

# شناسایی آنومالی در ترافیک دریایی بر اساس داده‌های مکانی - زمانی سیستم شناسایی خودکار (AIS)

مجتبی گودرزی، \* مهدی شعبانی\*\*

تاریخ دریافت: ۱۴۰۰/۹/۲۷ تاریخ پذیرش: ۱۴۰۰/۱۰/۱۱ نوع مقاله: پژوهشی

## چکیده

نظارت بر ترافیک دریایی یکی از جنبه‌های مهم ایمنی و امنیتی به خصوص در مسیرهای دریایی شلوغ و ترافیک در نزدیکی اسکله‌هاست. در این تحقیق، مدل شناسایی آنومالی دو بخشی در ترافیک دریایی پیشنهاد شده است. در بخش اول، الگوی معمولی خط سیر و مسیر کشتی‌ها را که براساس قوانین IMO جمع‌آوری و در دسترس عموم قرار گرفته‌اند، با استفاده از الگوریتم‌های خوشه‌بندی و براساس چگالی در داده‌کاوی شناسایی شده است. مهم‌ترین حسن خوشه‌بندی بر اساس چگالی در این مرحله، شناسایی الگوی حرکتی کشتی‌ها با توجه به قوانین و شرایط موجود در منطقه دریایی است. در بخش دوم، مدل پیشنهادی با توجه به الگوی شناخته‌شده در بخش اول و با استفاده از رویکرد براساس نقطه، جابه‌جایی غیرعادی کشتی‌ها در دریا شناسایی شده است. برای خوشه‌بندی و شناسایی رفتار آنومال از طول و عرض جغرافیایی، سرعت و جهت در هر نقطه از مسیر طی‌شده به وسیله کشتی‌ها استفاده شده است. رویکرد شناسایی آنومالی براساس نقطه قابل استفاده در سیستم‌های نظارتی سیستم شناسایی خودکار (AIS) از نوع بلادرنگ نیز است. در این تحقیق، مرحله شناسایی جابه‌جایی آنومال قابلیت شخصی‌سازی از طرف تصمیم‌گیرندگان نظارتی را دارد که در نتیجه، میزان دقت شناسایی آنومالی قابل تنظیم است.

**واژگان کلیدی:** شناسایی آنومالی، خوشه‌بندی براساس چگالی، کاوش در داده‌های خط سیر، نظارت دریایی.

---

\* گروه کامپیوتر، دانشکده مهندسی برق، کامپیوتر و مکانیک، دانشگاه ایوان کی، سمنان

m.goodarzi1001@gmail.com

\*\* گروه کامپیوتر، دانشکده فنی و مهندسی، دانشگاه تهران جنوب، تهران، ایران.

## مقدمه

با توجه به رشد روزافزون سیستم‌های هوشمند حمل و نقل و سرویس‌های مبتنی بر مکان که از داده‌های مکانی- زمانی بهره می‌برند، در سال‌های اخیر، پایگاه‌های داده مکانی- زمانی<sup>۱</sup> با حجم انبوهی از داده‌های مکانی- زمانی اشیا متحرک به وجود آمده است (عباسی‌فرد و نادری، ۲۰۱۷) (مازیم‌پاکا و تیمپاف<sup>۲</sup>، ۲۰۱۶).

افزایش استفاده از سیستم‌هایی که از مکان آگاهی دارند<sup>۳</sup> و ذخیره‌سازی داده‌های مربوط به جابه‌جایی اشیا و افراد در پایگاه داده‌های مستقل و همچنین رشد روش‌های مربوط به شناسایی جابه‌جایی باعث به وجود آمدن داده‌هایی با نام داده خط سیر مکان<sup>۴</sup> شده است. اگر در هنگام ذخیره این داده‌ها بعد زمان مهم باشد، نوع جدیدی از داده با نام داده‌های خط سیر مکانی-زمانی<sup>۵</sup> به وجود می‌آید.

امنیت، پیشگیری و آگاهی به موقع از وقوع حادثه و جرم به عنوان یک موضوع مهم در تمام دوران‌ها به حساب می‌آید. افزایش تعداد عملیات‌های تروریستی و ترس به وجود آمده از آن در بین مردم، این مسئله را به یک موضوع خیلی مهم تبدیل کرده است.

امروزه ۹۰ درصد تجارت جهانی به وسیله حمل و نقل دریایی انجام می‌شود (وسپ، ویسننتینی، بریان و براکا<sup>۶</sup>، ۲۰۱۲). پس در نتیجه، این حجم از کاربرد امنیت و ایمنی بسیار مهم است. مدل‌سازی رفتار کشتی‌ها به افزایش قدرت سیستم‌های نظارتی برای کنترل امنیت و افزایش ایمنی در دنیای دریایی بسیار کمک می‌کند. رفتارهای آنومال در دنیای دریانوردی شامل: ایست‌های غیرمنتظره، انحراف از مسیرهای معمول و تنظیم‌شده، تعارض سرعت و جهت با رفتارهای معمول و تعیین‌شده و ... می‌شود. این رفتارها می‌تواند به ریسک‌های خطرناکی مانند: تصادفات دریایی، قاچاق دریایی، دزدی دریایی و عدم انجام دستور به دلیل مستی مرتبط باشد. برای جلوگیری از این ریسک‌ها برخی منابع داده مانند: سیستم شناسایی اتوماتیک<sup>۷</sup>، رادار دهانه ترکیبی<sup>۸</sup>، رادار با فرکانس بالا موج سطحی<sup>۹</sup>، سنسور مادون سرخ<sup>۱۰</sup>، ویدئو و گزارش‌های هوشمند مورد استفاده قرار می‌گیرند (کاظمی، ابگری و ری‌من<sup>۱۱</sup>، ۲۰۱۳).

1. Spatio-temporal
2. Mazimpaka & Timpf
3. Location-aware
4. Spatial trajectory data
5. spatio-temporal trajectory
6. Vespe, Visentini, Bryan, & Braca
7. Automatic Identification System (AIS)
8. synthetic aperture radar
9. high frequency surface wave radars
10. infra-red sensors
- 11 Kazemi, Abghari, Lavesson, Johnson, & Ryman

داده‌های استفاده‌شده در این تحقیق، بر اساس سیستم خودکار شناسایی<sup>۱</sup> است. سیستم شناسایی خودکار یک سیستم ردیابی خودکار برای شناسایی و تعیین مکان کشتی‌ها و تبادل اطلاعات بین کشتی‌های نزدیک به هم است. این سیستم بر اساس ایستگاه‌های زمینی و ماهواره‌ها فعالیت می‌کند. استفاده از این سیستم، از IMO<sup>۲</sup> از سال ۲۰۰۴ برای افزایش امنیت، ایمنی و افزایش کارایی دریانوردی مورد استفاده قرار گرفته است. این سیستم در جنوا در سال ۱۹۴۸ با نام اولیه<sup>۳</sup> IMCO تأسیس شد؛ سپس در سال ۱۹۸۲، تغییر نام به IMO داد. دفتر مرکزی IMO در شهر لندن در بریتانیا است و از سوی ۱۷۱ کشور سازمان ملل مورد پذیرش قرار گرفته است (IMO، ۲۰۱۸). IMO استاندارد جهانی برای کنترل، افزایش ایمنی و امنیت کشتیرانی در جهان است (بال و ستلایت<sup>۴</sup>، ۲۰۱۳).

بر اساس آیین‌نامه تعریف‌شده در IMO همه کشتی‌های با تناژ ناخالص ۳۰۰ تن و کشتی‌های باربری با تناژ ناخالص ۵۰۰ تن و تمام کشتی‌های مسافربری باید مجهز به سیستم شناسایی اتوماتیک باشند. بیش از چهارصد هزار کشتی در سراسر دنیا مجهز به این سیستم شده‌اند. پیام‌های سیستم شناسایی خودکار به صورت اتوماتیک پخش<sup>۵</sup> می‌شود. به صورت تخمینی می‌توان گفت روزانه بیش از ۸۰ میلیون پیام انتشار می‌یابد. دو دسته از (اطلاعات ایستا<sup>۶</sup> و پویا) از طریق فرستنده‌های سیستم شناسایی خودکار انتشار می‌یابند. اطلاعات ایستا باید در فرستنده گیرنده سیستم شناسایی خودکار برنامه‌ریزی شود. این اطلاعات شامل: شماره هویت خدمات دریایی همراه<sup>۷</sup> (MMSI)، نشانه تماس کشتی<sup>۸</sup>، نوع کشتی و ابعاد کشتی می‌شود. اطلاعات پویا که به صورت اتوماتیک به روزرسانی می‌شود، شامل: مکان کشتی (طول<sup>۹</sup> و عرض<sup>۱۰</sup> جغرافیایی)، جهت دریایی<sup>۱۱</sup> (COG)، سرعت دریایی<sup>۱۲</sup> (SOG)، عنوان<sup>۱۳</sup>، نرخ چرخش<sup>۱۴</sup> (ROT)، زمان در UTC<sup>۱</sup> و سایر اطلاعات ناوبری می‌شود.

