

کشف ناهنجاری رفتار تراکنشی وابسته به کسب و کار با استفاده از تحلیل سری‌های زمانی

نویسندگان

فاطمه کاوه یزدی*
علیرضا بادامچی
شبکه الکترونیکی پرداخت کارت (شاپرک)

9th Annual Conference on
Electronic Banking &
Payment Systems



شرکت ملی انفورماتیک



بانک مرکزی جمهوری اسلامی ایران



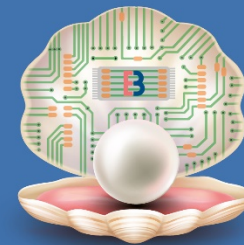
پژوهشکده پولی و بانکی
بنک مرکزی جمهوری اسلامی ایران



نهمین همایش سالانه
بانکداری الکترونیک
و نظام‌های پرداخت

(ارزش آفرینی دیجیتالی)

تهران، مرکز همایش‌های بین‌المللی برج میلاد - ۱ و ۲ اسفند ۱۴۰۱



مروری بر مطالب

- معرفی مسئله کشف ناهنجاری در نظام‌های پرداخت
 - معرفی مسئله کشف ناهنجاری در شبکه پرداخت
- معرفی دادگان مورد استفاده برای کشف ناهنجاری
- معرفی سامانه کشف ناهنجاری تراکنشی کسب‌وکار
 - سامانه کشف ناهنجاری بدون سرپرست
 - نتایج ارزیابی چهارچوب پیشنهادی
 - سامانه عملیاتی کشف ناهنجاری با سرپرست



رفتارشناسی بازیگران نظام‌های پرداخت

- بازیگران اصلی شبکه‌های پرداخت

- صادرکنندگان کارت

- دارندگان کارت

- فروشندگان کالا و ارائه‌دهندگان خدمات

- ارائه‌دهندگان خدمات پرداخت

- دستاوردهای بررسی رفتار بازیگران شبکه‌ی پرداخت

- رده‌بندی کیفی خدمات صادرکنندگان کارت و ارائه‌دهندگان خدمات پرداخت

- رده‌بندی تجاری فروشندگان کالا و ارائه‌دهندگان خدمات

- تعیین اعتبار دارندگان کارت و رتبه‌بندی کاربران

- پیش‌بینی ترک بازار توسط هر یک از بازیگران عرصه پرداخت

- کشف هر نوع الگوی سواستفاده و کلاهبرداری

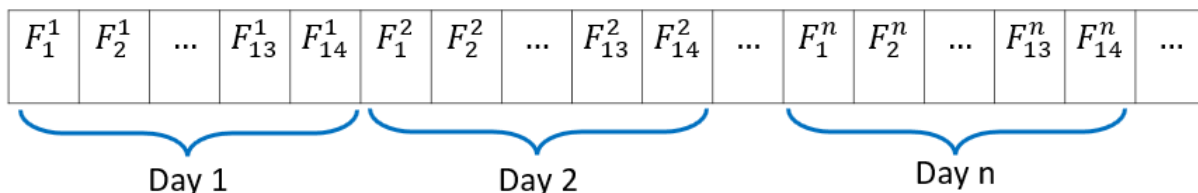
استخراج الگوهای تکرارشونده و کشف رفتار مشابه ~ استخراج موارد ناهنجار (غیرمشابه با سایرین)

