyor ise mavi, limandan ayrılıyor ise kırmızı ile gösterilmektedir. Anomali belirleme uygulamaları limana doğru hareket eden kritik bölgedeki gemiler alınarak yapılmıştır. Bu aşamadan sonra yanlızca limana giriş yapan veriler filtrelenmiştir (Şekil 5.8).

5.4.3 ST-DBSCAN ile Anomali Belirleme

Bu bölümde ST-DBSCAN algoritmasının anomali belirlemede kullanımını araştıracağız. Üzerinde çalışılan veri gerçek hayat verisidir. Gemilerin hareketleri olağan normal davranış olarak kabul edilmiştir. Charleston limanı için tehlike arz edebilecek durumlar, gemilerin limana girmek ve limandan çıkmak için kullanılan kanala çok yakın seyir etmeleri ya da kanala geç girmeleri olabilir. Belirlenen kanalın dışındaki seyir durumlarında karaya oturma riski vardır. Bu bağlamda anomali senaryosu yaratılmış ve ST-DBSCAN algoritması ile tespit edilmeye çalışılmıştır. ST-DBSCAN algoritması zamansal-mekansal veriye duyarlı olduğu için bu alandaki performansı tartışılacaktır.



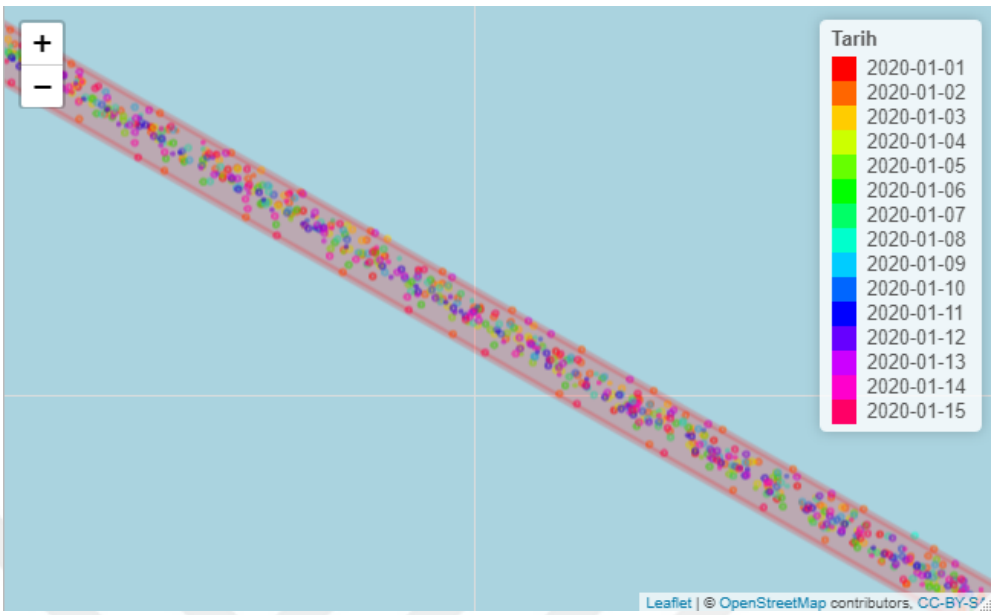
Şekil 5.8 Kritik bölgede liman doğrultusunda hareket eden gemilerin kanal içindeki dağılımları

```
paldata <- colorFactor(palette = rainbow(100), domain = kesitG$ Date)
m<-leaflet(kesitG) %>%
addTiles() %>%
addCircles(lat = LAT, lng = LON,weight = 2,radius = SOG,
label = paste(LAT, LON, NewCOG,MMSI), color= paldata(Date) ) %>%
addLegend("topright", pal = paldata, values = Date,
title = " Tarih ", labFormat = labelFormat(suffix = " " ),opacity = 2)
m
m %>%
addPolygons(
lng = c(-79.591750,-79.590133,-79.844233,-79.846283),
lat = c(32.616750, 32.619167,32.740250 , 32.738050),
color = "red",opacity = 0.1)
```

Şekil 5.8'de kritik bölgede liman doğrultusunda hareket eden gemilerin trafik akışının bir bölümü yer almaktadır. Her renk farklı bir günü temsil etmektedir. Bu kanal sınırları referans alınarak anomali gemi davranışları oluşturulmuştur (Şekil 5.9).

Şekil 5.8'de olağan trafik akışı yer almaktadır. Bu durum referans alınarak şekil 5.9'de oluşturulan anomali davranış senaryosu sonucu gerçekleştirilen trafik akışı yer almaktadır. Seçilen gemilerin enlem ve boylam değerlerine müdahale edilerek kanal

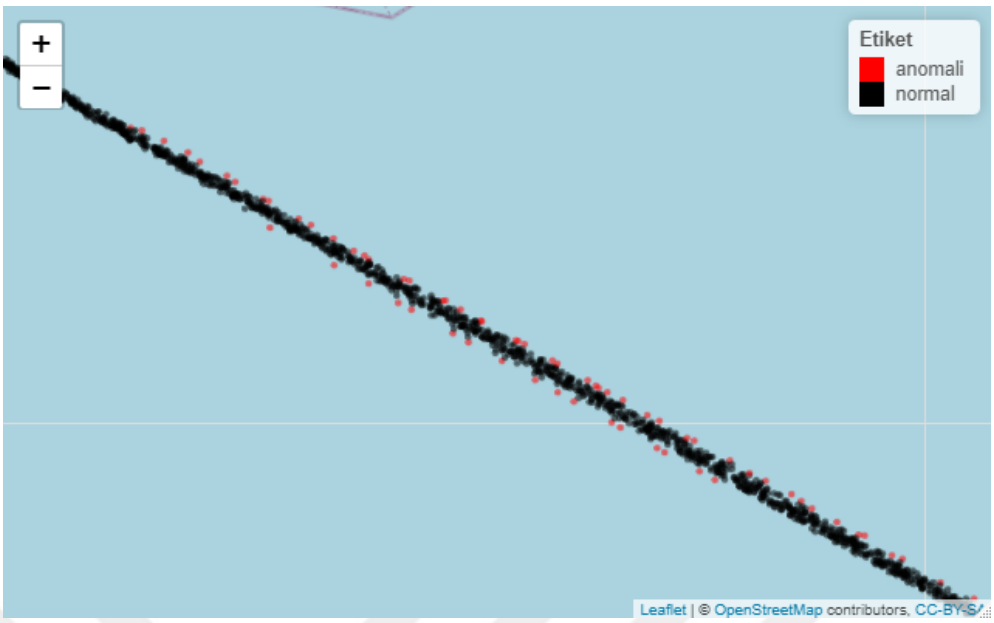
sınırına daha yakın hareket etmeleri sağlanmıştır.



