

## بهبود زمان اجرای روش قواعد انجمنی در داده‌کاوی به کمک الگوریتم جستجوی فاخته

پیام عبدالمحمدی<sup>۱</sup>، رهام فراهانی<sup>۲</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد مهندسی کامپیوتر نرم‌افزار، دانشگاه آزاد اسلامی واحد اراک، اراک، ایران

<sup>۲</sup> مدیر گروه کارشناسی ارشد مهندسی کامپیوتر نرم‌افزار، دانشگاه آزاد اسلامی واحد اراک، اراک، ایران

---

### چکیده

هدف ما بررسی روش مناسب داده‌کاوی جهت استخراج قوانین از مجموعه داده‌ی و بررسی مزایای استفاده از الگوریتم فاخته جهت استخراج قواعد انجمنی و بررسی میزان زمان اجرای الگوریتم فاخته نسبت به الگوریتم ژنتیک است که در این کار، یک الگوریتم جدید مبتنی بر جستجوی فاخته برای پنهان کردن قوانین ارتباط حساس<sup>۱</sup> با هزینه زمانی بهتر و درعین حال کنترل بهتر اثرات جانبی قوانین غیر حساس<sup>۲</sup> ارائه شد. ما با بهره‌گیری از الگوریتم فاخته و با استفاده از بازگشت در تابع هدف آن موفق به دستیابی به این امر شدیم. با هدف نشان دادن برتری رویکرد ما، الگوریتم ژنتیک به‌عنوان معیار ارزیابی انتخاب شد. تجزیه و تحلیل تجربی نشان داد که عملکرد الگوریتم فاخته با توجه به زمان اجرا بهتر از ژنتیک بود.

**واژه‌های کلیدی:** داده‌کاوی، قواعد انجمنی، الگوریتم فاخته CUCKOO، مقایسه زمان اجرای الگوریتم، الگوریتم ژنتیک.

---

<sup>۱</sup> Sensitive

<sup>۲</sup> Non-sensitive

## ۱- مقدمه

داده‌کاوی به‌عنوان یک روش تأثیرگذار برای افزایش تصمیم‌گیری در نظر گرفته می‌شود و چندین سازمان و مؤسسه از آن برای استخراج اطلاعات مفید و کشف الگوهای پنهان در مجموعه داده‌ها استفاده می‌کنند. این امر داده‌کاوی را به یک روند عملکردی امیدوارکننده تبدیل کرده است. روش کاوی قانون انجمن (ARM) یک روش داده‌کاوی مهم است که داده‌های متداول مشترک را از پایگاه‌های داده تراکنشی و رابطه‌ای استخراج می‌کند و همبستگی‌های پنهان بین آن‌ها را آشکار می‌کند. در مجموعه‌های از تراکنش‌ها، ARM بر یافتن ارتباط بین اقلام مکرر تمرکز می‌کند و محققان را قادر می‌سازد تا فراوانی یک مجموعه اقلام را بسته به فراوانی آیتم دیگری در یک تراکنش پیش‌بینی کنند. روش ARM به‌طور گسترده در چندین برنامه مورد استفاده قرار گرفته است، از جمله حفظ حریم خصوصی [۱۰]، تجزیه و تحلیل داده‌های معاملات سبد بازار [۱۱]، توصیه [12]، مراقبت‌های بهداشتی [13]، پیش‌بینی، الگویابی در وب. مرور [14]، رتبه‌بندی اسناد متنی [15] و شناسایی خطر و مجموعه داده‌های تراکنش در سال ۱۹۹۳ آغاز شد [16] و غیره. در این مقاله می‌خواهیم با استفاده از استخراج قوانین ارتباط از مجموعه داده‌های تراکنش زمان اجرای قواعد انجمنی در داده‌کاوی به کمک الگوریتم جستجوی cuckoo یا الگوریتم فاخته نسبت به الگوریتم ژنتیک را بررسی و مقایسه کنیم.