1 Automatic Identification Systems (AIS)

2. International Maritime Organization (IMO)

3. Inter-Governmental Maritime Consultative Organization

4. Ball & Satellite

5. Broadcast

6. Static

7. Maritime Mobile Service Identity (MMSI)

8. vessel call sign

9. Longitude

10. Latitude

11. Course Over Ground

12. Speed Over Ground

13. heading

14. Rate of Turn

در این تحقیق، هم اطلاعات ایستا و هم داده‌های پویا برای شناسایی آنومالی در داده‌های خط سیر دریایی مورد استفاده قرار گرفته‌اند. همچنین IMO خطوط ناوبری که باید کشتی‌ها در آنها حرکت کنند، به خصوص قوانینی که در باره جهت‌های مختلف در یک منطقه از دریا برای کشتی‌ها را که معروف به طرح جداسازی ترافیک<sup>۲</sup> (TSS) است، کنترل می‌کند. اولین مسئول طرح اجباری ترافیکی مربوط به طرح‌های جداسازی ترافیک<sup>۳</sup> است. کنترل اینکه کشتی‌ها این قوانین را رعایت می‌کنند، نمی‌تواند تمام نیاز سیستم‌های نظارت دریایی را پاسخ دهد. TSS می‌تواند خطوط جغرافیایی برای کشتیرانی را تأمین کند اما پاسخی برای سرعت در هر منطقه ندارد؛ البته بعضی از مسیرها برای خودشان حداکثر سرعت را قرار داده‌اند ولی همچنان این مسئله نیاز به تحقیق دارد.

### پیشینه تحقیق

خوشه‌بندی فرایندی است که در آن مجموعه‌ای از اشیای فیزیکی و یا انتزاعی را به کلاس‌هایی از اشیا مشابه به یکدیگر تقسیم می‌کند. با توجه به اینکه در این تحقیق، خط سیر می‌تواند دارای انواع اشکال باشد، از الگوریتم خوشه‌بندی بر اساس چگالی استفاده شده است.

DBSCAN<sup>۴</sup> در سال ۱۹۹۶ از سوی مارتین و همکاران ارائه شد. ایده اصلی این الگوریتم آن است که هر نقطه در یک خوشه باید دارای حداقل نقطه همسایه در اطراف خود باشد (دزیکوفسکی و همکاران<sup>۵</sup>، ۱۹۹۶). این الگوریتم دارای دو پارامتر ورودی شعاع همسایگی<sup>۶</sup> و حداقل نقاط همسایگی<sup>۷</sup> است. Eps به معنای شعاع و حداقل نقاط همسایگی به معنای حداقل تعداد همسایه برای منطقه چگال است.

الگوریتم خوشه‌بندی DBSCAN دارای دو مزیت اساسی برای خوشه‌بندی خط سیرهای دریایی است. اولین مزیت این است که این الگوریتم می‌تواند خوشه‌ها را با هر شکلی تشخیص دهد و همچنین می‌تواند داده‌های خارج از محدوده را نیز جداسازی کند. دومین مزیت آن است که برخلاف الگوریتم خوشه‌بندی K-Means نیاز نیست که کاربر تعداد خوشه‌ها را از قبل مشخص کند (روسو<sup>۸</sup>، ۱۹۸۷).

1. Coordinated Universal Time
2. Traffic Separation Schemes
3. <http://www.imo.org/OurWork/Safety/Navigation/Pages/ShipsRouteing.aspx>
4. Density-based spatial clustering of applications with noise
5. Daszykowski et al
6. Eps
7. MinPts
8. Rousseeuw

مهم‌ترین محدودیت الگوریتم خوشه‌بندی برای مسئله این تحقیق این است که این الگوریتم توانایی خوشه‌بندی مکانی را در خود دارد. با توجه به اینکه در مسئله تحقیق سرعت و جهت حرکت در نقطه برای ما مهم است، این محدودیت باید حل شود تا بتوان با استفاده از اطلاعات مکانی، سرعت و جهت، الگوی ترافیک نرمال خط سیر کشتی‌های در حال حرکت را به دست آورد. برای رفع این محدودیت تحقیقات مختلفی در سراسر جهان انجام شده است که در ادامه تعدادی از آنها بیان خواهد شد.

روش  $DB-SMOT^1$  یک روش خوشه‌بندی بر اساس تغییر جهت در حداقل محدوده زمانی است (راک، تایمز، البیورا، آلوارس، و بوگورنی<sup>۲</sup>، ۲۰۱۰). در این الگوریتم اگر دارای توالی تعدادی نقطه باشیم که دارای تغییرات جهت زیاد باشند اما این تغییرات بین اولین نقطه و آخرین نقطه کمتر از آستانه زمانی تعیین شده باشد، این توالی در خوشه نهایی در نظر گرفته نمی‌شود. هدف اصلی این الگوریتم پیدا کردن مکان‌های مد نظر با توجه به خط سیر طی شده به وسیله یک شناور ماهی‌گیری است. این روش با استفاده از ایده موجود در الگوریتم خوشه‌بندی DBSCAN برای نقاطی که دارای تغییر جهت بیشتر از آستانه هستند، مناسب است اما برای شناسایی آنومالی با در نظر گرفتن سرعت تناسبی ندارد. روش دیگری به نام  $CB-SMOT^3$  از سوی پالما و همکاران پیشنهاد شد که در آن الگوریتم سرعت را در نظر گرفت (پالما، بوگورنی، کوچپرز، و آلوارس<sup>۴</sup>، ۲۰۰۸) اما این الگوریتم از سرعت به عنوان یک آستانه استفاده کرده است و پس از آن با در نظر گرفتن نقاط با سرعت پایین خوشه‌بندی را انجام داده است و مکان‌های مد نظر را پیدا کرده است. یکی از مزیت‌های شیوه خوشه‌بندی و شناسایی آنومالی بر اساس نقطه<sup>۵</sup> این است که آنالیز نقاط به گذشته وابسته نیست (لیو، متوین و سیدو<sup>۶</sup>، ۲۰۱۴).

هدف اصلی شناسایی آنومالی شناسایی الگویی در داده است که با آنچه انتظار می‌رود، تطابق ندارد و این الگو با نام‌های آنومالی و یا خارج از محدوده معرفی می‌شود (مهان، هوآنگ، مهروتا<sup>۷</sup>، ۲۰۱۷). شناسایی آنومالی در داده‌های خط سیر دریایی به دو دسته اصلی مدل‌سازی استاتیک<sup>۸</sup>

1. Direction Based Stops and Moves of Trajectories
2. Rocha, Times, Oliveira, Alvares, & Bogorny
3. Clustering-Based Stops and Moves of Trajectories
4. Palma, Bogorny, Kuijpers, & Alvares
5. Point-based
6. Liu, Matwin, & Sydow
7. Mehrotra, Mohan, & Huang
8. Statistical

(لژامر<sup>۱</sup>، ۲۰۰۸) (د وریس و وان سامرن<sup>۲</sup>، ۲۰۱۲) و مدل‌های پیش‌بینی‌کننده<sup>۳</sup> (پالوتا، وسپاس و براین<sup>۴</sup>، ۲۰۱۳) (لین، نول، هایوارد، و بیانی<sup>۵</sup>، ۲۰۱۰) تقسیم می‌شود. ایده عمومی در تکنیک‌های مدل‌سازی استاتیک این است که مدل رفتارهای نرمال در پایگاه داده را شناسایی می‌کنند، سپس از استنتاج استاتیک برای شناسایی وضعیت نقاط دیده‌نشده در مدل استفاده می‌کنند (چندولا، بانرجی و کومار<sup>۶</sup>، ۲۰۰۹). ایده عمومی در تکنیک‌های مدل‌سازی پیش‌بینی‌کننده این است که پس از یادگیری الگوی نرمال در داده، نرمال یا آنومال بودن حرکت یک کشتی را پیش‌بینی می‌کنند؛ سپس این پیش‌بینی را با الگوی نرمال مقایسه می‌کنند و در آخر آنومالی یا نرمال بودن نقطه را شناسایی می‌کنند. البته نوع دیگری از رویکردها برای شناسایی آنومالی وجود دارد که آنومالی را بر اساس قوانین از قبل تعیین شده تعیین می‌کنند. در یک تحقیق که از سوی روی و همکاران<sup>۷</sup> انجام شده است و به دنبال شناسایی آنومالی بوده است. محدودیتی که در این نوع از سیستم‌ها وجود دارد، این است که به قانونی که از قبل تعریف شده است بسیار وابسته هستند. باید سعی کرد برای دنیای دریا که فقط خطوط دریایی TSS از سوی IMO تعیین شده است و سرعت و جهت در این خطوط دلخواه است، فکر کرد تا در مدل پیشنهادی به این مسئله کاملاً توجه شود.

