

ارائه مدلی ترکیبی مبتنی بر خوشه‌بندی و قواعد انجمنی برای شناسایی رفتارهای متقلبانه در تراکنش‌های بانکی

- عبدالله عشقی
- مهرداد کارگری
- مصطفی جاویده
- حامد میرشک
- علی محمد نادری

پژوهشکده پولی و بانکی
بانک مرکزی جمهوری اسلامی ایران

شرکت ملی انفورماتیک
بانک مرکزی جمهوری اسلامی ایران

هفتمین همایش سالانه بانکداری الکترونیک و نظام‌های پرداخت

نوآوری، بازیگران جدید و کارآیی در کسب و کار مالی

www.ebps.ir

تقلب

هر نوع سوء استفاده مالی (برداشت، خرید، واریز، جعل و ...) از حساب های بانکی افراد بدون اطلاع آنها و یا فعالیت هایی که منجر به زیان رسانی به بنگاه مالی شود، تقلب محسوب می شود.



کشف تقلب

فرایندی که منجر به شناسایی تقلب شود.

هدف مطلوب

- قبل از اتمام تراکنش
- بعد از اتمام تراکنش

موانع

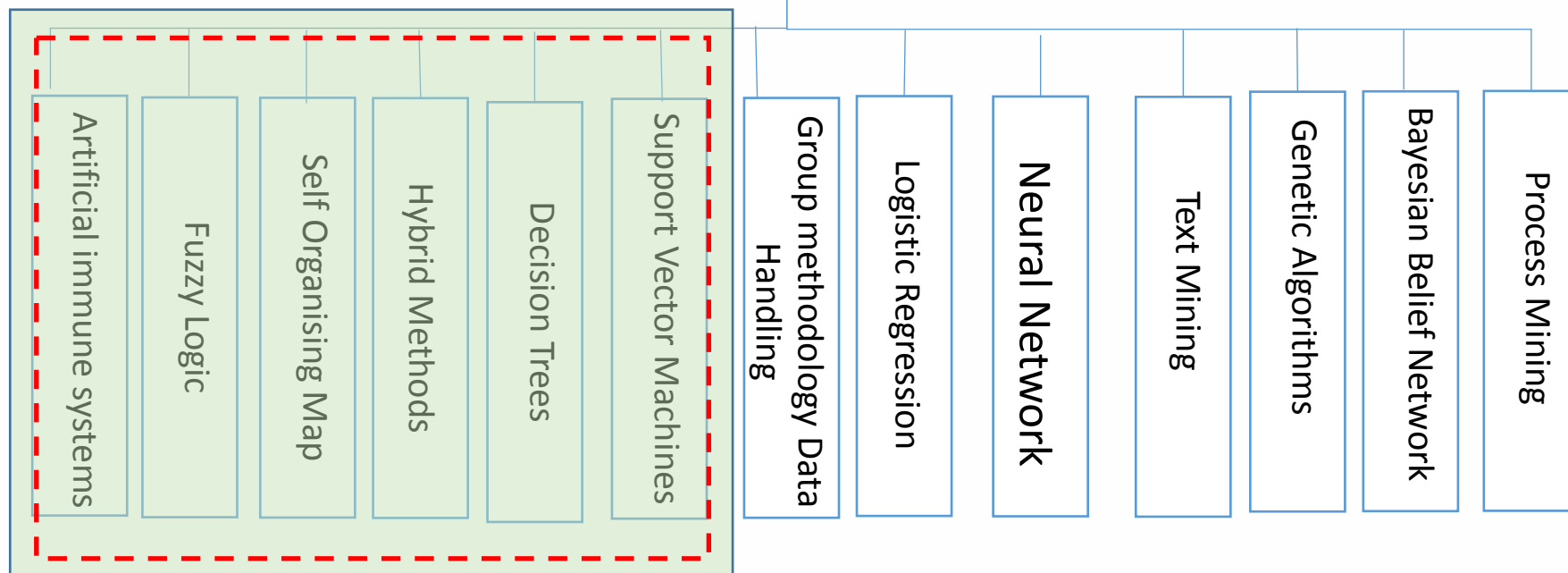
- پیچیده بودن فرایند کشف تقلب (پیدا کردن نوک سوزن در انبار گاه)
- نیاز به سرعت بالا برای پاسخ گویی
- تعداد هشدارهای اشتباه

تکنیک های کشف تقلب

Supervised Methods ?

Unsupervised Methods ?

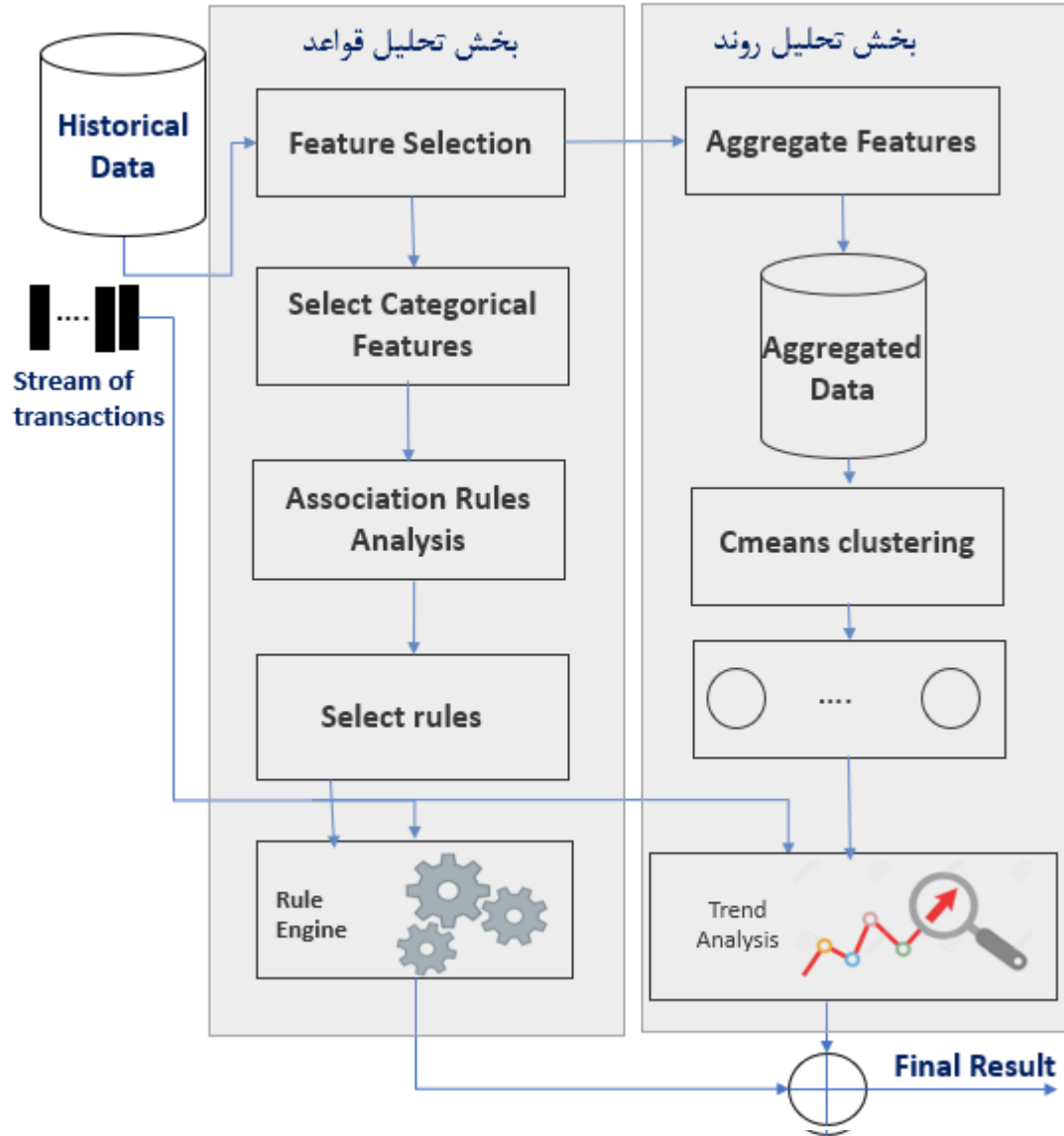
Combinational Methods ?