Şekil 5.9 Oluşturulan Anomali Davranışlar

```
paldata <- colorFactor(palette = rainbow(15), domain = deneme$ Date)
m<-leaflet(deneme) %>%
addTiles() %>%
addCircles(lat = LAT, lng = LON, weight = 2, radius = SOG,
label = paste(LAT, LON, NewCOG,MMSI), color = paldata(Date)) %>%
addLegend("topright", pal = paldata, values = Date,
title = "Tarih",
labFormat = labelFormat(suffix = " " ), opacity = 2)
m

paldata <- colorFactor(palette = c("red", "black"),domain = deneme$ label)
m<-leaflet(deneme) %>%
addTiles() %>%
addCircles(lat = LAT, lng = LON, weight = 3,
label = paste(LAT, LON, NewCOG, MMSI), color = paldata(label)) %>%
addLegend("topright", pal = paldata, values = label,
title = "Etiket", labFormat = labelFormat(suffix = " " ), opacity = 2)
m
```

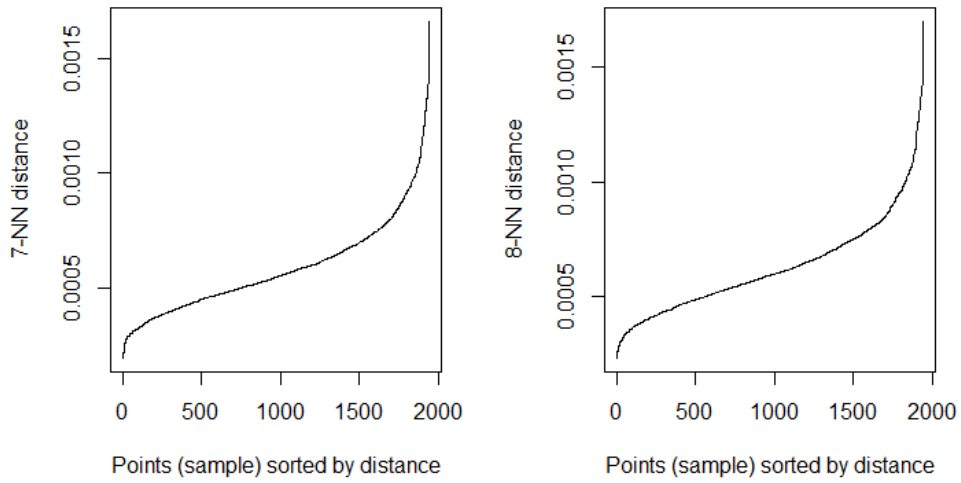


Şekil 5.10 Normal ve anomali noktaların dağılımı

Veride bulunan 73 kargo ve tanker gemisinden 7 tanesinin rotası ile oynanmıştır. Şekil 5.10'da normal ve anomali noktaların dağılımı yer almaktadır. Normal noktalar siyah ile, anomali noktalar kırmızı ile temsil edilmektedir. Kanal sınırlarına yakın hareket eden gemileri ST-DBSCAN algoritmasının tespit etmesi beklenmektedir. ST-DBSCAN algoritmasını uygulamadan önceki adım parametre değerlerinin belirlenmesidir.

5.4.4 Parametre Belirleme

ST-DBSCAN algoritmasının uygulanabilmesi için eps1, eps2, MinPts parametrelerinin değerleri belirlenmelidir. MinPts için bu değer $ln(n)$ ile hesaplanabilir. Burada n verilerin boyutudur. Verilerimiz için MinPts değeri $ln(1944) \cong 7.5725$ olarak hesaplanmıştır. Eps1 mesafe parametresinin seçiminde bölüm 4.2'de açıklanan yaklaşım kullanılmıştır. Öklit mesafe ölçüsü kullanılarak enlem ve boylam değerleri için k-en yakın komşu grafiği oluşturulmuştur. Burada ki k değeri MinPts değerine eşittir. MinPts değeri yaklaşık 7.57 olduğundan k=7 ve k=8 için k-nn grafikleri çizdirilmiştir.