### روش تحقیق

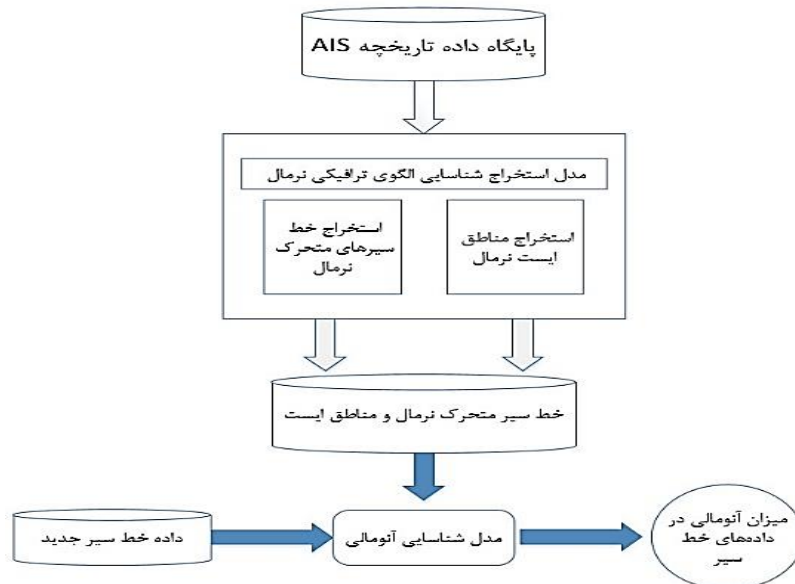
کشتی‌ها در نقاط مختلف دریا دارای الگو<sup>۸</sup> حرکتی متفاوتی هستند. برای مثال، کشتی‌های باری شاید در وسط دریا با سرعت بالا به صورت مستقیم مسیر را طی کنند؛ این در حالی است که در تنگه‌ها و نزدیکی تأسیسات دریایی، اغلب جهت را با توجه به سرعت پایین تنظیم می‌کنند. در نتیجه، همین تغییرات اسپیس‌کپیتر<sup>۹</sup> مدلی با نام ایست و حرکت<sup>۱۰</sup> را معرفی کرد (اسپیس‌کپیتر و همکاران<sup>۱۱</sup>، ۲۰۰۸). پس در نتیجه، همین مدل، خط سیر<sup>۱۲</sup> به عنوان یک دنباله از نقاط حرکت و

1. Laxhammar
2. De Vries & Van Someren
3. predictive
4. Pallotta, Vespe, & Bryan
5. Lane, Nevell, Hayward, & Beaney
6. Chandola, Banerjee, & Kumar
7. Roy et al
8. Pattern
9. Spaccapietra
10. stops and moves
11. Spaccapietra et al
12. Trajectory

ایست‌ها معرفی شده است. در تعدادی از تحقیقات، داده‌های دریایی با آستانه سرعت حرکت به دو قسمت ایست و حرکت تقسیم شده‌اند.

مدل ایست و حرکت دارای دو مزیت اصلی است. اولین مزیت آن است که مکان‌هایی که در آنها ایست انجام می‌شود، معمولاً دارای ارزش اطلاعات مکانی هستند و لازم است به صورت جداگانه نسبت به آنها مطالعه انجام شود. دومین مزیت داده‌های ایست آنجاست که این داده‌ها دارای قابلیت نوپز شدن بالا برای داده‌های متحرک هستند. دلیل این مسئله آن است که این داده‌ها حاوی اطلاعات حرکتی نیستند. در کنار دو مزیت بالا مزیت دیگری که از جداسازی داده‌های متحرک و ایست می‌توان به وضوح دید، این است که این جداسازی باعث کاهش پیچیدگی آنالیز مکانی به وسیله الگوریتم انتخابی می‌شود (پالوتا و همکاران، ۲۰۱۳).

در شکل یک مدل پیشنهادی برای شناسایی آنومالی در داده‌های خط سیر دریایی را می‌بینید. در این تحقیق، از مدل حرکت، ایست با آستانه سرعت ۵،۰ نات<sup>۱</sup> استفاده شده است. دلیل انتخاب این آستانه فقط تجربه است. اکثر رکوردهای قسمت داده‌های ایست دارای سرعت صفر هستند و این عدد این آستانه‌گذاری تأثیر زیادی در نتیجه خوشه‌بندی ندارد.



شکل ۱: چارچوب پیشنهادی برای شناسایی آنومالی در داده‌های خط سیر دریایی

### استخراج الگو متحرک نرمال<sup>۱</sup>

برای استخراج الگو حرکتی نرمال از داده‌های متحرک سیستم شناسایی خودکار<sup>۲</sup> از الگوریتم خوشه‌بندی DBSCANSD استفاده شده است. در ادامه این الگوریتم خوشه‌بندی و دلایل استفاده شرح داده خواهد شد. الگوریتم خوشه‌بندی DBSCANSD<sup>۳</sup> با روش خوشه‌بندی بر اساس چگالی و با در نظر گرفتن سرعت و جهت حرکت در هر نقطه کار می‌کند. همان طور که از اسم این الگوریتم مشخص است، این الگوریتم توسعه یافته الگوریتم خوشه‌بندی DBSCAN است. مسئله اصلی که این الگوریتم به دنبال پاسخ دادن به آن است این است که بتوان شناورهای مختلف را که دارای سرعت و جهت‌های مختلف در مکان‌های یکسان هستند، خوشه‌بندی کند. برای مثال، ممکن است یک کشتی ماهی‌گیری در یک مکان خاص دارای سرعت نابرابر با یک کشتی باری باشد. حتی دو کشتی یکسان ممکن است با داشتن دو سرعت برابر ولی دارای جهت‌گیری متفاوت نسبت به هم باشند. مثلاً کشتی حمل سوخت زمانی که دارای سوخت است، سرعتش کاملاً با زمانی که خالی است متفاوت است.

الگوریتم DBSCANSD به روز رسانی شده الگوریتم DBSCAN است که دو فاکتور حداکثر تفاوت سرعت<sup>۴</sup> و حداکثر تفاوت جهت<sup>۵</sup> بین دو نقطه همسایه را به الگوریتم اضافه کرده است. در حقیقت در این الگوریتم اینکه دو نقطه از نظر مکانی نزدیک باشند کافی نیست و این دو نقطه باید دارای سرعت<sup>۶</sup> و جهت<sup>۷</sup> شبیه به هم باشند. بنابراین، DBSCANSD مفهوم شعاع همسایگی<sup>۸</sup> که در الگوریتم DBSCAN موجود است را به این مفهوم تغییر داده است.

در پایگاه داده D مسیرهای متحرک در یک منطقه مشخص شعاع همسایگی<sup>۹</sup> نقطه P که با E

نشان داده شده است برابر است با:

$$N_{\epsilon}(p) = \{ \emptyset D \mid \text{dist}(p, q) < \epsilon \text{ and } |p.SOG - q.SOG| < \text{MaxSpd} \\ \text{and } |p.COG - q.COG| < \text{MaxDir} \}$$

- 
1. Normal Moving
  2. Automatic Identification Systems (AIS)
  3. Density-Based Spatial Clustering of Applications with Noise considering Speed and Direction
  4. maximum speed variance (MaxSpd)
  5. maximum direction variance (MaxDir)
  6. COG (Course Over Ground)
  7. SOG (Speed Over Ground)
  8. Eps-neighborhood
  9. Eps-neighborhood

منظور از فاصله  $dist(p, q)$  در این الگوریتم فاصله جغرافیایی هیورسان<sup>۱</sup> است (ون بروملن<sup>۲</sup>، ۲۰۱۳) (ونس<sup>۳</sup>، ۲۰۰۲). دلیل استفاده از این معیار فاصله این است که با توجه به داده‌های مکانی-زمانی که در سطح زمین اتفاق می‌افتند، ضروری است انحنای سطح کره زمین را در هنگام محاسبه فاصله بین دو نقطه در نظر گرفت.