داده‌کاوی، یک تکنولوژی نوظهور است که از ابزارها و روش‌های مدل‌سازی و تجزیه و تحلیل آماری، الگوریتم‌های ریاضی و متدهای یادگیری ماشین برای کشف الگوهای معتبر و ناشناخته در مجموعه داده‌های حجیم استفاده می‌کند. هرچند این تکنولوژی دوران نوباوگی خود را طی می‌کند، اما شرکت‌ها و سازمان‌های بسیاری از جمله خرده‌فروشی‌ها، بانک‌ها، مراکز درمانی، کارخانجات تولیدی، ارتباطات راه دور و مؤسسات دولتی از ابزارها و روش‌های داده‌کاوی برای تحلیل داده‌هایشان و کشف اطلاعات و دانش مفید از آن‌ها استفاده می‌کنند [۱۷]. داده‌کاوی اطلاعاتی را از پایگاه داده‌ها استخراج می‌کند که از طریق کوئری‌ها و گزارش‌گیری‌ها قابل دستیابی نیستند.

رشد انفجاری داده‌های ذخیره شده در پایگاه داده‌ها، نیاز به تکنولوژی‌های جدید که بتوانند حجم عظیم داده‌ها را هوشمندانه به دانش مفید تبدیل کنند را پدید آورده است. داده‌کاوی به معنای یافتن نیمه خودکار الگوهای پنهان در مجموعه داده‌های<sup>۳</sup> موجود می‌باشد [۱۸]. این تکنولوژی با دیگر روش‌های تحلیل داده که سیستم، مقادیر اولیه را می‌گیرد و خود، الگوهایی را تولید می‌کند، متفاوت است. داده‌کاوی توسط ابزارهای الگوریتمیک، الگوها، تغییرات، آنومالی‌ها، قوانین و ساختارهای مهم آماری و رویدادها را از مجموعه داده‌های عظیم استخراج می‌کند [۱۹]. می‌توان گفت که داده‌کاوی در جهت کشف اطلاعات پنهان و روابط موجود در بین داده‌های فعلی و پیش‌بینی موارد نامعلوم و یا مشاهده نشده عمل می‌کند.

برای انجام عملیات کاوش لازم است قبلاً روی داده‌های موجود پیش پردازش‌هایی انجام گیرد. عمل پیش پردازش اطلاعات خود از دو بخش کاهش اطلاعات و خلاصه‌سازی و کلی‌سازی داده‌ها تشکیل شده است. کاهش اطلاعات عبارت است از تولید یک مجموعه کوچک‌تر، از داده‌های اولیه که تحت عملیات داده‌کاوی نتایج تقریباً یکسانی با نتایج داده‌کاوی روی اطلاعات اولیه به دست دهد [۱۸]. پس از انجام عمل کاهش اطلاعات و حذف خصایص غیر مرتبط نوبت به خلاصه‌سازی و کلی‌سازی داده‌ها می‌رسد. داده‌های موجود در بانک‌های اطلاعاتی معمولاً حاوی اطلاعات در سطوح پایینی هستند، بنابراین خلاصه‌سازی مجموعه بزرگی از داده‌ها و ارائه آن به صورت یک مفهوم کلی اهمیت بسیار زیادی دارد. کلی‌سازی اطلاعات، فرآیندی است که تعداد زیادی از رکوردهای یک بانک اطلاعاتی را به صورت مفهومی در سطح بالاتر ارائه می‌نماید.

