Financial Research Journal
Print ISSN: 1024-8153
Online ISSN: 2423-5377

Reducing Fraud Detection Costs in Credit Card Transactions: An Information Fusion Approach

MohammadReza Sadeghi Moghadam*
*Corresponding Author, Associate Prof., Department of Industrial Management, Faculty of Industrial Management and Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: rezasadeghi@ut.ac.ir

MohammadReza Mehregan
Prof., Department of Industrial Management, Faculty of Industrial Management and Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: mehregan@ut.ac.ir

Nila Bahrambeig
MSc., Department of Industrial Management, Faculty of Industrial Management and Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: nila.bahrambeyk@ut.ac.ir

Abstract

Objective

Most companies and organizations use e-commerce to gain productivity and efficiency in their services and products in different areas such as credit cards, telecommunications, health insurance, car insurance, etc. Due to the growing volume of credit card transactions and the various methods of fraud and cheating on these cards, the demand for detecting fraud in this area is also increasing. Considering the various solutions and algorithms presented to reduce the cost of fraud detection in credit card transactions in the literature, the purpose of this research is to present a combined and optimal method to reduce the cost of fraud detection in credit card transactions for financial systems, using the fusion of heterogeneous classification and clustering algorithms at the decision-making level by using two fusion methods: Probabilistic fusion and Dempster-Shafer Evidence theory.

Citation: Sadeghi Moghadam, MohammadReza; Mehregan, MohammadReza & Bahrambeig, Nila (2025). Reducing Fraud Detection Costs in Credit Card Transactions: An Information Fusion Approach. *Financial Research Journal*, 27(2), 324-353. https://doi.org/10.22059/FRJ.2024.338715.1007300 (*in Persian*)

Financial Research Journal, 2025, Vol. 27, No.2, pp. 324- 353
Published by University of Tehran, Faculty of Management
https://doi.org/10.22059/FRJ.2024.338715.1007300
Article Type: Research Paper

Received: February 13, 2022
Received in revised form: October 19, 2024
Accepted: November 03, 2024
Published online: May 21, 2025

Methods

This study utilizes a transaction dataset from a Brazilian bank, covering two months from July 14 to September 12, 2004. Fraud detection performance is evaluated using a cost function derived from both a supervised learning approach, namely, a neural network, and an unsupervised method, the K-Means clustering algorithm. Drawing on established fraud detection metrics in the literature, the study adopts the cost function introduced by Gadi as the benchmark. Recognizing the high cost associated with using a single algorithm, the study implements two information fusion techniques—Dempster-Shafer evidence theory and probabilistic fusion—to reduce detection costs. Both fusion methods operate at the decision level, integrating heterogeneous outputs from the supervised and unsupervised models.

Results

Depending on the algorithms implemented, using only one algorithm to obtain an acceptable cost function can be very costly. While using a fusion approach can have a significant impact on cost reduction. The findings indicate that probabilistic fusion significantly outperforms the Dempster-Shafer evidence theory in minimizing the cost function. Specifically, probabilistic fusion achieves a 21.4% cost reduction compared to the artificial neural network and a 35.8% reduction relative to the K-Means algorithm. The results of this study were finally compared with a paper in which this dataset was first used with the artificial immune system (AIS) algorithm and showed a significant cost reduction.

Conclusion

In this study, two classification and clustering algorithms were utilized, and their fusion at the decision level was implemented to demonstrate that combined methods reduce the cost function more effectively than the use of individual algorithms. Furthermore, it was shown that probabilistic fusion yields a lower cost in detecting fraud within financial systems compared to the Dempster-Shafer Evidence theory. This finding is considered significant for banks and financial institutions aiming to develop effective fraud detection systems.

Keywords: Fusion, Dempster-shafer evidence theory, Fraud detection, Credit cards, Artificial neural network.

تحقیقات مالی
شاپا چاپی: ۱۰۲۴-۸۱۵۳
شاپا الکترونیکی: ۲۴۲۳-۵۳۷۷

کاهش هزینهٔ کشف تقلب در تراکنش‌های کارت‌های اعتباری: با رویکرد هم‌جوشی اطلاعات

محمدرضا صادقی مقدم *
* نویسنده مسئول، دانشیار، گروه مدیریت صنعتی، دانشکده مدیریت صنعتی و فناوری، دانشکدگان تهران، دانشگاه تهران، تهران، ایران.
رایانامه: rezasadeghi@ut.ac.ir

محمدرضا مهرگان
استاد، گروه مدیریت صنعتی، دانشکده مدیریت صنعتی و فناوری، دانشکدگان تهران، دانشگاه تهران، تهران، ایران. رایانامه: mehregan@ut.ac.ir

نیلا بهرام‌بیگ
کارشناس ارشد، گروه مدیریت صنعتی، دانشکده مدیریت صنعتی و فناوری، دانشکدگان تهران، دانشگاه تهران، تهران، ایران. رایانامه: nila.bahrambeyk@ut.ac.ir

چکیده

هدف: امروزه اکثر شرکت‌ها و سازمان‌ها، تجارت الکترونیک را برای به‌دست‌آوردن بهره‌وری در خدمات و محصولات خود، در زمینه‌هایی مانند کارت اعتباری، مخابرات، بیمهٔ درمانی، بیمهٔ خودرو و غیره به‌کار گرفته‌اند. از طرفی، با توجه به حجم رو به رشد تراکنش‌های کارت‌های اعتباری و انواع روش‌های کلاه‌برداری و تقلب در این کارت‌ها، تقاضا برای کشف تقلب در این حوزه نیز افزایش یافته است. با توجه به انواع راه‌کارها و الگوریتم‌های ارائه‌شده برای کاهش هزینهٔ تقلب در تراکنش‌های کارت‌های اعتباری، هدف از این پژوهش، ارائه روشی ترکیبی و بهینه، برای کاهش هزینهٔ تشخیص تقلب در تراکنش‌های کارت‌های اعتباری، با استفاده از هم‌جوشی الگوریتم‌های ناهمگن طبقه‌بندی و خوشه‌بندی در سطح تصمیم‌گیری است.

روش: این پژوهش روی داده‌های یک مجموعه تراکنش‌های بانک برزیلی در بازهٔ زمانی دو ماهه، از ۱۴ جولای ۲۰۰۴ تا ۱۲ سپتامبر همان سال انجام شده است. در این مقاله با استفاده از شبکهٔ عصبی مصنوعی، به‌عنوان یک رویکرد با سرپرستی و الگوریتم خوشه‌بندی K نزدیک‌ترین همسایه، به‌عنوان یک رویکرد بدون سرپرستی، تابع هزینه را به‌دست آوردیم. همچنین با توجه به شاخص‌های مختلف کشف تقلب که تاکنون در ادبیات معرفی شده، شاخص هزینهٔ کشف تقلب انتخاب و بر اساس این تابع هزینه که نخستین‌بار توسط گادی و همکارانش (۲۰۰۸) معرفی شد، به سنجش این شاخص‌ها پرداختیم. از آنجا که استفاده از تنها یک الگوریتم هزینهٔ زیادی دارد، به‌منظور کاهش آن، هم‌جوشی الگوریتم‌ها به دو روش نظریهٔ گواه دمستر – شفر و هم‌جوشی احتمالی پیشنهاد شده است. هر دو روش هم‌جوشی در سطح تصمیم استفاده شده و ورودی‌های ناهمگن، از دو رویکرد با سرپرستی و بدون سرپرستی با هم ترکیب شده‌اند.

