Geographical Urban Planning Research Quarterly

Online ISSN: 2423-7779

Journal Homepage: www.jurbangeo.ut.ac.ir

Mining frequent spatial patterns of urban activities
The Case study a Banks and financial and credit institutions of Tehran

Hassanali Faraji Sabokbar [1]

1. *(Corresponding Author) Department of Human Geography, Faculty of Geography, University of Tehran, Tehran, Iran*
Email: hfaraji@ut.ac.ir

Article Info

Article type:
Research Article

Article History:
Received:
24 August 2023
Received in revised form:
25 November 2023
Accepted:
29 December 2023
Available online:
25 January 2024

Keywords:
Spatial Data Mining
,Frequent Pattern Mining (FPM),
Frequent itemset,
Itemset.

ABSTRACT

Usually, geographical complications have a kind of adhesion to each other, and one complication attracts other complications. In other words, these collections are placed together so that as a result of gathering and settling in the vicinity, they complement each other and meet the needs of customers. The composition of such collections is influenced by economic, social, cultural, political, planning, physical factors, etc. To carry out this research, the location of branches and urban financial and credit institutions of Tehran was used, 3773 branches were used as a location for collecting data and making a collection of items. 56 classes of urban complications were considered for data analysis and extraction, and all complications that were within 200 meters of branches were extracted and then analyzed using a priori method. In this research, a set of rules were extracted that put different types of activities together. These rules show a better understanding of compatible activities, complementary activities, and previous and previous links of different types of activities.

Cite this article: Faraji Sabokbar, H. (2023). Mining frequent spatial patterns of urban activities The Case study a Banks and financial and credit institutions of Tehran. *Geographical Urban Planning Research Quarterly, 11* (4), 49-65.
http://doi.org/10.22059/JURBANGEO.2023.347390.1727

© The Author (s).

Publisher: University of Tehran Press

Extended Abstract

Introduction

The subject of geography is the study of the distribution and spatial arrangement of geographic objects. Objects are not usually randomly distributed, but often have specific spatial patterns. Some geographic phenomena are interdependent and often complement each other, while some activities support other activities. Understanding these spatial connections and recognizing patterns helps us better understand the spatial distribution of elements.

Activities are attracted to each other through the links they have. These links address each other's needs through backward and forward communications and may generate demand for various services. Bank branches, by attracting resources and providing various credit services, engage in service provision and activities. They prefer to be located in places with high resource attraction, such as proximity to large commercial complexes, market corridors, hospitals, and other service centers (business factors). Therefore, the distribution of bank branches is somewhat dependent on other activities, especially those with higher financial turnover. In addition, many urban services are somewhat complementary or competitive to each other. Moreover, in commercial corridors in cities, they play an important role in attracting customers. Depending on the scale of activities and the volume of businesses, they can have penetration areas with different extents.

Methodology

Machine learning based on association rules (Apriori algorithm) has been used for research purposes to explore recurring or repetitive patterns. It has been used to analyze branches of banks and financial institutions in Tehran. For the research, the data is divided into two categories: data related to banks and financial institutions, which constitute the basket or set of items. To this end, 3773 bank branches (including Export, National, Maskan banks, etc.) were selected, and maps and information layers were generated for them. The second part is about urban incidents in Tehran. The information was categorized, and their maps were collected and entered into the system. 81066 urban incidents were investigated, including healthcare facilities, public places, city squares, etc. To extract information items about the branches, first, a 200-meter radius was set for each branch. In the next step, the incidents that occurred within this range were extracted and entered into a matrix.

The research findings indicate that the distribution and dispersion of bank branches in Tehran are not uniform. The central parts of Tehran have the highest number of branches. The northern part of Tehran has been affected by the placement of offices and service centers along Valiasr Street, pulling them away from the center towards the north. The southern part of Tehran is mostly occupied by government or government-private banks (such as Bank Melli, Maskan, Mellat, Sepah, and Tejarat). The western part of Tehran has a low number of branches.

Results and discussion

According to the research findings, the highest frequencies in the range of bank branches are related to insurance, restaurants, official document offices, pharmacies, etc., while the lowest frequency is related to hospitals and emergency centers. The set of recurrent items and association rules contain valuable information for spatial analysis and decision-making in urban areas. For example, in the above table, the first row indicates the association between hospitals (Ph) and insurance offices (Ic) with a support of 52%, and the next item is the official document office (Np). This means that if there are hospitals and insurance offices in the vicinity of a bank branch, there will also be an official document office. Another rule (1618961) with a support of 1 and a confidence of 70% includes the pharmacy (Ph) as the antecedent and physiotherapy, clinics, radiology, medical diagnostic laboratories, newspaper offices, and government facilities as the consequent. In other words, when there is a pharmacy within a 200-meter radius of a bank branch, we expect these 7 items to be located as well. The rule (3086569) with a support of 1 and a confidence of 70% includes the official

document office (Np) as the antecedent and marriage and divorce offices (Md), insurance offices (Ic), travel agencies (Ta), government facilities (G), restaurants (R), and medical diagnostic laboratories (La) as the consequent. Therefore, we expect the types of services listed in the consequent to be present where there are official document offices. Another rule with three antecedents and three consequents, a support of 5%, and a confidence of 67% has the pharmacy (Ph), clinic (Cl), and hospital (Ho) as the antecedents, and medical diagnostic laboratories (La), emergency centers (Uc), and radiology (Ra) as the consequents. These items are of the same kind and indicate that the presence of a hospital, pharmacy, and clinic as the antecedent attracts complementary services and centers such as medical diagnostic laboratories, emergency centers, and radiology. In other words, there is a link and correlation between these activities in urban areas, and exploring and discovering such relationships can provide a better understanding of urban activities and processes, as well as the connection and interaction between services. It offers guidelines for urban planning to gain a better understanding of economic and social activities in the city, identify complementary or competing services, discover urban processes, and facilitate intelligent urban management. Current models of planning usually rely on literature and theories about cities, while the discovery of activity and service distribution patterns and their associations and dependencies can provide a more realistic understanding of urban connections.

Conclusion

In this study, the spatial distribution and establishment of geographical features within the branches of banks were analyzed, and it was determined what types of activities dominate within the vicinity of each branch. Based on this, spatial linkages between the features could be identified, and complementary and supplementary geographical features could be identified accordingly. For future research, it is suggested that these patterns be examined locally to extract the background effects on the patterns. In this study, Tehran as a whole has been considered uniformly, but in future research, spatial differences need to be addressed, geographical analysis should be added, and other algorithms for exploring recurring patterns should be introduced. Additionally, attention should be paid to other urban features besides banks.

Funding

There is no funding support.

Authors' Contribution

Authors contributed equally to the conceptualization and writing of the article. All of the authors approved thecontent of the manuscript and agreed on all aspects of the work declaration of competing interest none.

Conflict of Interest

Authors declared no conflict of interest.

Acknowledgments

We are grateful to all the scientific consultants of this paper.

فصلنامه پژوهش‌های جغرافیای برنامه‌ریزی شهری
شاپا الکترونیکی: 2423-7779
Journal Homepage: www.jurbangeo.ut.ac.ir

کاوش الگوهای پرتکرار فضایی فعالیت‌های شهری
مطالعه موردی: بانک‌ها و مؤسسات مالی و اعتباری شهر تهران

حسنعلی فرجی سبکبار [1]

۱ – نویسنده مسئول، گروه جغرافیایی انسانی، دانشکده جغرافیا، دانشگاه تهران، تهران، ایران. رایانامه: hfaraji@ut.ac.ir

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

تاریخ دریافت:
۱۴۰۲/۰۶/۰۲
تاریخ بازنگری:
۱۴۰۲/۰۹/۰۴
تاریخ پذیرش:
۱۴۰۲/۱۰/۰۸
تاریخ چاپ:
۱۴۰۲/۱۱/۰۵

واژگان کلیدی:
داده‌کاوی فضایی،
کاوش الگوهای پرتکرار،
مجموعه اقلام پرتکرار،
مجموعه اقلام

چکیده

معمولاً عوارض جغرافیایی دارای نوعی چسبندگی به هم هستند و یک عارضه در کنار خود سایر عوارض را جذب می‌کند. به عبارتی این مجموعه‌ها در کنار هم قرار می‌گیرند تا در نتیجه تجمع و استقرار در مجاورت هم مکمل یکدیگر بوده و نیازهای مشتریان را تأمین نمایند. ترکیب چنین مجموعه‌هایی تحت تأثیر عوامل اقتصادی، اجتماعی، فرهنگی، سیاستی، برنامه‌ریزی، کالبدی و.. قرار دارند. برای انجام این تحقیق از موقعیت شعب و مؤسسات مالی و اعتباری شهری تهران استفاده‌شده است، ۳۷۷۳ شعبه به‌عنوان موقعیتی برای جمع‌آوری داده‌ها و ساخت مجموعه اقلام استفاده شد. ۵۶ کلاس عارضه شهری برای تحلیل و استخراج داده‌ها در نظر گرفته شد و کلیه عوارضی که در محدوده ۲۰۰ متری شعب قرار داشتند استخراج شدند و سپس با استفاده از روش اپریوری مورد تحلیل قرار گرفتند. در این تحقیق مجموعه‌ای از قواعد استخراج شد که نحوه قرارگیری انواع فعالیت‌ها را در کنار یکدیگر قرار می‌دهند این قواعد به درک بهتر فعالیت‌های سازگار، فعالیت‌های مکمل، پیوندهای پسین و پیشین انواع فعالیت‌ها را نشان می‌دهد.