<sup>3</sup>Dataset

## مقالات و کارهای قبلی:

## جدول ۱. خلاصه ای از مطالعات صورت گرفته توسط محققین داخلی

شماره	مطالعه	یافته ها
[۱]	شمشیری و همکاران (۱۳۹۰)	پژوهشی تحت عنوان، شناسایی و کشف الگوهای خرید مشتریان لوازم خانگی از طریق داده کاوی انجام دادند
[۲]	خدیور و حامدی (۱۳۹۴)	پژوهشی تحت عنوان، ارائه یک الگو ترکیبی داده کاوی با استفاده از قواعد انجمنی و خوشه بندی برای تعیین استراتژی تخفیف دهی، مطالعه موردی شرکت پخش پگاه انجام دادند
[۳]	غفوری پور و میرزامؤمن (۱۳۹۷)	پژوهشی تحت عنوان استفاده از ارتباطات بین برچسبها در ایجاد زنجیره رده بندها برای بهبود رده بندی چند برچسبی انجام دادند
[۴]	انصاری و همکاران (۱۳۹۷)	پژوهشی تحت عنوان، کشف الگوی دسترسی کاربران و تحلیل تراکنشهای امانت منابع اطلاعاتی با استفاده از تکنیک قوانین انجمنی در داده کاوی (مطالعه موردی: کتابخانهها و مراکز اطلاع رسانی دانشگاه علوم پزشکی همدان) انجام دادند
[۵]	میرهاشمی و همکاران (۱۳۹۷)	پژوهشی تحت عنوان، استخراج قواعد انجمنی جهت تحلیل رفتار آبخوان . انجام دادند Apriori شهرستان قزوین توسط الگوریتم
[۶]	غریب و همکاران (۱۳۹۸)	پژوهشی تحت عنوان، ارائه الگوی ترکیبی داده کاوی با استفاده از قواعد انجمنی و خوشه بندی جهت شناسایی الگوهای غالب رفتار مشتریان مطالعه موردی: بانک انصار انجام دادند
[۷]	المعلگی و ارکوک (۲۰۱۴)	برای قوانین انجمنی Apriori پژوهشی تحت عنوان، الگوریتم بهبود یافته انجام دادند
[۸]	یاسیر و همکاران (۲۰۱۹)	با استفاده از Schistosomiasis پژوهشی تحت عنوان، تشخیص عوامل استخراج قوانین انجمنی انجام دادند

در این تحقیق دو هدف دنبال می شود که عبارتند از:

## هدف اصلی:

مقایسه زمان اجرای قواعد انجمنی در داده کاوی به کمک الگوریتم جستجوی cuckoo یا الگوریتم فاخته نسبت به الگوریتم ژنتیک

## هدف فرعی:

۱. بررسی روش مناسب داده کاوی جهت استخراج قوانین از مجموعه داده ی

۲. بررسی مزایای استفاده از الگوریتم فاخته جهت استخراج قواعد انجمنی

۳. بررسی میزان زمان اجرای الگوریتم فاخته

۴. بررسی میزان زمان اجرای الگوریتم ژنتیک

### سؤالات تحقیق

۱. به کارگیری الگوریتم فاخته به همراه روش کاهش بعد در داده‌های ورودی چه تأثیری در نتایج به دست آمده وارد می‌نماید؟
۲. زمان اجرای الگوریتم فاخته نسبت به سایر روش‌ها چگونه است؟
۳. کدام روش‌های داده‌کاوی جهت استخراج قوانین و دانش در مجموعه داده‌ی پژوهش مناسب‌تر هستند؟

### روش تحقیق

یک مجموعه داده تراکنش شامل یک سری  $n$  تراکنش  $t_1, t_2, t_n, \dots$  است. هر تراکنش مجموعه‌ای از آیتم‌های  $I = \{i_1, i_2, \dots, i_m\}$  است، در حالی که تراکنش یک جفت است که از یک مجموعه اقلام و یک شناسه تراکنش منحصر به فرد  $(TID, X)$  تشکیل شده است. در صورتی که  $Supp(X) \geq T_s$  که در آن  $T_s$  آستانه پشتیبانی تعریف شده توسط داده‌کاوی تعریف شده است، مورد  $X$  مکرر در نظر گرفته می‌شود.  $Supp(X)$  پشتیبان مجموعه آیتم‌ها است که تعداد دفعاتی را که مجموعه آیتم‌ها  $(X)$  در مجموعه داده رخ می‌دهد را نشان می‌دهد. یک قانون تداعی به صورت  $X \Rightarrow Y$  ارائه می‌شود که در آن  $X$  و  $Y$  دو مجموعه مجزا از آیتم‌ها هستند. قانون  $X \Rightarrow Y$  نشان می‌دهد که وجود  $X$  نشان دهنده حضور  $Y$  دیگری در یک تراکنش یکسان با اطمینان خاص است. برای نشان دادن این وضعیت در یک سناریوی واقعی، از مجموعه داده‌های تراکنش یک سوپرمارکت به‌عنوان مثال استفاده می‌کنیم، جایکه لیست خرید مشتری یک تراکنش است. اگر مشتری «نان» و «کره» می‌خرد، مطمئناً «شیر» می‌خرد - قانون ارتباط احتمالی {نان، کره}  $\Rightarrow$  شیر است. به‌طور دقیق‌تر،  $Y \Rightarrow X$  تنها در صورتی به‌عنوان یک قانون ارتباط در نظر گرفته می‌شود که  $conf(X \cup Y) \geq T_c$  و  $Supp(X \cup Y) \geq T_s$  باشد. ما  $conf(X \cup Y)$  را به‌عنوان اطمینان  $X \Rightarrow Y$  تعریف می‌کنیم که نسبت حمایت قاعده نسبت به سمت چپ  $conf(X \cup Y) = Supp(X \cup Y) / Supp(X)$  است.  $T_c$  آستانه اطمینان تعیین شده توسط داده‌کاوی را نشان می‌دهد. مقادیر آستانه اطمینان و پشتیبانی  $T_s$  و  $T_c$  معمولاً با دسته تراکنش، استفاده از نتایج استخراج و اندازه مجموعه مرتبط هستند.