در ادامه الگوریتم خوشه‌بند DBSCANSD شرح داده خواهد شد. این الگوریتم دارای پنج پارامتر، شعاع همسایگی<sup>۴</sup>، حداقل نقاط همسایه<sup>۵</sup>، حداکثر تفاوت جهت<sup>۶</sup>، پایگاه داده M و حداکثر تفاوت سرعت<sup>۷</sup> است. پایگاه داده M لیست تمام نقاط متحرک در تمامی خط سیرهاست. Eps برابر است با فاصله در دسترس<sup>۸</sup> و حداقل نقاط همسایه برابر با حداقل نقاط در دسترس<sup>۹</sup> (دقیقاً همان پارامترهای ورودی الگوریتم خوشه‌بندی DBSCAN) است. پارامتر حداکثر تفاوت جهت برابر است با حداکثر تفاوت جهت<sup>۱۰</sup> و حداکثر تفاوت سرعت برابر با حداکثر تفاوت سرعت دو نقطه مجاور<sup>۱۱</sup> است. الگوریتم نقطه شروع را به صورت تصادفی انتخاب می‌کند و سپس با توجه به حداقل نقاط همسایه، همسایه هر نقطه را مشخص می‌کند. نقطه به عنوان نقطه‌هسته<sup>۱۲</sup> تنظیم می‌شود و نقاط همسایه در خوشه این نقطه قرار می‌گیرند. این فرایند تا زمانی ادامه دارد که همه نقاط مشاهده شوند و در نهایت، خوشه‌ها در خروجی قرار می‌گیرند (لیو و همکاران<sup>۱۳</sup>، ۲۰۱۴).

#### انتخاب پارامترهای ورودی DBSCANSD

برای انتخاب میزان حداقل نقاط همسایه به صورت تجربی عمل شده است؛ البته نباید این پارامتر ورودی را خیلی کوچک و یا خیلی بزرگ در نظر گرفت. برای تعیین پارامتر ورودی شعاع همسایگی از روش استفاده شده در (رحما و سیتانگنگ<sup>۱۴</sup>، ۲۰۱۶) (الباتا و آشور<sup>۱۵</sup>، ۲۰۱۳)

1. Haversine formula
2. Van Brummelen
3. Veness
4. Eps
5. MinPts
6. MaxDir
7. MaxSpd
8. Reachable distance
9. Minimum number of points
10. Maximum direction variance
11. Maximum speed variance
12. CorePoint
13. Liu et al
14. Rahmah & Sitanggang
15. Elbatta & Ashour

(کاسامبرا<sup>۱</sup>، ۲۰۱۸) استفاده شده است. با توجه به حداقل نقاط همسایه انتخابی در الگوریتم خوشه‌بندی، نمودار فاصله  $Knn^2$  رسم شده است و با استفاده از روش موجود در مقاله (ساتوپا، البرچتفر و رگوان<sup>۲</sup>، ۲۰۱۱) نقطه‌ای که دارای بیشترین تغییر در نمودار<sup>۴</sup> است، پیدا شده است. این نقطه برابر با مقدار شعاع<sup>۵</sup> مناسب در مسئله این تحقیق است.

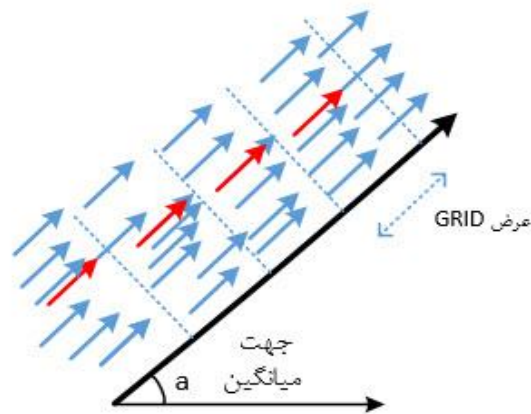
برای انتخاب پارامتر ورودی سرعت و جهت نیز از نمودار فاصله  $Knn$  به عنوان راهنما برای شروع استفاده شده است ولی در نهایت، این دو پارامتر ورودی به صورت تجربی تعیین شده است. باید برای تعیین این دو پارامتر ورودی به کل مسیرها و جهت‌ها در یک TSS با توجه به قوانین IMO توجه کرد. توجه به این نکته ضروری است که اگر این دو پارامتر خیلی کوچک تعیین شود، می‌تواند زمینه‌ای را فراهم کند تا بسیاری از نقاط آنومال شناسایی شود.

### بردار گرانش

اگر چه خوشه‌بندی به وسیله DBSCANSD می‌تواند الگو حرکتی نرمال شناورها را نشان دهد ولی مقایسه نقاط موجود در یک خط سیر با کل نقاط موجود در یک خوشه دارای پیچیدگی زمانی  $O(n * m)$  است که در این فرمول  $n$  برابر با تعداد نقاط موجود در یک خوشه و  $m$  برابر با مجموعه نقاط موجود در یک خط سیر جدید است. این پیچیدگی زمانی با بزرگ شدن پایگاه داده، بزرگ و بزرگ‌تر می‌شود. برای کاهش این پیچیدگی زمانی بهترین راه‌حل نماینده‌سازی<sup>۶</sup> از خوشه‌هاست.

در این تحقیق، از ایده بردار گرانش<sup>۷</sup> (GV) موجود در مقاله (لیو و همکاران، ۲۰۱۴) استفاده شده است. این ایده از ایده مرکزی‌سازی و پاکت‌نامه<sup>۸</sup> که برای مدل‌سازی و یادگیری خط سیر تصویری طراحی شده است، سرچشمه گرفته است (موریس و تریودی<sup>۹</sup>، ۲۰۰۸). شکل ۲ مثالی از نحوه محاسبه بردار گرانش در یک خوشه است. همان طور که در شکل ۲ می‌بینید تمام بردارهای آبی رنگ متعلق به یک خوشه هستند. همان طور که در شکل مشخص است، تمام نقاط موجود در

- 
1. Kassambara
  2. Knn-dist-plot
  3. Satopaa, Albrecht, Irwin, & Raghavan
  4. Kneedle point
  5. Eps
  6. Labeling
  7. Gravity Vecto.r (GV)
  8. Centroid And Envelope
  9. Morris & Trivedi



شکل ۲: بردار گرانش در یک خوشه

یک خوشه دارای جهت<sup>۱</sup> شبیه به یکدیگر هستند. سرعت<sup>۲</sup> نقاط موجود در یک خوشه نیز با توجه به استفاده از الگوریتم DBSCANSD برای خوشه‌بندی دارای تفاوت کمینه است. در شکل ۲ خوشه به ۴ قسمت تقسیم شده است؛ پس در نتیجه، این خوشه دارای ۴ بردار گرانش است. بردار گرانش برداری است که شامل پنج ویژگی: سرعت میانگین<sup>۳</sup>، جهت میانگین<sup>۴</sup>، عرض جغرافیایی میانگین<sup>۵</sup>، طول جغرافیایی میانگین<sup>۶</sup> و فاصله میانه<sup>۷</sup> است. یک بردار گرانش را می‌توان به صورت زیر نمایش داد:

$$GV_i = \langle COG_{avg}, SOG_{avg}, LAT_{avg}, LON_{avg}, D_{median} \rangle$$

### استخراج مناطق توقف عادی

در مناطق ایست، مانند بنادر و اسکله‌ها و مناطقی از دریا که ایستادن برای ماهیگیری و... ممکن است، انواع مختلف کشتی‌ها لنگر انداخته باشند و یا کشتی‌ها با سرعت بسیار پایین در حال ورود و حرکت باشند. در نتیجه، جهت و سرعت حرکت برای محاسبه ریسک برخورد ویژگی مهمی نیستند. تنها در صورتی می‌توان خطر تصادف را بالا دانست که شناورها با سرعت بالا در مکان‌هایی که

1. COG
2. SOG
3. average SOG
4. average COG
5. average Latitude
6. average Longitude
7. Median Distance

مخصوص ایست هستند، عبور کنند. پس در این قسمت، ما باید مناطقی که سرعت پایین و یا صفر است را با استفاده از تاریخچه مکانی موجود در آن مناطق شناسایی کنیم. پس از شناسایی این مناطق می‌توانیم حرکت‌های با سرعت پایین در مناطق نامناسب را شناسایی کنیم. با توجه به عدم نیاز به در نظر گرفتن سرعت و جهت حرکت در پایگاه داده از جنس ایست و فقط با در نظر گرفتن موقعیت مکانی شناور با استفاده از الگوریتم خوشه‌بندی HDBSCAN به منظور خوشه‌بندی استفاده شده است (مولوی، زیمک و سندر<sup>۱</sup>، ۲۰۱۵) (هیلی و مسی‌النس<sup>۲</sup>، ۲۰۱۷).