استناد: فرجی سبکبار، حسنعلی. (۱۴۰۲). کاوش الگوهای پرتکرار فضایی فعالیت‌های شهری مطالعه موردی: بانک‌ها و مؤسسات مالی و اعتباری شهر تهران. *پژوهش‌های جغرافیای برنامه‌ریزی شهری*، ۱۱ (۴)، ۶۵-۴۹.

http://doi.org/10.22059/JURBANGEO.2023.347390.1727

© نویسندگان

ناشر: مؤسسه انتشارات دانشگاه تهران

مقدمه

موضوع علم جغرافیا، مکان است و به مطالعه پراکندگی عوارض و اشیا فضایی می‌پردازد. توزیع و پراکندگی عوارض جغرافیایی معمولاً درسته فضا، تصادفی نبوده و دارای بافت و الگویی فضایی خاص هستند و بین پدیده‌ها و اشیا جغرافیایی رابطه و همبستگی فضایی وجود دارد (Hesse & Markus,2004; Fotheringham,1994; Fotheringham,2000; Egidi,2020). شناسایی و کشف الگوها و روابط به درک و شناخت بهتر ما از فضا می‌انجامد در نتیجه شناخت و کشف الگوهای فضایی ما می‌توانیم به علل شکل‌گیری چنین الگوهایی، پی ببریم و از آن‌ها در برنامه‌ریزی و توسعه شهری و منطقه‌ای بهره ببریم (Rocha,2018; Haining,2018; Cox,2019). معمولاً عوارض جغرافیایی دارای نوعی چسبندگی به هم هستند و یک عارضه در کنار خود سایر عوارض را جذب می‌کند برای مثال یک در کنار بیمارستان یا درمانگاه، مجموعه‌ای از مطب‌های پزشکان، داروخانه‌ها، رادیولوژی، آزمایشگاه، رستوران، اغذیه، گل‌فروشی، سوپرمارکت و ...، قرار می‌گیرند و یا پوشاک که معمولاً چندین مغازه مکمل مانند پارچه‌فروشی، خیاطی، لباس‌فروشی، کفش و پوشاک قرار می‌گیرد. این مجموعه‌ها در کنار هم قرار می‌گیرند تا در نتیجه تجمع و استقرار در مجاورت هم مکمل یکدیگر باشند و نیازهای مشتریان را تأمین نمایند. ترکیب چنین مجموعه‌هایی تحت تأثیر عوامل اقتصادی، اجتماعی، فرهنگی، سیاستی، برنامه‌ریزی، کالبدی و.. قرار دارند. لذا درک چنین مجموعه‌هایی در سبد شهری هر فعالیت می‌تواند در شناخت فرایند، ساختار، سازمان فضایی شهرها به‌گونه‌ای دیگر مؤثر باشد، روش‌های مختلفی برای شناخت الگوها مطرح می‌باشند. با توسعه فناوری اطلاعات، داده‌کاوی و مهم‌تر از همه علم داده امروزه یادگیری ماشین و داده‌کاوی به پردازش و کشف چنین الگوهایی می‌پردازند. در این تحقیق به موضوع الگویابی همایندی بانک‌ها و مؤسسات مالی و اعتباری شهر تهران پرداخته می‌شود. بانک‌ها مؤسسات بخش خصوصی و دولتی می‌باشند، که پیوند نزدیکی با خصوصیات منطقه‌ای که در آن مستقر می‌شود دارد لذا به‌عنوان مطالعه موردی این تحقیق انتخاب‌شده است تا بهتر بتواند این موضوع را مدل نمایند.

مبانی نظری

شهرها موجوداتی پویا هستند، پویایی شهر تحت تأثیر مجموعه‌ای از عوامل اقتصادی، اجتماعی، سابقه تاریخی، ساختار و ترکیب جمعیت و جریانات کلان بین‌المللی، ملی، محلی و ... قرار دارد. این عوامل از طریق تعامل و ارتباطات با یکدیگر در شکل‌گیری و توسعه و یا افول شهری مؤثر می‌باشند. فعالیت‌ها از طریق پیوندهایی که با یکدیگر دارند به‌سوی یکدیگر جذب می‌شوند. این پیوند از طریق ارتباطات پسین و پیشین به تأمین نیازهای یکدیگر می‌پردازند و ممکن است به تولید تقاضا برای انواع خدمات بپردازند. شعب بانک از طریق جذب منابع و ارائه انواع خدمات اعتباری به ارائه خدمت و فعالیت می‌پردازند و بیشتر تمایل دارند در مکان‌هایی استقرار یابند که جذب منابع بالا است همانند مجاورت با مجتمع‌های بزرگ تجاری، راسته‌های بازار، بیمارستان‌ها و سایر مراکز خدماتی (عوامل بنگاه نگاری) لذا توزیع شعب بانک به‌نوعی وابسته به سایر فعالیت‌ها است به‌ویژه فعالیت‌هایی که گردش مالی بیشتری دارند. علاوه بر آن بسیاری از خدمات شهری، به‌نوعی مکمل و یا رقیب یکدیگر می‌باشند. علاوه بر آن در شهرها راسته‌های تجاری خود عامل مهمی در جذب مشتری می‌باشند و با توجه به مقیاس فعالیت‌ها و توده (حجم) کسب‌وکارها می‌تواند حوزه نفوذی با وسعت‌های مختلف داشته باشد.

یادگیری ماشین

یادگیری ماشین، شاخه‌ای از هوش مصنوعی و محاسبات هوشمند است که به‌طورکلی به توانایی ماشین به‌منظور تقلید از شیوه رفتار انسان هوشمند طراحی‌شده‌اند (Coyle,2023; Shavlik,1990). همان‌گونه که ما در نتیجه تکرار یک‌چیز

را تجربه می‌کنیم و یادمی گیریم ماشین‌ها هم می‌توانند از طریق کاوش و پردازش داده‌ها پدیده‌ها و اشیا این الگوها را کشف کنند و از آن‌ها برای پیش‌بینی، تخمین، دسته‌بندی استفاده کنند. تکنیک‌های مبتنی بر یادگیری ماشین امروز به‌عنوان موتور محرکه عصر جدید (به‌اصطلاح ابر داده) شناخته می‌شوند (El Naqa & Murphy 2015). تکنیک‌های مبتنی بر یادگیری ماشین با موفقیت در زمینه‌های مختلف از تشخیص الگو، بینایی کامپیوتر، مهندسی فضاپیما، مالی، سرگرمی و زیست‌شناسی محاسباتی تا کاربردهای زیست پزشکی و پزشکی، آموزش و همچنین مطالعات جغرافیایی و برنامه‌ریزی فضایی استفاده‌شده‌اند (Shah,2022; Bhatia:2024; Romero:2010; Jafarzadeh:2017; Bin:2014; Liao, 2010).

انواع متعددی از سیستم‌های یادگیری ماشین ارائه‌شده‌اند که می‌توانیم آن‌ها برحسب نوع و مقدار نظارت انسان در فرایند یادگیری را به چهار دسته با نظارت، بدون نظارت، نیمه نظارتی و یادگیری تقویتی تقسیم‌بندی کنیم. الگوریتم‌ها برای کشف و ارائه ساختار در داده‌ها به تدبیر خود عمل می‌کنند (Zhou,2021; Jordan,2015; El Naqa,2015; Hsieh,2020).

مهم‌ترین روش‌های بدون نظارت عبارت‌اند از

-خوشه‌بندی: از ابزارهای مفید در علم داده محسوب می‌شود که گزینه‌ها را در گروه‌هایی قرار می‌دهد که حداکثر شباهت را نسبت به مرکز خوشه و حداکثر اختلاف را با سایر گروه‌ها دارد (Sinaga,2020)

-یادگیری قوانین انجمنی: یادگیری قوانین انجمنی از جمله روش‌های داده‌کاوی است که هدف آن کشف روابط بین متغیرهای موجود در پایگاه‌های داده و ارائه آن‌ها به‌صورت مجموعه‌ای از قوانین است.

-بصری سازی و کاهش ابعاد: برای مسائل با ابعاد بزرگ کاربرد دارد و به‌جای تحلیل داده‌های با حجم زیاد، داده‌هایی استفاده شود که بیشترین اثرگذاری را دارند و ابعاد مسئله را کاهش داد.

در این تحقیق از روش apriori استفاده می‌شود که یکی از روش‌های یادگیری قوانین انجمنی است که به کاوش پیوندهای فضایی بین اشیا فضایی می‌پردازیم (Appice,2003:2005) لازم است توجه داشت، کشف قوانین انجمنی که موضوع تحقیق حاضر می‌باشد یکی از موضوعات موردعلاقه در تحقیقات داده‌کاوی محسوب می‌شود (Malerba,2003)

کاوش الگوهای مکرر و یا پرتکرار

روش اپریوری از جمله روش‌های کاوش الگوهای مکرر و یا پرتکرار می‌باشد که به دنبال کشف پیوندها، ارتباطات و همبستگی در مجموعه‌ای از داده‌های مشخص است لذا ابتدا به موضوع و ادبیات کاوش الگوهای مکرر می‌پردازیم و سپس بر اساس آن روش اپریوری تبیین خواهد شد. در آغاز برای درک بهتر موضوع به تعریف چند مفهوم می‌پردازیم:

مجموعه اقلام، به تعدادی از اقلام (اشیا فضایی) که با هم در یک مجموعه قرار می‌گیرند مجموعه اقلام [1]می‌گویند، اگر مجموعه‌ای حاصل گرد هم آیی، K شیء یا قلم باشد، به آن مجموعه اقلام k تایی گویند. به‌طورمعمول مجموعه اقلام از دو یا چند قلم تشکیل می‌شود، زمانی که مجموعه‌ای از اقلام به‌صورت تکراری اتفاق افتد، به‌عنوان **مجموعه اقلام مکرر**[2] نامیده می‌شود.

کاوش الگوهای پرتکرار[3]**:** الگوریتم کاوش الگوهای مکرر یکی از مهم‌ترین تکنیک‌های داده‌کاوی برای کشف روابط بین آیتم‌های مختلف در یک مجموعه داده است. این روابط در قالب قوانین انجمن نشان داده می‌شوند. به یافتن نظم و بی‌نظمی‌ها در داده‌ها کمک می‌کند.