اگرچه ARM موثرترین روش داده برای استخراج الگوهای پنهان و ارتباط بین موارد است، استخراج این اطلاعات بدون نقض حریم خصوصی یک کار چالش برانگیز است. اطلاعات خصوصی که معمولاً در مجموعه داده پنهان است، میتواند با انجام یک تحلیل متقابل از مجموعه داده‌های مرتبط افشا شود. یکی از عوارض جانبی اصلی این فرآیند، افشای اطلاعات حساس است؛ بنابراین حریم خصوصی به یک مسئله حیاتی که جامعه تحقیقاتی داده‌کاوی با آن مواجه است تبدیل شده است. امروزه کاربرد قواعد انجمنی، بسیار گسترده شده است و الگوریتم‌های زیادی نیز جهت استخراج قوانین وابستگی ارائه شده‌اند. در این پژوهش از الگوریتم جستجوی cuckoo یا الگوریتم فاخته استفاده می‌شود. الگوریتم فاخته توسط یانگ و دب<sup>۴</sup>، توسعه یافته و یک الگوریتم فرا ابتکاری جدیدی است که از مزاحمت جوجه‌های در محظور از برخی گونه‌های کوکو که بر روی تخم‌های خود در آشیانه‌ی سایر پرندگان خوابیده‌اند، برای حل مسائل بهینه‌سازی الهام گرفته است. قابلیت‌های الگوریتم فاخته هم فضای

<sup>4</sup> Yang & Deb

گسسته و هم در فضای پیوسته است و الگوریتم بهینه سازی فاخته دارای توانایی همگرایی بسیار سریع دارد [۲۰]؛ از طرف دیگر اهمیت آن در حل مسائل بهینه سازی این انگیزه را ایجاد کرده است که در این پژوهش به ارائه ایده هایی جهت کاوش در قواعد انجمنی با استفاده از الگوریتم فاخته پردازیم. هدف از این پژوهش ارائه روشی بهینه با استفاده از الگوریتم فاخته بهبود فاکتورهای قواعد انجمنی با هدف نشان دادن برتری رویکرد ما، الگوریتم ژنتیک به عنوان معیار انتخاب شد. بیست و چهار قانون ارتباط در جدول ۱ نمایش داده شده است. این قوانین نتیجه یک الگوریتم داده کاوی پیش بینی است. دیگر الگوریتم های محبوب ARM عبارتند از FP-Growth [۹] و Eclat.

## جدول ۲. قوانین انجمنی

اطمینان	قواعد انجمنی
80%	$\{l\} \Rightarrow \{m\}$
80%	$\{l\} \Rightarrow \{n\}$
80%	$\{n\} \Rightarrow \{l\}$
80%	$\{n\} \Rightarrow \{m\}$
75%	$\{p\} \Rightarrow \{m\}$
100%	$\{r\} \Rightarrow \{m\}$
100%	$\{t\} \Rightarrow \{n\}$
75%	$\{o\} \Rightarrow \{p\}$
75%	$\{p\} \Rightarrow \{o\}$
75%	$\{o\} \Rightarrow \{q\}$
75%	$\{q\} \Rightarrow \{o\}$
75%	$\{p\} \Rightarrow \{q\}$
75%	$\{q\} \Rightarrow \{p\}$
80%	$\{l\} \Rightarrow \{m, n\}$
80%	$\{n\} \Rightarrow \{l, m\}$
100%	$\{l, m\} \Rightarrow \{n\}$
100%	$\{l, n\} \Rightarrow \{m\}$
100%	$\{m, n\} \Rightarrow \{l\}$
75%	$\{o\} \Rightarrow \{p, q\}$
75%	$\{p\} \Rightarrow \{o, q\}$
75%	$\{q\} \Rightarrow \{o, p\}$
100%	$\{o, p\} \Rightarrow \{q\}$
100%	$\{o, q\} \Rightarrow \{p\}$
100%	$\{p, q\} \Rightarrow \{o\}$