استناد: صادقی مقدم، محمدرضا؛ مهرگان، محمدرضا و بهرام‌بیگ، نیلا (۱۴۰۴). کاهش هزینهٔ کشف تقلب در تراکنش‌های کارت‌های اعتباری: با رویکرد هم‌جوشی اطلاعات. *تحقیقات مالی*، ۲۷(۲)، ۳۲۴– ۳۵۳.

تحقیقات مالی، ۱۴۰۴، دوره ۲۷، شماره ۲، صص. ۳۲۴– ۳۵۳
ناشر: دانشکدهٔ مدیریت دانشگاه تهران
نوع مقاله: علمی پژوهشی

تاریخ دریافت: ۱۴۰۰/۱۱/۲۴
تاریخ ویرایش: ۱۴۰۳/۰۷/۲۸
تاریخ پذیرش: ۱۴۰۳/۰۸/۱۳
تاریخ انتشار: ۱۴۰۴/۰۲/۳۱
doi: https://doi.org/10.22059/FRJ.2024.338715.1007300

یافته‌ها: با توجه به الگوریتم‌های اجرا شده، استفاده از تنها یک الگوریتم برای به‌دست‌آوردن تابع هزینهٔ قابل قبول، می‌تواند بسیار پُرهزینه باشد. در حالی‌که استفاده از رویکرد هم‌جوشی، می‌تواند در کاهش هزینه تأثیر بسزایی داشته باشد. هم‌جوشی احتمالی، در مقایسه با نظریهٔ گواه دمستر- شفر کاهش هزینهٔ چشمگیری داشته است که هر دو این الگوریتم‌ها، در سطح تصمیم به‌کار رفته‌اند. هم‌جوشی احتمالی نسبت به شبکهٔ عصبی مصنوعی، کاهش هزینه‌ای معادل ۴/۲۱ درصد و نسبت به الگوریتم K نزدیک‌ترین همسایه، کاهش هزینه‌ای معادل ۸/۳۵ درصد داشته است. نتیجهٔ این مطالعه، در نهایت با مقاله‌ای که اولین بار این مجموعه داده در آن با الگوریتم سیستم ایمنی مصنوعی به‌کار رفته است، مقایسه شده و کاهش هزینهٔ چشمگیری را نشان داده است.

نتیجه‌گیری: در این مطالعه با استفاده از دو الگوریتم طبقه‌بندی و خوشه‌بندی و هم‌جوشی آن‌ها در سطح تصمیم، نشان دادیم که روش‌های ترکیبی در مقایسه با استفادهٔ هر یک از الگوریتم‌ها به تنهایی، کاهش هزینهٔ بیشتری خواهند داشت. همچنین هم‌جوشی احتمالی در مقایسه با نظریهٔ گواه دمستر ـ شفر، هزینهٔ کمتری برای کشف تقلب در سیستم‌های مالی دارد که این نتیجه درخور توجهی برای بانک‌ها و مؤسسه‌های مالی است تا یک سیستم کشف تقلب خوب بسازند.

کلیدواژه‌ها: هم‌جوشی، نظریهٔ گواه دمستر ـ شفر، کشف تقلب، کارت‌های اعتباری، شبکهٔ عصبی مصنوعی.

مقدمه

امروزه با گسترش فنّاوری، استفاده از خدمات الکترونیکی در دنیای کسب‌وکار رشد چشمگیری داشته است. در ایران نیز در سال‌های اخیر، استفاده از کارت‌های الکترونیکی در صنعت بانکداری رایج شده است. اگرچه آمار دقیقی از میزان تقلب در کارت‌های اعتباری معتبر در داخل کشور وجود ندارد، با توجه به رشد استفاده از این کارت‌ها، پیش‌بینی می‌شود که مسئلهٔ تقلب در تراکنش‌های کارت‌های اعتباری، به یکی از چالش‌های اصلی بانکداری الکترونیک در کشور تبدیل شود[۱]. با توجه به اینکه در دههٔ گذشته، کارت الکترونیکی پرداخت، پُرکاربردترین راه برای انتقال وجه بوده، تقلب در این حوزه نیز رشد کرده است.

در سال‌های اخیر، اغلب شرکت‌ها و سازمان‌ها تجارت الکترونیک را برای به‌دست‌آوردن بهره‌وری در خدمات و محصولات خود در زمینه‌هایی مانند کارت اعتباری، مخابرات، بیمهٔ درمانی، بیمهٔ اتومبیل، حراج آنلاین و غیره به‌کار گرفته‌اند (عبدالله، معروف و زینال[۲]، ۲۰۱۶). با توجه به حجم زیاد تراکنش‌ها در هر روز و از طرف دیگر، به‌دلیل اینکه در صورت موفقیت، کلاه‌بردار در زمان کوتاهی، مبلغ چشمگیری به‌دست می‌آورد، کلاه‌برداران به‌شدت به تقلب در کارت‌های بانکی تمرکز کرده‌اند و آمار آن رو به افزایش است؛ به‌گونه‌ای که سالانه، میلیاردها دلار خسارت به کسب‌وکارهای مختلف در سراسر جهان زده می‌شود (تریپاتی و پاواسکار[۳]، ۲۰۱۲).

مسئلهٔ تقلب در کارت‌های اعتباری موضوعی بسیار پُرهزینه برای بانک‌ها و سایر مؤسسه‌های مالی صادرکننده است. تشخیص اینکه تراکنش انجام شده توسط مشتری بوده یا کلاه‌بردار، کاری بسیار پُرهزینه است (گادی، وانگ و دو لاگو[۴]، ۲۰۰۸). از طرفی عدم تشخیص به‌موقع تقلب، هزینهٔ بسیار زیادی را برای مؤسسه‌ها دارد و این موضوع ضرورت ایجاد یک سیستم کشف تقلب را بیش از پیش نشان می‌دهد. طی سال‌های اخیر، انواع تقلب مانند تقلب کارت‌های پرداخت، پول‌شویی و... توجه زیادی را به خود جلب کرده‌اند. طبق گزارش بانک مرکزی اروپا، در سال ۲۰۱۲، تنها در اتحادیهٔ اروپا، ارزش کل تقلب به ۱ میلیارد و ۳۳ میلیون یورو رسیده که رشد ۱۴/۸ درصد را نسبت به سال ۲۰۱۱ نشان می‌دهد (بانسن، آودا، استوجانویک و اوترستن[۵]، ۲۰۱۶). گزارش تقلب آنلاین از منبع سایبر (۲۰۱۷) میانگین تلفات کلاه‌برداری سالانه را در بین کانال‌های سفارش مختلف متمایز می‌کند. ۰/۹ درصد از درآمد سالانهٔ تجارت الکترونیک، به‌دلیل کلاه‌برداری در پرداخت از طریق کانال فروشگاه وب در آمریکای شمالی از بین می‌رود. این تعداد برای کانال موبایل ۰/۸ درصد و سفارش‌های تلفن ـ پُست الکترونیکی ۰/۳ درصد است.