HDBSCAN الگوریتم خوشه‌بندی سلسله‌مراتبی است که مانند DBSCAN براساس چگالی کار می‌کند. زمانی باید از این الگوریتم استفاده شود که منطقه مورد نظر دارای تغییر چگالی زیاد است. در داده‌های ایستا چگالی کاملاً متغیر و نابرابر است. از طرف دیگر، دو منطقه (خوشه) ایست می‌توانند در فاصله نزدیک به هم باشند و نباید یک خوشه در نظر گرفته شوند. الگوریتم‌هایی مانند DBSCAN با توجه به عدم توجه به تفاوت چگالی نمی‌توانند این دو منطقه را به درستی تشخیص دهند. مزیت دیگر این الگوریتم خوشه‌بندی آن است که تنها نیازمند یک پارامتر ورودی است.

### شناسایی آنومالی در خط سیرهای جدید

در این قسمت نحوه شناسایی آنومالی یک خط سیر با توجه به نتایج خوشه‌بندی در مدل تقسیم‌بندی پایگاه داده به دو قسمت ایست و متحرک شرح داده خواهد شد. شیوه شناسایی آنومالی پیشنهادی در این مقاله با رویکرد براساس نقطه است و این یعنی برای خارج از محدوده بودن هر نقطه از خط سیر به صورت مجزا عمل می‌شود. این رویکرد باعث می‌شود مدل پیشنهادی قابل استفاده به صورت بلادرنگ باشد.

### شناسایی خارج از محدوده در داده‌های ایستا

برای شناسایی خارج از محدوده در داده‌های ایست از الگوریتم شناسایی خارجی از محدوده GLOSH استفاده شده است (کزامبرا<sup>۳</sup>، ۲۰۱۸). GLOSH چگالی نقطه را با چگالی در خوشه یا زیر خوشه (اگر زیر خوشه‌ای وجود داشته باشد) مقایسه می‌کند. نقطه‌ای که دارای چگالی خیلی پایین باشد و با هیچ خوشه‌ای سازگار نباشد، خارج از محدوده تشخیص داده می‌شود (موریس و تریودی<sup>۴</sup>، ۲۰۰۸).

- 
1. Campello, Moulavi, Zimek, & Sander
  2. McInnes & Healy
  3. Kassambara
  4. Morris & Trivedi

### شناسایی خارج از محدوده در داده‌های متحرک

از دو معیار فاصله  $RDD(R_{relative})$  و  $CDD(D_{cosine})$  برای شناسایی آنومالی در نقاط در حال حرکت استفاده شده است. RDD برابر با نسبت فاصله نقاط با مرکز به فاصله میانه<sup>۳</sup> است. فاصله میانه برابر با میانه لیست فاصله مکانی میانه هر قسمت از خوشه با تمام نقاط موجود در آن قسمت از خوشه است. هر خوشه می‌تواند دارای چند GV باشد (هیلیارد، سوزا و همکاران<sup>۴</sup>، ۲۰۱۵).

$$D_{relative} = \frac{Distance((P_i.Lat, P_i.Lon), (GV.Lat, GV.lon))}{GV.MedianDistance}$$

در CDD از سرعت و جهت استفاده شده است. این فاصله بین نقطه هدف  $P_i$  و بردار گرانش GV و برابر است با:

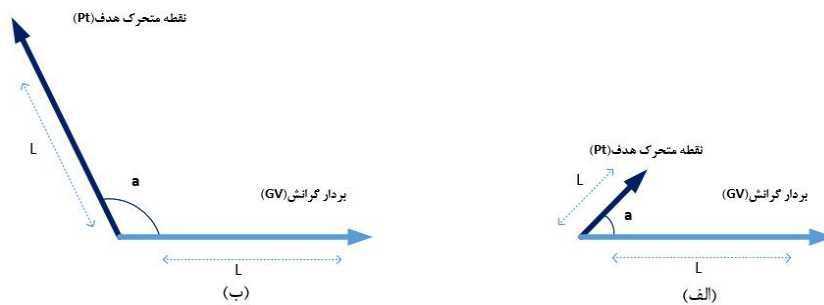
$$D_{cosine} = \cos \theta * \frac{\min(P_i.SOG, GV.SOG)}{\max(P_i.SOG, GV.SOG)}$$

در فرمول بالا  $\theta$  برابر است با زاویه بین COG نقطه هدف و بردار گرانش است. CDD به دنبال آن است که زاویه ( $180^\circ$ ) بین نقطه هدف و بردار گرانش (که برابر است با مقدار تفاوت COG نقطه و بردار گرانش) را با تفاوت سرعت آنها ترکیب کند.

شکل ۳ شامل دو حالت را که در آن COG و SOG در نظر گرفته شده است و آنومالی اتفاق می‌افتد، نمایش می‌دهد. در شکل ۳ (ب) هر دو بردار دارای طول L برابر هستند ولی زاویه بین آنها با هم بسیار زیاد است. این حالت می‌تواند به معنی این باشد که شناور در یک خط دریایی نرمال ولی در خلاف جهت حرکت می‌کند. این حالت می‌تواند باعث برخورد شود.

در شکل ۳ (الف) سرعت شناور بسیار پایین‌تر از سرعت GV است. این در حالتی است که میزان تفاوت COGها کم است و این یعنی جهت شناور طبیعی است. در چنین شرایطی، شناور با سرعت پایین ریسک برخورد در دریا را بالا می‌برد. در چنین شرایطی CDD می‌تواند هم سرعت و هم جهت را برای شناسایی آنومالی در نظر بگیرد.

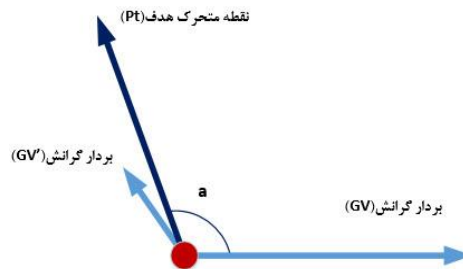
- 
1. Relative Division Distance
  2. Cosine Division Distance
  3. Median distance
  - 4 Liu, Souza, Hilliard, & Matwin



شکل ۳: دو حالت آنومال پس از در نظر گرفتن OGCOG و S

یکی از ایده‌های جایگزین به جای CDD محاسبه تفاوت سرعت و جهت به صورت جداگانه است. این ایده می‌تواند سبب افزایش اشتباهات سیستم شود. شکل ۴ شامل یکی از سناریوهای آنومال این ایده است. اگر یک نقطه در مکانی قرار داشته باشد که دو GV به آن نزدیک باشند (مثلاً بین دو خوشه)، در این حالت، یکی از GV ها به شما شرایط را طبیعی نشان می‌دهد؛ این در صورتی است که GV دیگر این نقطه را آنومال نشان می‌دهد.