اگر قانون  $\{l\} \Rightarrow \{m, n\}$  به عنوان یک قاعده حساس انتخاب شود، آنگاه مورد قربانی (n) خواهد بود زیرا در مقایسه با (m) پشتیبانی کمتری دارد که در آن  $\alpha(n) = 5$  و  $\alpha(m) = 6$  است.

## توصیف دیتاست

در این تحقیق برای انجام آزمایشها از سه مجموعه داده استفاده شده است. جدول ۲ شرح مجموعه دادهها را نشان می دهد: شطرنج، قارچ و موبایل. دو مورد از مجموعه دادهها از مخزن یادگیری ماشین ایزوین هستند. این مجموعه دادهها در اصل در مرجع تولید و شرح داده شدند. در مجموعه داده شطرنج، شش متغیر ویژگی و یک متغیر کلاس وجود دارد. مجموعه داده دوم مجموعه داده Mushroom است که شامل 8124 رکورد است که هر کدام به یک قارچ منفرد اشاره دارد. ستون برجسب طبقه بندی قارچ را براساس دو دسته نشان می دهد:

"خوراکی" و "سمی". 22 ستون دیگر ویژگی های مختلف قارچ مانند رنگ، شکل، زیستگاه، شکل شنل، اندازه آبشش و رنگ آبشش هستند. مجموعه داده سوم مجموعه داده موبایل است که این مجموعه داده عظیمی است، بنابراین 5000 تراکنش تصادفی انتخاب شده را برای کار روی آن انتخاب می کنیم.

این مجموعه داده دارای 16 ستون است که هر کدام یک ویژگی شامل اطلاعات مکان، زمان، آدرس IP، عملکرد شبکه و غیره است.

## جدول ۳. توصیف دیتاست

Dataset	No. of Items	No. of Transactions	Transaction Length
Chess	75	3196	37
Mushroom	119	8124	23
Mobile	5250	5000	16

## جدول ۴

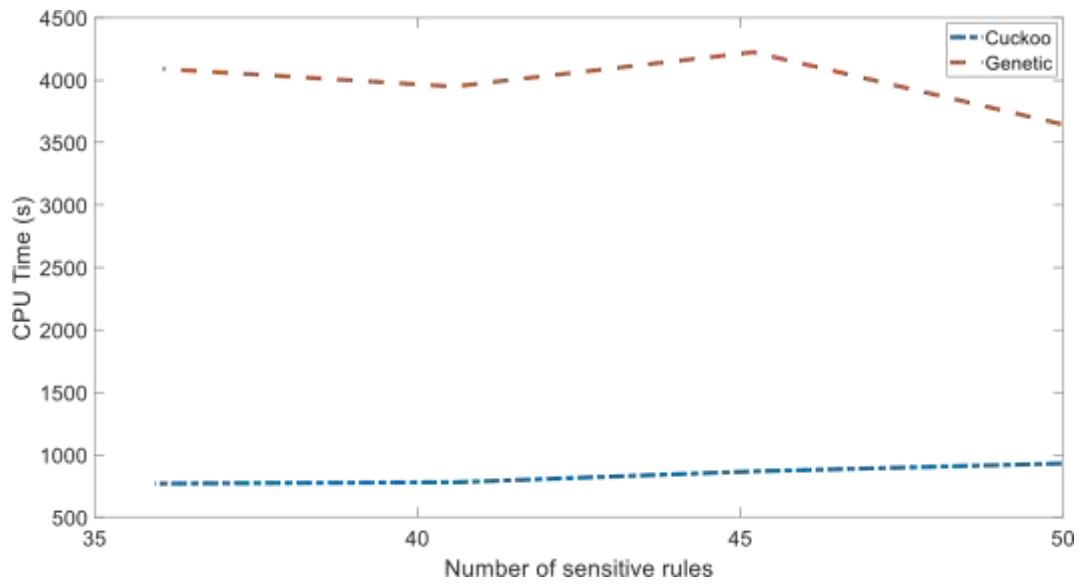
دیتاست	حداقل اطمینان	تعداد قوانین انجمنی	تعداد قوانین حساس
	94%	7177	72
Chess	95%	6145	62
	96%	4435	45
	97%	4081	41
	45%	4949	50
Mushroom	50%	4402	45
	55%	3937	40