1. Itemset
2. Frequent itemset
3. Frequent Pattern Mining (FPM)

روش پژوهش

قوانین انجمنی[1]**:** مفهوم قوانین انجمنی، اولین بار توسط اگراوال، نلنسکی، سوامی[2] ارائه شد و برای کشف قوانین نهفته در داده‌های اولیه استفاده می‌شود، مطالعات مربوط به قوانین انجمنی در حوزه‌های مختلف علمی موردتوجه قرارگرفته است. فرض کنید، $I = \{i_1, i_2, \dots, i_m\}$ مجموعه اقلام باشد، و D مجموع پایگاه‌های داده‌های مکانی است، هر مجموعه دارای یک شناسه است. T می‌توان مجموعه‌ای از اقلام تعریف کرد، و $T \subset I$ است و D را می‌توان مجموع‌های متشکل از T تعریف کرد. طبق تعریف، X مجموعه‌های خاص از I است اگر $X \subset T$ باشد ، بنابراین T شامل X نیز هست. تعداد اقلام، موجود در مجموعه به‌عنوان اقلام شناخته می‌شود حالتی مانند $X \Rightarrow Y$ ، شرایط $X \subset T$ و $X \subset I$ و $X \cap T = \emptyset$ را برآورده نماید. پشتیبان یک قانون انجمنی برای $X \Rightarrow Y$ را می‌توان اینگونه می‌توان

$$Supp(X \Rightarrow Y) = \frac{|\{T: X \cup Y \cup T, T \in D\}|}{|D|}$$

که به‌صورت$(X \Rightarrow Y)$پشتیبان مشخص می‌شود.

مقدار اطمینان[3] یک قانون انجمنی برای $X \Rightarrow Y$ را می‌توان این‌گونه نگاشت:

$$Confidence(X \Rightarrow Y) = \frac{|\{T: X \cup Y \cup T, T \in D\}|}{|\{T: X \subset T, T \in D\}|}$$

که به‌صورت$(X \Rightarrow Y)$اطمینان مشخص می‌شود از این شاخص برای تعیین اینکه یک عارضه چند بار در مجموعه داده ظاهر می‌شود، استفاده می‌شوند و مجموعه‌ایی از عوارض جغرافیایی X (شعب بانک) است که شامل Y (عوارض شهری مانند بیمارستان‌ها) می‌شود (Angeline,2013; Li,2011; Andriani,2020).

بالابری Lift بالابری به‌صورت زیر تعریف می‌شود:

$$Lift(X \Rightarrow Y) = \frac{Supp(X \Rightarrow Y)}{supp(X) * supp(Y)}$$

برای تحلیل داده‌های بالابری اگر مقدار آن برابر با ۱ باشد اشاره به آن دارد که احتمال اینکه مقدم و تالی مستقل باشند را مشخص می‌کند و زمانی که این دو رویداد مستقل از هم باشند، هیچ قاعده‌ای را نمی‌توان برای این دو نگاشت. اگر این مقدار بزرگ‌تر از ۱ باشد به این معنا است که وابستگی وجود دارد و از قواعد ایجادشده می‌توان برای تخمین استفاده کرد و درصورتی‌که کمتر از یک باشد نشان می‌دهد که اقلام جایگزین یکدیگر می‌شوند (Kavitha,2020).

رابطه بالا را می‌توان به‌صورت زیر فرموله بندی ریاضی نمود

$$Supp(A) = \frac{\text{تعداد عارضه } A \text{ در محدوده شعب}}{\text{تعدادکل شعب}}$$

$$confidenct(A \Rightarrow B) = \frac{Supp(A \cup B)}{Supp(A)}$$

$$Lift(A \Rightarrow B) = \frac{Supp(A \cup B)}{Supp(A) * Supp(B)}$$

به‌طور کل کاوش قوانین انجمنی یک فرایند دومرحله‌ای محسوب می‌شود:

یافتن کلیه مجموعه اقلام مکرر: در این مرحله مجموعه اقلامی استخراج می‌شوند که تکرار آن‌ها حداقل برابر با مقدار

1. Association Rules
2. Agrawal, Hnielinski and Swami
3. confidence

کمینه پشتیبان Min_support باشد.

تولید قوانین انجمنی قوی از مجموعه اقلام مکرر: این قوانین باید دارای حداقل مقدار پشتیبان و اطمینان باشند.

الگوریتم Apriori – الگوریتم‌های الگوی مکرر

همان‌گونه که اشاره شد الگوریتم Apriori اولین الگوریتمی بود که برای کاوش مجموعه آیتم‌های مکرر پیشنهاد شد. بعدها توسط R Agarwal و R Srikant بهبود یافت و به Apriori معروف شد. این الگوریتم از دو مرحله "پیوستن" و "هرس" برای کاهش فضای جستجو استفاده می‌کند. این‌یک رویکرد تکراری برای کشف متداول‌ترین مجموعه اقلام است.

احتمال اینکه مورد I مکرر نباشد این است که:

P(I) < حداقل آستانه پشتیبانی، پس I مکرر نیست.

P (I+A) < حداقل آستانه پشتیبانی، سپس I+A مکرر نیست، جایی که A نیز به مجموعه آیتم‌ها تعلق دارد.

اگر یک مجموعه آیتم مقدار کمتری از حداقل پشتیبانی داشته باشد، تمام ابر مجموعه‌های آن نیز کمتر از حداقل پشتیبانی می‌شوند و بنابراین می‌توان آن‌ها را نادیده گرفت. به این ویژگی، خاصیت ناهمنوا[1] می‌گویند.

مراحل انجام‌شده در الگوریتم Apriori داده‌کاوی عبارت‌اند از:

مرحله پیوستن[2]: این مرحله با افزودن آیتم‌ها به مجموعه آیتم‌های تعریف می‌شود.

مرحله هرس[3]: این مرحله تعداد آیتم‌های پایگاه داده اسکن می‌شود. اگر آیتم کاندید حداقل حمایت را نداشته باشد، نادر و کم تکرار تلقی شده و از مجموعه اقلام حذف می‌شود. این مرحله برای کاهش اندازه مجموعه آیتم‌های کاندید انجام می‌شود.

مراحل Apriori

الگوریتم Apriori یکی از روش‌های کاوش انجمنی قواعد است لذا شبیه به روش‌های کاوش انجمنی از چند مرحله تشکیل می‌شود که هدف آن یافتن متداول‌ترین مجموعه آیتم‌ها در پایگاه داده است (Harun,2017). این تکنیک داده‌کاوی مراحل پیوستن و هرس را به‌طور مکرر دنبال می‌کند تا زمانی که متداول‌ترین مجموعه موارد به دست آید. حداقل آستانه پشتیبانی در مسئله داده‌شده یا توسط کاربر در نظر گرفته‌شده است.

مرحله ۱: در اولین تکرار الگوریتم، هر آیتم به‌عنوان کاندیدای مجموعه‌ای در نظر گرفته می‌شود. الگوریتم تعداد دفعات هر مورد را شمارش می‌کند.

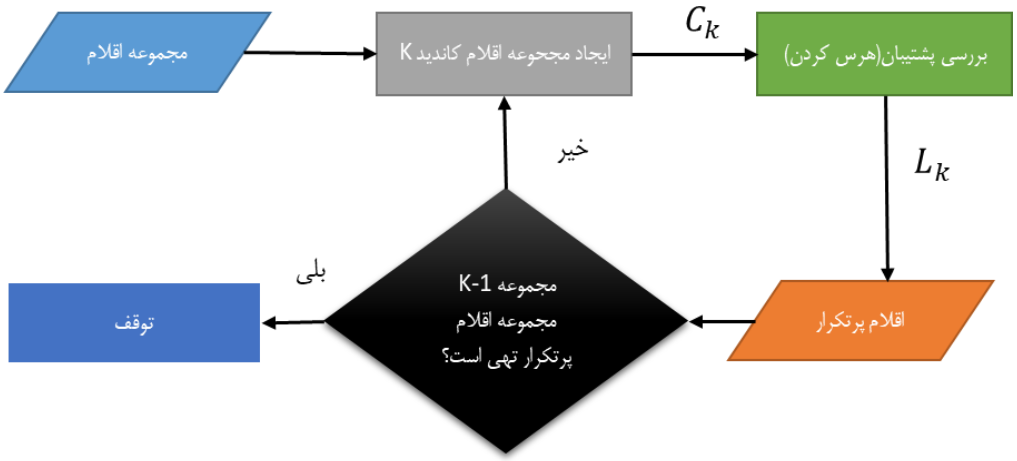
مرحله ۲: حداقل پشتیبانی وجود داشته باشد، min_sup (مثلاً ۲). مجموعه ۱ – مجموعه مواردی که وقوع آن‌ها حداقل مقدار را برآورده می‌کند تعیین می‌شود. فقط آن دسته از کاندیداهایی که تعداد آن‌ها بیش از یا مساوی min_sup باشد، برای تکرار بعدی جلو می‌روند و بقیه هرس می‌شوند.

مرحله ۳: در مرحله بعد، آیتم‌های مکرر ۲ موردی با min_sup کشف می‌شوند. برای این کار در مرحله پیوستن، مجموعه ۲ موردی با تشکیل یک گروه ۲ تایی با ترکیب آیتم‌ها با خودش تولید می‌شود.

مرحله ۴: نامزدهای ۲ موردی با استفاده از مقدار آستانه min-sup هرس می‌شوند. اکنون جدول دارای ۲ مجموعه آیتم با min-sup است.