برای تعیین میزان آستانه CDD و RDD برای تشخیص نرمال یا آنومال بودن یک نقطه از خط سیر در ابتدا میزان حداقل RDD و حداکثر CDD هر نقطه از خط سیرهای موجود در پایگاه داده آزمون به دست می‌آید. پس از آن میزان RDD و CDD کل با استفاده از نمودار هیستوگرام RDD و CDD همه نقاط موجود در پایگاه داده آزمون تعیین می‌شود



شکل ۴: اگر نقطه بین دو خوشه قرار بگیرد

### الگوریتم شناسایی آنومالی

برای شناسایی میزان آنومالی یک مسیر جدید مراحل زیر را طی می‌کنیم:

۱. در ابتدا نقاط خط سیر موجود در پایگاه داده را با توجه به ترتیب موجود در پایگاه داده و MMSI ثبت شده شناورها مرتب می‌شود؛
۲. در مرحله دوم، با توجه به آستانه سرعت تنظیم‌شده در هنگام خوشه‌بندی در مورد ایستا یا متحرک بودن هر نقطه موجود در خط سیرها<sup>۱</sup> تصمیم گرفته می‌شود؛
۳. در مرحله سوم، وضعیت نرمال یا آنومال بودن هر نقطه از خط سیر مشخص می‌شود؛ اگر نقطه جزء داده‌های ایستا باشد، با استفاده از فایل باینری خروجی HDBSCAN خوشه آن تعیین می‌شود و اگر عضو خوشه‌ای نبود، آن نقطه به عنوان خارج از محدوده<sup>۲</sup> ثبت می‌شود اما اگر نقطه دارای سرعت بالاتر از سرعت آستانه بود، با استفاده از بردار گرانش و دو مقیاس فاصله RDD و CDD، طبیعی و یا خارج از محدوده بودن آن نقطه تعیین می‌شود. اگر حداقل RDD آن نقطه در مقایسه با GV های مختلف بیشتر از آستانه تعیین‌شده از سوی کاربر باشد و یا اینکه دارای حداکثر CDD کمتر از آستانه تعیین‌شده از طرف کاربر باشد، خارج از محدوده شناسایی می‌شود؛
۴. در مرحله آخر تعداد نقاط خارج از محدوده هر خط سیر با هم جمع می‌شود و در نهایت، با درصدگیری نسبت به کل نقاط همان خط سیر، میزان آنومالی آن خط سیر تعیین می‌شود.

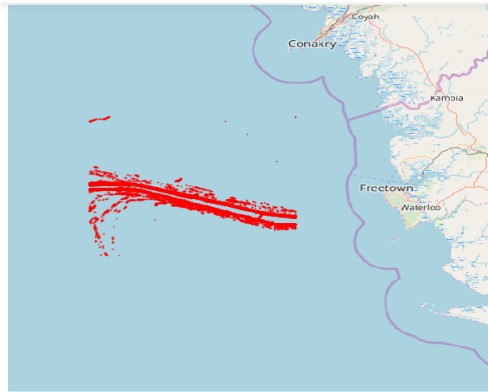
### ارزیابی مدل پیشنهادی

برای پیاده‌سازی مدل پیشنهادی از زبان برنامه‌نویسی پایتون<sup>۳</sup> و برای افزایش کارایی و سرعت محاسبات از سایتون<sup>۴</sup> استفاده شده است.

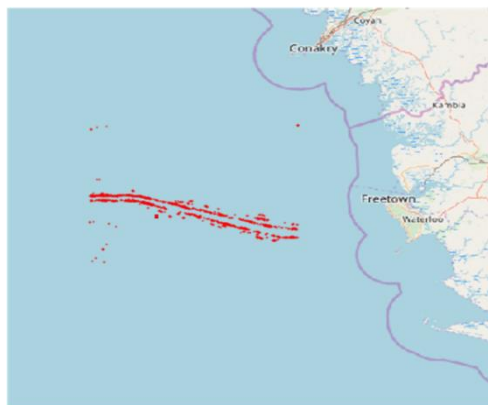
برای ارزیابی مدل پیشنهادی، از پایگاه داده گردآوری شده در مقاله (لیو و همکاران<sup>۵</sup>، ۲۰۱۴) استفاده شده است. در این پایگاه داده در هنگام ساخت کاملاً به قوانین IMO و TSS توجه شده است. با توجه به نیاز، از ۳۴۰۰۰ سطر اول از این پایگاه داده استفاده شده است. برای نزدیک کردن شرایط ارزیابی‌ها به شرایط دنیای واقعی، این پایگاه داده به دو قسمت آموزش و آزمون تقسیم شده است.

---

1. Trajectory  
2. OUTLIER  
3. <https://www.python.org/>  
4. <https://cython.org/>  
5. Liu et al



شکل ۵: مسیر رفت و آمد شناورها در پایگاه داده

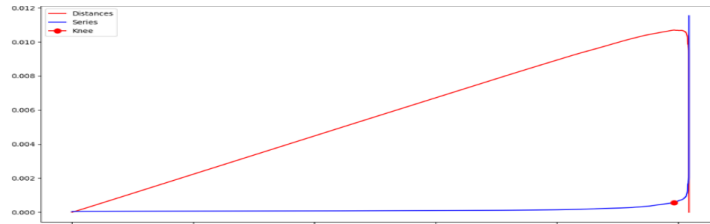


شکل ۶: مسیر رفت و آمد شناورها مربوط به قسمت  
آزمون پایگاه داده یک پایگاه داده یک

۳۲۰۰۰ سطر اول از این پایگاه داده را برای آموزش و ۲۰۰۰ سطر آخر از این پایگاه داده برای آزمون استفاده شده است. نقاط موجود در پایگاه داده آزمون در هنگام آزمون برای شناسایی الگوی نرمال متحرک و ایستا استفاده نشده است. همان‌طور که در شکل ۵ مشاهده می‌شود، در این پایگاه داده شناورها در دو خط دریایی مجاور هم در حال حرکت هستند. در شکل ۶ مسیر رفت و آمد کشتی‌ها در قسمت آزمون را می‌بینید.

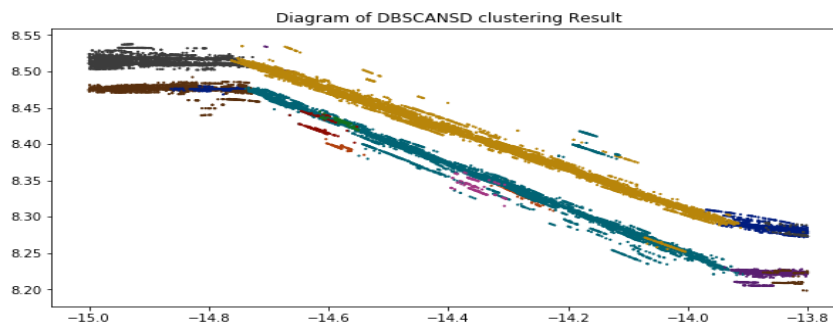
### خوشه‌بندی داده‌های متحرک

تعداد حداقل نقاط همسایه<sup>۱</sup> برابر با ۵۰ تعیین شد. مقدار شعاع<sup>۲</sup> در DBSCANSD با استفاده از نمودار فاصله knn برابر با ۳۶۰۰ متر تعیین شده است. در شکل ۷ این نمودار را می‌بینید. نقطه قرمز در شکل بهترین مکان برای میزان Eps است؛ البته کمی بالاتر یا پایین‌تر از این مقدار ایرادی ندارد. از آنجا که در هنگام محاسبه فاصله‌ها از رادیان استفاده شده است، در آخر باید مقدار خروجی را در شعاع زمین ضرب کنیم تا مقدار واقعی را به متر به دست آوریم. میزان تفاوت سرعت برابر با ۱,۵ و میزان تفاوت جهت برابر با ۲,۵ تنظیم شده است.



شکل ۷: نمودار فاصله Knn برای محاسبه میزان Eps در پایگاه داده

با توجه به تنظیمات انجام‌شده، الگوریتم خوشه‌بندی DBSCANSD داده‌ها را در ۲۰ خوشه مختلف خوشه‌بندی کرده است. در شکل ۸ نتیجه این خوشه‌بندی را به صورت رنگ‌بندی شده می‌بینید. با توجه به میزان عرض GRID تعیین‌شده که برابر یک درصد تنظیم کرده‌ایم، ۲۳۴ نقطه به عنوان GV انتخاب شده‌اند. میزان عرض GRID کاملاً در تعداد این نقاط تأثیرگذار است. این مقدار به صورت تجربی به دست آمده است.