Credit Card Fraud

(Phua et al., 2012),(Ngai et al., 2011) (West & Bhattacharya, 2016)

چارچوب سیستم کشف قلب



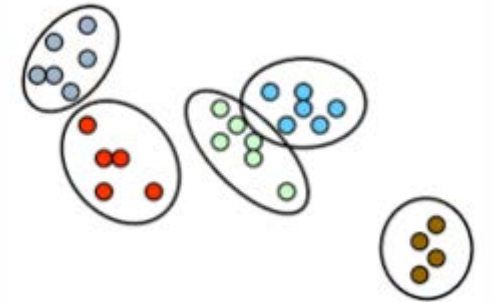
۱- بخش تحلیل قواعد (Supervise)

۲- بخش تحلیل روند (Unsupervised)

Association Rules



Clustering



قواعد انجمنی

تحلیل انجمنی (وابستگی)، یکی از الگوریتم‌های داده‌کاوی است که با استفاده از آن می‌توان وابستگی بین یک یا چند مشخصه را کشف کرد و به این ترتیب می‌توان روابط بین آنها را کمی سازی نمود

مثال: اگر شخصی مرد/زن از طریق کانال موبایل در ساعات آخر روز در صنف کتاب‌فروشان تراکنشی داشته باشد، آنگاه مقدار این تراکنش کمتر از X ریال است

الگوریتم‌ها:

- روش Apriori
- روش Eclat
- روش FP-growth

قواعد انجمنی - رویکرد دو مرحله ای

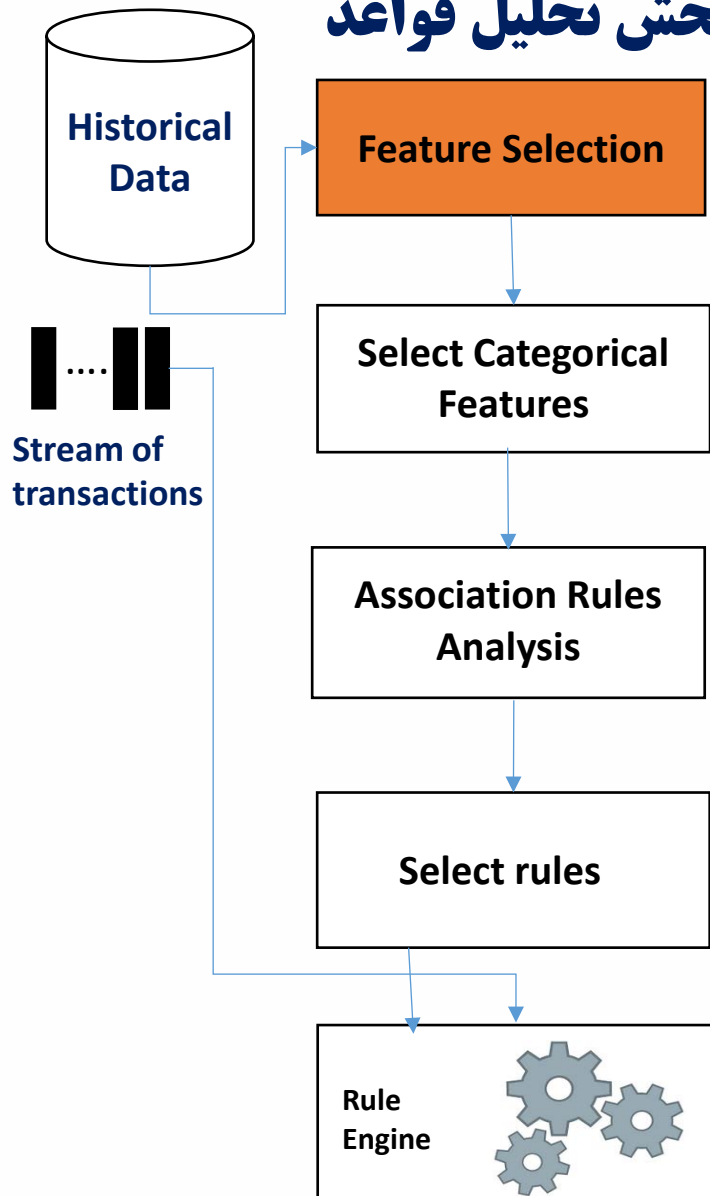
1- Frequent Itemset Generation support > minsupport

2- Rule Generation high confidence rules



متغیرهای اصلی

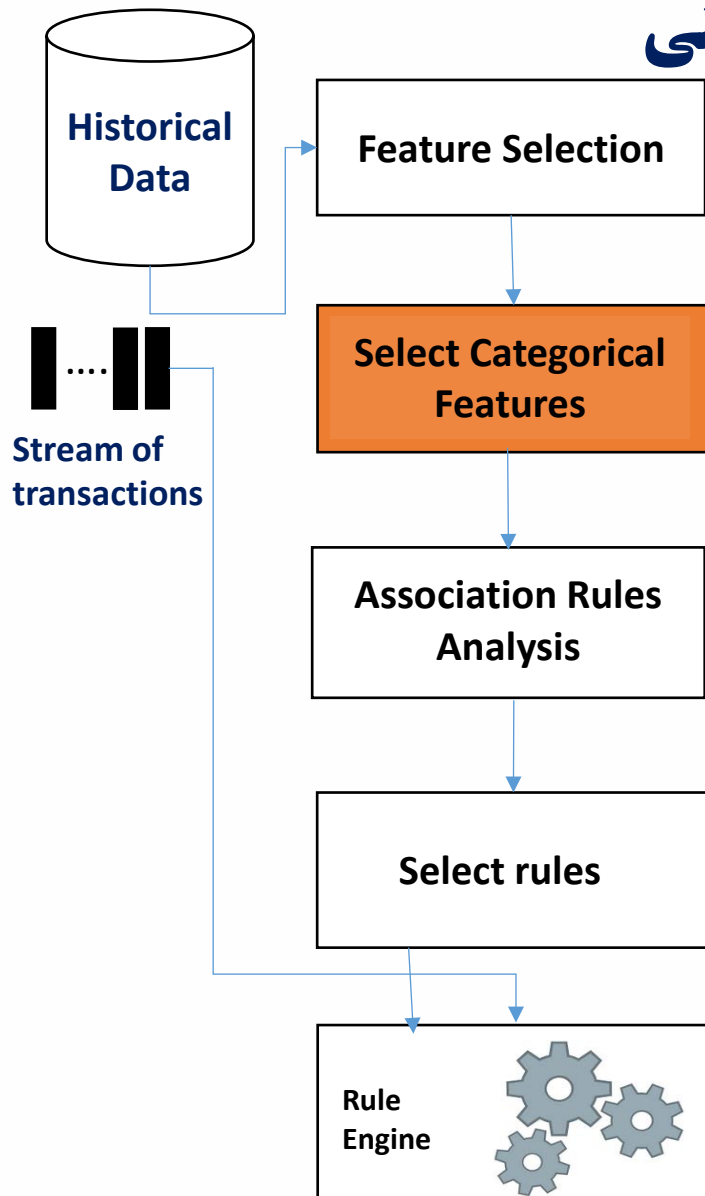
بخش تحلیل قواعد



نام مشخصه	توضیحات
1 Transaction ID	مشخصه تراکنش
2 Time	زمان و تاریخ انجام تراکنش
3 Account number	مشخصه شناسایی حساب مشتری
4 Card number	مشخصه شناسایی کارت
5 Transaction type	و ...ATM،POs کانال انجام تراکنش مثلاً
6 Entry mode	تراکنش با حضور کارت یا بدون کارت
7 Amount	مقدار تراکنش
8 Merchant code	کد شناسایی پذیرنده
9 Merchant group	نوع و صنف پذیرنده
10 Gender	جنسیت دارنده کارت
11 Age	سن دارنده کارت
12 Bank	بانک صادرکننده کارت