1. Antimonotone
2. Join Step
3. Prune Step

مرحله ۵: تکرار بعدی با استفاده از مرحله پیوستن و هرس، ۳ مجموعه آیتم را تشکیل می‌دهد. این تکرار از خاصیت ضد یکنواخت پیروی می‌کند که در آن زیرمجموعه‌های ۳ آیتمی، یعنی زیرمجموعه‌های ۲ آیتمی هر گروه در min_sup قرار می‌گیرند. اگر همه زیرمجموعه‌های ۲ موردی مکرر باشند، سوپرست مکرر خواهد بود در غیر این صورت هرس می‌شود.



شکل ۱. مراحل پردازش الگوریتم اپریوری

مرحله ۵ ساخت ۴ آیتم با پیوستن ۳ آیتم به خودش و هرس اگر زیرمجموعه آن معیارهای min_sup را نداشته باشد، دنبال می‌شود. الگوریتم زمانی متوقف می‌شود که مجموعه اقلام موردنظر به دست آید.

در این تحقیق از داده‌های فضایی برای تحلیل استفاده‌شده است، داده‌کاوی فضایی که هدف این تحقیق می‌باشد، فرآیند کشف الگوهای فضایی موردعلاقه و قبلاً ناشناخته، اما بالقوه مفید از داده‌های فضایی بزرگ است (Yoo,2012). این الگوهای فضایی عبارت‌اند از داده‌های پرت فضایی۱، گسستگی‌های۲ فضایی، مدل‌های پیش‌بینی مکانی، خوشه‌های فضایی و الگوهای همایند فضایی است (Mehta,2018).

داده‌های جغرافیایی اغلب خصوصیات وابستگی فضایی را نشان می‌دهند. وابستگی فضایی به معنی روند مشاهده اشیا واقع نزدیک یکدیگر در فضایی جغرافیایی است که شباهت یا عدم شباهت‌های جغرافیایی را با درجه بالانشان می‌دهد (Miller,2009). این وابستگی بر اساس دیدگاه توبلر است که به‌عنوان قانون اول جغرافیا مشهور می‌باشد و بر اساس آن در فضای جغرافیایی همه‌چیز به یکدیگر مربوط می‌باشند، اما چیزهایی نزدیک‌ترند نسبت به اشیا جغرافیایی دورتر بیشتر به هم مرتبط می‌باشند (Miller,2004; Goodchild,2009; Toble,1970).

داده‌کاوی مکانی به استخراج روابط مکانی و کشف روابط بین اشیا فضایی و نیز شناسایی الگوهای فضایی است که به‌وضوح در مجموعه داده‌های مکانی ذخیره نشده‌اند (Shekhar,2001; Eick,2008). این موضوع کاربرد گسترده‌ای در زمینه‌های مختلف مانند سیستم‌های اطلاعات جغرافیایی، ژئومارکتینگ، کاوش پایگاه داده، تصویربرداری پزشکی، پردازش تصویر، کنترل ترافیک، مطالعات محیطی پیداکرده است. چالشی که برای داده‌کاوی مکانی وجود دارد، اکتشاف تکنیک‌های کارآمد داده‌کاوی فضایی به دلیل حجم زیاد داده‌های مکانی و روش‌های دسترسی مکانی است (Priya,2011). همان‌طور که قبال نیز اشاره‌شده یکی از موضوعات داده‌کاوی فضایی الگوهای همایندی فضایی است که نشان‌دهنده مجموعه‌ای از

1. Spatial outliers
2. Discontinuities

عوارض و اشیا جغرافیایی است که اغلب با هم در مجاورت فضایی مشاهده می‌شوند.

بانک‌ها و مؤسسات مالی و اعتباری، ارائه انواع خدمات مالی و اعتباری هستند، توزیع و پراکندگی آن‌ها در سطح شهر یکنواخت توزیع نشده است و عوامل متعددی در شکل‌گیری الگوهای استقرار مؤسسات مالی و اعتباری تأثیر دارد که آن‌ها را به سه دسته می‌توان تقسیم کرد: ۱- عوامل سیاستی، ۲-عوامل ژئودموگرافیک، ۳-عوامل فرموگرافیک یا بنگاه نگاری، عامل اول بیشتر بانک‌های دولتی دنبال می‌کنند و ارائه خدمات به تمام جمعیت می‌باشد، عدالت دسترسی به خدمات مالی و اعتباری، و...، دسته دوم عوامل ژئودموگرافیک می‌باشد که به ساختار و ترکیب جمعیت مربوط می‌شود، بخشی از مشتریان بانک‌ها از افراد حقیقی تشکیل می‌دهد، نوع سرویس‌ها و خدماتی که توسط این بخش از مشتریان موردتقاضا می‌باشد، تحت تأثیر ویژگی‌های جمعیت است، برای مثال گروه‌های شغلی، گروه‌های سنی، گروه‌های تحصیل‌کرده، هرکدام رفتار فضایی خاص خود را دارند. اما بخش دیگر مشتریان بانک‌ها را کسب‌وکارها، بنگاه‌های اقتصادی و مؤسسات و دستگاه‌های اجرایی و دولتی تشکیل می‌دهد، این بخش از جامعه به‌عنوان بنگاه نگاری نامبرده می‌شود. در این تحقیق به بررسی کسب‌وکارهای و بنگاه‌های تجاری و نیز مؤسسات دولتی و بخش عمومی موردبررسی قرارگرفته است.

برای تحقیق دودسته داده داریم، داده‌های مربوط به بانک‌ها و مؤسسات مالی که تشکیل‌دهنده سبد و یا مجموعه اقلام می‌باشند. بدین منظور ۳۷۷۳ شعبه بانک (شامل بانک‌های صادرات، ملی، مسکن و ...) انتخاب‌شده‌اند و برای آن‌ها نقشه و لایه اطلاعات تولید شد. بخش دوم عوارض شهری تهران می‌باشد، اطلاعات دسته‌بندی شدند و نقشه‌های آن‌ها جمع‌آوری و وارد سیستم شده است و ۸۱۰۶۶ عارضه شهری موردبررسی قرارگرفته است که شامل مراکز بهداشت و درمان، اماکن عمومی، میدان‌ها شهری و... می‌شود.

آماده‌سازی داده‌ها: برای ساخت مجموعه اقلام[1] از ابزارهای تحلیل فضایی در محیط ArcGIS pro استفاده‌شده است. بدین منظور قطعه کدی با arcpy نوشته شد و با استفاده از ابزار SpatialJoin_analysis در شعاع ۲۰۰ متری شعب بانک، عوارضی که واقع می‌شدند شناسایی و استخراج شدند (شکل شماره ۲). این داده‌ها تبدیل به جدول شد و سپس با هم تلفیق شدند، بعد از تلفیق فضایی، ماتریس باینری استخراج شد، در آن عوارضی که در محدوده شعبه قرار می‌گرفت مقدار ۱ و در غیر این صورت مقدار صفر اختصاص می‌یافت، ابزار تحلیلی تعداد عوارض را بر می‌گرداند که تعداد تبدیل به مقادیر باینری شد، فرایند آماده‌سازی داده‌ها و تشکیل ماتریس در محیط ArcGIS Pro انجام‌شده است.

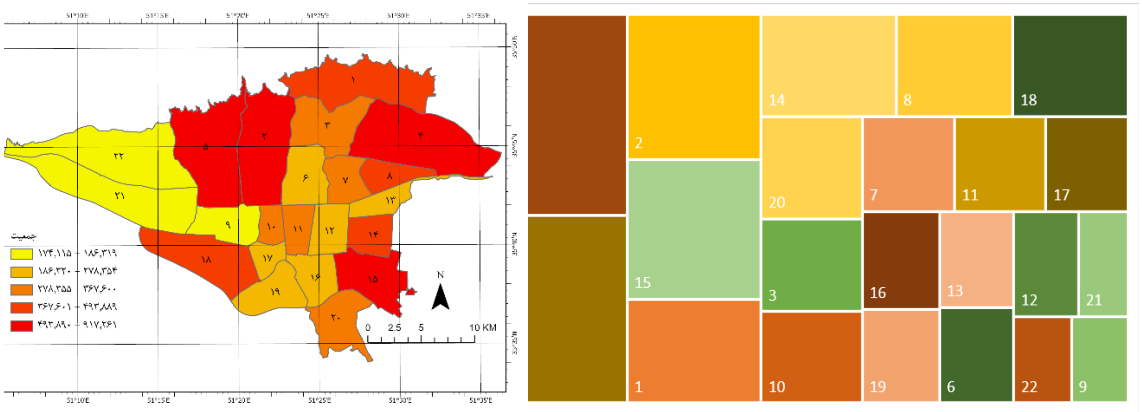
شکل ۲. نمونه انتخاب عوارض شهری واقع در شعاع ۲۰۰ متر شعب بانک‌ها

1. Item Set

داده‌ها به محیط پایتون فراخوانی شدند و سپس به فرمت مناسب برای ورود به بسته تحلیل اپریوری تبدیل شد. برای تحلیل الگوریتم اپریوری، از بسته apriori-python 1.0.4 استفاده‌شده است.