Şekil 5.11 Eps1 parametre seçimi K-NN grafiği

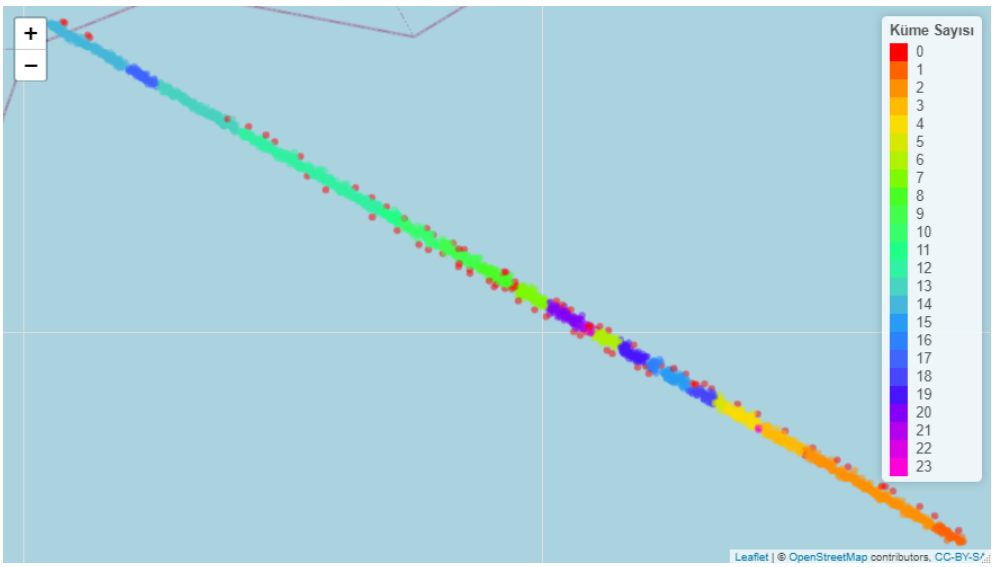
Eps1 parametresi belirlenirken k en yakın komşu komşu grafiğinde bir dirsek noktası aranmaktadır. Bu nokta ve yakınındaki değerler parametre değeri olarak alınarak algoritma çalıştırılabilir. Dirsek noktasından -pozitif yönde- büyük seçilen değerlerde küme sayılarının artması beklenmektedir. Optimum küme sayısı için farklı eps1 değerleri seçilerek denemeler yapılabilir.

Eps2 parametresi belirlenirken önceki çalışmalar gözden geçirilmiştir (Datlıca ve Çakıt, 2020; Chimwayi ve Anuradha, 2018). Eps2 uzamsal olmayan parametre, yani zaman değişkeni için uzaklık parametresidir. Eps2 değeri saniye cinsinden alınmıştır. Değer 1285421 olarak belirlenmiştir. Algoritma, farklı eps1 ve MinPts parametreleri ile uygulanmıştır. Optimum parametre değerleri belirlenmeye çalışılmıştır.

Tablo 5.4 ST-DBSCAN algoritmasının farklı parametre değerleri için performansının değerlendirilmesi

Eps1	Eps2	MinPts	Küme Sayısı	Anomali Sayısı	F1 Puanı
0.00090	1285421	10	4	23	0.4554455
0.00085	1285421	10	6	30	0.537037
0.00080	1285421	10	12	57	0.637037
0.00075	1285421	10	22	118	0.6020408
0.00070	1285421	10	35	180	0.5193798
0.00090	1285421	9	3	20	0.4081633
0.00085	1285421	9	3	25	0.4660194
0.00080	1285421	9	10	41	0.6218487
0.00075	**1285421**	**9**	**17**	**78**	**0.6794872**
0.00070	1285421	9	29	128	0.5728155
0.00090	1285421	8	2	19	0.3917526
0.00085	1285421	8	2	23	0.4356436
0.00080	1285421	8	4	35	0.5663717
0.00075	**1285421**	**8**	**13**	**54**	**0.6666667**
0.00070	**1285421**	**8**	**23**	**82**	**0.6875000**
0.00090	1285421	7	2	17	0.3578947
0.00085	1285421	7	2	21	0.4040404
0.00080	1285421	7	4	32	0.5454545
0.00075	1285421	7	8	49	0.6141732
0.00070	**1285421**	**7**	**16**	**70**	**0.6756757**
0.00090	1285421	6	1	15	0.3225806
0.00085	1285421	6	2	19	0.371134
0.00080	1285421	6	4	27	0.4761905
0.00075	1285421	6	6	43	0.6280992
0.00070	**1285421**	**6**	**8**	**59**	**0.6569343**

Tablo 5.4'de yer alan farklı eps1 ve MinPts değerleri ile algoritma çalıştırılmıştır. Oluşan küme sayıları ve tespit edilen anomali gözlem sayıları tabloda yer almaktadır. Performans ölçüm kriteri olarak seçilen F1 puanı değerleride kaydedilmiştir. F1 puanı 0 ile 1 arasında değer almaktadır. En yüksek F1 puanını veren eps1 ve MinPts değerleri işaretlenmiştir. Bu parametre değerleri arasından en yüksek F1 puanını veren eps1= 0.00070, eps2= 1285421 ve MinPts= 8 parametreleri optimum parametre olarak seçilmiştir. Küme numarası 0 olan değerler herhangi bir kümeye dahil olmamış noktalardır. Bu noktalar anomali olarak etiketlenmiştir.



Şekil 5.12 ST-DBSCAN ile kümeleme sonuçları (eps1= 0.00070, eps2= 1285421, MinPts= 8)

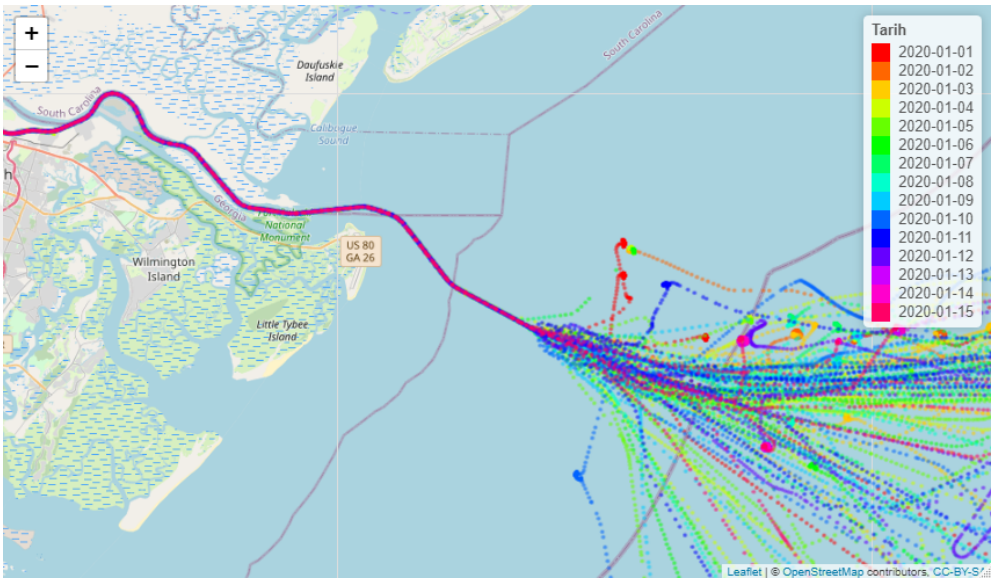
Tablo 5.5'de eps1= 0.00070, eps2= 1285421 ve MinPts= 8 parametre değerleri ile çalıştırılan ST-DBSCAN algoritmasının sonuçları yer almaktadır. Şekil 5.12'de kümeleme sonuçlarına görsel olarak yer verilmiştir. Toplam 78 anomali gözlemden 55 tanesi doğru tahmin edilmiştir.