کشف ناهنجاری تراکنشی کسب و کاری

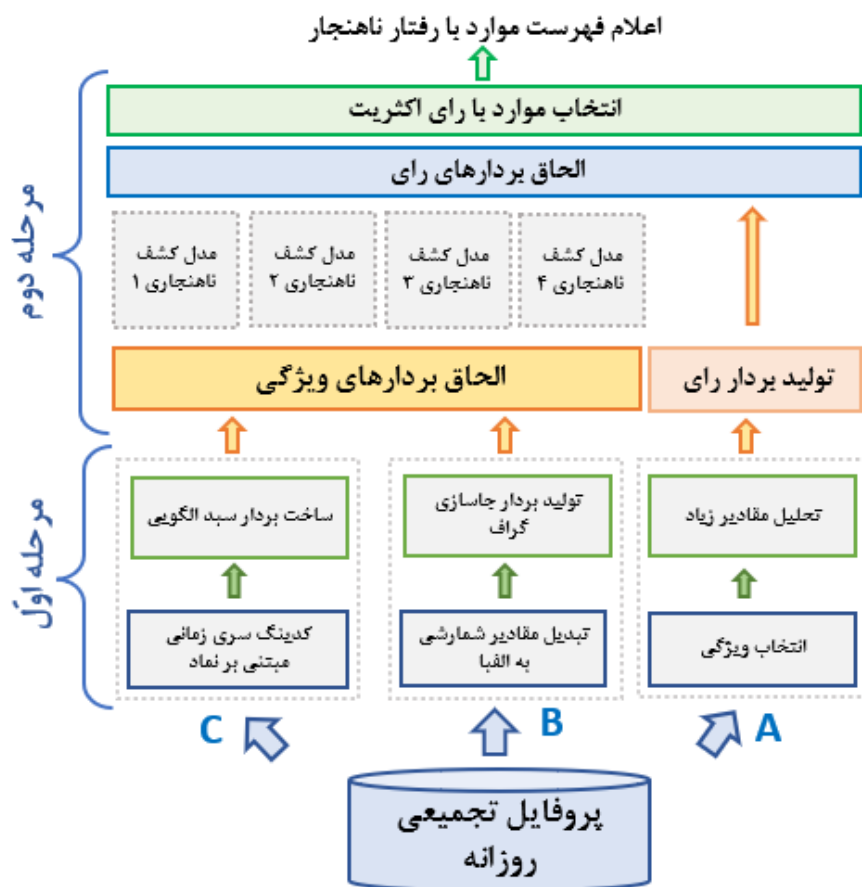
- پذیرنده: صاحب هر کسب و کار و بهره‌بردار از پایانه پرداخت.
 - پذیرندگان پایانه فیزیکی در شاپرک، بالغ بر نه میلیون پذیرنده
 - هر پذیرنده می‌تواند از یک یا چند پایانه بهره‌برداری کند.
 - هر پذیرنده یک کسب و کار را از بین گروه‌های کسب و کاری ممکن انتخاب می‌کند.
- گروه کسب و کاری: یک یا چند صنف با حوزه کاری مرتبط که در یک ساختار سلسله‌مراتبی موضوعی و در قالب ۲۸۱ گروه دسته‌بندی شده‌اند.
- هر گروه کسب و کاری معمولاً از نظر تراکنشی سباهت رفتاری دارند و رفتارهای آنها در یک یا چند خوشه یکسان قابل دسته‌بندی هستند.
 - بروز هر گونه ناهنجاری می‌تواند به معنای زنگ خطری برای تخلف و یا افت کیفیت باشد.

دادگان مورد استفاده

- دادگان مورد استفاده شامل ۱۴ ویژگی، از سه دسته به شرح زیر است:
- تأخر زمانی (Recency)
 - یک ویژگی عددی
- تعدادی (Frequency)
 - هشت ویژگی شمارشی
- پولی (Monetary)
 - شش ویژگی عددی



ساختار سامانه کشف ناهنجاری با سرپرست



• مرحله اول

– شامل سه ستون پرداز داده موازی

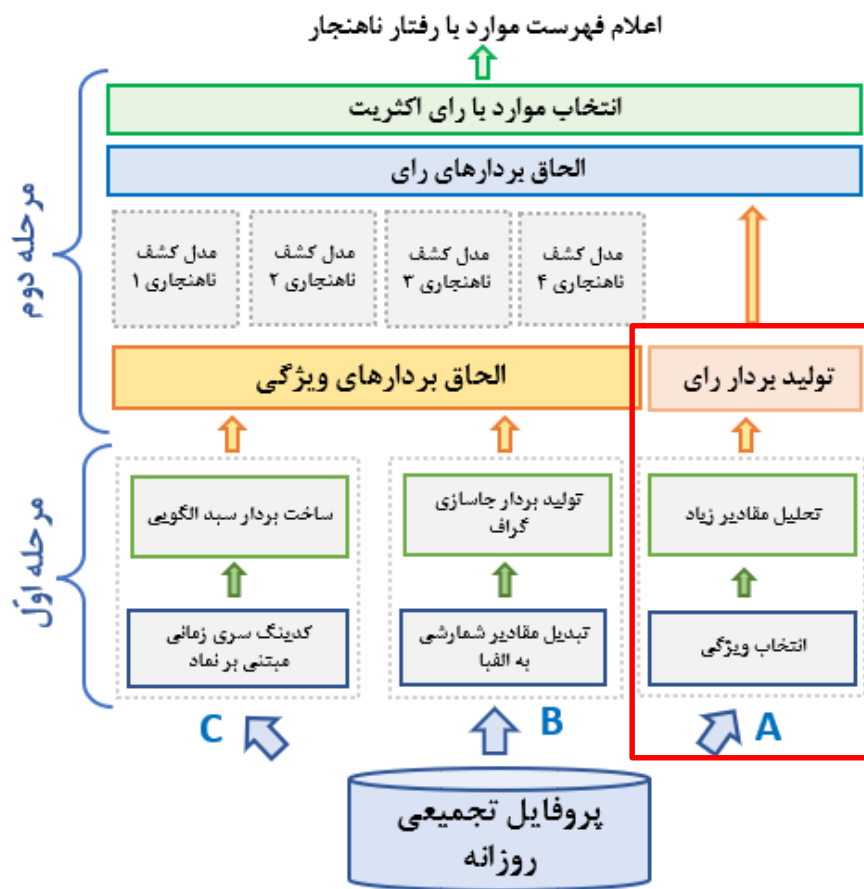
- ورودی: سری‌های زمانی ۱۴ گانه
- خروجی: بردارهای ویژگی