طبق آمار ارائه شده، در اردیبهشت سال ۹۹، شبکهٔ پرداخت الکترونیک کشور در سال ۱۳۹۸، افزون بر ۲۶ میلیارد و ۷۱۶ میلیون تراکنش در ابزارهای کارت‌خوان، ابزار پذیرش اینترنتی و موبایلی به ارزش ۳۱ هزار و ۹۳۹ میلیارد ریال

---

۱. پایگاه آگاه‌سازی و اطلاع‌رسانی از جرایم اقتصادی داخل کشور. دسترسی در: https://www.iranhoshdar.ir

2. Abdallah, Maarof & Zainal
3. Tripathi & Pavaskar
4. Gadi, Wang & do Lago
5. Bahnsen, Aouada, Stojanovic & Ottersten

انجام داده است[۱]. همچنین بر اساس مندرجات جدول تراکنش‌های بهمن ۱۳۹۸، به نسبت ماه مشابه سال ۱۳۹۷ از نظر تعدادی ۱۷/۲۵ درصد و از نظر ریالی ۳۵/۹۴ درصد رشد داشته است[۲].

جدول ۱. مقایسهٔ آمار سالانه تراکنش‌های شاپرک در سال‌های ۱۳۹۷ و ۱۳۹۸

نوع	۱۳۹۷	۱۳۹۸	درصد تغییرات
تعداد تراکنش	۱،۹۷۲،۰۲۱،۷۷۷	۲،۳۱۲،۲۰۹،۸۲۷	۱۷/۲۵ درصد
مبلغ تراکنش (میلیون ریال)	۲،۳۸۵،۰۸۳،۰۶۳	۳،۲۴۲،۳۰۸،۳۹۷	۳۵/۹۴ درصد

این حجم تراکنش و مبالغ جابه‌جاشده، بدون وجود سامانهٔ کشف تقلب فعال در کشور و تحقیقات اندک در این زمینه، می‌تواند بستری مناسب برای کلاه‌برداری فراهم و افراد و سازمان‌ها را با زیان روبه‌رو کند (وثوق، تقوی فرد و البرزی، ۱۳۹۳).

در صنعت بانکداری نیز به‌دنبال رشد و توسعهٔ بانکداری نوین، پیشرفت‌های فناوری اطلاعات و در دسترس بودن امکانات کامپیوتری پیشرفته به‌منظور ذخیره داده‌ها، حجم عظیمی از داده‌ها در دسترس تصمیم‌گیرندگان قرار دارد که با توجه به وضعیت رقابتی موجود، تصمیم‌گیری سریع، امکان تبدیل فرصت‌ها و تهدیدها به فرصت‌های طلایی، صنعت بانکداری را به‌سمت استفاده از تکنیک‌های داده‌کاوی[۳] ترغیب می‌کند. با توجه به اینکه تشخیص قانونی بودن یا نبودن تراکنش‌ها به‌راحتی ممکن نیست، کم‌هزینه‌ترین و بهترین راه کشف تقلب با استفاده از الگوریتم‌های ریاضی است. الگوریتم‌های استفاده شده، به‌طور عمده در چهار گروه مدل‌های با سرپرستی[۴]، بدون سرپرستی[۵]، نیمه سرپرستی[۶] و ترکیبی قرار می‌گیرند. هریک از این الگوریتم‌ها با درصدی خطا، به مسئله کشف تقلب پاسخ می‌دهند. یکی از روش‌های کاهش هزینه در این مسئله، ترکیب نتایج الگوریتم‌های گوناگون با یکدیگر است تا طبقه‌بندی و تصمیم‌گیری با دقت بیشتری انجام شود. روش‌های ریاضی متفاوتی نیز برای این ترکیب ارائه شده است که از آن جمله می‌توان به هم‌جوشی احتمالی[۷] و هم‌جوشی به‌روش نظریهٔ گواه[۸] اشاره کرد. این پژوهش درصدد پاسخ به سؤال‌های زیر است؛ از این رو فرضیه ندارد:

- آیا هزینهٔ کشف تقلب با یک الگوریتم بیشتر است یا با ترکیب الگوریتم‌ها؟
- ترکیب هزینه‌های الگوریتم‌ها با رویکرد نظریهٔ گواه بیشتر است یا هم‌جوشی احتمالی؟

هدف از این پژوهش کاهش هزینهٔ کشف تقلب با هم‌جوشی در داده‌ها، از روش دمسترـ شفر و مقایسهٔ نتیجهٔ آن با هم‌جوشی احتمالی در سطح تصمیم‌گیری است.

---

۱. پایگاه خبری اقتصادگردان. بازیابی از: : http://eghtesadgardan.ir

۲. گزارش اقتصادی سالیانه شاپرک (۱۳۹۹).

3. Data mining
4. Supervised
5. Unsupervised
6. Semi-supervised
7. Probabilistic fusion
8. Dempster-Shafer Evidence Theory

پیشینهٔ نظری پژوهش

تقلب یک جرم است که هدف آن، کسب منفعت با ابزارهای غیرقانونی است و در اقتصاد، قانون و حتی ارزش‌های اخلاقی انسان تأثیر منفی چشمگیری دارد (عبدالله و همکاران، ۲۰۱۶).