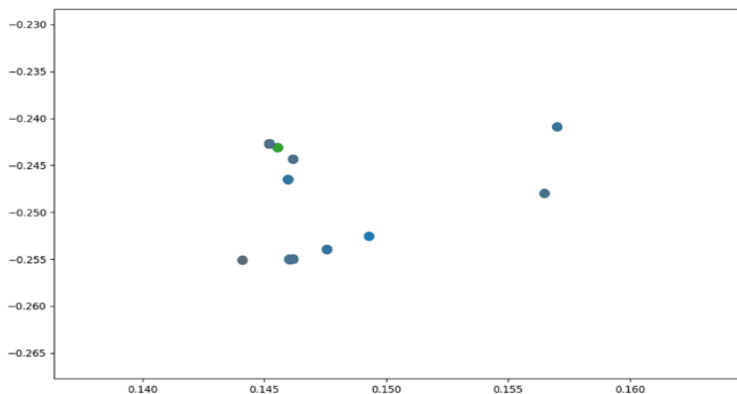


شکل ۸: تصویر رنگ‌بندی نتیجه خوشه‌بندی DBSCANSD برای نقاط متحرک قسمت آموزش

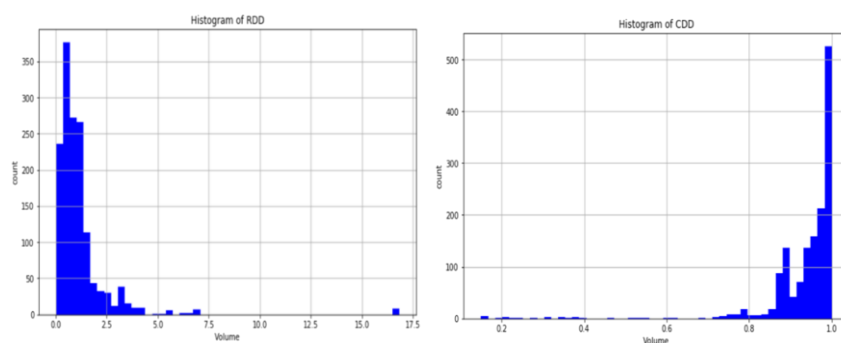
1. MinEps
2. Eps

### خوشه‌بندی نقاط ایستا

برای خوشه‌بندی داده‌های ایستا با الگوریتم خوشه‌بندی HDBSCAN میزان حداقل اندازه خوشه برابر با ۵۰ تنظیم شده است. بهتر است که این مقدار برابر با حداقل نقطه همسایه<sup>۱</sup> در الگوریتم DBSCAN باشد. همان‌طور که در شکل ۹ می‌بینید، HDBSCAN نقاط ایستا در چهار خوشه خوشه‌بندی کرده است.



شکل ۹: شکل نتایج خوشه‌بندی HDBSCAN برای نقاط ایستا قسمت آموزش



شکل ۱۰: نمودار هیستوگرام حداقل RDD و حداکثر CDD نقاط موجود در داده‌های آموزش

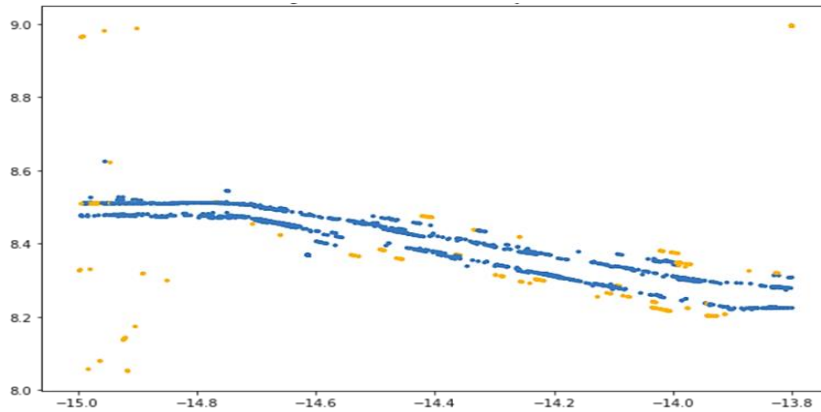
### 1. MinPts

### نتایج شناسایی آنومالی

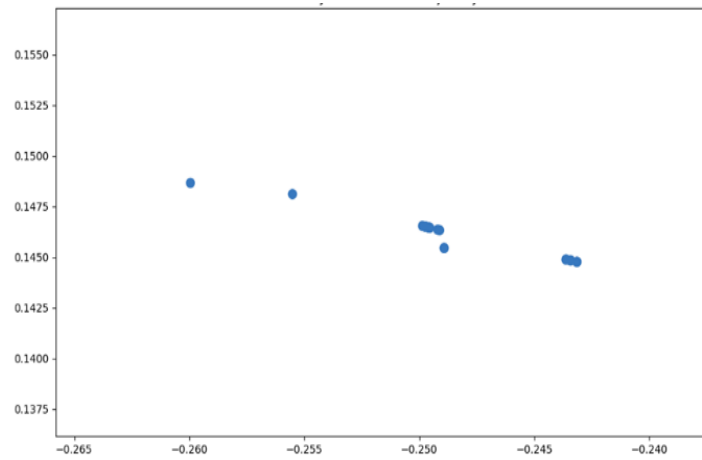
پس از بررسی نمودارهای هیستوگرام CDD و RDD، میزان آستانه CDD را برابر با ۰.۸ و میزان آستانه RDD را برابر با ۲ تنظیم شده است.

در شکل ۱۱ نتیجه خروجی میزان آنومالی در بین تمام نقاط را می‌بینید. رنگ آبی نشان نقاطی است که نرمال تشخیص داده شده‌اند.

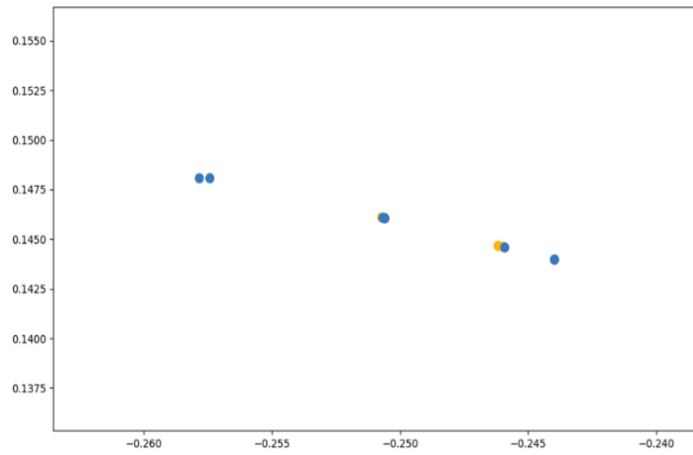
در ادامه تعدادی از خط سیرهای طی شده به وسیله کشتی‌های مختلف و میزان آنومالی تشخیص داده شده برای هر مسیر را می‌بینید. رنگ آبی نقاطی از هر خط سیر است که نرمال هستند.



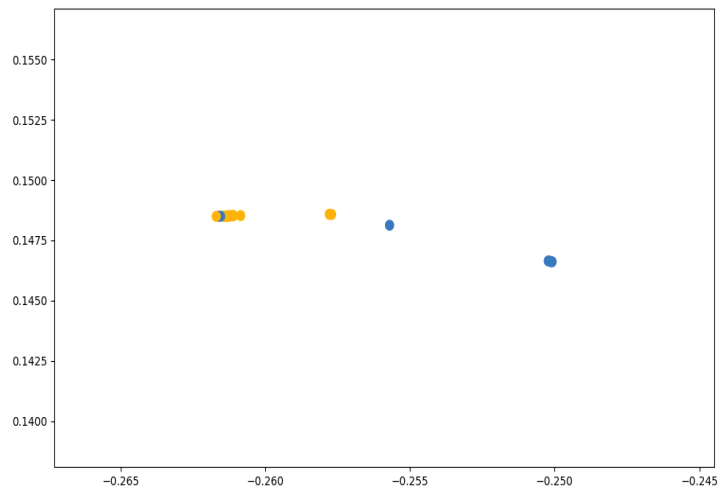
شکل ۱۱: نتایج خروجی شناسایی آنومالی در قسمت آزمون پایگاه داده