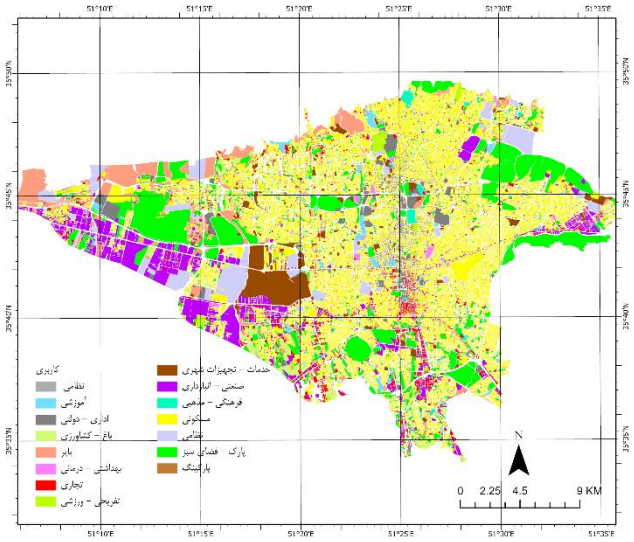
محدوده موردمطالعه

محدوده موردمطالعه شهر تهران می‌باشد، این شهر پایتخت کشور محسوب می‌شود، بزرگ‌ترین شهر غرب آسیا محسوب می‌شود، بر اساس آخرین سرشماری نفوس و مسکن در سال ۱۳۹۵ جمعیت تهران ۸۶۷۹۹۵۰ نفر می‌باشد که در مساحتی حدود ۷۰ هزار کیلومتر مستقر می‌باشند الگوی پراکندگی جمعیت در تهران در شکل شماره ۳ نشان داده‌شده است توزیع جمعیت در مناطق تهران یکسان نیست. این عرصه به کاربری‌های مختلف اختصاص‌یافته است، شکل شماره ۴ توزیع کاربری‌های شهری تهران را نشان می‌دهد. توزیع و پراکندگی کاربری و فعالیت‌های شهری تهران، تحت تأثیر عوامل مختلف اقتصادی، اجتماعی، محیط و بستر طبیعی آن قرار دارد. بخش عمده فعالیت‌های تجاری تهران در مرکز تهران (منطقه ۱۲) مستقر می‌باشد، صنایع در بخش جنوب تهران، در منطقه ۲۱ تهران مستقرمی باشند. الگوی توزیع خدمات و فعالیت‌ها درعین‌حالی که ناهمگون توزیع‌شده‌اند از یک نظم فضایی تبعیت می‌کنند.



ب الف

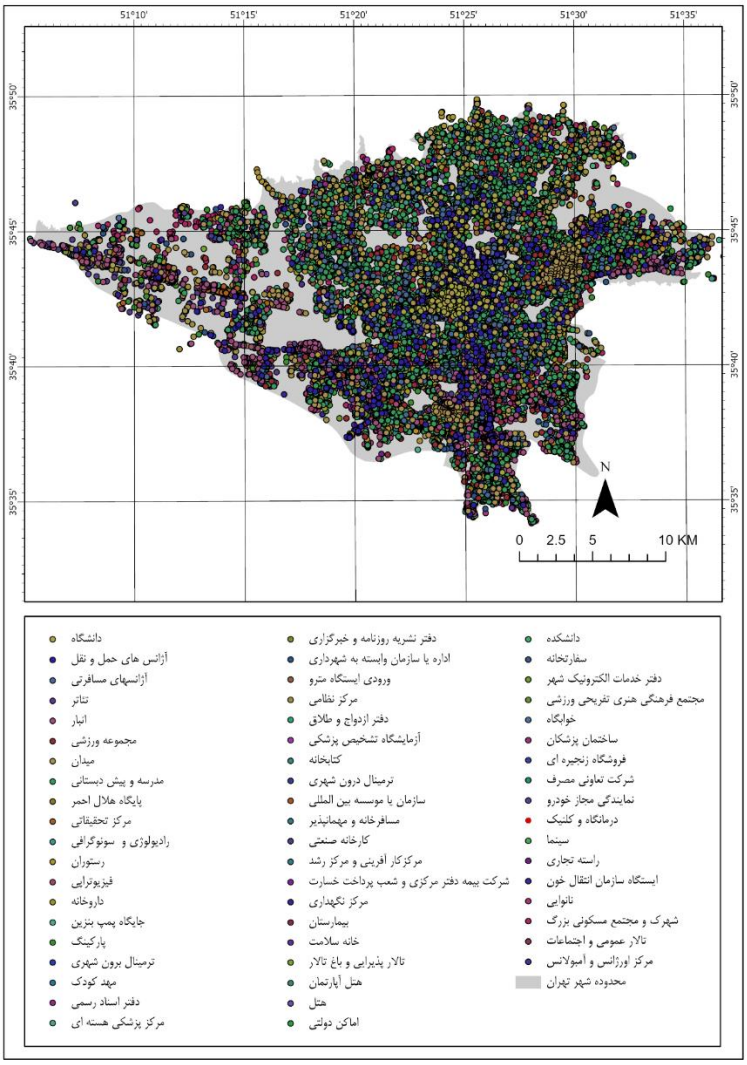
شکل ۳. توزیع جمعیت مناطق تهران (الف– نمودار توزیع جمعیت، ب–نقشه توزیع جمعیت)



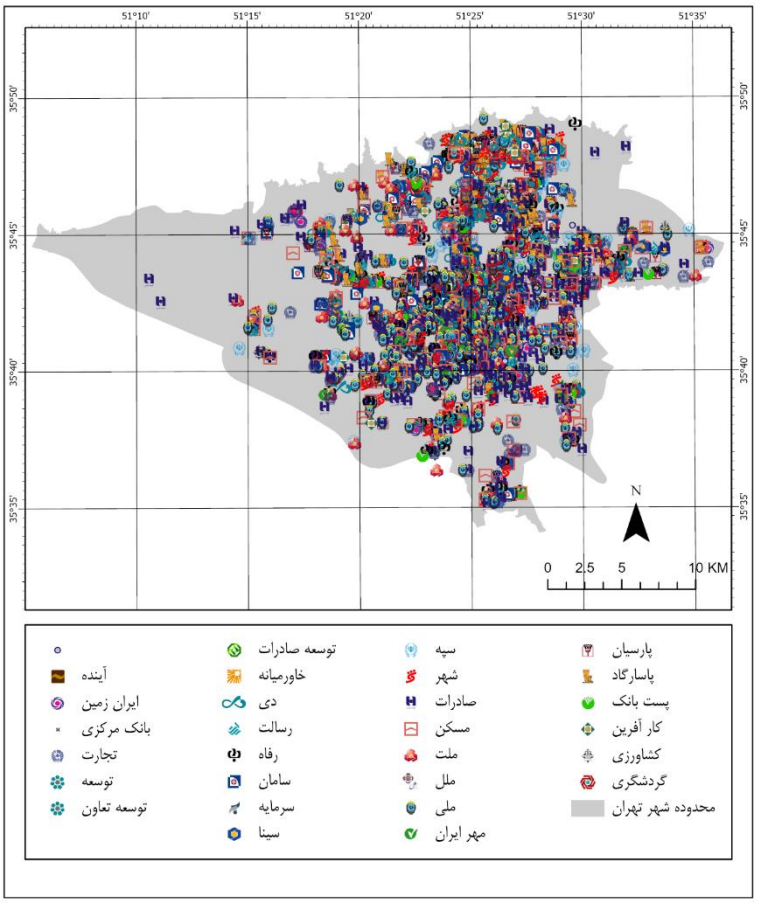
شکل ٤. الگوی کاربری اراضی شهر تهران

یافته‌ها

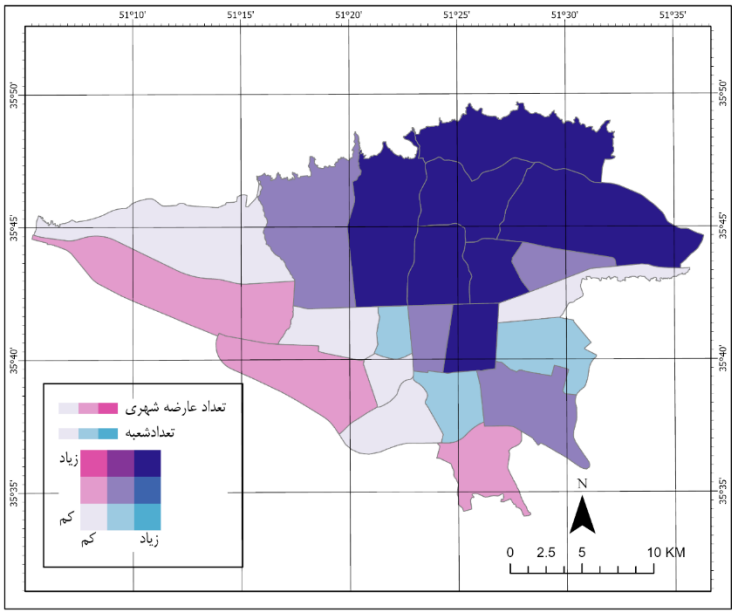
برای تحلیل الگوهای پرتکرار فضایی در سطح شهر تهران که در پیرامون شعب بانک‌ها مورد تحلیل قرارگرفته است از داده‌های پراکندگی عوارض شهری استفاده‌شده است. شکل شماره ۵ الگوی توزیع فضایی معیارهای استفاده‌شده را نشان می‌دهد که فشردگی آن‌ها در سطح تهران متعادل نیست همچنین از موقعیت شعب بانک‌های مستقر در شهر تهران به‌عنوان واحد پایه برای تلفیق و اندازه‌گیری استفاده‌شده است شکل شماره ۶ الگوی پراکندگی شعب را در سطح تهران نشان می‌دهد، همان‌طوری که در شکل دیده می‌شود، توزیع شعب بانک‌ها در تهران متعادل نیست، بخش‌های مرکزی تهران با بالاترین تعداد شعب همراه است، بخش شمالی تهران تحت تأثیر استقرار ادارات و مراکز خدماتی در امتداد خیابان ولی‌عصر از مرکز به سمت شمال کشیده شده‌اند، جنوب شهر تهران بیشتر بانک‌های دولتی یا دولتی–خصوصی (مانند بانک ملی، مسکن، ملت، سپه، تجارت) استقراریافته‌اند، غرب تهران تعداد شعب کم می‌باشد.. شکل شماره ۷ الگوی پراکندگی شعب و عوارض شهری را نشان می‌دهد که حکایت از عدم توازن توزیع عوارض شهری و پراکندگی شعب بانک دارد. در شمال تهران، تعداد شعب و عوارض زیاد می‌باشد که با رنگ آبی نشان داده‌شده است و در مقابل جنوب و جنوب غرب تعداد شعب و عوارض شهری کم می‌باشد.