Final Result

آماده سازی داده ها برای تحلیل انجمنی



جنسیت	کد			
زن	1	0	0	0
مرد	0	0	0	1

سن	کد			
نوجوان	1	0	0	0
جوان	0	1	0	0
میانسال	0	0	1	0
مسن	0	0	0	1

کانال	کد			
موبایل	1	0	0	0
اینترنت	0	1	0	0
ATM	0	0	1	0
POS	0	0	0	1
کیوسک	1	1	0	0

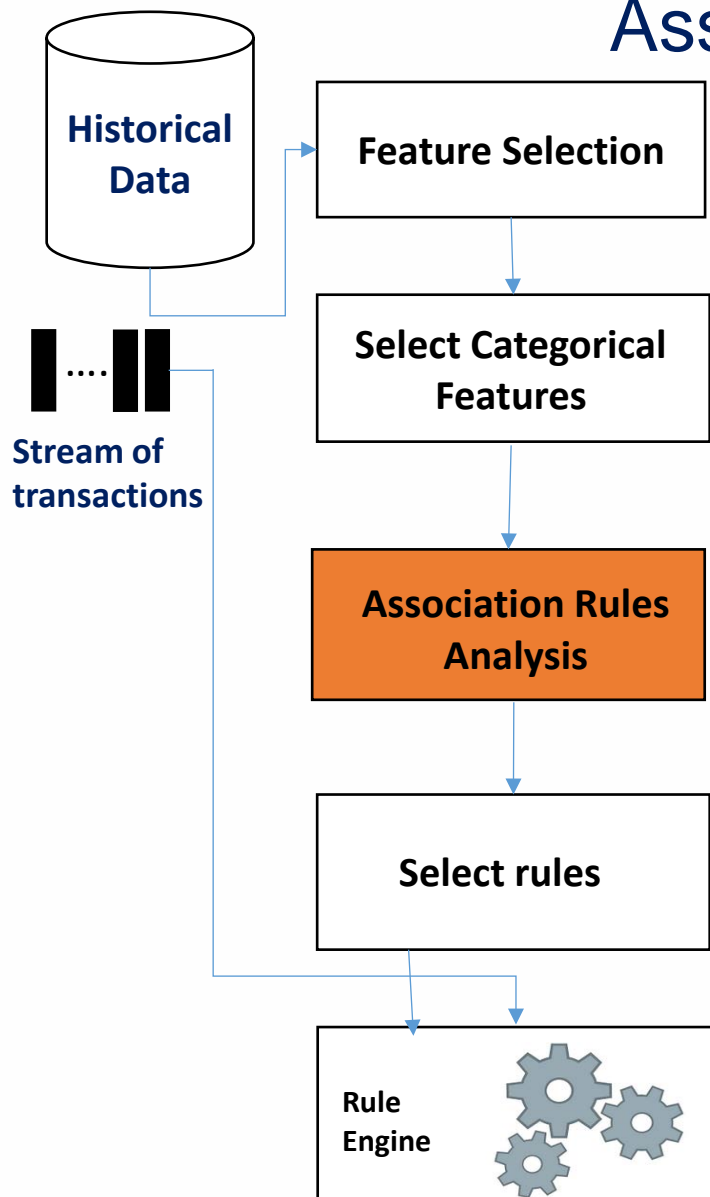
نوع پذیرنده	کد			
نوع ۱	1	0	0	0
نوع ۲	0	1	0	0
نوع ۳	0	0	1	0
نوع ۴	0	0	0	1

زمان	کد			
۰ تا ۷	1	0	0	0
۷ تا ۱۳	0	1	0	0
۱۳ تا ۲۱	0	0	1	0
۲۱ تا ۰	0	0	0	1

Final Result

Association Rules- APRIORI اعمال

- یک الگوریتم کلاسیک برای یادگیری قوانین وابستگی است.
- ورودی این الگوریتم مجموعه‌ای از مجموعه آیت‌ها است.
- این الگوریتم تلاش می‌کند تا زیرمجموعه‌هایی از آیت‌ها را که حداقل بین C مجموعه آیت‌ها مشترک است بیابد.



```
Apriori(T, ε)
L1 ← { large 1-itemsets }
k ← 2
while Lk-1 ≠ ∅
  Ck ← {c | c = a ∪ {b} ∧ a ∈ Lk-1 ∧ b ∈ ∪ Lk-1 ∧ b ∉ a}
  for transactions t ∈ T
    Ct ← {c | c ∈ Ck ∧ c ⊆ t}
    for candidates c ∈ Ct
      count[c] ← count[c] + 1
  Lk ← {c | c ∈ Ck ∧ count[c] ≥ ε}
  k ← k + 1
return ∪k Lk
```

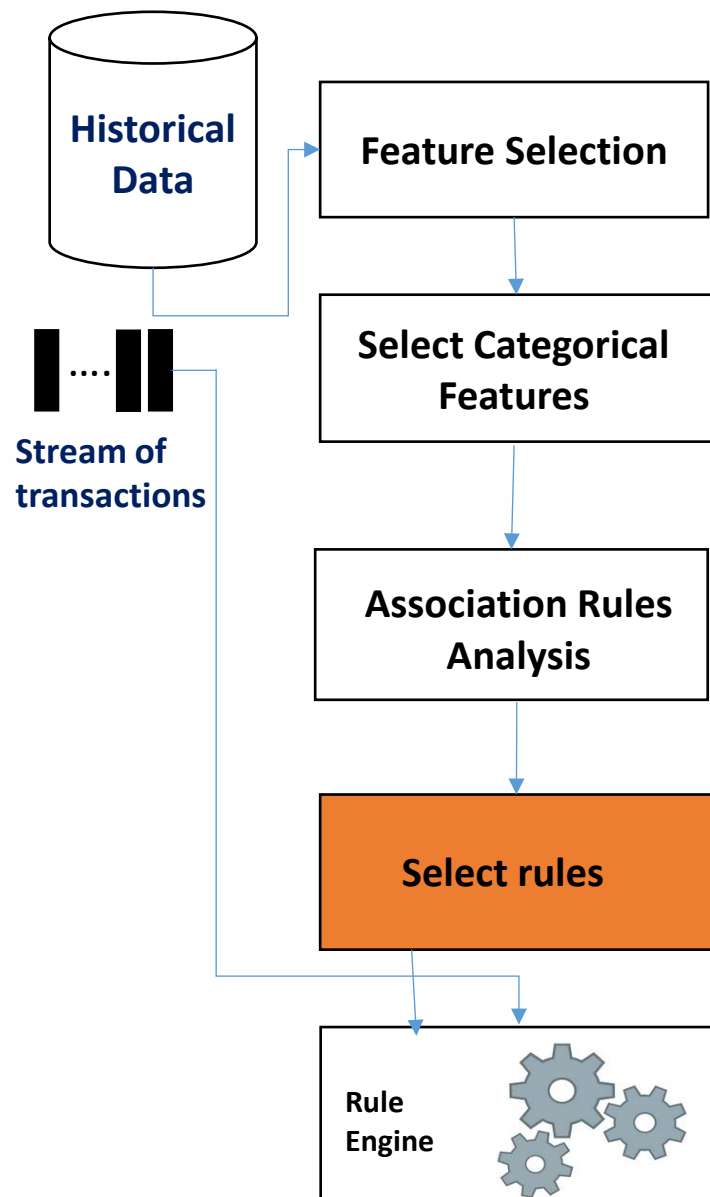
نمونه ای از قواعد انجمنی استخراج شده

"اگر مردی نوجوان در ساعت ۲۲ شب از طریق کانال موبایل تراکنشی و بر روی پذیرنده نوع ۴ تراکنشی انجام دهد که تراکنش قبلی آن نه از آن کانال و نه از همان پذیرنده باشد مقدار تراکنش کم خواهد بود."