• مرحله دوم

– دو ستون کشف ناهنجاری

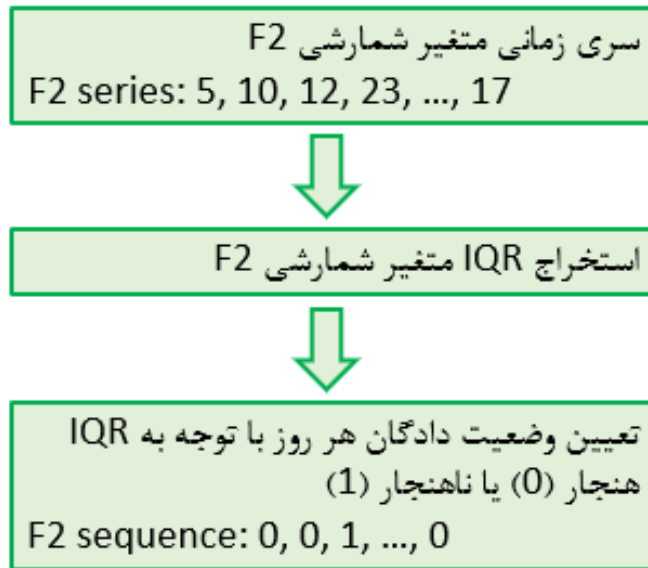
- ورودی: بردارهای ویژگی پردازش شده
- خروجی: بر چسب ناهنجار/ هنجار برای هر نمونه

خط لوله A



خط لوله A

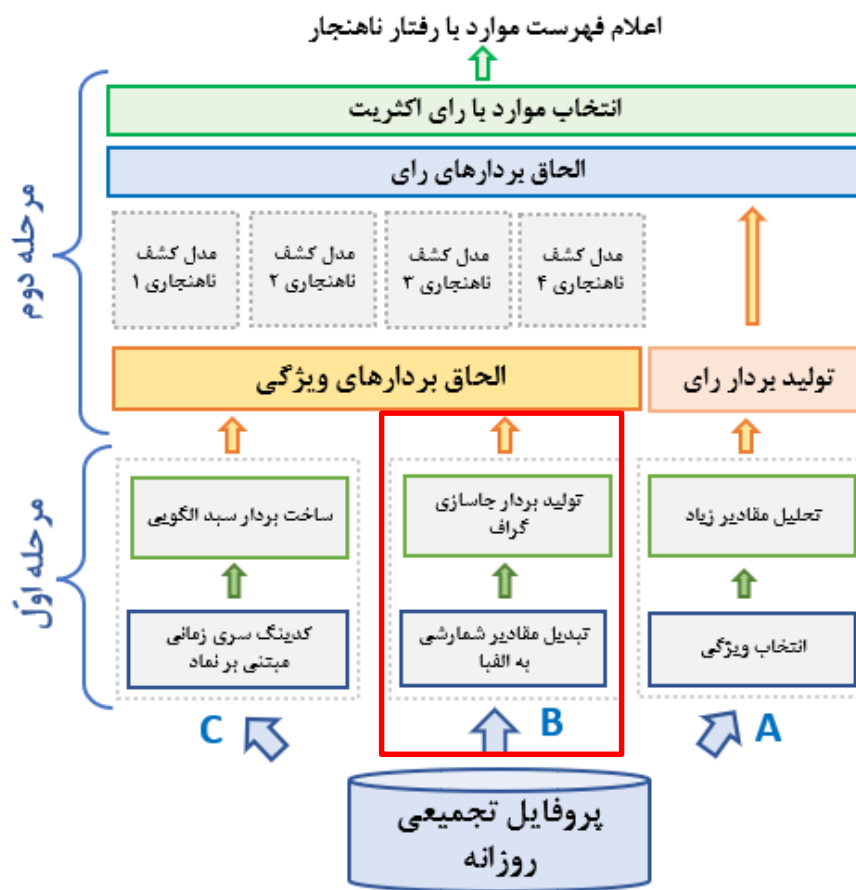
- مراحل خط لوله (A)