Tablo 5.5 Charleston Limanı Yanılgı matrisi

		GERÇEK		
		Anomali	**Normal**	**Toplam**
TAHMİN	**Anomali**	*55*	*27*	82
	Normal	*23*	*1839*	1646
	Toplam	78	1866	

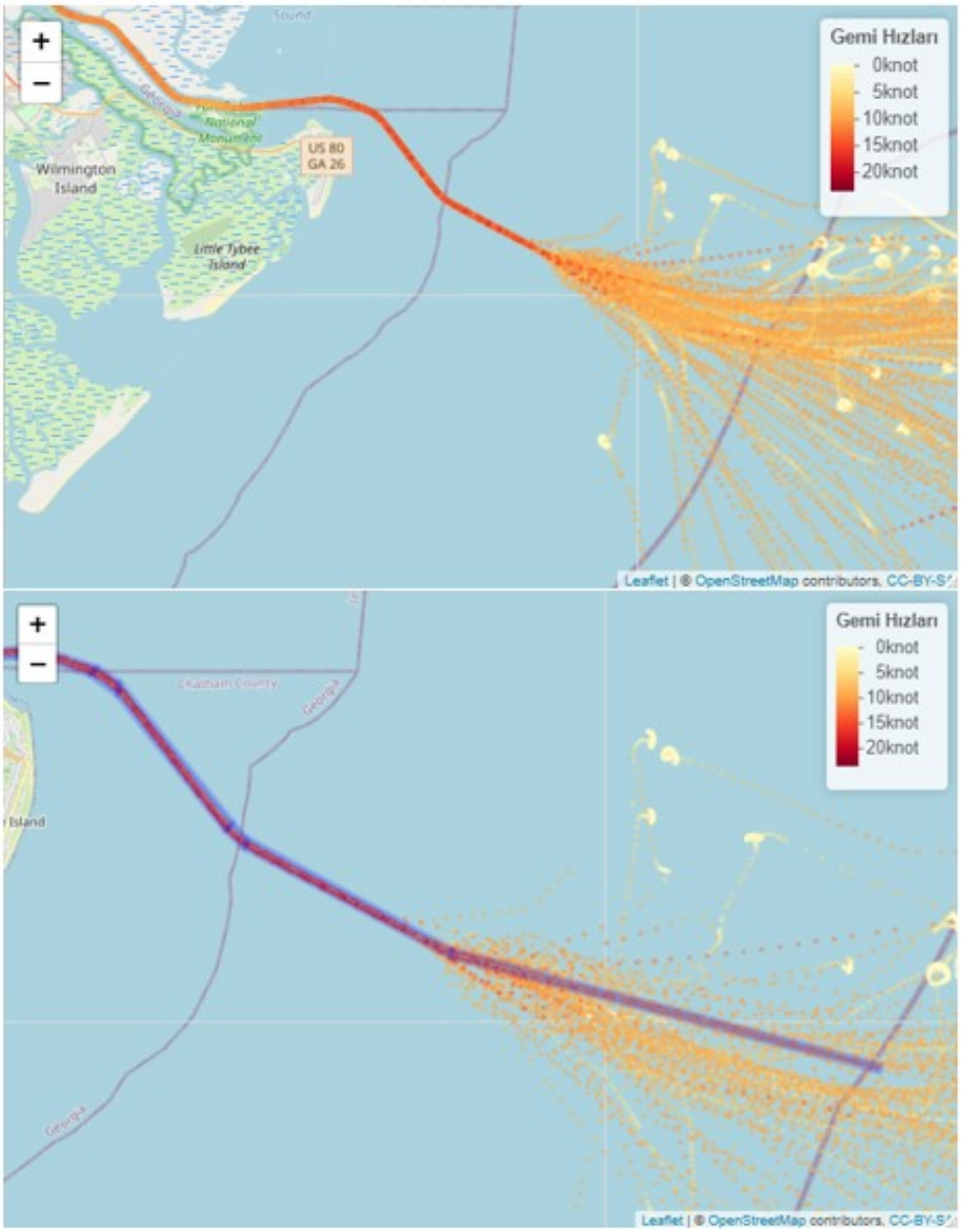
5.5 Savannah Limanı

Savannah Limanı, ABD'nin Georgia eyaletinde bulunmaktadır. Kuzey Amerikanın önemli ve işlek limanlarındandır. Charleston Limanı üzerinden detaylı olarak gösterilen tüm adımlar Savannah Limanı içinde uygulanmıştır. Bu bölümde grafiklerin altında R kodlarına yer verilmeyecektir. Charleston Limanı için kullanılan R kodları Savannah Limanı içinde kullanılmıştır.