Frequent itemset OR Rule

مرد	نوجوان	ساعت ۲۲	موبایل	پذیرنده نوع ۴	عدم تکرار کانال	عدم تکرار پذیرنده	مقدار کم
-----	--------	---------	--------	---------------	-----------------	-------------------	----------

انتخاب قواعد



برای انتخاب قواعد:

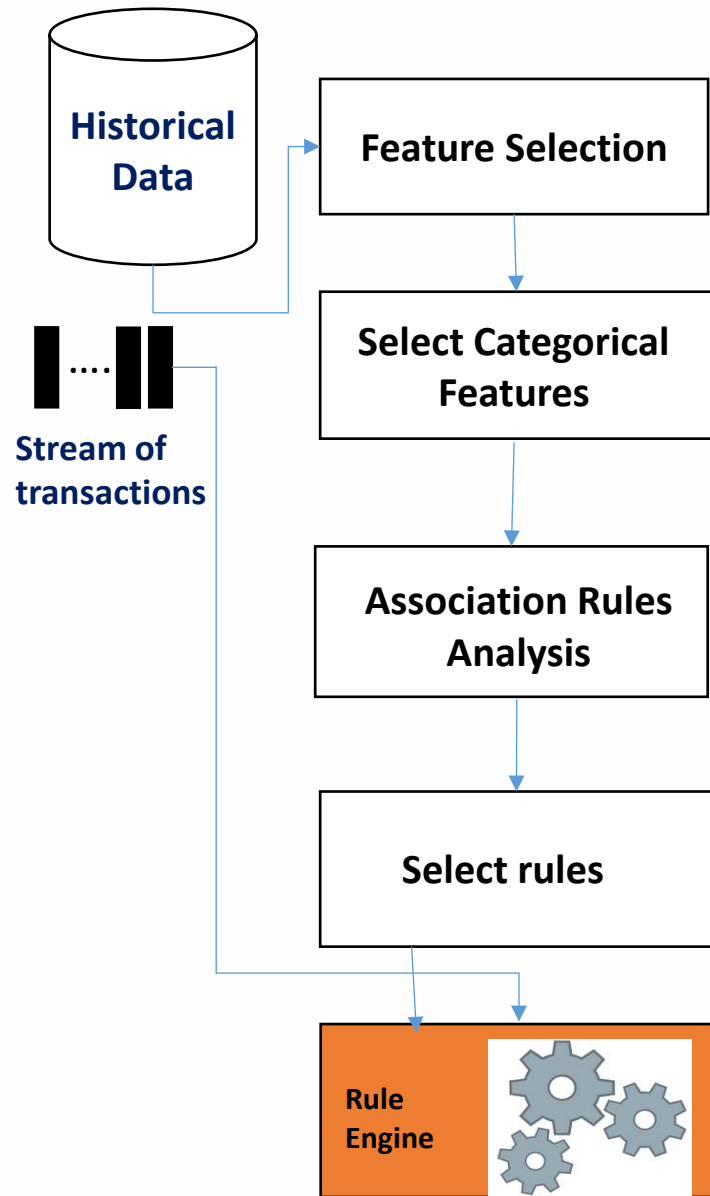
۱- قواعد با confidence بالا

۲- تست قواعد با مجموعه داده تست غیر متقلبانه: هر قاعده ای که به اشتباه

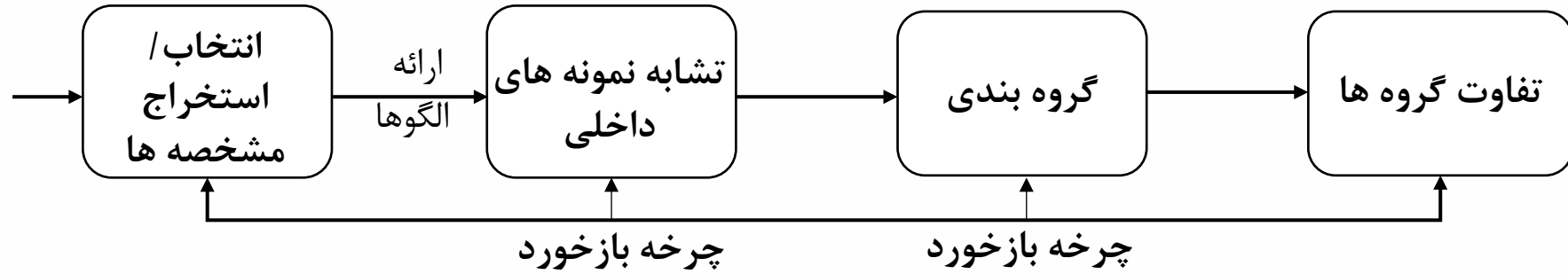
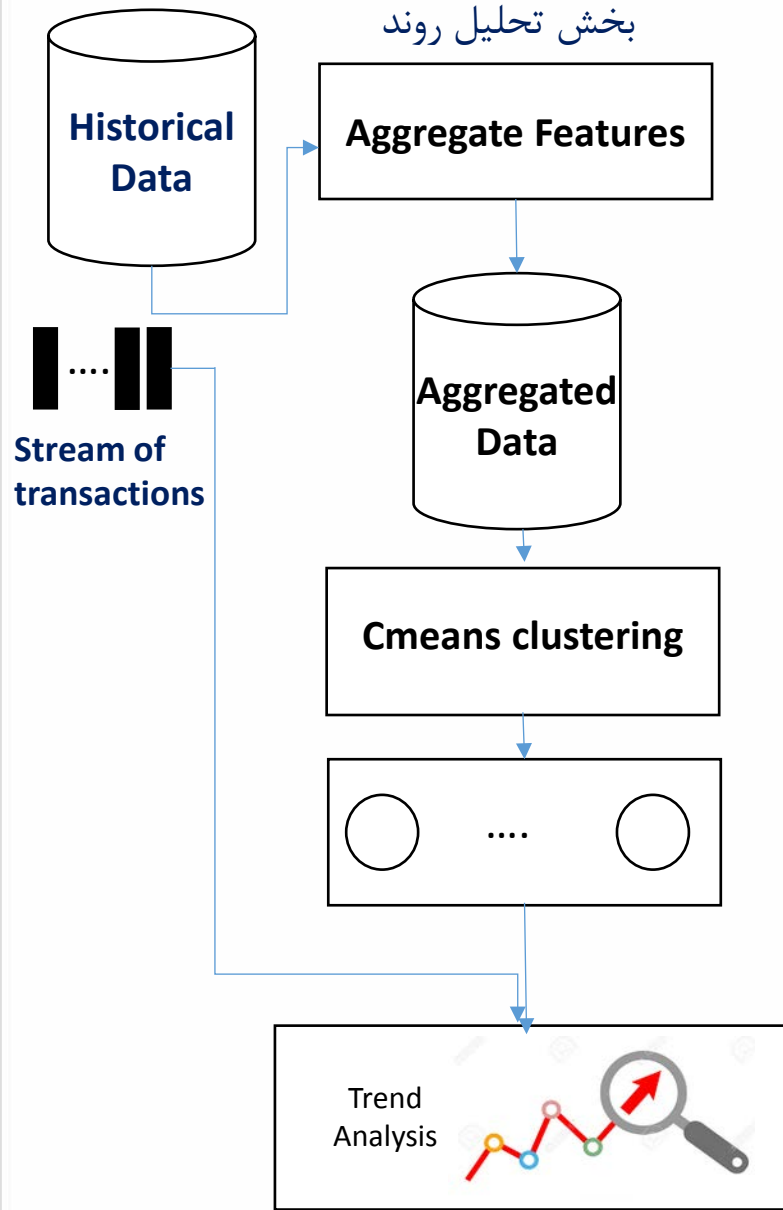
تراکنش غیرمتقلبانه ای را متقلبانه شناسایی کند، حذف می شود

موتور قواعد

اعمال قواعد مربوط به هر تراکنش و تعیین منطبق بودن یا نبودن تراکنش با قاعده یا قاعده های انتخابی برای آن و محاسبه نتیجه بخش قواعد