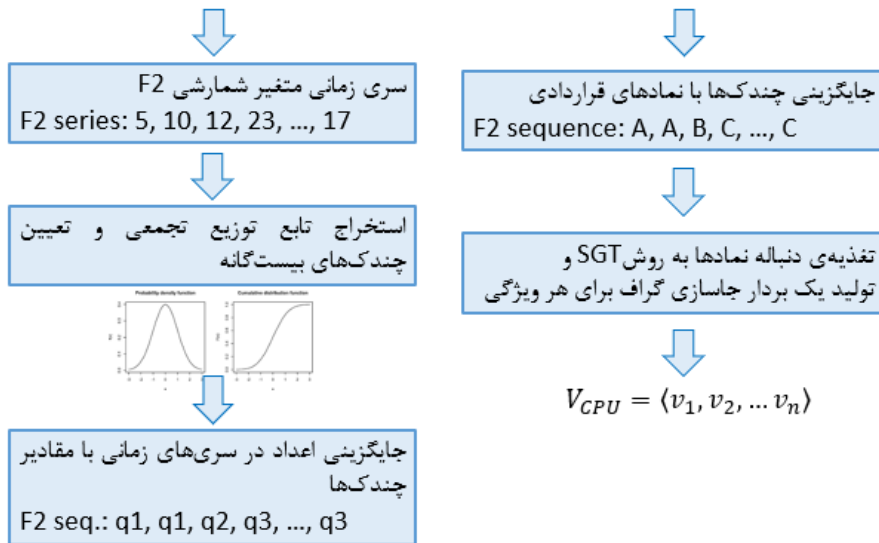
- اجرای روش تحلیل مولفه‌های اصلی (PCA)
- انتخاب ویژگی‌های متناظر با بالاترین مقادیر ویژه
- محاسبه بازه IQR برای مقادیر هر ویژگی انتخابی
 $InterQuartile\ Range\ (IQR) = Q_3 - Q_1$
- محاسبه حد بالا و پایین ناهنجاری (قانون تاکی)
 $Higher\ Range = Q_3 - 1.5 \times IQR$
 $Lower\ Range = Q_1 - 1.5 \times IQR$
- اعطای برچسب هنجار (zero) و ناهنجار (one) به هر نمونه در هر روز

مراحل اعمال شده بر روی سری زمانی مقادیر n روز متوالی از تعداد تراکنش خرید در خط لوله A.

خط لوله B



خط لوله B

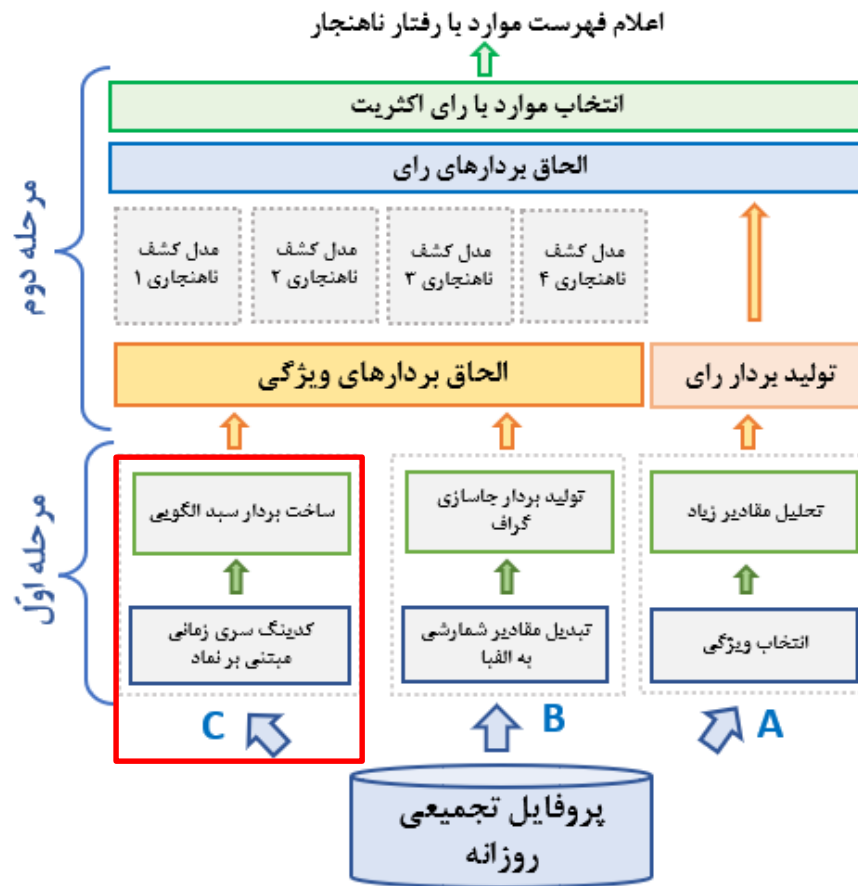


• مراحل خط لوله (B)