شکل ۵. توزیع عوارض شهری موردمطالعه



شکل ٦. توزیع فضایی شعب بانک‌ها و مؤسسات مالی و اعتباری



شکل ۷. مقایسه عوارض شهری و شعب بانک‌ها در سطح مناطق تهران

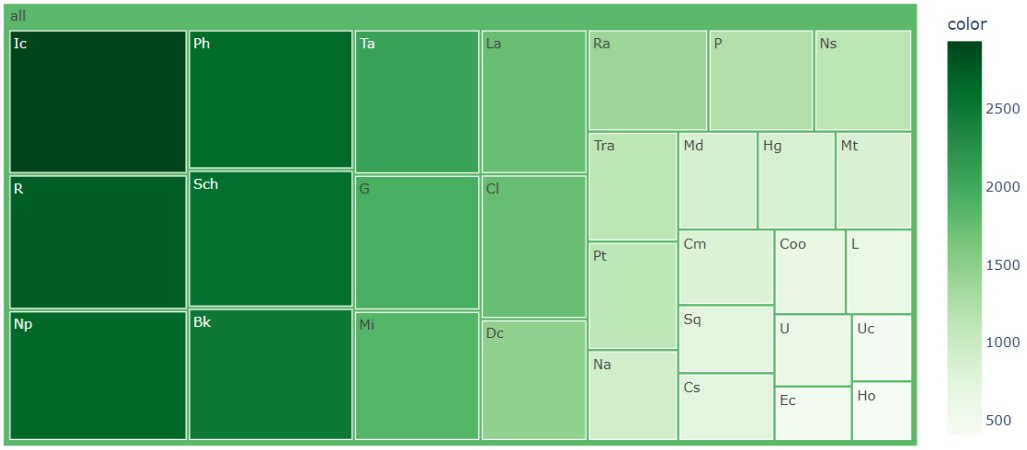
	Do	Dc	Cs	Coo	Cm	Cl	C	Bu_Feature	Bs	Bk	...	Rcs	U	Tra	Ta	T	Sto	St	Sq	Sch	Uc
0	1	1	1	1	0	1	0	0	0	1	...	0	1	1	1	0	0	0	1	1	1
1	1	1	1	1	0	1	0	0	0	1	...	0	1	1	1	0	0	0	1	1	1
2	1	1	0	0	0	1	1	0	1	0	...	0	1	1	1	0	0	0	0	1	1
3	1	1	1	0	1	1	0	0	0	1	...	0	0	0	1	0	0	0	1	1	1
4	1	1	0	0	0	1	1	0	1	0	...	0	1	1	1	0	0	0	0	1	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3733	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3734	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3735	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3736	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3737	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

شکل ۸. ماتریس باینری مکانی اقلام فضایی در محدوده شعب

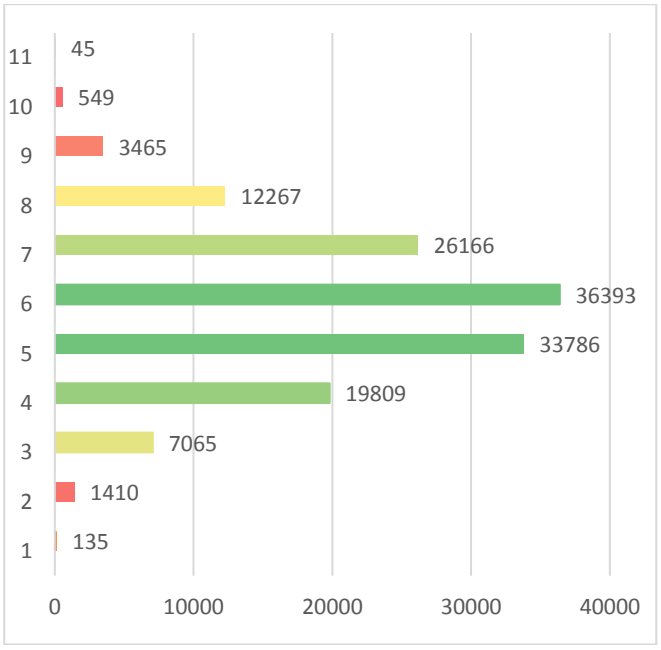
شکل شماره ۸ ماتریس باینری مکانی، اقلام فضایی برحسب شعب بانک در سطح تهران (۳۷۳۷ شعبه موردمطالعه که تعداد عوارض در شعاع ۲۰۰ متری بیش از ۲۰ می‌باشد) را نشان می‌دهد. در سطر مجموعه اقلام هر شعبه می‌باشد و ستون‌ها نام عوارض جغرافیایی که در محدوده شعبه واقع‌شده می‌باشد. بر اساس یافته‌های تحلیلی عوارض موجود در محدوده هر شعبه استخراج و کدگذاری (۰، ۱) شده است، در جدول شماره ۱ عارضه با بیشترین تواتر و شکل شماره ۹ حجم و ترکیب عوارض را در محدوده شعب بانک‌ها در تهران نشان می‌دهد. بیشترین فراوانی مربوط به بیمه، رستوران، دفتر اسناد رسمی، داروخانه و... و کمترین فراوانی مربوط به بیمارستان و مراکز اورژانس می‌باشد.

جدول ۱. توزیع عوارض پرتکرار در محدوده شعب بانک‌ها و مؤسسات مالی و اعتباری

نوع عارضه	Ic	R	Np	Ph	Sch	Bk	Ta	G	Mi	La
	بیمه	رستوران	دفتر اسناد رسمی	داروخانه	مدرسه	نانوایی	آژانس مسافرتی	اماکن دولتی	مراکز نظامی	آزمایشگاه تخصصی
فراوانی رخداد	۲۹۳۶	۲۷۴۷	۲۶۵۱	۲۶۲۷	۲۵۸۴	۲۴۹۵	۲۰۷۷	۱۹۴۳	۱۸۶۴	۱۷۵۱



شکل ۹. فراوانی عوارض پرتکرار شهری



طول	تعداد عارضه
۱	۱۳۵
۲	۱۴۱۰
۳	۷۰۶۵
۴	۱۹۸۰۹
۵	۳۳۷۸۶
۶	۳۶۳۹۳
۷	۲۶۱۶۶
۸	۱۲۲۶۷
۹	۳۴۶۵
۱۰	۵۴۹
۱۱	۴۵

جدول ۲. تعداد عوارض در مجموعه اقلام برحسب طول اقلام

پس آماده‌سازی داده‌ها، زمان اجرای مدل اپریوری است که برای یافتن الگوهای پرتکرار می‌توان از آن استفاده کرد. برای اجرای مدل نیاز به تنظیمات اولیه می‌باشیم، اولین نکته تعیین حداقل مقدار پشتیبان می‌باشد، برای اجرای مدل حداقل سطح پشتیبان را ۰/۰۵(۵٪) در نظر می‌گیریم، به عبارتی در مدل فقط از اقلامی استفاده خواهد شد که تعداد دفعات وقوع آن حداقل ۵٪ باشد.

نتایج تحلیل مجموعه اقلام با بالاترین پشتیبانی را نشان می‌دهند. بالاترین پشتیبان با ۷۸/۵٪ مربوط به دفاتر بیمه ((IC، سپس رستوران‌ها ((R با ۷۳/۵٪، در مرتبه بعد دفتر اسناد رسمی (Np) با ۷۰/۱٪ و داروخانه با ۷۰/۰۴٪ و... قرار می‌گیرد. به عبارتی در ۷۸/۵ درصد مجموعه اقلام دفاتر بیمه، در ۷۳,۵ درصد مجموعه اقلام رستوران وجود دارد. در این گروه مجموعه اقلام با حداقل پشتیبان بیش از ۵٪ دارای; کمترین طول ۱(به عبارتی مجموعه اقلام دارای یک‌قلم) برابر با ۴۵ می‌باشد، بیشترین طول نیز ۱۱ قلم می‌باشد که تعداد آن برابر با ۱۵ می‌باشد جدول شماره ۳ توزیع مجموعه اقلام پرتکرار را نشان می‌دهد.

جدول ۳. مشخصات مجموعه اقلام پرتکرار

طول مجموعه اقلام	نمونه آیتم‌های مجموعه اقلام	پشتیبان	تعداد
1	(Ic)	0.785	24
2	(R, Ic)	0.635	380
3	(R, Ic, Np)	0.525	2473
4	(R, Ph, Ic, Np)	0.433	8798
5	(Np, Ph, Sch, Ic, R)	0.344	20101
6	(Np, Ph, Sch, Ic, Ta, R)	0.263	32568
7	(Np, Ph, Sch, Ic, Ta, R, G)	0.188	40479
8	(Np, Ph, Sch, Ic, Cl, Ta, R, La)	0.143	44364
9	(Np, La, Ph, Sch, Ic, Cl, Ta, R, G)	0.108	46457
10	(Np, La, Ph, Sch, Cl, Ic, Ta, Ra, R, G)	0.085	46978
11	(Np, La, Ph, Sch, Cl, Ic, Ta, Dc, Ra, R, G)	0.062	47019

جدول شماره ۴ نمونه‌ای از مجموعه‌های پر اقلام را نشان می‌دهد کارکرد این جدول این است که می‌توانیم تشخیص

دهیم پیرامون شعب بانک‌ها در تهران چه مجموعه اقلامی بیشترین تکرار را دارند. در مجموعه اقلام با طول بیشتر از ۱ نشان می‌دهد که معمولاً چند قلم عارضه شهری با هم ظاهر می‌شوند برای مثال دفتر اسناد رسمی، دفاتر بیمه، رستوران‌ها و... معمولاً در مراکز تجاری و اداری بیشتر مکان‌یابی می‌شود که در مجاورت آن معمولاً شعب بانک نیز وجود دارد.