بخش تحلیل روند



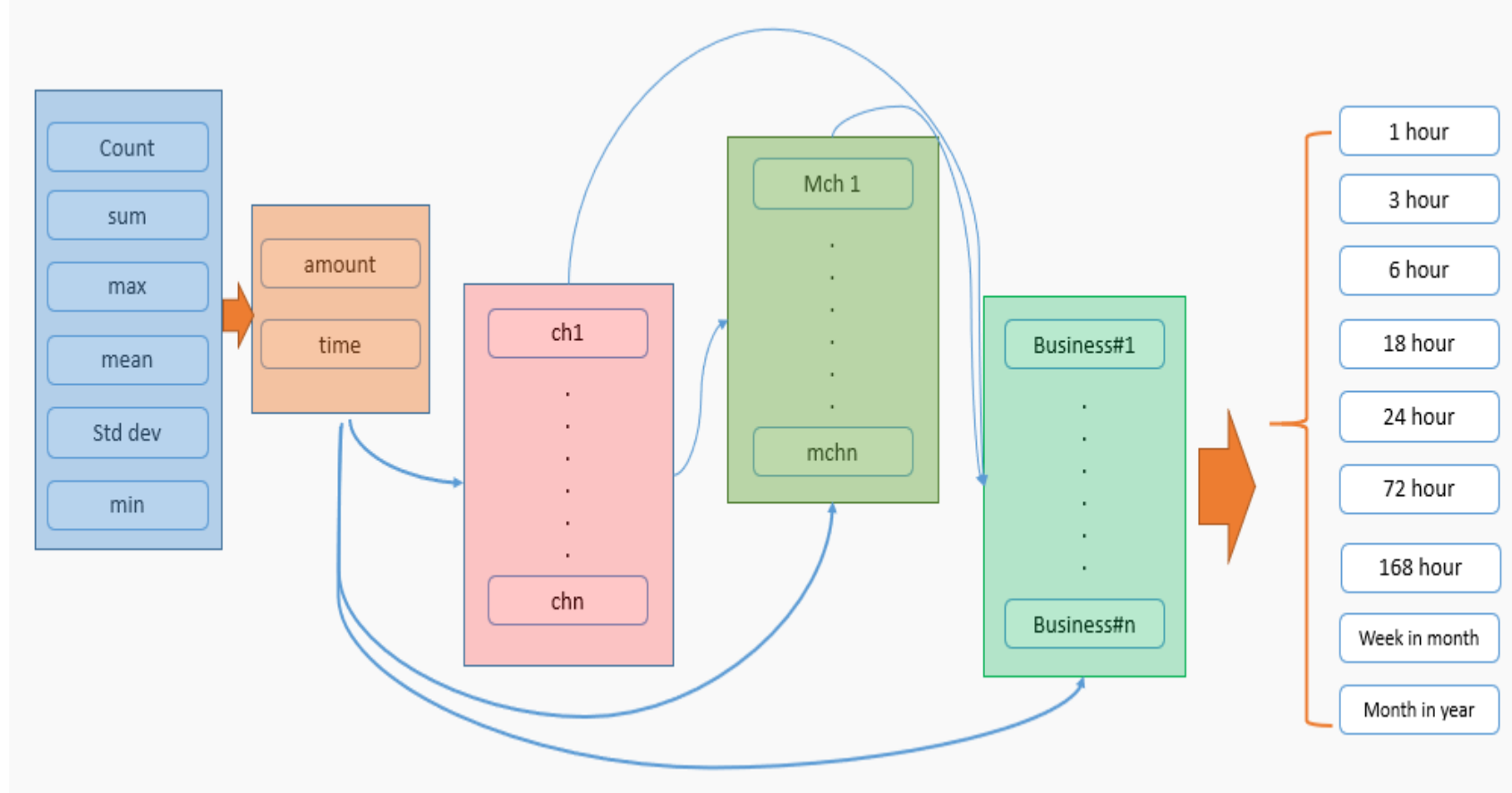
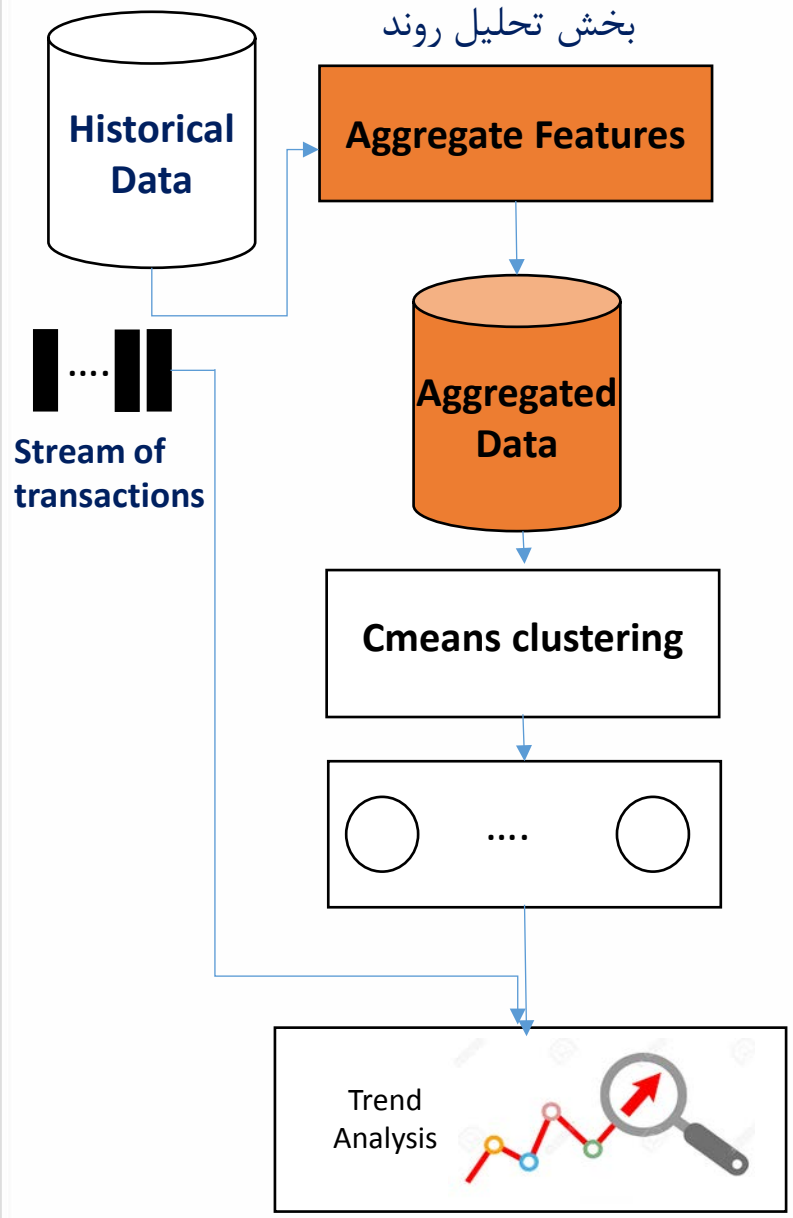
- مبتنی بر مرکزیت
- مبتنی بر پیوستگی
- مبتنی بر چگالی
- مفهومی
- مبتنی بر یک تابع هدف

الگوریتم های خوشه بندی:

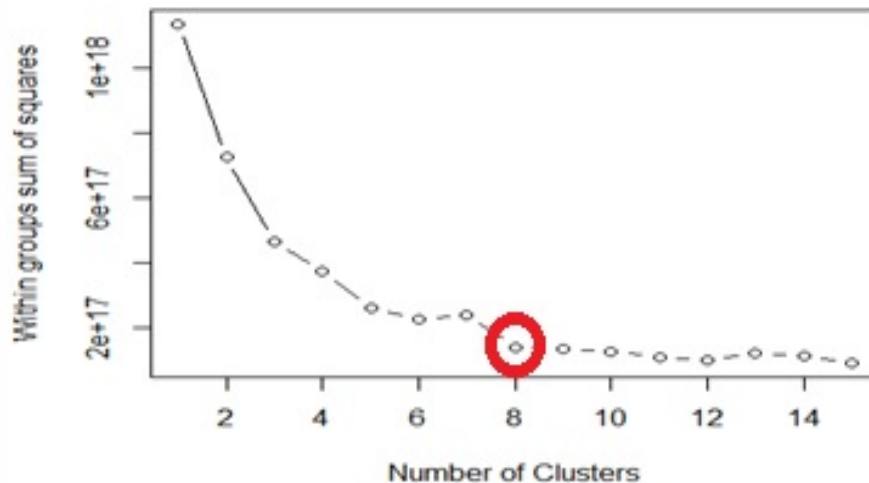
- ۱- الگوریتم های بخش بندی
- ۲- الگوریتم های سلسله مراتبی