- استخراج چندکهای ۲۰ گانه
- تبدیل مقادیر شمارشی خام به چندک مربوطه
- جایگزینی چندکها با نمادهای قراردادی
- استفاده از روش جاسازی SGT [۱] برای تبدیل دنباله‌های نمادها به بردارهای ویژگی عددی

مراحل اعمال شده بر روی سری زمانی مقادیر n روز متوالی از تعداد تراکنش خرید (به عنوان نمونه) در خط لوله B.

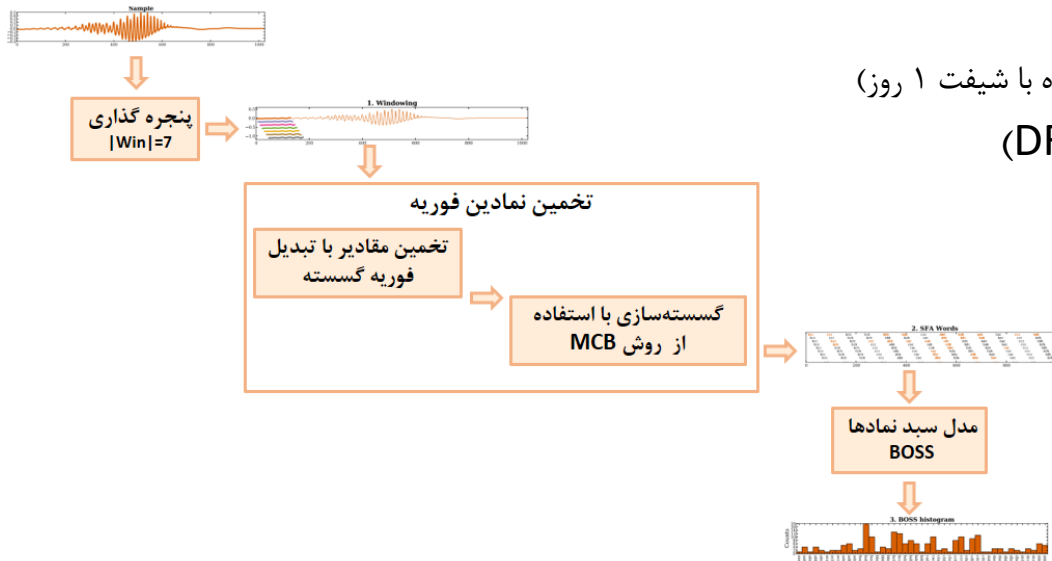
خط لوله C



خط لوله C

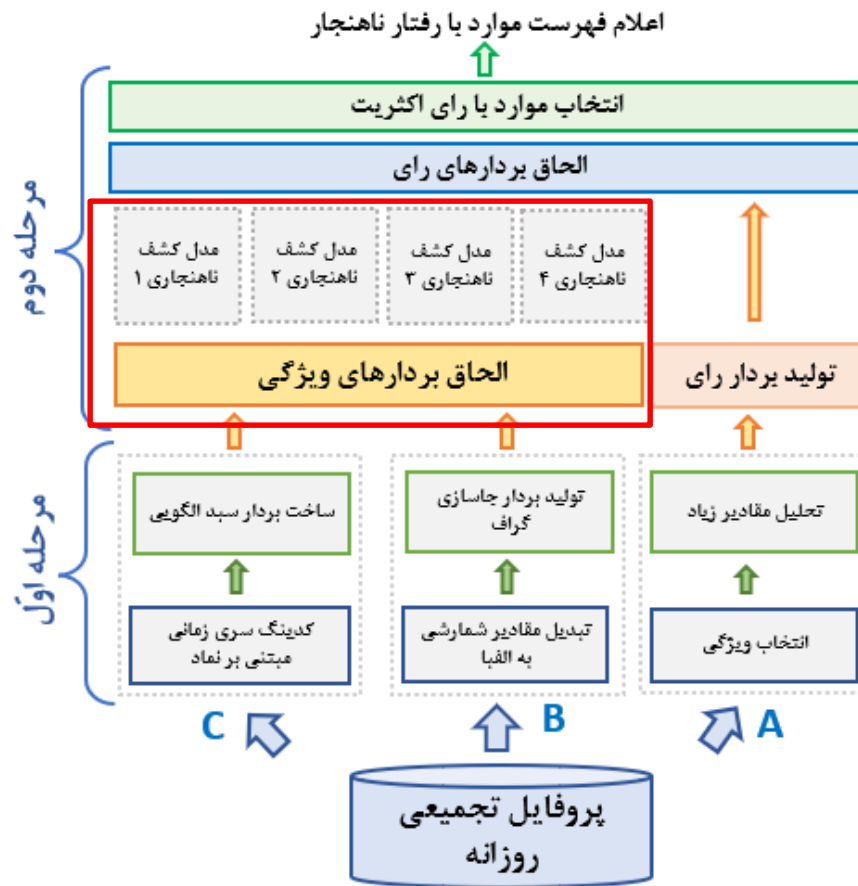
• مراحل خط لوله (C)