Şekil 5.13 Savannah Limanı trafik akışı

Şekil 5.13'de Savannah Limanı iki haftalık kargo ve tanker gemilerine ait trafik akışı görselleştirilmiştir. Her renk ayrı bir günü temsil etmektedir. Bu şekil gemilerin limana çevresindeki davranışlarını davranışları incelemek amacıyla elde edilmiştir. Tüm gemiler, belirli bir noktadan sonra limana girmek ya da limandan ayrılmak için benzer davranışı sergilemektedir. Belirli bir yolu takip ederek limana bu yoldan uğramaktadırlar. Navionics uygulaması üzerinden liman incelendiğinde, bu durumun su derinliği ile ilgili olduğu, bu yolu takip etmeyen kargo ve tanker gemilerinin karaya oturma tehlikesi ile karşılaşabileceği görülmüştür.



Şekil 5.14 Savannah Limanı gemilerin hız davranışı

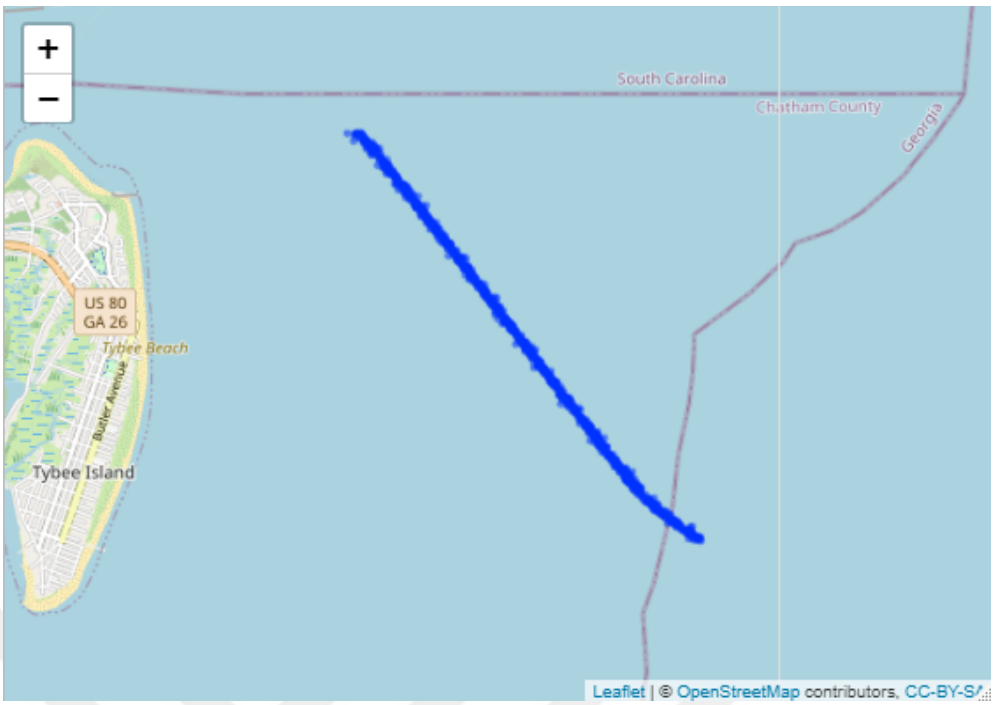
Navionics uygulaması üzerinden liman haritası elde edilmiş ve limana uğramak isteyen gemilerin takip etmesi gereken yol şekil 5.14'de hız grafiğinin üzerine saydam, mavi bir kanal olarak eklenmiştir. Kanalın genişliği yaklaşık 175 metre uzunluğu ise yaklaşık 30 kilometredir. Böylelikle gerçekleşen trafik akışı kanala girme açısından incelenebilmiştir.

Gemilerin takip etmeleri gereken yola girmeden ya da zamanında girmemiş ise hızlarını oldukça düşürdükleri, daha sonra bu yola girdikleri görülmektedir. Bu yolun dışına çıkılması durumunda karaya oturma tehlikesi bulunmaktadır. Ancak veride gerçekleşen durumda gemilerin bu kanala sonradan girdikleri görülmektedir. Bu durum su derinliğinin ilerideki alanlara göre kısmen daha derin olmasıdır. Bölgeyi iyi tanıyan kaptanların su derinliğinin kritik seviyelere geldiği ana kadar açıktan seyir etmeye devam ettikleri düşünülebilir. Bir çok gemi bu davranışı sergilemektedir. Bu nedenle liman için kritik bölge başlangıcı tüm gemilerin kanala girdikleri son nokta olarak belirlenmiştir.