بخش تحلیل روند

بخش تحلیل روند

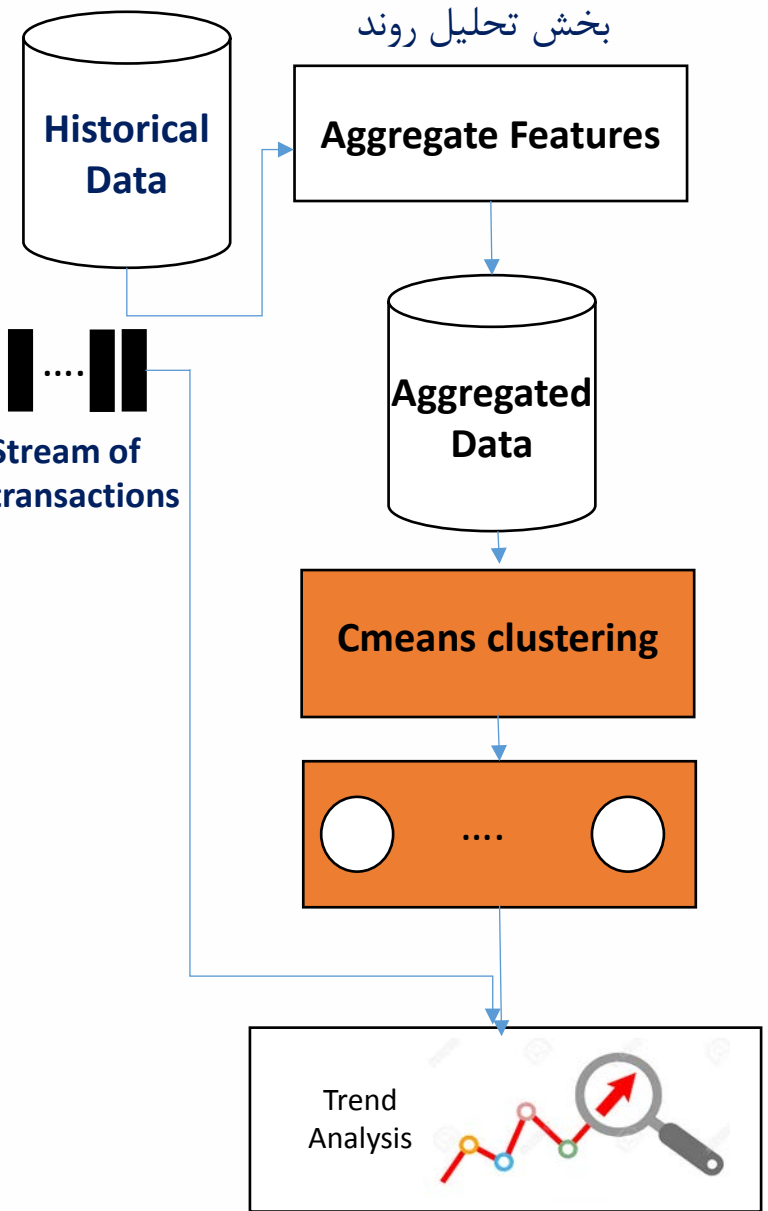


انتخاب تعداد خوشه ها

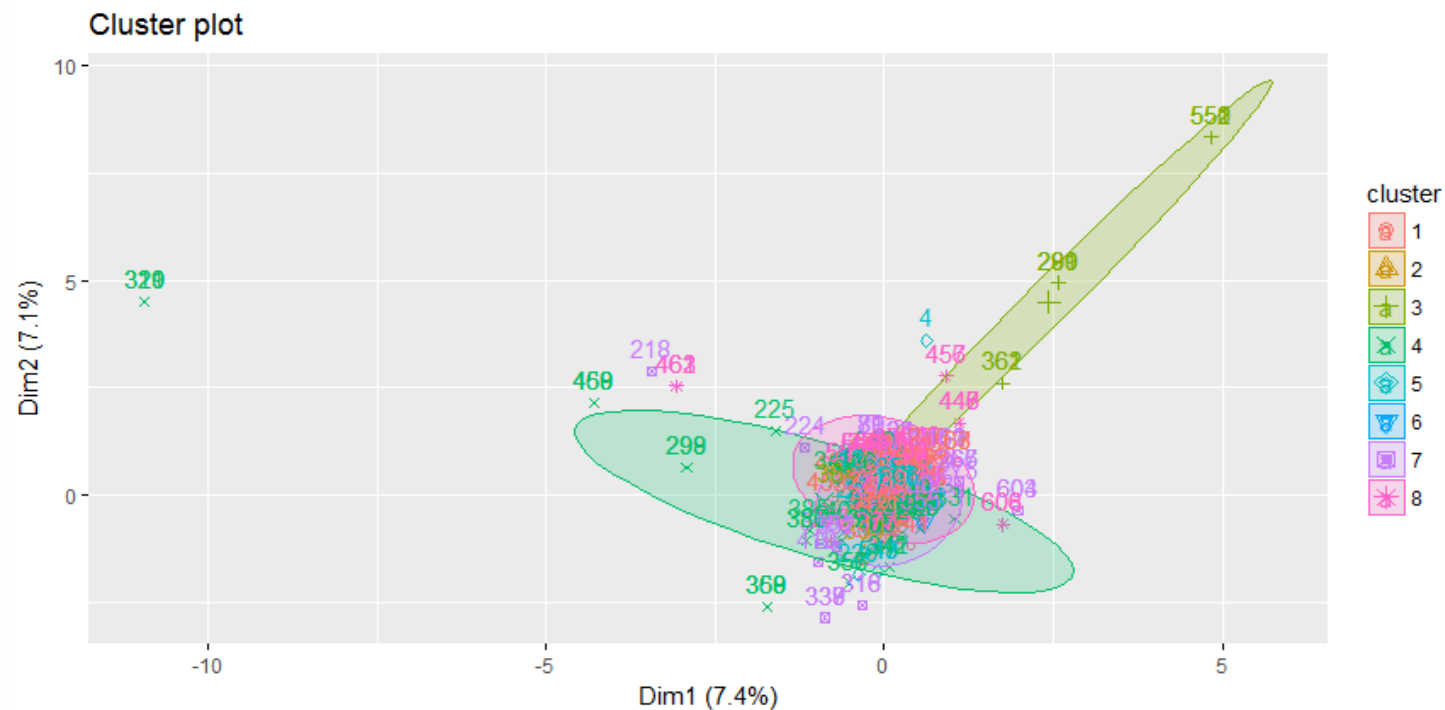
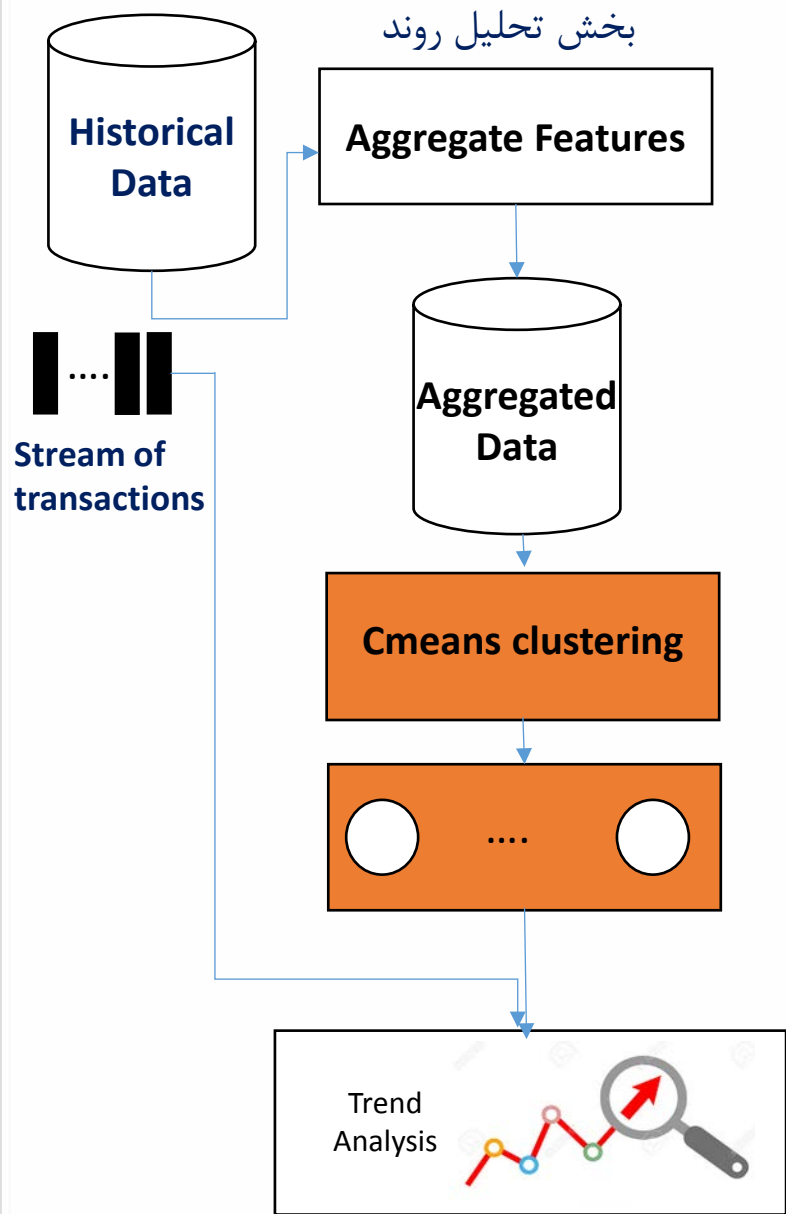