	60%	3535	36
	40%	2092	20
	50%	1984	19
<b>Mobile</b>	60%	1801	18
	70%	1580	15
	80%	1333	13
	90%	1246	12

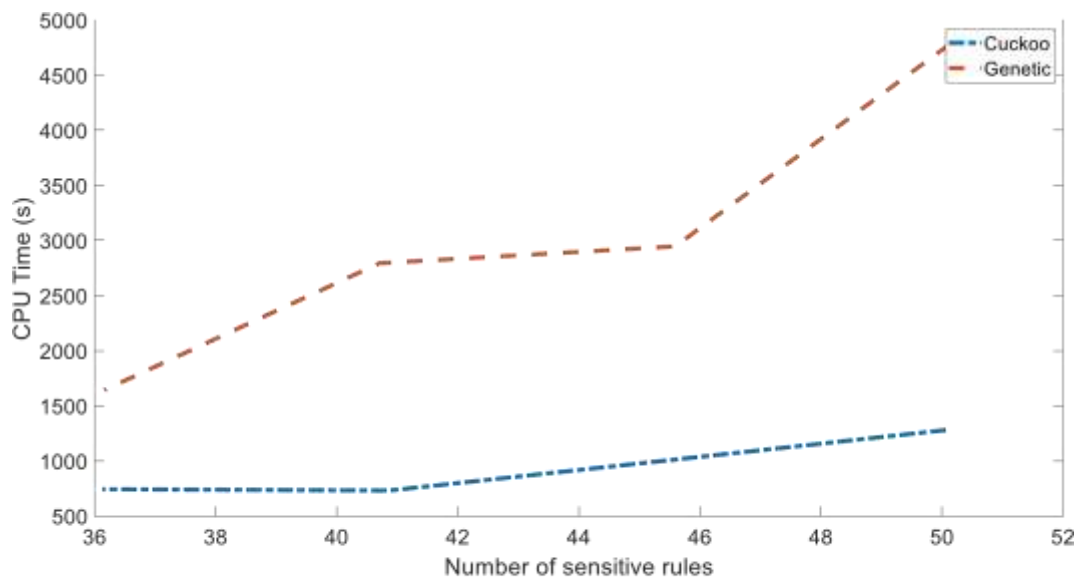
#### مدل ارایه شده: جدول ۴. تنظیمات تجربی مجموعه داده‌ها

ابتدا، الگوریتم داده‌کاوی برای این مجموعه داده‌ها با استفاده از مقادیر مختلف آستانه‌های پشتیبانی و اطمینان اعمال شد. سپس، 1 درصد از قوانین ارتباط حاصل به‌طور تصادفی به‌عنوان قوانین حساس انتخاب شدند. در نهایت از دو الگوریتم برای پنهان کردن داده‌های حساس استفاده شد و نتایج زیر تولید شد. ما چندین آزمایش برای هر مجموعه داده انجام دادیم. در هر آزمایش، تعداد قوانین حساس به‌طور مداوم 1 درصد بود و هر بار به‌طور تصادفی انتخاب شدند. جدول ۳ آستانه‌های اطمینان مختلف را در مرحله استخراج قوانین انجمن نشان می‌دهد.

نمودارهای زیر مزیت کارایی رویکرد الگوریتم جستجوی فاخته‌ی پیشنهادی در این تحقیق را در مقایسه با الگوریتم ژنتیک نشان می‌دهد که زمان اجرای دو الگوریتم در برابر آستانه اطمینان برای مجموع‌های داده ترسیم شده است. زمان اجرا مدت زمانی است که الگوریتم برای اجرای یک برنامه صرف می‌کند. این آستانه‌های اطمینان افزایشی منجر به کاهش تعداد قوانین حساس می‌شود و واضح است که زمان اجرای کمتری برای الگوریتم پیشنهادی در مقایسه با الگوریتم ژنتیک برای همه مجموعه داده‌ها در همه شرایط مورد نیاز است.