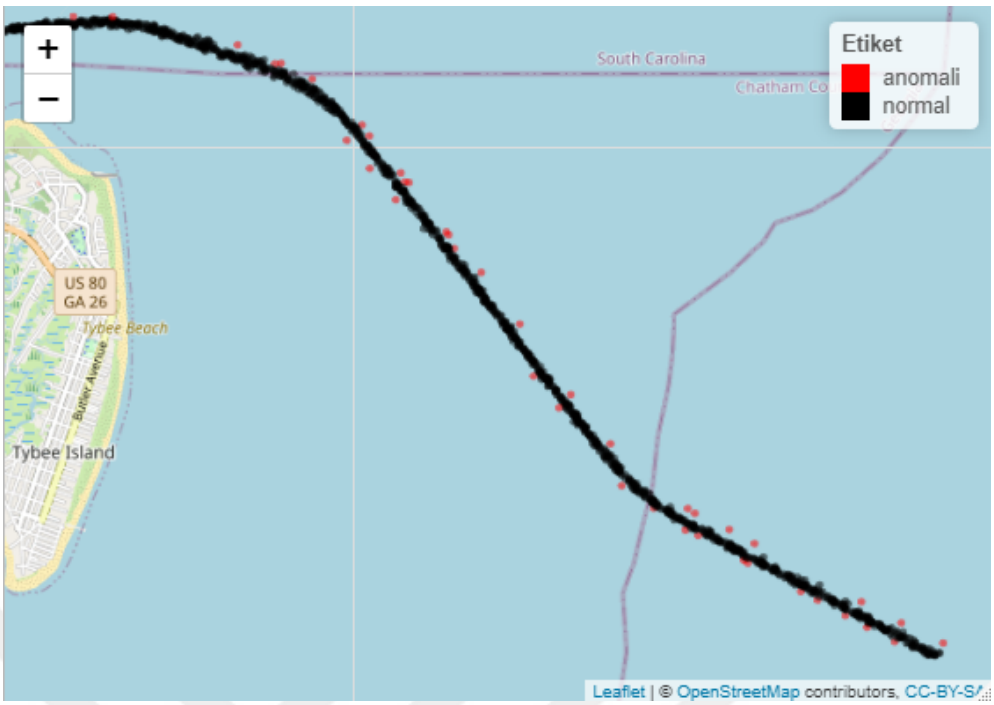
Şekil 5.15 Savannah Limanı kritik bölge

Şekil 5.15'da liman için seçilen kritik bölge yer almaktadır. Anomali senaryosu oluşturulurken, bu alana girdikten sonra kanal sınırlarına çok yakın hareket eden gemilerin anomali davranış sergiledikleri kabul edilecektir. Bu alan üzerinde anomali senaryosu oluşturulduktan sonra ST-DBSCAN algoritması ile anomali belirleme uygulamaları yapılacaktır. Bu bölgeden sadece limana doğru hareket eden gemilerin verileri kullanılmıştır. Detayları ilerleyen paragraflarda anlatılmıştır.



Şekil 5.16 Savannah Limanı trafik yön ayrımı için kesit alınan bölge

Savannah limanından trafik akışı yön ayrımı için Şekil 5.16'de görülen doğrusal alan kesit alınmıştır. Bu kesit içerisinde yer alan tüm gemiler bu limana uğramıştır. Bu alandaki gemiler, COG ve heading değerlerine bakılarak limana gidiş ve limandan ayrılış olarak sınıflandırılmıştır. Bu sınıflandırmanın detayları bölüm 5.4'de detaylı olarak açıklanmıştır. Savannah Limanın'da Charleston Limanın'dan farklı olarak kritik bölge ve trafik yön ayrımı yapılması için kullanılan bölgeler birbirinden farklıdır. Trafik yön ayrımı yapılacak olan bölgenin doğrusal olması gereklidir. Ancak kritik bölge için böyle bir zorunluluk yoktur. Anomali senaryosu oluşturulurken limana giriş doğrultusunda hareket eden gemiler alınmıştır.



Şekil 5.17 Savannah Limanı anomali senaryosu

Oluşturulan anomali gemi davranışları Şekil 5.17'de görselleştirilmiştir. ST-DBSCAN algoritmasının kırmızı ile temsil edilen noktaları ayırt etmesi beklenmektedir. ST-DBSCAN algoritmasının parametre belirleme adımları Charleston Limanı bölüm 5.4.4'de açıklandığı gibi belirlenmiştir.

ST-DBSCAN için parametre seçimi, eps1 değerleri 0.0006 ile 0.0007 aralığında 0.00005 birim arttırılarak MinPts değeri ise 6 ile 9 arasında birer birim arttırılarak denemeler ile yapılmıştır. Tablo 5.6 en yüksek F1 puanını veren modellerin eps1 değerleri 0.0006 olduğu görülmüştür. Bu tablodan yola çıkılarak eps1 değerlerinin daha hassas bir aralıkta ayarlanıp daha iyi sonuçlar elde edilmiştir (Tablo 5.7).

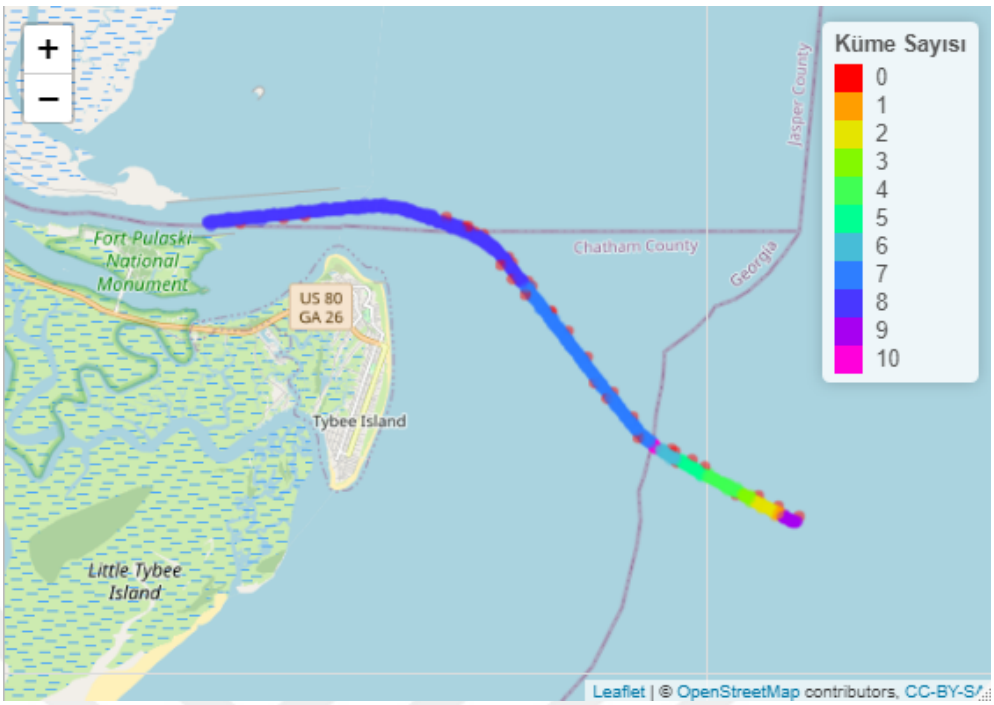
Tablo 5.6 ST-DBSCAN parametre seçimi

Eps1	Eps2	MinPts	Küme Sayısı	Anomali Sayısı	F1 Puanı
0.0006	1189891	6	3	30	**0.7631579**
0.00065	1189891	6	2	27	0.7123288
0.0007	1189891	6	1	22	0.6470588
0.0006	1189891	7	6	31	**0.7721208**
0.00065	1189891	7	3	27	0.7123288
0.0007	1189891	7	1	22	0.6470588
0.0006	1189891	8	12	31	**0.7792208**
0.00065	1189891	8	5	27	0.7123288
0.0007	1189891	8	1	22	0.6470588
0.0006	1189891	9	19	35	**0.7901235**
0.00065	1189891	9	13	27	0.7123288
0.0007	1189891	9	4	22	0.6470588