شکل ۱۲: خط سیر کشتی با شماره ۷۰۹ که کاملاً نرمال تشخیص داده شده است



شکل ۱۳: خط سیر طی شده کشتی با شماره ۵۹۴ که ۲۷ درصد آنومال تشخیص داده شده است

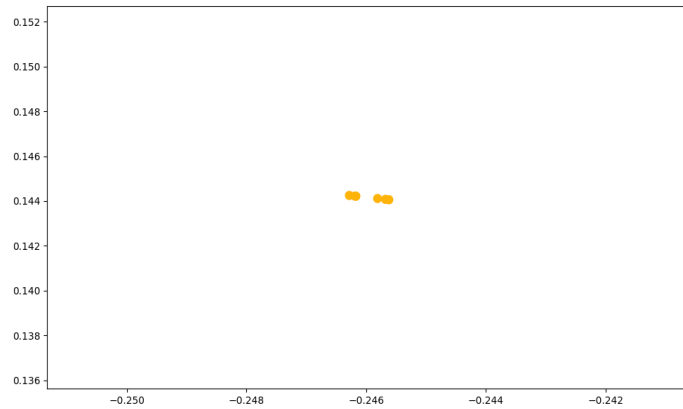


شکل ۱۴: خط سیر طی شده کشتی با شماره ۸۴۷ که دارای ۷۴ درصد آنومالی شناسایی شده است

## نتیجه‌گیری

در این مقاله، ما در ابتدا با توجه به آستانه سرعت تعیین شده پایگاه داده خود را با تبعیت از مدل ایست و حرکت به دو قسمت متحرک و ایست تقسیم کرده‌ایم؛ سپس با استفاده از الگوریتم خوشه‌بندی بر اساس چگالی DBSCANSD برای نقاط متحرک و HDBSCAN برای نقاط ایستا، توانستیم الگوی نرمال حرکتی در پایگاه داده با اطلاعات مکانی-زمانی سیستم شناسایی خودکار (AIS) را شناسایی کنیم.

برای خوشه‌بندی داده‌های متحرک از طول و عرض جغرافیایی، سرعت و جهت ذخیره شده از حرکت کشتی‌ها در هر نقطه از خط سیرشان استفاده شده است. برای انتخاب پارامتر ورودی Eps در الگوریتم خوشه‌بندی DBSCANSD از نمودار فاصله Knn استفاده شده است. برای انتخاب سایر پارامترهای ورودی مورد نیاز به صورت تجربی عمل شده است. برای خوشه‌بندی و شناسایی مکان‌های ایست نرمال در محدوده دریایی پایگاه داده از طول و عرض جغرافیایی استفاده شده است.



شکل ۱۵: خط سیر طی شده کشتی با شماره ۴۸۶ که ۱۰۰ درصد انومال شناسایی شده است.

استفاده از الگوریتم خوشه‌بندی بر اساس چگالی باعث شد تا بتوانیم هر خط سیر با هر الگویی را بتوانیم خوشه‌بندی کنیم.

مدل شناسایی آنومالی در این تحقیق از نوع استاتیک و با استفاده از رویکرد براساس نقطه است. این رویکرد قابل استفاده در سیستم‌های شناسایی آنومالی به صورت بلادرنگ است. از دو فاصله RDD و CDD در نقاط متحرک به عنوان معیار طبیعی (نرمال) یا خارج از محدوده بودن هر نقطه از نقاط دارای سرعت بالاتر از آستانه تعیین شده استفاده شده است. برای تنظیم مقدار این دو پارامتر

از نمودار هیستوگرام حداقل RDD و حداکثر CDD نقاط موجود در داده‌های آزمون استفاده شده است. برای شناسایی نقاط ایست خارج از محدوده از الگوریتم شناسایی خارج از محدوده GLOSH استفاده شده است. تنظیم‌پذیر بودن RDD و CDD برای داده‌های متحرک سبب شده است که میزان دقت مدل پیشنهادی قابل تنظیم باشد که این امر باعث کاربردی بودن این تحقیق در دنیای واقعی می‌شود. استفاده از محاسبات توزیع‌شده یکی از موضوعاتی است که می‌تواند باعث افزایش سرعت الگوریتم‌های خوشه‌بندی در هنگام به دست آوردن الگوی حرکتی نرمال شود.

## منابع

- Abbasifard, M. R., & naderi, hassan. (2017). Indexing the past and current position of moving objects in large-scale dataset. *Journal of Geospatial Information Technology*, 5(2), 141–162. <https://doi.org/10.29252/jgit.5.2.141>
- Ball, H., & Satellite, A. I. S. (2013). *Satellite AIS for Dummies*. Mississauga, ON: Wiley.
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 160–172). Springer.
- Campello, R. J. G. B., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(1), 5.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 15.
- Daszykowski, M., Walczak, B., Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, pp. 226–231). <https://doi.org/10.1016/B978-044452701-1.00067-3>
- De Vries, G. K. D., & Van Someren, M. (2012). Machine learning for vessel trajectories using compression, alignments and domain knowledge. *Expert Systems with Applications*, 39(18), 13426–13439.
- Elbatta, M. T. H., & Ashour, W. M. (2013). A dynamic method for discovering density varied clusters. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(1), 123–134.
- IMO. (2018). About IMO. <https://doi.org/10.1007/bf02532644>
- Kassambara, A. (2018). DBSCAN: Density-Based Clustering Essentials - Datanovia. Retrieved February 15, 2019, from <https://www.datanovia.com/en/lessons/dbscan-density-based-clustering-essentials/>
- Kazemi, S., Abghari, S., Lavesson, N., Johnson, H., & Ryman, P. (2013). Open data for anomaly detection in maritime surveillance. *Expert Systems with Applications*, 40(14), 5719–5729.
- Lane, R. O., Nevell, D. A., Hayward, S. D., & Beaney, T. W. (2010). Maritime anomaly detection and threat assessment. In *Information Fusion (FUSION), 2010 13th Conference on* (pp. 1–8). IEEE.
- Laxhammar, R. (2008). Anomaly detection for sea surveillance. In *Information Fusion, 2008 11th International Conference on* (pp. 1–8). IEEE.
- Liu, B., de Souza, E. N., Matwin, S., & Sydow, M. (2014). Knowledge-based clustering of ship trajectories using density-based approach. In *2014 IEEE International Conference on Big Data (Big Data)* (pp. 603–608). IEEE. <https://doi.org/10.1109/BigData.2014.7004281>
- Liu, B., Souza, E. N. De, Hilliard, C., & Matwin, S. (2015). Ship Movement

- Anomaly Detection Using Specialized Distance Measures. *2015 18th International Conference on Information Fusion, Fusion 2015*, 1113–1120. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84960498152&partnerID=40&md5=5ec29d27f2707d1e546d7d47d68b008a>
- Mazimpaka, J. D., & Timpf, S. (2016). Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science*, 2016(13), 61–99.
- McInnes, L., & Healy, J. (2017). Accelerated Hierarchical Density Based Clustering. In *Data Mining Workshops (ICDMW), 2017 IEEE International Conference on* (pp. 33–42). IEEE.
- Mehrotra, K. G., Mohan, C. K., & Huang, H. (2017). *Anomaly Detection Principles and Algorithms*. Springer.
- Morris, B. T., & Trivedi, M. M. (2008). A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8), 1114.
- Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy*, 15(6), 2218–2245.
- Palma, A. T., Bogorny, V., Kuijpers, B., & Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. In *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 863–868). ACM.
- Rahmah, N., & Sitanggang, I. S. (2016). Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra. *IOP Conference Series: Earth and Environmental Science*, 31(1). <https://doi.org/10.1088/1755-1315/31/1/012012>
- Rocha, J. A. M. R., Times, V. C., Oliveira, G., Alvares, L. O., & Bogorny, V. (2010). DB-SMoT: A direction-based spatio-temporal clustering method. In *Intelligent systems (IS), 2010 5th IEEE international conference* (pp. 114–119). IEEE.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops* (pp. 166–171). IEEE. <https://doi.org/10.1109/ICDCSW.2011.20>
- Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., & Vangenot, C. (2008). A conceptual view on trajectories. *Data & Knowledge Engineering*, 65(1), 126–146.
- Van Brummelen, G. (2013). *Heavenly mathematics: the forgotten art of spherical trigonometry*. Princeton University Press. Retrieved from

<https://books.google.com/books?id=0BCCz8Sx5wkC&pg=PR7>

Veness, C. (2002). Calculate Distance and Bearing between Two Latitude/Longitude Points Using Haversine Formula in JavaScript. *MIT Open Source*, 1–39. Retrieved from <https://www.movable-type.co.uk/scripts/latlong.html>

Vespe, M., Visentini, I., Bryan, K., & Braca, P. (2012). Unsupervised learning of maritime traffic patterns for anomaly detection.