جدول ٤. جدول نمونه ۱۰ قاعده الگوی تکرار فضایی

طول تالی	طول مقدم	بالابر	اطمینان	پشتیبان	پشتیبان تالی	پشتیبان مقدم	تالی	مقدم
1	2	1.23	87%	52%	71%	60%	(Np)	(Ph, Ic)
2	1	1.23	73%	52%	60%	71%	(Ph, Ic)	(Np)
2	1	1.21	66%	47%	55%	71%	(R, Ph)	(Np)
1	2	1.21	86%	47%	71%	55%	(Np)	(R, Ph)
2	1	1.21	57%	44%	47%	79%	(Np, Ta)	(Ic)
1	2	1.22	87%	44%	71%	51%	(Np)	(Ic, Ta)
2	1	1.27	80%	44%	63%	56%	(Ic, Np)	(Ta)
2	1	1.22	63%	44%	51%	71%	(Ic, Ta)	(Np)
1	2	1.27	70%	44%	56%	63%	(Ta)	(Ic, Np)
1	2	1.21	95%	44%	79%	47%	(Ic)	(Np, Ta)

در مرحله بعد پیدا کردن ارتباط و پیوند بین اقلام مکانی می‌باشد، پس از اجرای الگوریتم اپریوری و یافتن اقلام پرتکرار، حال زمان آن رسیده است تا قوانین انجمنی به دست آید. از قوانین انجمنی می‌توانیم اطلاعات و دانش مرتبط به اینکه چه اقلامی به‌طور مؤثرتر باهم استقرار می‌یابند را کشف کنیم به عبارتی مجموعه‌هایی که هم آیند هستند و مکمل یکدیگر محسوب می‌شوند را شناسایی می‌کنیم.

قواعد مجموعه اقلام پرتکرار و قوانین انجمنی، حاوی اطلاعات بسیاری مفیدی برای تحلیل و تصمیم‌گیری مکانی در محدوده‌های شهری می‌باشند. برای نمونه در جدول فوق ردیف اول مقدم ph(بیمارستان) و Ic (دفاتر خدمات بیمه) با ۵۲ درصد پشتیبان است و تالی دفتر اسناد رسمی (Np) است به عبارتی اگر در محدوده شعبه بیمارستان و دفتر خدمات بیمه وجود داشته باشد، دفتر اسناد رسمی نیز مستقر خواهد بود. یا قاعده دیگر (۱۶۱۸۹۶۱) با اطمینان ۱ و پشتیبان تالی ۰/۷، تعداد اقلام مقدم ۱ و تالی ۶ است، در این قاعده مقدم Ph (داروخانه) است و تالی آن فیزیوتراپی، درمانگاه، رادیولوژی، آزمایشگاه تشخیص طبی، دفتر روزنامه، اماکن دولتی است به عبارتی مجموعه اقلام پرتکرار شامل این هفت‌قلم می‌شود و زمانی که در محدوده ۲۰۰ متری شعبه بانک، داروخانه مستقر باشد تا ۷۰ درصد انتظار داریم که این ۷ قلم نیز مکان‌یابی شده باشند. خوانش قاعده (۳۰۸۶۵۶۹) با اطمینان ۱ و پشتیبانی تالی ۷۰٪ مقدم شامل دفتر اسناد رسمی (Np) است و تالی دفتر ازدواج و طلاق (Md)، دفتر بیمه (Ic)، آژانس مسافرتی ((Ta، اماکن دولتی (G)، رستوران (R)، آزمایشگاه تشخیص طبی (La) است لذا انتظار داریم، در مکان‌هایی که دفاتر اسناد رسمی وجود داشته باشد، انواع خدمات مجموعه اقلام فهرست تالی فوق نیز مستقر باشد. یک قاعده دیگر با سه مقدمه و سه تالی، و پشتیبان ۵٪ و اطمینان ۶۷٪ که بالاترین مقدار بالابر را دارد، مقدم شامل مجموعه اقلام پرتکرار بیمارستان (Ho)، داروخانه (Ph)، کلینیک (Cl) و تالی شامل آزمایشگاه تشخیص طبی ((La، اورژانس (Uc)، رادیولوژی (Ra) است. این مجموعه اقلام دقیقاً از یک جنس می‌باشند و نشان‌دهنده آن می‌باشد که حضور بیمارستان، داروخانه و کلینیک به‌عنوان مقدم، خدمات و مراکز مکمل شامل آزمایشگاه تشخیص طبی، مرکز اورژانس و رادیولوژی را جذب خود می‌نماید و به عبارتی نوعی پیوند و همبستگی بین این مجموعه فعالیت‌ها در مجموعه‌های شهری وجود دارد و کاوش و کشف چنین ارتباطاتی به‌خوبی در شناخت فعالیت‌های شهری، فرایندهای شهری و نیز پیوند و ارتباط بین خدمات را مشخص می‌سازد، در واقع برای برنامه‌ریزی شهری،

دستورالعملی ارائه می‌دهد تا شناخت بهتری نسبت به فعالیت‌های اقتصادی، اجتماعی شهری به وجود آید، خدمات مکمل و یا رقیب را شناخت، فرایندهای شهری را کشف کرد و بر اساس آن به مدیریت هوشمندانه شهری دست پیدا کرد. در مدل‌های جاری، برنامه‌ریزی معمولاً بر اساس ادبیات و نظریه‌ها در خصوص شهر اظهارنظر می‌شود کشف الگوهای توزیع فعالیت‌ها و خدمات و همبستگی و وابستگی‌های آن‌ها می‌تواند شناخت واقع‌گرایانه‌تر نسبت به پیوندهای شهری ارائه دهد.

نتیجه‌گیری

داده‌کاوی فضایی، مجموعه از روش‌شناسی را معرفی می‌کند تا به کمک آن الگو، بافت، توزیع، پیچیدگی‌ها و نظم نهفته در مجموعه عوارض شهری را کاوش و شناسایی کرد. توسعه شهری، پیچیدگی مسائل شهری و در هم تنیدگی آن‌ها لزوم کاربرد روش‌های هوشمند را مطرح می‌سازد، کشف الگوهای پرتکرار فضایی، یکی از روش‌های مفید برای شناسایی الگوهایی بشمار می‌رود که ما را درک و شناخت شهری کمک می‌کند. پدیده‌های شهری دارای ارتباط پسین و پیشین با یکدیگر هستند و مجموعه اقلام شهری پرتکرار، بسته‌ها و الگوهایی را معرفی می‌کند تا ارتباط و پیوستگی انجمنی عوارض را شناخت. در این مقاله یکی از پرکاربردترین روش‌های داده‌کاوی در کشف الگوهای پرتکرار یعنی اپریوری معرفی شد و به کمک این روش توانستیم مجموعه اقلام پرتکرار و درنهایت شناسایی عوارضی که با یکدیگر ظهور می‌یابند را شناسایی کنیم و پایگاه قواعد الگوهای ایجادشده می‌تواند راهنمایی برای شناسایی، راهنمایی برای شناخت ارتباطات و پیوندهای بین عوارض شهری باشد و مجموعه‌ای از اقلام فضایی پرتکرار در شهر تهران را مشخص کند. برای کشف این الگوها موقعیت بانک‌ها و مؤسسات اعتباری را به‌عنوان نمونه انتخاب شد و شعاع ۲۰۰ متری شعب جستجو شد و فعالیت‌هایی که وجود داشتند فهرست شدند. پس از آماده‌سازی داده‌ها، به کمک الگوریتم اپریوری به تحلیل الگوهای پرتکرار فضایی در محدوده شعب پرداخته شد. موضوع بصری سازی فضایی، بخشی از این الگوها و قواعد را به تصویر می‌کشند این الگوها از طریق مشخص کردن رابطه بین فعالیت‌های شهری حاوی اطلاعات کاربردی برای مدیریت و برنامه‌ریزی شهری در سطوح مختلف می‌باشند. اینکه که مشخص شود چه کاربری‌هایی باهم می‌آیند ما را به درک بهتر همزیستی فعالیت‌ها هدایت می‌کند، اینکه چه فعالیت‌ها و کاربری‌هایی در کنار هم مستقرشده‌اند و این الگوها در بخش‌های مختلف شهر همواره تکرار می‌شود. آنچه در الگوها نهفته است حاصل فرایندهای شهری است که در آن عوامل مختلف دموگرافیک، اقتصادی، اجتماعی و ... نقش داشته‌اند و حوزه‌های مختلف می‌توانند مورداستفاده قرار گیرند در حوزه اقتصاد شهری، پیوندهای اقتصادی بین فعالیت‌های سازگار مشخص می‌شود، اینکه چه فعالیت‌هایی می‌توانند مکمل هم باشند و یکدیگر را جذب می‌کنند. برای نمونه بر اساس یکی از قواعد استخراج‌شده، در محدوده استقرار شعب بانک صادرات شرکت‌های بیمه قرار دارند، (به عبارتی بین استقرار محل شرکت بیمه و محل شعبه بانک صادرات رابطه وجود دارد)، این مقدم قاعده است، تالی و نتیجه چنین بیان می‌شود، در مکان‌هایی که شرکت‌های بیمه مکان‌یابی شده‌اند، آژانس‌های مسافرتی، دفترخانه ثبت‌اسناد رسمی نیز وجود خواهد داشت. کشف چنین الگوهایی به روابط و پیوندهای بین این فعالیت‌ها اشاره دارد و مدیران بنگاه‌های اقتصادی در شهر (مانند دفتر اسناد رسمی) می‌توانند تشخیص دهند که یک گزینه برای مکان‌یابی مجاورت با شرکت‌های بیمه، بانک‌ها، شرکت‌های مسافرتی است که مدنظر قرار دهند سایر قواعد نیز برای اغلب فعالیت‌ها دارای چنین قواعدی هستند عشورنژاد و همکاران (۱۳۹۶)، مدلی برای ارائه مکان بهینه استقرار شعب بانک‌ها در تهران پرداخته‌اند، آن‌ها ۱۷ عارضه شهری را مطالعه کرده‌اند، در این تحقیق کل شعب بانک‌های تهران (۳۷۷۳ شعبه) و ۵۷ قلم اطلاعاتی (بالغبر ۸۰ هزار عارضه) موردبررسی قرار گرفت، در قوانین استخراجی بیشترین اطمینان و پشتیبان مربوط به سینما، پارک، دبیرستان،

دانشگاه می‌باشد، قواعد تحقیق حاضر فیزیوتراپی، شرکت‌های بیمه، دفتر اسناد رسمی را به‌عنوان الگوهای پرتکرار را مشخص می‌کند، امکان مقایسه نتایج کار ضعیف است زیرا ورودی‌های دو مقاله، شیوه پردازش متفاوت می‌باشد ولی در دو مقاله دفاتر اسناد رسمی به‌عنوان یکی از عوامل معرفی‌شده‌اند.