طبق تعریف انجمن خبرگان بازرسان تقلب (ACFE)، مفهوم تقلب عبارت است از: «استفاده از جایگاه یک نفر برای کسب منفعت شخصی از راه سوءاستفاده یا استفادهٔ نامناسب عمدی از منابع سازمان یا دارایی‌های آن» (عبدالله و همکاران، ۲۰۱۶). همچنین تقلب مالی را می‌توان فریب جنایی با هدف کسب سود بیشتر مالی به‌حساب آورد (اویمی، آدتومبی و الوادر[1]، ۲۰۱۷). تمام سیستم‌های تکنولوژیکی که شامل پول و خدمات می‌شوند، ازجمله مخابرات، بیمهٔ درمانی، بیمهٔ اتومبیل، سیستم حراج آنلاین، مراقبت‌های سلامت، خدمات عمومی و بانکداری، می‌توانند با اقدام‌های جعلی درگیر شوند (عبدالله و همکاران، ۲۰۱۶؛ دومان و ازلیک[2]، ۲۰۱۱). بر اساس نظر کمیتهٔ بازل در سرپرستی بانکداری، تقلب به دو دسته تقلب‌های داخلی (شغلی) و تقلب‌های خارجی تقسیم‌بندی می‌شود (عبدالله و همکاران، ۲۰۱۶). کلاه‌برداری مالی را «اشتباه عمدی برداشت ارزش‌های خالص مالی، برای افزایش سودآوری و فریب سهام‌داران» تعریف می‌کنند، درحالی‌که تقلب تراکنشی، خراب‌کاری در دارایی‌های سازمان است (البشراوی[3]، ۲۰۱۶). تراکنش‌های کارت‌های اعتباری، به دو دستهٔ فیزیکی و مجازی تقسیم می‌شوند. در تراکنش‌های فیزیکی، دارندهٔ کارت، هنگام خرید به‌صورت فیزیکی کارت خود را ارائه می‌دهد. در این حالت، برای تقلب، باید سارق کارت را به‌صورت فیزیکی برباید؛ اما در حالت مجازی، تنها برخی از اطلاعات کارت مانند شمارهٔ کارت، تاریخ انقضا، کد امنیتی و... به سرقت می‌رود. معمولاً هنگام سرقت به‌شیوهٔ مجازی، دارندهٔ کارت در جریان دزدیده شدن اطلاعات نیست (فوآ، لی، اسمیت و گیلر[4]، ۲۰۰۵). به‌طورکلی در صنعت پرداخت، تقلب در کارت زمانی اتفاق می‌افتد که کسی اطلاعات کارت شما را سرقت می‌کند تا بدون اجازهٔ شما خرید کند (جارگوفسکی و همکاران[5]، ۲۰۱۸).

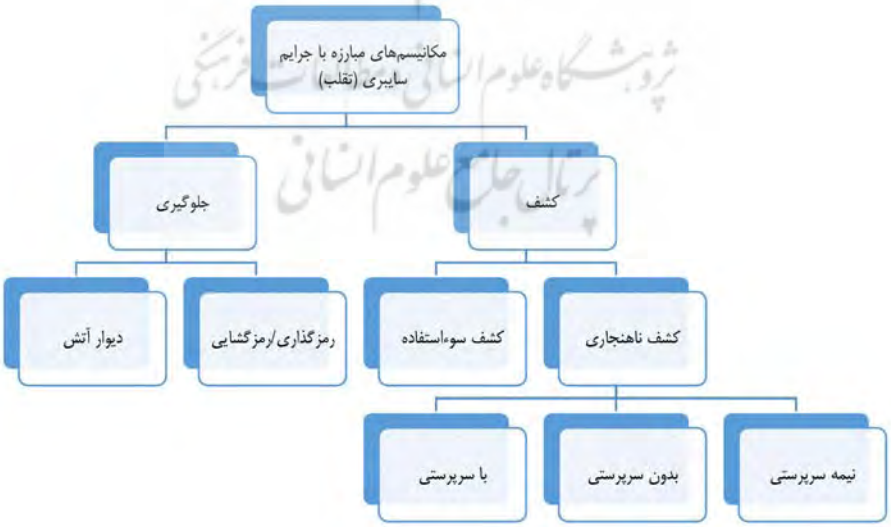
کشف تقلب به هر تلاشی برای تشخیص رفتار متقلبانه در سیستم‌هایی اطلاق می‌شود که از طریق آن متقلب، می‌تواند اقدام به تحصیل مال کند (آرال، گوونیر، سابونکوگلو و آکار[6]، ۲۰۱۲). سلطانی کشف تقلب را عمل تشخیص و متوقف‌کردن آن در سریع‌ترین زمان ممکن، یعنی قبل از انجام تراکنش تعریف می‌کند (حلوایی و اکبری[7]، ۲۰۱۴). بهاتیا نیز تشخیص تقلب کارت‌های اعتباری را به‌عنوان یک مسئلهٔ داده‌کاوی مطرح می‌کند که هدف آن، طبقه‌بندی صحیح تراکنش‌ها به دو دستهٔ قانونی و تقلب‌آمیز است (بهاتیا، بجاج و هزاری[8]، ۲۰۱۶). سرورنژاد کشف تقلب را شناسایی فعالیت‌های تقلب‌آمیز معدود، از بین تراکنش‌های متعدد قانونی در اسرع وقت معرفی می‌کند (سرورنژاد، ابراهیمی آتانی و

1. Awoyemi, Adetunmbi & Oluwadare
2. Duman & Ozcelik
3. Albashrawi
4. Phua, Lee, Smith & Gayler
5. Jurgovsky et al.
6. Aral, Güvenir, Sabuncuoğlu & Akar
7. Halvaiee & Akbari
8. Bhatia, Bajaj & Hazari

منجمی[1]، ۲۰۱۶) و به‌زعم کومار، کشف تقلب، به‌معنای نظارت بر عملکرد کاربران به‌منظور تخمین، کشف و جلوگیری از عملکرد ناخواسته است (کومار و رائو[2]، ۲۰۱۳).

کشف تقلب برای کاهش تأثیر تراکنش‌های تقلب‌آمیز بر ارائهٔ خدمات، هزینه‌ها و شهرت شرکت‌ها، بسیار حیاتی است (حلوایی و اکبری، ۲۰۱۴). اجرای راه‌حل‌های مؤثر کشف تقلب، مهم‌ترین مسئله برای همهٔ مؤسسه‌های صادرکنندهٔ کارت اعتباری و مدیریت تراکنش‌های آنلاین است تا به‌طور هم‌زمان ضرر را کاهش و اعتماد مشتریان را افزایش دهند (فیوره، دسانتیس، پرلا، زانتی و پالمیری[3]، ۲۰۱۹). به‌دست‌آمدن الگوی تقلب به‌طور خودکار، مشخص‌کردن «احتمال تقلب» برای هر مورد و در نتیجه، اولویت‌بندی بررسی موارد مشکوک و کشف انواع تقلب جدید که قبلاً شناسایی نشده بودند، از مزایای سیستم‌های خودکار کشف تقلب هستند (عبدالله و همکاران، ۲۰۱۶). سیستم‌های کشف تقلب، در حالت پیشرفته به شناسایی الگوهای مشکوک تراکنش‌های ثبت‌شده که در آن، تراکنش‌های غیرقانونی با تراکنش‌های قانونی ترکیب شده، با استفاده از تجزیه‌وتحلیل پیچیده و تکنیک‌های داده‌کاوی می‌پردازند و روی داده‌ها طبقه‌بندی باینری انجام می‌دهند و تراکنش‌ها را به دو طبقه قانونی و تقلب‌آمیز تفکیک می‌کنند (فیوره و همکاران، ۲۰۱۹).