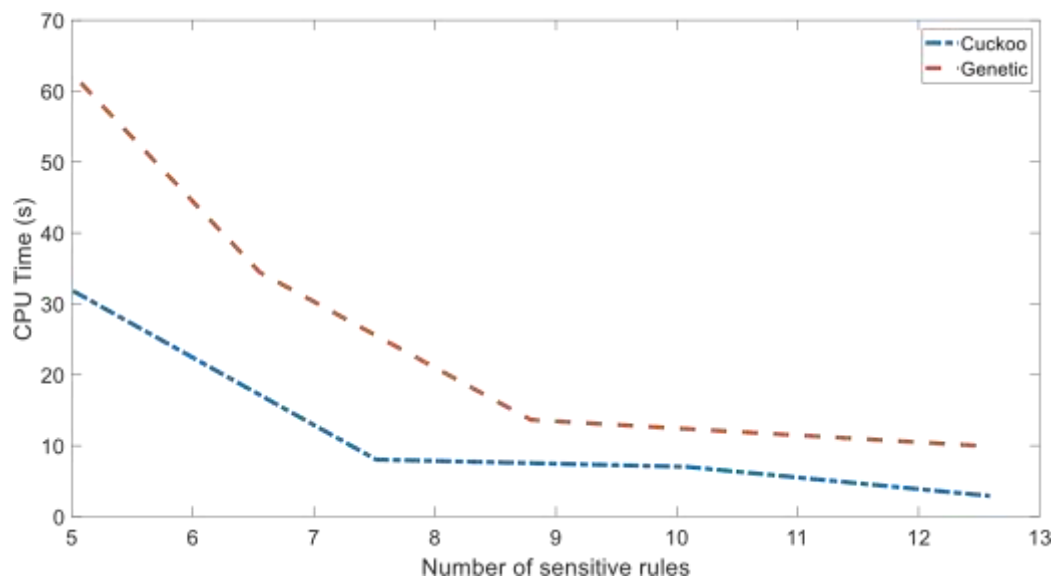


نمودار ۱



(ب)  
نمودار ۲





نمودار ۳

نمودارهای فوق مقایسه عملکرد زمان اجرا روی (الف) مجموعه داده قارچ، (ب) مجموعه داده شطرنج؛ و (ج) مجموعه داده موبایل.

آزمایشها نشان می دهند که الگوریتم پیشنهادی در جنبه زمانی بهتر از الگوریتم ژنتیک عمل می کند. بهبود عملکرد در جدول ۵ توضیح داده شده است.

جدول ۵

بهبودی	Mushroom	Chess	Mobile	Max
زمان	81%	70%	48%	81%

### نتیجه گیری

در این کار، یک الگوریتم جدید مبتنی بر جستجوی فاخته برای پنهان کردن قوانین ارتباط حساس با هزینه زمانی بهتر ارائه شد. ما با بهره گیری از الگوریتم فاخته و با استفاده از بازگشت در تابع هدف آن موفق به دستیابی به این امر شدیم. با هدف نشان دادن برتری رویکرد ما، الگوریتم ژنتیک به عنوان معیار انتخاب شد. تجزیه و تحلیل تجربی نشان داد که عملکرد الگوریتم فاخته با توجه به زمان اجرا، سودمندی و دقت بهتر از ژنتیک بود.