- پنجره گذاری با طول ۷ و شیفت ۱ (هر هفته یک پنجره با شیفت ۱ روز)
- تخمین مقادیر با تبدیل فوریه گسسته (DFT)
- گسسته سازی داده ها با روش MCB [۲]
- تولید سبد نماد (BOSS) و هیستوگرام مربوطه

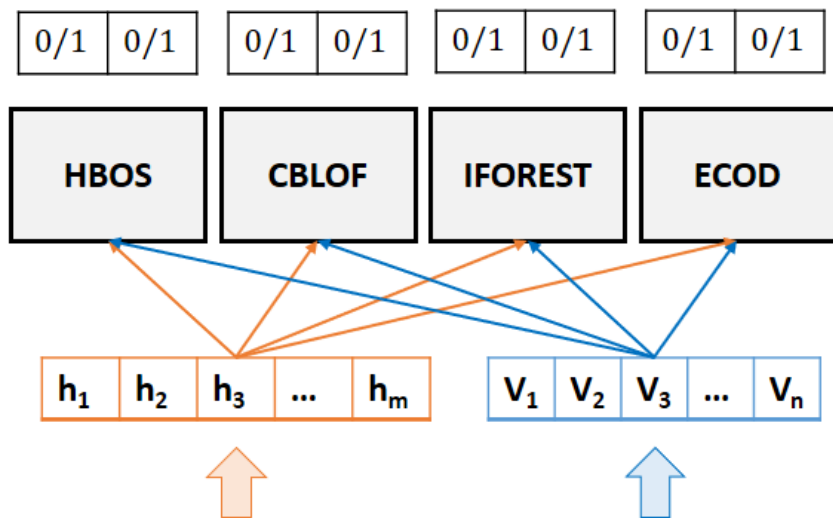


مراحل تبدیل سری های زمانی مقادیر عددی (غیرشمارشی) به بردارهای عددی.