Tablo 5.7 ST-DBSCAN Parametre seçimi-2

Eps1	Eps2	MnPts	Küme Sayısı	Anomali Sayısı	F1
0.00058	189891	6	5	32	0.7948718
0.00058	**189891**	**7**	**10**	**33**	**0.8101266**
0.00058	**189891**	**8**	**16**	**33**	**0.8101275**
0.00058	189891	9	27	38	0.7857143
0.00062	189891	6	3	27	0.7123288
0.00062	189891	7	4	27	0.7123288
0.00062	189891	8	10	28	0.7297297
0.00062	189891	9	17	30	0.7368421

Tablo 5.7 incelendiğinde maksimum F1 puanını veren iki model göze çarpmaktadır. F1 puanlarının oldukça birbirine yakın olması ve tespit edilen anomali sayılarının eşit olması nedeniyle küme sayısı az olan model seçilmiştir. Bu modele ait parametre değerleri eps1 = 0.00058, eps2 = 189891 ve MinPts = 7 'dir.



Şekil 5.18 Savannah Limanı kümeleme sonuçları (eps1 = 0.00058, eps2 = 189891, MinPts = 7

Tablo 5.8 Savannah Limanı Yanılgı matrisi

		GERÇEK		
		Anomali	**Normal**	**Toplam**
TAHMİN	**Anomali**	*32*	*1*	33
	Normal	*14*	*1632*	1646
	Toplam	46	1633	

Tablo 5.8'de ST-DBSCAN algoritması eps1 = 0.00058, eps2 = 189891 ve MinPts = 7 parametre değerleri için yanılgı matrisi yer almaktadır. Liman için hazırlanan anomali senaryosunda 46 adet anomali nokta oluşturulmuştur. Uygulanan kümeleme algoritması sonucunda bu noktalardan 32 tanesi doğru tahmin edilmiştir. Şekil 5.18'de bu parametre değerlerine ait kümeleme algoritması sonuçları görsel olarak yer almaktadır. Her küme farklı bir renk ile temsil edilirken anomali noktalar kırmızı renk ile temsil edilmiştir.

BÖLÜM ALTI
SONUÇLAR

Küresel ticarette deniz taşımacılığı büyük bir rol oynamaktadır. Deniz trafiğinin her geçen gün artan hacmi ve ticaret için önemi, bu alandaki tehditleri en aza indirgemek için çeşitli gözetim tekniklerinin ve trafik analizlerinin geliştirilmesine neden olmuştur. Bu alanda kullanılan en önemli tekniklerden biri de anomali tespitidir. Deniz trafiğinde anomali tespiti, deniz taşıtlarının olağan dışı davranışlarını veri madenciliği yöntemleriyle belirlemeyi amaçlayan bir çalışma alanıdır. Deniz trafiğinde olağan ve olağan dışı davranışları tanımlamak için genellikle AIS verileri kullanılmaktadır. AIS verileri bu alandaki en önemli veri kaynağıdır.

Bu çalışmada zamansal-mekansal kümeleme algoritmalarından ST-DBSCAN algoritmasının deniz trafiği analizinde anomali gemi davranışı tespiti için kullanımı araştırılmıştır. Deniz trafiği analizinde kümeleme yöntemleri ile anomali tespiti yapan çalışmalar genellikle mekansal kümeleme algoritmalarını kullanmışlardır. AIS verileri zamansal-mekansal özellikler içerdiğinden dolayı zamansal-mekansal bir kümeleme algoritması olan ST-DBSCAN kullanılmıştır. Tezde, AIS sistemi tanıtılmış AIS verilerini oluşturan değişkenler ayrıntılı olarak incelenmiştir. Ayrıca AIS verilerinin R programlama dili ile görselleştirilmesine detaylı olarak yer verilmiştir. Deniz trafiğinde anomali gemi davranışı tespit süreci görsel içeriklerle desteklenmiştir.

Kuzey Amerika'da bulunan Charleston ve Savannah Limanları üzerinde yapılan uygulamalarda ST-DBSCAN algoritmasının, gemilerin konumlarında meydana gelen anomali davranışlarını tespit edebildiği görülmüştür. Algoritma, olağan davranışların gerçekleştiği alanları kümeleyebilirken, olağan davranıştan uzak noktaları ayırabilmiştir. ST-DBSCAN algoritmasının kullanımında en büyük zorluk optimum parametrelerin belirlenmesinde yaşanmıştır. Seçilen parametre uzayındaki farklı parametre kombinasyonları denenerek F1 puanına göre anomali belirleme performansı değerlendirilmiştir. En uygun parametre değerleri seçilebildiğinde ST-DBSCAN algoritmasının AIS verileri ile deniz trafiğinde anomali gemi

davranışlarını tespit edebildiği görülmüştür. Bu çalışma, bildiğimiz kadarıyla, ST-DBSCAN algoritmasının AIS verileri ile anomali belirleme konusunda ilk çalışmadır. Gelecek çalışmalarda ST-DBSCAN için en uygun parametrelerin seçimi ve farklı zamansal-mekansal kümeleme algoritmalarının AIS verilerinde kullanımı konularında çalışılması planlanmaktadır.

KAYNAKLAR

Ansari, M. Y., Ahmad, A., Khan, S. S., Bhushan, G., ets al. (2020). Spatiotemporal clustering: a review. *Artificial Intelligence Review*, *53*(4), 2381–2423.