نمونه ای از اعمال خوشه بندی بر تراکش های ورودی

	1	2	3	4	5	6	7	8
202	0.07565578	0.07531216	0.57214645	0.07238493	0.07542144	0.06665178	0.05931105	0.0031164164
551	0.01166859	0.01169947	0.01431993	0.01185935	0.01169018	0.01092711	0.01072896	0.9171064155
286	0.17097992	0.16761352	0.05434246	0.14874751	0.16863930	0.16793644	0.11864531	0.0030955349
615	0.12767577	0.12173979	0.03175306	0.09552415	0.12348869	0.30507318	0.19339987	0.0013454877
654	0.07867602	0.07445176	0.02253146	0.05685035	0.07568711	0.43822453	0.25260020	0.0009785617
32	0.06750322	0.06525850	0.03552227	0.05511506	0.06592094	0.22636271	0.48238600	0.0019312987

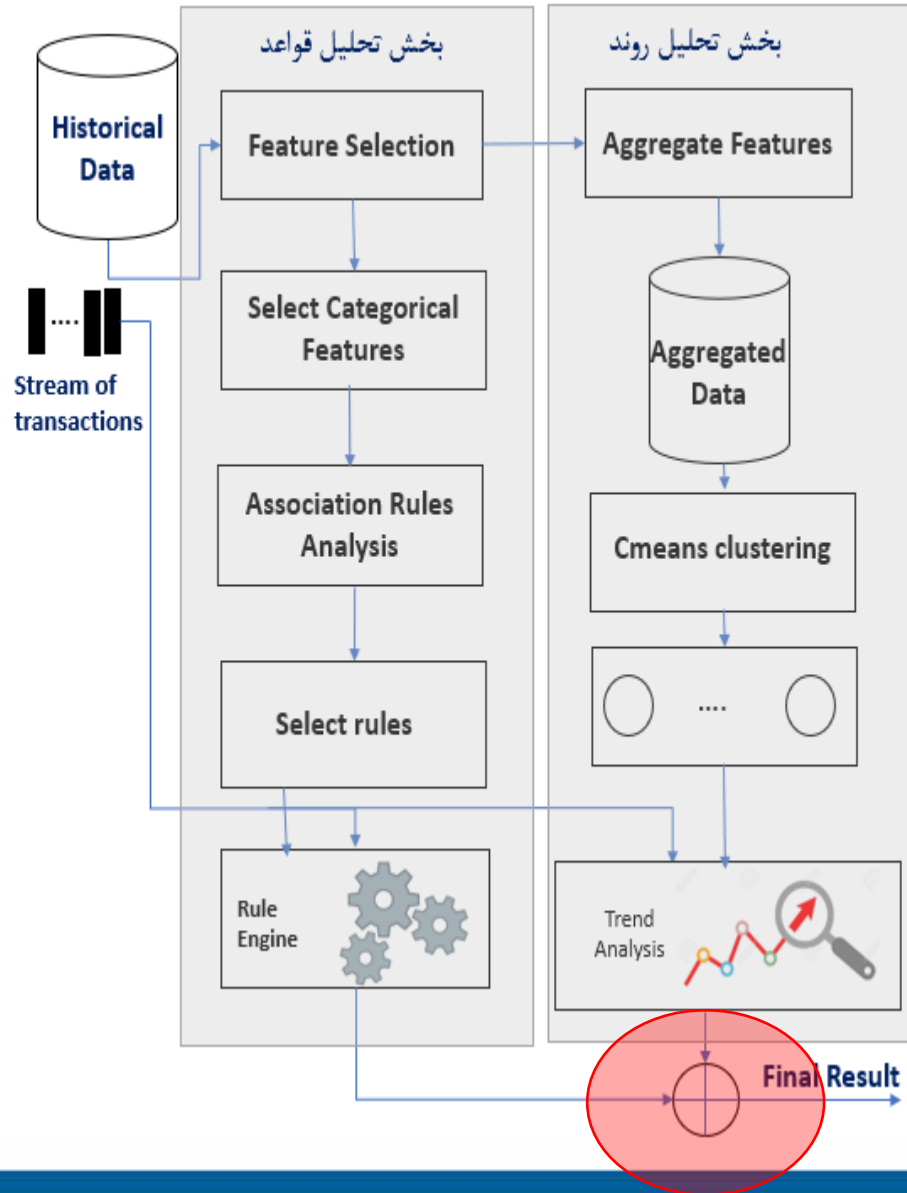


نمونه ای از اعمال خوشه بندی بر تراکنش های ورودی



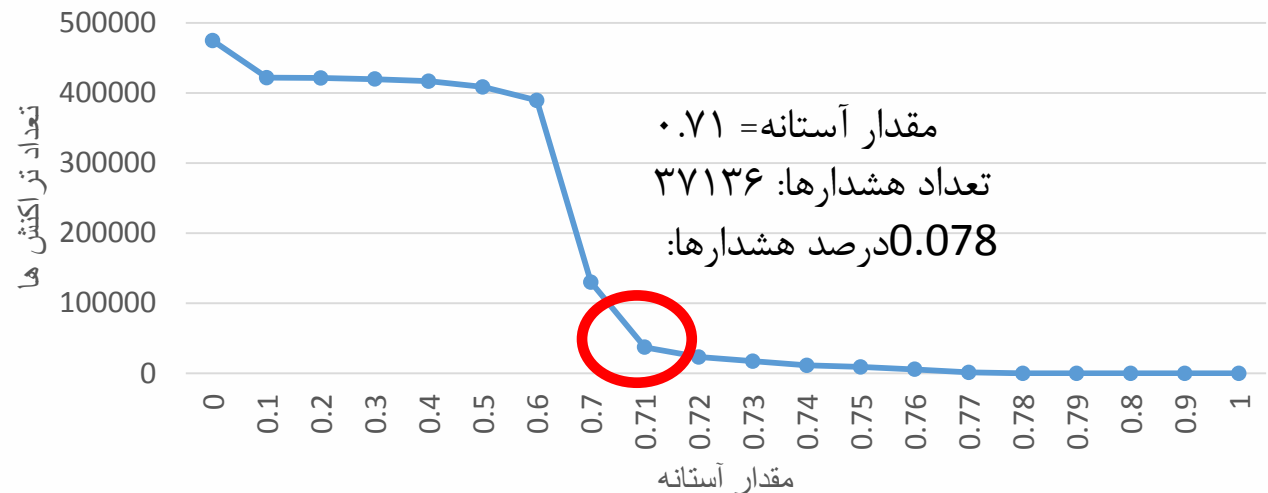
	1	2	3	4	5	6	7	8
202	0.07565578	0.07531216	0.57214645	0.07238493	0.07542144	0.06665178	0.05931105	0.0031164164
551	0.01166859	0.01169947	0.01431993	0.01185935	0.01169018	0.01092711	0.01072896	0.9171064155
286	0.17097992	0.16761352	0.05434246	0.14874751	0.16863930	0.16793644	0.11864531	0.0030955349
615	0.12767577	0.12173979	0.03175306	0.09552415	0.12348869	0.30507318	0.19339987	0.0013454877
654	0.07867602	0.07445176	0.02253146	0.05685035	0.07568711	0.43822453	0.25260020	0.0009785617
32	0.06750322	0.06525850	0.03552227	0.05511506	0.06592094	0.22636271	0.48238600	0.0019312987