الحاق بردارهای ویژگی و کشف ناهنجاری

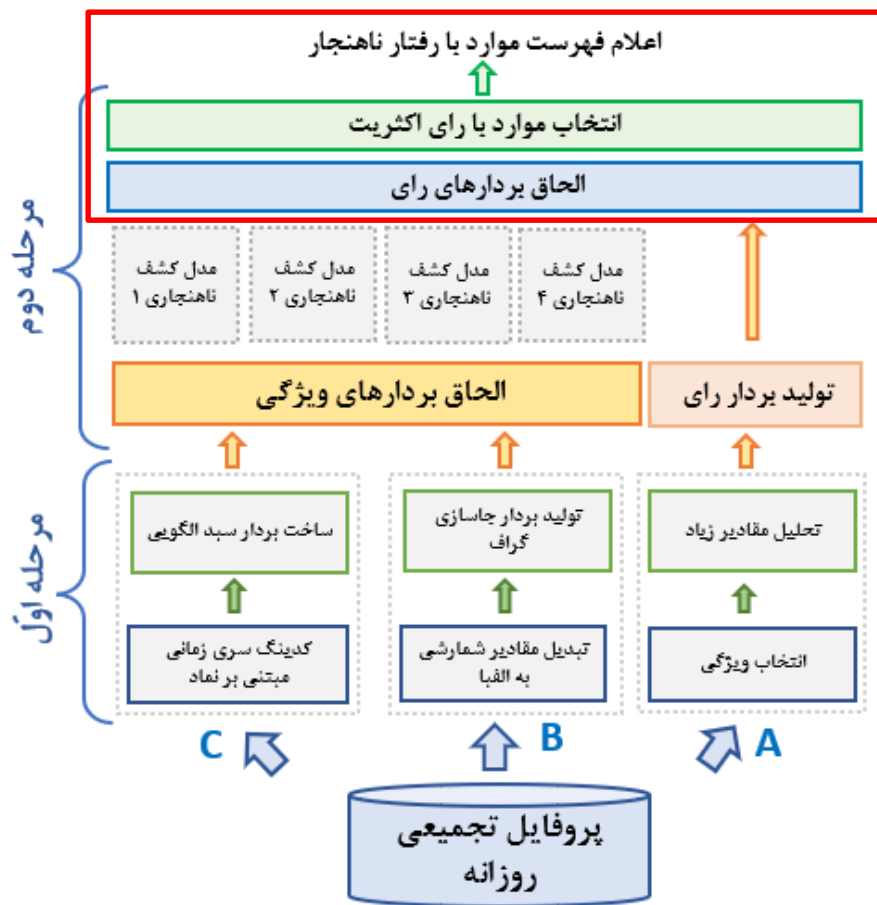


الحاق بردارهای ویژگی و کشف ناهنجاری



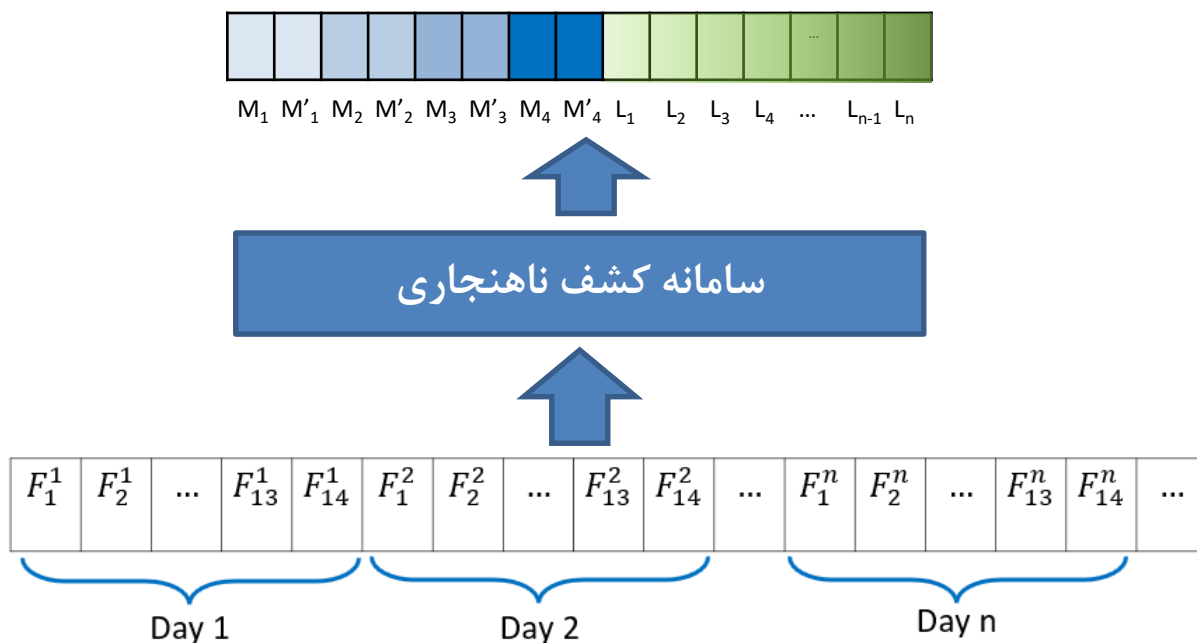
- مراحل ادغام خروجی دو کانال B و C
 - دریافت بردارهای ویژگی هر دو کانال
 - تغذیه روش‌های کشف ناهنجاری با بردارها به تفکیک کانال
 - تولید رای مبتنی بر رفتار هنجار یا ناهنجار
 - مجموعاً ۸ رای (۴ رای برای خروجی هر کانال)
 - تولید بردارهای رای شامل ۸ رای دودویی

الحاق بردارهای رای



الحاق بردارهای رای و تعیین رای اکثریت

$$Vote = (|F| + 8) - \left[\sum_{\substack{m \in \{HBOS, ISOFOREST, \\ ECOD, CBLOF\} \\ L \in \{B, C\}}} \mathbb{I}[V_{method, line} = 1] + \frac{Active\ Days}{n} \times \sum_{F \in Selected\ Feature} \mathbb{I}[V_A^F = 1] \right]$$



ارزیابی سامانه کشف ناهنجاری

- داده آزمون: ۲۵۰۰ نمونه از ۵ صنف نیازمند مجوز
 - تکنیک نمونه گیری: لایه‌بندی شده با ۱۰ لایه
 - تعداد نمونه از هر صنف: ۵۰۰ نمونه
 - نویز: ۵۰۰ نمونه از ۲۷۶ صنف باقیمانده
 - روش انتخاب موارد نویز: تصادفی
- معیار ارزیابی: معیار صحت وزن‌دار

$$Accuracy = \sum_{Class_i} \frac{\# \text{ Records in Class}_i}{\text{Total \# of Records}} \times \frac{\# \text{ Correct Predictions}}{\text{Total \# of Predictions}}$$