یک سیستم مفید کشف تقلب، باید توانایی حل چالش‌ها را به‌منظور دستیابی به بهترین عملکرد داشته باشد (سرورنژاد و همکاران، ۲۰۱۶). حضور هریک از این چالش‌ها، به هشدارهای نادرست زیاد، دقت تشخیص پایین و تشخیص کُند منجر می‌شود (عبدالله و همکاران، ۲۰۱۶). اصلی‌ترین این چالش‌ها عبارت‌اند از: داده‌های نامتقارن، اهمیت طبقه‌بندی‌های اشتباه متفاوت، داده‌های هم‌پوشانی شده، ناسازگاری، هزینهٔ کشف تقلب، عدم وجود معیارهای استاندارد و زمان‌بر بودن ثبت داده‌های جدید. سازوکارهای تشخیص و پیشگیری تقلب به‌منظور مبارزه با تقلب، در شکل ۱ مشاهده می‌شود.



شکل ۱. سازوکار تشخیص و پیشگیری تقلب

1. Sorournejad, Ebrahimi Atani & Monadjemi
2. Kumar & Rao
3. Fiore, De Santis, Perla, Zanetti & Palmieri

جلوگیری، رویکردی پیشگیرانه است؛ اما هدف از کشف تقلب، توقف آن در کوتاه‌ترین فاصلهٔ زمانی ممکن پس از وقوع است (وثوق و همکاران، ۱۳۹۳).

سلطانی و سرورنژاد رویکردهای کشف تقلب را به دو دستهٔ کلی طبقه‌بندی کرده‌اند: تجزیه‌وتحلیل که به‌دنبال مقداری بین رفتار قانونی و تقلب‌آمیز می‌گردد و تجزیه‌وتحلیل رفتار کاربر (رویکرد دیفرانسیلی) که تلاش می‌کند تغییر شدید رفتار کاربر را شناسایی کند. آن‌ها رویکرد اول را رویکردی با سرپرستی می‌نامند که مبتنی بر تشخیص سوءاستفاده است و رویکرد دوم را بدون سرپرستی و منطبق بر رفتار کاربر در طول زمان تعریف می‌کنند که بر اساس تشخیص ناهنجاری است (حلوایی و اکبری، ۲۰۱۴؛ سرورنژاد و همکاران، ۲۰۱۶). سلطانی و اکبری با روش AIRS[1] که بر مبنای الگوریتم سیستم ایمنی مصنوعی است، توانستند تابع هزینهٔ تعریف شده روی این مجموعه داده را تا ۸۵ درصد کاهش دهند. این روش در عین حال دقت را ۲۵ درصد افزایش و در مقایسه با الگوریتم پایه زمان را ۴۰ درصد کاهش داد (حلوایی و اکبری، ۲۰۱۴).

به طور کلی، رویکردهای کشف تقلب به سه دسته اصلی رویکردهای باسرپرستی، رویکردهای بدون سرپرستی و رویکردهای نیمه سرپرستی تقسیم می‌شوند. اصلی‌ترین الگوریتم‌های طبقه‌بندی رویکرد باسرپرستی عبارت‌اند از: شبکهٔ عصبی مصنوعی[2]، K نزدیک‌ترین همسایه[3]، درخت تصمیم[4]، بیز ساده[5]، ماشین بردار پشتیبان[6] (عبدالله و همکاران، ۲۰۱۶). الگوریتم‌های خوشه‌بندی مانند K نزدیک‌ترین همسایه و الگوریتم‌های کاهش ابعاد مانند تجزیه‌وتحلیل مؤلفه اصلی[7]، از الگوریتم‌های رویکرد بدون سرپرستی هستند (عبدالله و همکاران، ۲۰۱۶). در رویکرد نیمه سرپرستی، ابتدا از رویکردهای بدون سرپرستی برای ایجاد یک دانش اولیه در خصوص توزیع مشاهدات استفاده می‌شود، پس‌ازآن از یک روش با سرپرستی برای استنتاج نهایی بهره‌گیری می‌شود (فوآ و همکاران، ۲۰۱۰).

منظور از روش‌های ترکیبی با سرپرستی، استفاده از رویکردهای ترکیبی با بیش از یک الگوریتم است که در آن دو الگوریتم با سرپرستی باهم استفاده می‌شوند. در ادبیات روش‌های معروف مانند شبکه‌های عصبی، درخت تصمیم یا شبکه‌های بیز، به‌صورت متوالی برای بهبود نتایج با یکدیگر ترکیب شده‌اند. سپس از یک ابر طبقه‌بند[8] برای ترکیب نتایج طبقه‌بندی استفاده می‌شود (آرال و همکاران، ۲۰۱۲). در حالت اول، معمولاً ترکیبی از نتایج طبقه‌بندها به‌عنوان یادگیری توأم صورت می‌پذیرد که منظور از آن، انجام نمونه‌گیری‌های تصادفی از یک منبع داده‌ای و اجرای طبقه‌بندی‌های مختلف با آن نمونه‌هاست. این روش برای افزایش صحت پیش‌بینی در مسائل پیچیده به‌کار می‌رود یا با دست‌کاری توزیع مجموعهٔ آموزش بر اساس عملکرد طبقه‌بندی قبلی انجام می‌شود (هداوندی، شهرابی و حیاشی[9]، ۲۰۱۵). نکتهٔ

1. Artificial Immune Recognition System
2. Artificial Neural Network (ANN)
3. K-Nearest Neighbor (KNN)
4. Decision Tree
5. Naïve Bayes
6. Support Vector Machine (SVM)
7. Principle Component Analysis (PCA)
8. Meta classifier
9. Hadavandi, Shahrabi & Hayashi

مهم در این روش، استفاده از یک منبع داده‌ای برای آموزش یک ابر طبقه‌بند بر اساس عملکرد چند طبقه‌بند است (گراوینا، آلینیا، قاسم‌زاده و فورتینو[۱]، ۲۰۱۷).

الگوریتم شبکه‌های مصنوعی یکی از معروف‌ترین روش‌های کشف تقلب است (چانگ، لای، سو، وانگ و کوه[۲]، ۲۰۰۶؛ براوس، لانگسدورف و هپ[۳]، ۱۹۹۹). در میان پژوهش‌های گزارش‌شده در خصوص کشف تقلب، به‌ویژه در کارت‌های اعتباری، بیشترین روش استفاده شده، شبکه‌های عصبی هستند (نای، هو، ونگ، چن و سون[۴]، ۲۰۱۱؛ شرلی[۵]، ۲۰۱۲).