ترکیب و تجمیع نتایج دو بخش

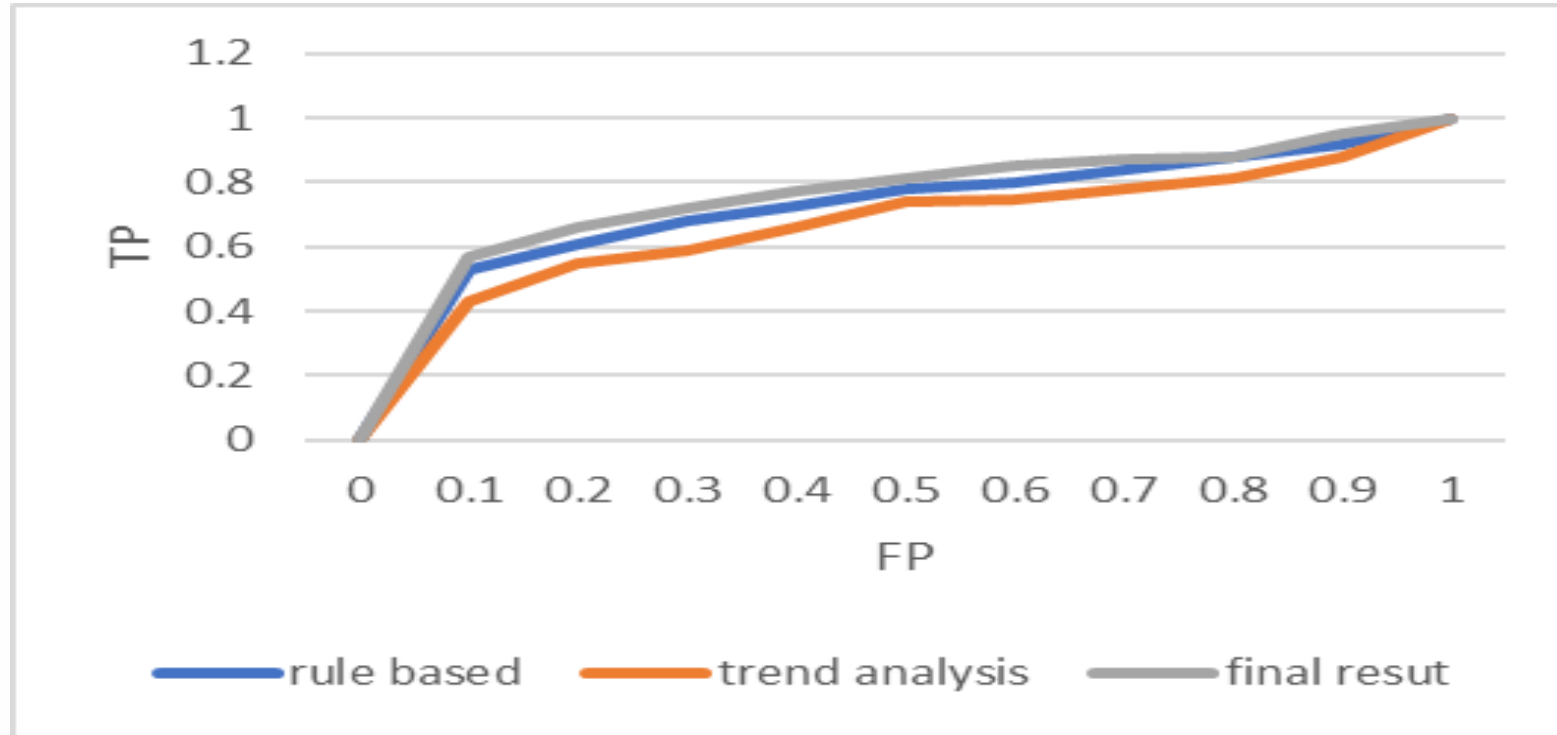


خروجی نهایی	خروجی بخش مبتنی بر قاعده	خروجی بخش تحلیل روند
کم ریسک	کم ریسک	کم ریسک
ریسک متوسط	کم ریسک	پر ریسک
ریسک متوسط	پر ریسک	کم ریسک
پر ریسک	پر ریسک	پر ریسک

تعداد تراکنش ها



نتایج



- [1] T. B. Joewono, B. A. Effendi, H. S. A. Gultom, and R. P. Rajagukguk, "Influence of Personal Banking Behaviour on the Usage of the Electronic Card for Toll Road Payment," *Transp. Res. Procedia*, vol. 25, pp. 4454-4471, Jan. 2017.
- [2] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Syst. Appl.*, vol. 51, pp. 134-142, Jun. 2016.
- [3] H. Hoehle, E. Scornavacca, and S. Huff, "Three decades of research on consumer adoption and utilization of electronic banking channels: A literature analysis," *Decis. Support Syst.*, vol. 54, no. 1, pp. 122-132, Dec. 2012.
- [4] V. Van Vlasselaer *et al.*, "APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions," *Decis. Support Syst.*, vol. 75, pp. 38-48, Jul. 2015.
- [5] A. Dal Pozzolo, O. Caelen, Y.-A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst. Appl.*, vol. 41, no. 10, pp. 4915-4928, Aug. 2014.
- [6] M. F. A. Gadi, X. Wang, and A. P. do Lago, "Credit Card Fraud Detection with Artificial Immune System," in *Artificial Immune Systems*, 2008, pp. 119-131.
- [7] R. J. Bolton, D. J. Hand, and D. J. H., "Unsupervised Profiling Methods for Fraud Detection," in *Proc. Credit Scoring and Credit Control VII*, 2001, pp. 5-7.
- [8] D. J. Hand, C. Whitrow, N. M. Adams, P. Juszczak, and D. Weston, "Performance criteria for plastic card fraud detection tools," *J. Oper. Res. Soc.*, vol. 59, no. 7, pp. 956-962, Jul. 2008.
- [9] N. Carneiro, G. Figueira, and M. Costa, "A data mining based system for credit-card fraud detection in e-tail," *Decis. Support Syst.*, vol. 95, pp. 91-101, Mar. 2017.
- [10] S. Wang, "A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research," in *2010 International Conference on Intelligent Computation Technology and Automation*, 2010, vol. 1, pp. 50-53.

- [11] N. A. L. Khac and M. T. Kechadi, "Application of Data Mining for Anti-money Laundering Detection: A Case Study," in *2010 IEEE International Conference on Data Mining Workshops*, 2010, pp. 577-584.
- [12] J. Wu, H. Xiong, and J. Chen, "COG: local decomposition for rare class analysis," *Data Min. Knowl. Discov.*, vol. 20, no. 2, pp. 191-220, Mar. 2010.
- [13] R. Liu, X. I Qian, S. Mao, and S. z Zhu, "Research on anti-money laundering based on core decision tree algorithm," in *2011 Chinese Control and Decision Conference (CCDC)*, 2011, pp. 4322-4325.
- [14] W. H. Chang and J. S. Chang, "Using clustering techniques to analyze fraudulent behavior changes in online auctions," in *2010 International Conference on Networking and Information Technology*, 2010, pp. 34-38.
- [15] L. Torgo and C. Soares, "Resource-bounded Outlier Detection Using Clustering Methods," in *Proceedings of the 2010 Conference on Data Mining for Business Applications*, Amsterdam, The Netherlands, The Netherlands, 2010, pp. 84-98.
- [16] L. Torgo and E. Lopes, "Utility-based Fraud Detection," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, Barcelona, Catalonia, Spain, 2011, pp. 1517-1522.
- [17] F. H. Glancy and S. B. Yadav, "A computational model for financial reporting fraud detection," *Decis. Support Syst.*, vol. 50, no. 3, pp. 595-601, Feb. 2011.
- [18] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning," *Inf. Fusion*, vol. 10, no. 4, pp. 354-363, Oct. 2009.
- [19] M. Singh and S. Raheja, "Credit Card Fraud Detection by Improving K-Means," vol. 2, no. 5, 2014.
- [20] "SAS Fraud management Report," 2015.
- [21] J. Kim, A. Ong, and R. E. Overill, "Design of an artificial immune system as a novel anomaly detector for combating financial fraud in the retail sector," in *The 2003 Congress on Evolutionary Computation, 2003. CEC '03*, 2003, vol. 1, p. 405-412 Vol.1.
- [22] U. Murad and G. Pinkas, "Unsupervised Profiling for Identifying Superimposed Fraud," in *Principles of Data Mining and Knowledge Discovery*, 1999, pp. 251-261.



با تشکر از توجه شما