برای تحقیقات آتی پیشنهاد می‌شود این الگوها به‌صورت محلی بررسی شوند تا اثرات زمینه بر الگوها استخراج شود. در این تحقیق کل تهران به‌صورت یکپارچه در نظر گرفته‌شده است و در تحقیقات بعدی به اختلافات فضایی لازم است توجه شود و بعد جغرافیایی نیز به تحلیل‌ها اضافه شود و همچنین الگوریتم‌های دیگر کاوش الگوهای پرتکرار معرفی شوند و نیز علاوه بر بانک‌ها به سایر عوارض شهری نیز پرداخته شود.

حامی مالی

این اثر حامی مالی نداشته است.

سهم نویسندگان در پژوهش

نویسنده در تمام مراحل و بخش‌های انجام پژوهش به تنهایی انجام گرفته است.

تضاد منافع

نویسنده اعلام می‌دارند که هیچ تضاد منافعی دررابطه‌با نویسندگی و یا انتشار این مقاله ندارند.

تقدیر و تشکر

نویسنده از همه کسانی که در انجام این پژوهش به ما یاری رساندند، به‌ویژه کسانی که کار ارزیابی کیفیت مقالات را انجام دادند، تشکر و قدردانی می‌نمایند.

منابع

عشورنژاد، غدیر؛ عشورنژاد، انیس و تومانیان، آرا. (۱۳۹۶). ارائه روش‌های جدید هرس کردن قوانین طبقه‌بندی انجمنی مورد کاوی: قوانین مرتبط با بهینگی مکان استقرار بانک‌های شهر تهران. *نشریه مهندسی نقشه‌برداری و اطلاعات مکانی*، *۸* (۲)، ۴۸-۳۹.

References

Ashournejad, Q., Ashournejad, A., & Toomanian, A. (2017). New Methods of Pruning Associative Classification Rules (Case Study: Rules Related to the Optimality of Location of Banks in Tehran City). *Geospatial Engineering Journal*, *8*(2), 39-48. [In Persian]

Andriani, S. P., (2020). Sales Display Re-layout Based on Analysis of Item Sets Pattern Using Combination of Complete Linkage Hierarchical Clustering Method and Association-Rule Method with Apriori Algorithm.

Angeline, D., (2013). Magdalene Delighta. "Association rule generation for student performance analysis using apriori algorithm. *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, 1(1), 12-16. https://doi.org/10.9756/sijcsea/v1i1/01010252

Appice, A. (2003). Discovery of spatial association rules in geo-referenced census data: A relational mining approach. *Intelligent Data Analysis*, 7(6), 541-566. https://doi.org/10.3233/ida-2003-7604

Appice, A. (2005). Mining and filtering multi-level spatial association rules with ARES." *International Symposium on Methodologies for Intelligent Systems*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11425274_36

Bhatia, P. (2024). *Machine Learning with Python*. Cambridge: Cambridge University Press.

Bin, R. (2014). Research on tourism service intelligent recommendation system based on apriori-MD algorithm." *Applied Mechanics and Materials*. Vol. 651. Trans Tech Publications Ltd. https://doi.org/10.4028/www.scientific.net/amm.651-653.1642

Cox, D., Isham, V., & Northrop, P. (2000). Statistical Modeling and Analysis of Spatial Patterns. In U. Dieckmann, R. Law, & J. Metz (Eds.), *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (Cambridge Studies in Adaptive Dynamics, pp. 65-88). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511525537.029

Coyle, D., Meyer, O., & Staschen-Dielmann, S. (Eds.). (2023). *A Deeper Learning Companion for CLIL: Putting Pluriliteracies into Practice*. Cambridge: Cambridge University Press.

Egidi, G. (2020). Towards local forms of sprawl: A brief reflection on mediterranean urbanization. *Sustainability*, 12(2), 582. https://doi.org/10.3390/su12020582

Eick, Ch. F. (2008). Finding regional co-location patterns for sets of continuous variables in spatial datasets. *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems.* https://doi.org/10.1145/1463434.1463472

El Naqa, I., & Martin J. M. (2015). *What is machine learning?." machine learning in radiation oncology*. Springer, Cham, 3-11. https://doi.org/10.1007/978-3-319-18305-3_1

Fotheringham, A. S., & Chris, B. (1999). Local forms of spatial analysis. *Geographical analysis*, 31 (4), 340-358. https://doi.org/10.1111/j.1538-4632.1999.tb00989.x

Fotheringham, A., Stewart, Chris B., & Martin Ch. (2000). *Quantitative geography: perspectives on spatial data analysis*. Sage.

Goodchild, M. F. (2009). *First law of geography*. *International encyclopedia of human geography*. Elsevier Inc. 179-182. https://doi.org/10.1016/b978-008044910-4.00438-7

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511754944

Harun, Nur A. (2017). The application of apriori algorithm in predicting flood areas. *management* 17 (18. https://doi.org/10.18517/ijaseit.7.3.1463

Hesse, M., & Jean-Paul, R. (2004). The transport geography of logistics and freight distribution. *Journal of transport geography*, 12(3), 171-184. https://doi.org/10.1016/j.jtrangeo.2003.12.004

Hsieh, Po-Ch. et al. (2020). Combination of acupoints in treating patients with chronic obstructive pulmonary disease: an apriori algorithm-based association rule analysis. *Evidence-Based Complementary and Alternative Medicine*. https://doi.org/10.1155/2020/8165296

Huang, Y., Shashi, Sh., & Hui, X. (2004). Discovering colocation patterns from spatial data sets: a general approach. *IEEE Transactions on Knowledge and data engineering 16*(12), 1472-1485. https://doi.org/10.1109/tkde.2004.90

Jafarzadeh, A. A., Mahdavi, A., & Jafarzadeh, H. (2017). Evaluation of forest fire risk using the Apriori algorithm and fuzzy c-means clustering. *Journal of forest Science*, 63(80, 370-380. https://doi.org/10.17221/7/2017-jfs

Jordan, Michael I., & Tom M. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Kavitha, M., & Subbaiah, S. (2020). Association rule mining using apriori algorithm for extracting product sales patterns in groceries. *Int. J. Eng. Res. Technol*, 8(03, 1-4.

Li, H. (2011). An improved multi-support Apriori algorithm under the fuzzy item association condition." *2011 International Conference on Electronics, Communications and Control (ICECC)*. IEEE,. https://doi.org/10.1109/icecc.2011.6067661

Liao, Sh-h., Yin-Ju, Ch., & Min-yi, D. (2010). Mining customer knowledge for tourism new product development and customer relationship management. *Expert Systems with Applications*, 37(6), 4212-4223. https://doi.org/10.1016/j.eswa.2009.11.081

Malerba, D. (2003). Mining spatial association rules in census data. *Research in Official Statistics. v5 i1*, 19-44.

Mehta, S. V., Shagun, S., & Dhaval, P. (2018). Spatial Co-location Pattern Mining-A new perspective using Graph Database." arXiv preprint arXiv:1810.09007

Miller, Harvey J. (2004). Tobler's first law and spatial analysis. *Annals of the association of American geographers, (94*)2, 284-289.

Miller, H. J., & Jiawei, H. (2011). *Geographic data mining and knowledge discovery*. CRC press.

Priya, G. (2009). Mining co-location patterns from spatial data using rulebased approach. *Journal of Global Research in Computer Science*, *2*(7), 58-61.

Rocha, J., & José, T. (2018). editors. *Spatial Analysis, Modelling and Planning*. IntechOpen. https://doi.org/10.5772/intechopen.81049

Romero, C. (2010). Mining rare association rules from e-learning data. *Educational Data Mining 2010*.

Shah, C. (2022). *A Hands-On Introduction to Machine Learning*. Cambridge: Cambridge University Press.

Shavlik, Jude W., Thomas, D., & Thomas, Glen D. eds. (1990). Readings in machine learning. Morgan Kaufmann.

Shekhar, Sh., & Yan, H. (2001). *Discovering spatial co-location patterns: A summary of results*. International symposium on spatial and temporal databases. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-47724-1_13

Sinaga, Kristina P., & Miin-Shen, Y. (2020). Unsupervised K-means clustering algorithm. *IEEE access* 8, 80716-80727. https://doi.org/10.1109/access.2020.2988796

Tobler, Waldo R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, *46*.sup1 234-240. https://doi.org/10.2307/143141

Verma, A., & Raman, K. (2022). Association Rule Generation using Pattern Mining Apriori Technique. *Journal Of Algebraic Statistics*, 13(2), 550-556.

Yoo, Jin S., & Mark, B. (2012). Mining spatial colocation patterns: a different framework. *Data Mining and Knowledge Discovery*, *24*(1), 159-194. https://doi.org/10.1007/s10618-011-0223-0

Zhou, Zhi-Hua. (2021). *Machine learning*. Springer Nature.