## منابع

- [1] شمشیری و همکاران (۱۳۹۰) پژوهشی تحت عنوان، شناسایی و کشف الگوهای خرید مشتریان لوازم خانگی از طریق داده کاوی انجام دادند.
- [2] خدیور و حامدی (۱۳۹۴) پژوهشی تحت عنوان، ارائه یک الگو ترکیبی داده کاوی با استفاده از قواعد انجمنی و خوشه‌بندی برای تعیین استراتژی تخفیف دهی، مطالعه موردی شرکت پخش پگاه انجام دادند.
- [3] غفوری پور و میرزامؤمن (۱۳۹۷) پژوهشی تحت عنوان استفاده از ارتباطات بین برجسبها در ایجاد زنجیره رده‌بندها برای بهبود رده‌بندی چند برجسی انجام دادند.
- [4] انصاری و همکاران (۱۳۹۷) پژوهشی تحت عنوان، کشف الگوی دسترسی کاربران و تحلیل تراکنش‌های امانت منابع اطلاعاتی با استفاده از تکنیک قوانین انجمنی در داده کاوی (مطالعه موردی: کتابخانه‌ها و مراکز اطلاع‌رسانی دانشگاه علوم پزشکی همدان) انجام دادند.
- [5] میرهاشمی و همکاران (۱۳۹۷) پژوهشی تحت عنوان، استخراج قواعد انجمنی جهت تحلیل رفتار آبخوان شهرستان قزوین توسط الگوریتم *Apriori* انجام دادند.
- [6] غریب و همکاران (۱۳۹۸) پژوهشی تحت عنوان، ارائه الگوی ترکیبی داده کاوی با استفاده از قواعد انجمنی و خوشه‌بندی جهت شناسایی الگوهای غالب رفتار مشتریان مطالعه موردی: بانک انصار انجام دادند.
- [7] المعلگی و ارکوک (۲۰۱۴) پژوهشی تحت عنوان، الگوریتم بهبود یافته *Apriori* برای قوانین انجمنی انجام دادند
- [8] یاسیر و همکاران (۲۰۱۹) پژوهشی تحت عنوان، تشخیص عوامل *Schistosomiasis* با استفاده از استخراج قوانین انجمنی انجام دادند

[9] Zeng, Y. , Yin, S. , Liu, J. , & Zhang, M. (2015). Research of Improved FP-Growth Algorithm in Association Rules Mining. *Sci. Program.* , 910281:1-910281:6.

[10] Gupta, M. ; Joshi, R. C. Privacy-Preserving Fuzzy Association Rules Hiding in Quantitative Data. *Int. J. Comput. Theory Eng.* **2009**, 1, 1793–8201

[11] Srivastava, N. ; Gupta, K. ; Baliyan, N. Improved Market Basket Analysis with Utility Mining. In Proceedings of the 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), Jaipur, India, 26–27 March 2018

[12] Liu, L. ; Yu, S. ; Wei, X. ; Ning, Z. An improved Apriori-based algorithm for friends recommendation in a microblog. *Int. J. Comput. Syst.* **2018**, 31, e3453.

[13] Hareendran, S. A. ; Chandra, S. V. Association Rule Mining in Healthcare Analytics. In *Data Mining and Big Data*; Springer: Cham, Switzerland, 2017; pp. 31–39.

[14] Rekik, R. ; Kallel, I. ; Casillas, J. ; Alimi, A. M. Assessing web sites quality: A systematic literature review by text and association rules mining. *Int. J. Inf. Manag.* **2018**, 38, 201–216

[15] Jabri, S. ; Dahbi, A. ; Gadi, T. ; Bassir, A. Ranking of text documents using TF-IDF weighting and association rules mining. In Proceedings of the 2018 4th International

Conference on Optimization and Applications (ICOA), Mohammedia, Morocco, 26–27 April 2018; pp. 1–6.

[16] Rehman, S. ; Sharma, A. Privacy-Preserving Data Mining Using Association Rule Based on Apriori Algorithm. *Int. Adv. Inform. Comput. Res.* **2017**, 712, 218–226.

[17] Rao V. S. , Vidyavathi S. , "Distributed Data Mining and Mining Multi-Agent Data," *IJCSE International Journal on Computer Science and Engineering*, vol. 2, pp. 1237-1244, 2010.

[18] Han J. , Kamber M. , "Data mining: Concepts and Techniques, 2nd Edition," *San Francisco ,CA, itd: Morgan Kaufmann*, 2005.

[19] Grossman R. , Kasif S. , Moore R. , Rocke D. , Ullman J. , "Data mining research: Opportunities and challenges," presented at the A Report of Three Workshops on Mining Large, Massive and Distributed Data, 1999.

[20] Mareli, M. , Twala,B. (2018). An adaptive Cuckoo search algorithm for optimisation, *Applied Computing and Informatics* 14(2), 107-115.