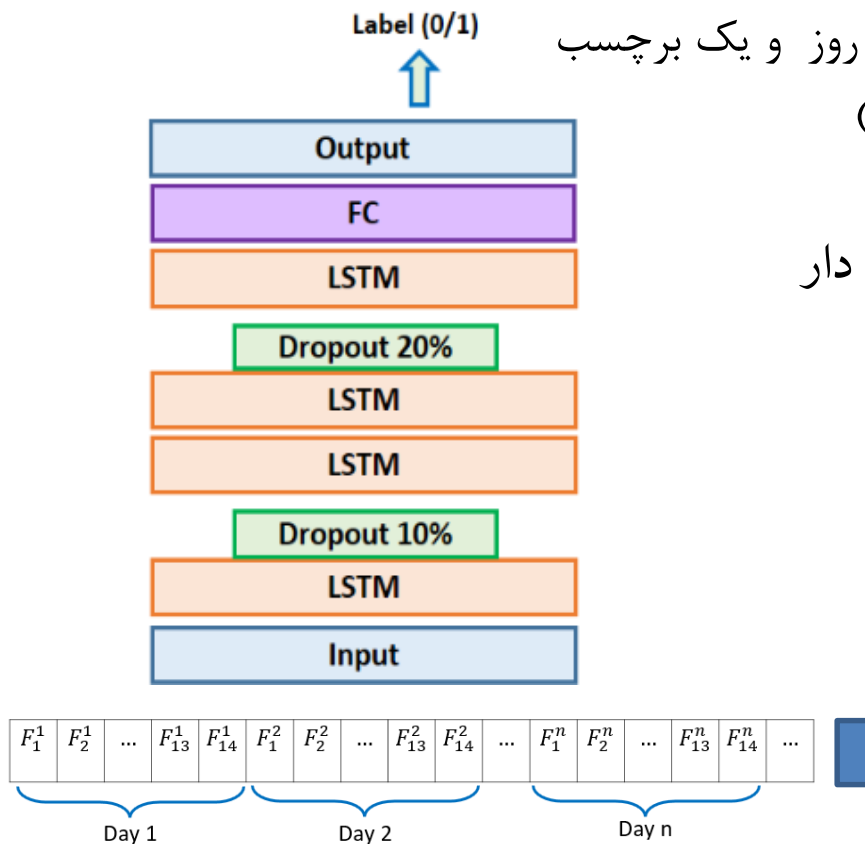
نتایج ارزیابی و جمع‌بندی

- نکات درخور توجه در نتایج ارزیابی‌ها
 - کارایی کم روش‌های KNN، CBLOF و OSVM
 - کارایی بالا برای روش‌های HBOS و Isolation Forest
 - هم‌خوانی رفتاری HBOS و ECOD

#	روش پیاده‌سازی شده	نوع صحت
۱	روش CBLOF (He et al., 2003)	۷۹/۸۳
۲	روش HBOS (Goldstein & Dengel, 2012)	۸۴/۰۳
۳	روش IFOREST (F. T. Liu et al., 2012)	۸۲/۲۷
۴	روش ECOD (Li et al., 2022)	۸۲/۲۳
۵	روش OSVM (Schölkopf et al., 2001)	۷۱/۵۲
۶	روش مبتنی بر KNN	۷۷/۴
۷	چهارچوب پیشنهادی	۸۶/۱۱

سامانه عملیاتی کشف ناهنجاری تراکنشی

- ساختار سامانه عملیاتی



- ورودی: بردارهای ویژگی‌های چهارده گانه در n روز و یک برچسب
- خروجی: برچسب مبنی بر هنجار/ناهنجار (0/1)
- تعداد پارامترهای قابل آموزش: 13222
- حجم داده مورد نیاز: 3 دوره داده برچسب دار
- زمان آموزش: 270 دقیقه

منابع

[1] Ranjan, C., Ebrahimi, S., & Paynabar, K. (2022). Sequence graph transform (SGT): a feature embedding function for sequence data mining. *Data Mining and Knowledge Discovery*, 36(2), 668–708. <https://doi.org/10.1007/s10618-021-00813-0>

[2] Schäfer, P. (2015). The BOSS is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, 29(6), 1505–1530. <https://doi.org/10.1007/s10618-014-0377-7>

[3]