Birant, D., ve Kut, A. (2007). ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data and Knowledge Engineering*, *60*(1), 208–221.

Bomberger, N. A., Rhodes, B. J., Seibert, M., ve Waxman, A. M. (2006). Associative learning of vessel motion patterns for maritime situation awareness. *2006 9th International Conference on Information Fusion*, 1–8.

Chandola, V., Banerjee, A., ve Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), 1–58.

Chimwayi, K. B., ve Anuradha, J. (2018). Clustering West Nile Virus Spatio-temporal data using ST-DBSCAN. *Procedia Computer Science*, *132*(Iccids), 1218–1227.

d'Afflisio, E., Braca, P., Millefiori, L. M., ve Willett, P. (2018). Maritime anomaly detection based on mean-reverting stochastic processes applied to a real-world scenario. *2018 21st International Conference on Information Fusion (FUSION)*, 1171–1177.

Datlıca, M., ve Çakıt, E. (2020). Değiştirilebilir konum süresine sahip takip cihazlarında kümeleme parametrelerinin tahmini ve anomali tespiti. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, *1*, 373–394.

Doğan, Y., Kart, Ö., Kundakçi, B., ve Nas, S. (2021). A Novel Approach for Knowledge Discovery from AIS Data: An Application for Transit Marine Traffic in the Sea of Marmara. *Advances in Electrical and Computer Engineering*, *21*(3), 73–80.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., ets al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *kdd*, vol. 96, 226–231.

Goldstein, M., ve Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLoS ONE*, *11*(4), 1–31.

Grech, M., Horberry, T., ve Koester, T. (2019). *Human Factors in the Maritime Domain*.

Handayani, D. O. D., Sediono, W., ve Shah, A. (2013). Anomaly detection in vessel tracking using support vector machines (svms). *2013 International Conference on Advanced Computer Science Applications and Technologies*, *IEEE*, 213–217.

Harati Mokhtari, A., Wall, A., Brooks, P., ve Wang, J. (2007). Automatic identification system (ais): Data reliability and human error implications. *Journal of Navigation*, *60*, 373 – 389.

Horn, S., Eisler, C., Dobias, P., ve Collins, J. (2016). Data requirements for anomaly detection. *Maritime Knowledge Discovery and Anomaly Detection Workshop*. https://bluehub.jrc.ec.europa.eu/KDAD/KDAD_Proceedings.pdf.

Izmirli, S., ve Yıldıztepe, E. (2021). Deniz trafiği analizi için ais verilerinin görselleştirilmesi. *Uluslararası Bilimsel Araştırmalar ve Yenilikçi Çalışmalar Sempozyumu"*. https://www.issris.org/_files/ugd/1dd905_284ad2cd33824c29b491987c223a4395.pdf.

Jha, V. S. (2020). Understanding the operation of" ais" in international maritime transactions. *Journal of Maritime Law and Commerce*, *51*(1), 65–77.

Jiacai, P., Qingshan, J., Jinxing, H., ve Zheping, S. (2012). An ais data visualization model for assessing maritime traffic situation and its applications. *Procedia Engineering*, *29*, 365–369.

Kaluza, P., Koelzsch, A., Gastner, M., ve Blasius, B. (2010). The complex network of global cargo ship movement. *Journal of the Royal Society, Interface / the Royal Society*, *7*, 1093–103.

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning*, vol. 1.

Kowalska, K., ve Peel, L. (2012). Maritime anomaly detection using gaussian process active learning. *2012 15th International Conference on Information Fusion*, *IEEE*, 1164–1171.

Lane, R. O., Nevell, D. A., Hayward, S. D., ve Beaney, T. W. (2010). Maritime anomaly detection and threat assessment. *2010 13th International conference on information fusion*, *IEEE*, 1–8.

Obradović, I., Miličević, M., ve Žubrinić, K. (2014). Machine learning approaches to maritime anomaly detection. *Naše more: znanstveni časopis za more i pomorstvo*, *61*(5-6), 96–101.

Pallotta, G., Vespe, M., ve Bryan, K. (2013). Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*, *15*(6), 2218–2245.

Pereira, R., Abreu, P. H., Polisciuc, E., ve Machado, P. (2019). A data visualization approach for intersection analysis using ais data. *VISIGRAPP (3: IVAPP)*, 208–215.

Riveiro, M., Pallotta, G., ve Vespe, M. (2018). Maritime anomaly detection: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(5), e1266.

Rong, H., Teixeira, A., ve Soares, C. G. (2020). Data mining approach to shipping route characterization and anomaly detection based on AIS data. *Ocean Engineering*, *198*(January), 106936.

Sidibé, A., ve Shu, G. (2017). Study of automatic anomalous behaviour detection techniques for maritime vessels. *Journal of Navigation*, *70*(4), 847–858.

Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., ve Huang, G.-B. (2017). Exploiting ais data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transactions on Intelligent Transportation Systems*, *19*(5), 1559–1582.

Varlamis, I., Tserpes, K., ve Sardianos, C. (2018). Detecting search and rescue missions from ais data. *2018 IEEE 34th international conference on data engineering workshops (ICDEW)*, *IEEE*, 60–65. https://www.computer.org/csdl/proceedings-article/icdew/2018/630601a060/12OmNwHhoLj.

Wang, Z., Hu, J., ve Fan, Q. (2018). Extracting the main routes and speed profiles between two locations from massive uncertain historical trajectories. *2018 International Symposium in Sensing and Instrumentation in IoT Era (ISSI)*, *IEEE*, 1–5.

Wang, Z., Yao, M., Meng, C., ve Claramunt, C. (2020). Risk assessment of the overseas imported COVID-19 of ocean-going ships based on AIS and infection data. *ISPRS International Journal of Geo-Information*, *9*(6).

Zhao, L., ve Shi, G. (2019). Maritime Anomaly Detection using Density-based Clustering and Recurrent Neural Network. *Journal of Navigation*, *72*(4), 894–916.