مائز و همکارانش در کشف تقلب کارت‌های اعتباری نتایج دو الگوریتم شبکهٔ عصبی و شبکه‌های بیزی را باهم مقایسه کردند و به این نتیجه رسیدند که شبکه‌های عصبی دقت بیشتری دارند (مائز، تویلس، وانشون وینکل و ماندریک[۶]، ۱۹۹۳). استولفو و همکارانش در سال ۱۹۹۷، از الگوریتم‌های درخت تصمیم و طبقه‌بند بیز برای مجموعه داده‌های متقارنی که نیمی از آن‌ها تقلب‌آمیز و نیم دیگر قانونی بودند، استفاده کردند و یک سیستم کشف تقلب برای استفاده از روش‌های ابر یادگیری ارائه دادند. در این پژوهش آن‌ها به بالاترین نرخ اعلان درست (TP)[۷] و پایین‌ترین نرخ اعلان نادرست (FP)[۸] دست یافتند (استولفو، فن، لی، پرودریمیدیس و چان[۹]، ۱۹۹۷).

به‌زعم کندو و همکارانش، هر دارندهٔ کارت رفتار خرید منحصربه‌فردی دارد که می‌توان بر اساس آن، یک پروفایل ایجاد کرد؛ اما استفاده از این منبع، در صورت تغییر رفتار دارندهٔ کارت ناکارآمد خواهد بود. بنابراین برای غلبه بر نقص ناکافی بودن منابع داده‌ای، از دو سیستم طبقه‌بندی بر اساس شبکه‌های بیزی استفاده کردند که یکی بر مبنای شباهت به رفتار قانونی مالک کارت و دیگری بر مبنای شباهت به رفتار تقلب‌آمیز بود. نتیجهٔ دو رویکرد با نظریهٔ گواه دمستر ـ شفر ترکیب شد و اعلان اشتباه پایین را گزارش داد (پانیگراهی، کوندو، سورال و ماجومدار[۱۰]، ۲۰۰۹). هرمزی و همکارانش در سال ۲۰۱۳، الگوریتمی بر مبنای سیستم ایمنی مصنوعی روی هادوپ با مدل برنامه‌ریزی کاهش نگاشت ارائه دادند که نتیجهٔ آن روی همین مجموعه داده‌ها، کاهش زمان آموزش بود (هرمزی، اکبری، هرمزی و سرگلزائی‌جوان[۱۱]، ۲۰۱۳).

لو و همکاران در سال ۲۰۱۵ با به‌کارگیری روش‌های طبقه‌بندی ایمنی مصنوعی و شبکهٔ ایمنی مصنوعی، عملکرد بهتری را نسبت به روش‌های درخت تصمیم، ماشین برداری پشتیبان، رگرسیون لجستیک[۱۲] و بیز ساده در داده‌های یک بانک پرتغالی گزارش کردند (لو، چو، چن و چنگ[۱۳]، ۲۰۱۵).

---

1. Gravina, Alinia, Ghasemzadeh & Fortino
2. Chang, Lai, Su, Wang & Kouh
3. Brause, Langsdorf & Hepp
4. Ngai, Hu, Wong, Chen & Sun
5. Sherly
6. Maes, Tuyls, Vanschoenwinkel & Manderick
7. True Positive
8. False Positive
9. Stolfo, Fan, Lee, Prodromidis & Chan
10. Panigrahi, Kundu, Sural & Majumdar
11. Hormozi, Akbari, Hormozi & Javan
12. Logistic Regression
13. Lu, Chu, Chen & Chang

روش جدید بر مبنای شبکهٔ عصبی که توسط قبادی و روحانی روی داده‌های بانک برزیلی ارائه شد، در مقایسه با سیستم ایمنی مصنوعی، کشف بیشتر و هزینهٔ کمتری را گزارش کرد (قبادی و روحانی[1]، ۲۰۱۶). وثوق و همکارانش با رویکرد تمرکز بر تراکنش‌های کارت‌ها، با استفاده از یک شبکهٔ عصبی چندلایه، موفق شدند تقلب را در تراکنش‌های بانکی به‌سرعت تشخیص دهند. آن‌ها در پژوهش خود از داده‌های واقعی استفاده کردند و عملکرد مطلوبی را ازلحاظ چهار شاخص نرخ اعلان درست، نرخ عدم اعلان نادرست، میانگین هندسی و آماره فیشر بتا گزارش دادند (وثوق و همکاران، ۱۳۹۳).

تحقیقات اویمی در سال ۲۰۱۷، روی داده‌های بانک اروپایی نشان داد که صحت الگوریتم بیز ساده در مقایسه با K نزدیک‌ترین همسایه و رگرسیون لجستیک بیشتر است؛ اما K نزدیک‌ترین همسایه، در سایر شاخص‌ها، از دو روش دیگر عملکرد بهتری دارد (اویمی و همکاران، ۲۰۱۷). فیوره نیز توانست با روش Generative Adversarial Networks به FP بالاتری دست یابد (فیوره و همکاران، ۲۰۱۹). کارنئیرو نیز بین الگوریتم‌های رگرسیون لجستیک، ماشین بردار پشتیبان و جنگل تصادفی[2]، بهترین عملکرد را از جنگل تصادفی به‌دست آورد (کارنئیرو، فیگویرا و کوستا[3]، ۲۰۱۷).

در سال ۲۰۱۸ نیز، آکیلا و ردی توانستند با استفاده از روش ریسک ناشی از نمونه‌گیری استنتاج بیزی حساس به هزینه[4]، تابع هزینهٔ مجموعه دادهٔ بانک برزیلی را ۱/۰۴ تا ۱/۵ بار کاهش دهند. این روش یک معماری جدید نمونه‌گیری است که یک کیسهٔ نمونهٔ محدود را شامل می‌شود و ریسک ناشی از نمونه‌گیری استنتاج بیزی به‌عنوان یادگیرندهٔ پایه است (آکیلا و ردی[5]، ۲۰۱۸). گوپتا، مالسا و گوپتا[6] (۲۰۱۷)، در مقاله‌ای مروری، بین روش‌های سیستم‌های ایمنی مصنوعی، مدل مخفی مارکوف، شبکهٔ عصبی، الگوریتم ژنتیک، درخت تصمیم و ماشین برداری پشتیبان مقایسه‌ای انجام داد و به این نتیجه رسید که سیستم ایمنی مصنوعی، مدل مخفی مارکوف، درخت تصمیم و شبکهٔ عصبی، در مقایسه با الگوریتم ژنتیک و ماشین بردار پشتیبان سرعت بیشتری دارند، درحالی‌که سیستم ایمنی مصنوعی و الگوریتم ژنتیک نسبت به سایر روش‌ها بسیار کم‌هزینه هستند.

مالینی با استفاده از الگوریتم K نزدیک‌ترین همسایه روی داده‌های واقعی، دقت و عملکرد بالایی به‌دست آورد (مالینی و پوشپا[7]، ۲۰۱۷).

ادیبی و شهرابی با استفاده از ترکیب شبکهٔ عصبی کوهونن و مدل خوشه‌بندی بردار پشتیبان، سیستمی برای کشف نفوذ در شبکه‌های کامپیوتری ارائه دادند. در این روش، نخست تراکنش‌ها توسط شبکهٔ کوهونن خوشه‌بندی شدند و در گام بعد، مرزهای نرمال و غیرنرمال مجموعهٔ به‌دست‌آمده با استفاده از خوشه‌بندی بردار پشتیبان مشخص شدند. این الگوریتم از لحاظ عملکرد و زمان اجرا در طبقه‌بندی، نتیجه خوبی نشان داده است (ادیبی و شهرابی[8]، ۲۰۱۵).

1. Ghobadi & Rohani
2. Random Forest
3. Carneiro, Figueira & Costa
4. Cost- sensitive Risk Induced Bayesian Inference Bagging
5. Akila & Reddy
6. Gupta, Malsa & Gupta
7. Malini & Pushpa
8. Adibi & Shahrabi

زری‌پور و همکارانش، ویژگی‌های الگوریتم‌های مختلف را بر اساس سه پارامتر اصلی هزینه، صحت و سرعت تشخیص دادند و طبق جدول زیر مقایسه کردند (زری‌پور، سیجا و عالم[1]، ۲۰۱۲).

جدول ۲. مقایسهٔ الگوریتم‌های مختلف داده‌کاوی

پارامتر	شبکهٔ عصبی	شبکهٔ بیزی	ماشین بردار پشتیبان	K نزدیک‌ترین همسایه	درخت تصمیم	سیستم فازی	ایمنی مصنوعی	الگوریتم ژنتیک	مدل مخفی مارکوف
سرعت کشف تقلب	سریع	سریع	آهسته	سریع	سریع	آهسته	سریع	سریع	سریع
صحت	بالا	بالا	متوسط	متوسط	متوسط	بسیار بالا	خوب	متوسط	پایین
هزینه	پُرهزینه	پُرهزینه	پُرهزینه	پُرهزینه	پُرهزینه	پُرهزینه	کم‌هزینه	ارزان	بسیار پُرهزینه

با توجه به جدول ۲، مشاهده می‌شود که هر یک از الگوریتم‌ها به تنهایی معایبی دارد و استفاده از رویکرد ترکیبی برای افزایش صحت تشخیص مناسب است.

گادی و همکارانش (۲۰۰۸) اطلاعات مربوط به تراکنش‌های کارت‌های اعتباری یک بانک برزیلی را معرفی کردند. آن‌ها از الگوریتم‌های شبکهٔ عصبی، شبکهٔ بیزی[2]، بیز ساده، سیستم ایمنی مصنوعی[3]، درخت تصمیم و الگوریتم ژنتیک[4] برای تحلیل این داده‌ها استفاده کردند. نتیجهٔ این پژوهش نشان داد که درخت تصمیم، سیستم ایمنی مصنوعی، شبکهٔ عصبی و شبکهٔ بیزی به‌طور مساوی بهترین الگوریتم‌ها هستند و درنهایت بیز ساده است. همچنین در این مقاله، روش محاسبهٔ تابع هزینه با توجه به مجموعه داده‌ها ارائه شده است. در مدل‌های مختلف کشف تقلب، شاخص‌های سنجش کارایی الگوریتم‌ها، صحت[5]، دقت[6]، فراخوانی[7]، شاخص F-measure و نمودار ROC است. گادی و همکارانش، علی‌رغم در نظر گرفتن تمام این شاخص‌ها، در صحبت با خبرگان برای جلوگیری از تقلب و مشاهدهٔ نتایج اولیهٔ نمودار ROC و دقت، دریافتند که اگر تابع هزینه‌ای را تعریف کنند که در آن برای تأیید هر تراکنش متوسط هزینهٔ یک دلار و برای هر تقلب تشخیص داده نشده، متوسط هزینه ۱۰۰ دلار را در نظر بگیرد، به نتایج کاربردی‌تری دست می‌یابند. این تابع هزینه توابع دقت و فراخوانی را در یک شاخص تجمیع می‌کند و عملکرد تابع را تنها در یک نقطهٔ کاربردی برش[8] ارزیابی می‌کند. این تابع هزینه به منظور شباهت بیشتر روش‌های استفاده شده برای یک امتیاز تقلب نسبت به نمودار ROC در نظر گرفته شده است که شاخص‌های مختلف را به‌طور همزمان مقایسه می‌کند.

1. Zareapoor, Seeja & Alam
2. Bayesian Network
3. Artificial Immune System (AIS)
4. Genetic Algorithm
5. Accuracy
6. Precision
7. Recall
8. Cut-off

با توجه به مزیت اصلی شاخص هزینه که در نظر گرفتن کلیهٔ شاخص‌ها در یک تابع هزینه است و همچنین، با توجه به اینکه عدم کشف تقلب برای بانک‌ها و مؤسسه‌های مالی بسیار هزینه‌بر است، در ایران نیز تابع هزینه اهمیت ویژه‌ای دارد؛ اما به‌دلیل اینکه در حال حاضر، در داخل کشور مجموعه داده‌ای برای تراکنش‌های کارت‌های اعتباری وجود ندارد، در این مطالعه از مجموعه دادهٔ بانک برزیلی استفاده کرده‌ایم، از تابع هزینهٔ محاسبه شده برای مجموعه داده‌های آن بانک، برای این تحقیق استفاده و نتایج را با تابع هزینهٔ مقالهٔ گادی (۲۰۰۸) مقایسه کردیم.

مسئلهٔ کشف تقلب دارای نقص ابهام[۱] (زارعی‌پور و شمس‌المعالی[۲]، ۲۰۱۵) و عدم اطمینان[۳] (پانیگراهی و همکاران، ۲۰۰۹) است که این نقایص به‌دلیل پیچیدگی رفتار متقلب و پراکندگی رخداد تقلب در میان حجم انبوه تراکنش‌های قانونی است. پژوهش‌های فراوانی انجام شده است تا بتوان نظریه‌های مختلف را با روش‌های مختلف به‌گونه‌ای باهم ترکیب کرد که بتوانند چند نوع نقص داده‌ای را هم‌زمان پوشش دهند و این تلاش درنهایت به ارائهٔ مفهومی با نام هم‌جوشی[۴] داده منجر شد. نظریه‌های ریاضی زیادی مانند نظریهٔ احتمال[۵]، نظریهٔ مجموعه فازی[۶]، نظریهٔ امکان[۷]، نظریهٔ مجموعهٔ ژولیده[۸] و نظریهٔ گواه دمستر ـ شفر برای پوشش این نقص‌ها به‌کار می‌روند. نظریه‌های موجود به‌تنهایی قادر نیستند مسائلی را حل کنند که چند نوع نقص اطلاعاتی دارند (گراوینا و همکاران، ۲۰۱۷)؛ به همین دلیل تعاریف متعددی از هم‌جوشی در ادبیات وجود دارد. بوسترم و همکاران تعریف جدیدی از هم‌جوشی با مضمون مطالعه روش‌های کارا برای تبدیل خودکار و یا نیمه‌خودکار اطلاعات از منابع و نقاط مختلف در زمان و بازنمایی آن به صورتی که پشتیبانی مؤثری برای انسان یا سیستم‌های تصمیم‌گیری فراهم آورد، ارائه دادند (بوسترم و همکاران[۹]، ۲۰۰۷).

گراوینا نیز هم‌جوشی داده را حوزه‌ای چندرشته‌ای می‌داند که ایده‌های مختلف را از حوزه‌هایی همچون پردازش سیگنال، نظریهٔ اطلاعات، استنتاج و برآورد آماری و هوش مصنوعی قرض گرفته است (گراوینا و همکاران، ۲۰۱۷). جاردات معتقد است که هم‌جوشی اطلاعات، باعث به وجود آمدن اطلاعات با کیفیت بالاتر و اطلاعات دقیق‌تر می‌شود (جاردات و همکاران[۱۰]، ۲۰۲۲). به‌زعم آهنگربهان و منتظر (۱۳۹۵)، اجرای هم‌جوشی داده، صحت و تشخیص، اعتماد و اطمینان را بهبود و ابهام را کاهش می‌دهد.

گراوینا هم‌جوشی را در سه سطح داده‌محور[۱۱]، ویژگی‌محور[۱۲] و تصمیم‌محور[۱۳] قرار می‌دهد. در سطح داده، فرض بر قابل‌اعتماد بودن اطلاعات و سیستم پردازش است و تمرکز بر الگوریتم‌های ترکیب اطلاعات همگن، به‌منظور دستیابی به

1. Ambiguity
2. Zareapoor & Shamsolmoali
3. Uncertainty
4. Fusion
5. Probability Theory
6. Fuzzy Set Theory
7. Possibility Theory
8. Rough Set Theory
9. Boström et al.
10. Jaradat et al.
11. Data Level
12. Feature Level
13. Decision Level

اطلاعات دقیق‌تر از منابع اولیه توسط منابع داده‌ای است. در سطح ویژگی، پدیده را با ویژگی‌های مختلف توصیف کرده و سپس آن‌ها را برای استنتاج کل پدیده باهم، ترکیب می‌کنیم. اما در سطح تصمیم که بالاترین سطح است، خروجی، یک تصمیم از بین چند تصمیم است که از منابع داده‌ای همگن یا غیرهمگن به‌دست می‌آید (گراوینا و همکاران، ۲۰۱۷).

استفاده از مقادیر احتمالی، یکی از رایج‌ترین روش‌ها برای مدل‌سازی اطلاعات ناکامل است. هم‌جوشی احتمالی، به‌طورکلی بر قانون بیز تکیه دارد. مهم است که توجه داشته باشیم برای ترکیب دو توزیع احتمالی، آن‌ها باید روی یک چارچوب تشخیص قرار گیرند (ژو، داوین، بوردس، ژاو و دناکس[1]، ۲۰۱۶).

نظریهٔ دمستر- شفر تفکر به‌عنوان توسعه‌ای از نظریهٔ بیز در نظر گرفته می‌شود و امکان قضاوت بر اساس اطلاعات غیردقیق را می‌دهد و کمک می‌کند تا درجهٔ احتمال به‌جای نقطه‌ای با بازه سنجیده شود. این نظریه، از مفهوم حدود بالایی و پایینی احتمال با به‌کارگیری یک نگاشت چندگانه سرچشمه می‌گیرد (ین[2]، ۱۹۹۰).

نظریهٔ گواه توانایی حل مسئله با چندین تصمیم‌گیر و انواع داده‌های ناهمگن را دارد. با توجه به اینکه این نظریه می‌تواند مسائلی را حل کند که در آن، بخشی از اطلاعات ازدست‌رفته و داده‌ها ناکامل یا نامطمئن هستند، به‌شدت مورد استقبال قرارگرفته است (آواشتی و چوهان[3]، ۲۰۱۱).

بائه و همکاران نظریهٔ گواه را با ارائهٔ الگوریتم کم‌هزینهٔ محاسباتی برای ساختارهای طراحی مهندسی، روی تحلیل کمی‌سازی داده‌های نامطمئن به‌کار گرفتند. الگوریتم آن‌ها روشی برای بهینه‌سازی و تقریبی برای ارزیابی توابع باور و مقبولیت را ارائه داده است (بائه، گراندهی و کانفیلد[4]، ۲۰۰۴). پوپسکو و همکارانش، از نظریهٔ گواه روی دادهٔ مبهم و ناکامل استفاده و با آن، الگوهای منفی و مثبت ورودی را مشاهده کردند (پوپسکو، لونا، ماراندا، وانسیا و تیوبه[5]، ۲۰۱۰).

هی و همکارانش از نظریهٔ گواه و هم‌جوشی بیزی برای تخمین سلامت و زمان مفید باقی‌مانده برای طول عمر باتری‌های لیتیومی استفاده کردند و توانستند زمان باقی‌ماندهٔ عمر باتری را پیش‌بینی کنند (هی، ویلارد، اوسترمن و پچ[6]، ۲۰۱۱). ژو و همکارانش نیز از هم‌جوشی احتمالی و نظریهٔ گواه دمستر ـ شفر در تشخیص فضاهای شهری استفاده کردند. ورودی مدل آن‌ها حسگرهای مختلف شهری بود (ژو و همکاران، ۲۰۱۶).

طبسیان و همکارانش از ترکیب یادگیری ماشین و نظریهٔ گواه، برای طبقه‌بندی در حالتی که برچسب داده‌ها ناقص و مبهم است، استفاده کردند. آن‌ها با استفاده از شبکهٔ عصبی، تخصیص باور پایه را انجام دادند. این مطالعه نشان داد که روش ترکیبی آن‌ها کارایی خوبی نسبت به استفادهٔ مجزا دارد (طبسیان، قادری و ابراهیم‌پور[7]، ۲۰۱۲). خطیبی و منتظر از ترکیب این دو نظریه برای ارزیابی بیماری‌های قلبی استفاده کردند که نتایج بهتری نسبت به روش‌های قبلی ارائه داده است (خطیبی و منتظر[8]، ۲۰۱۰). منصوری و همکارانش نیز با ترکیب نظریه‌های گواه و فازی، نتایج حاصل از دو سیستم

1. Xu, Davoine, Bordes, Zhao & Denœux
2. Yen
3. Awasthi & Chauhan
4. Bae, Grandhi & Canfield
5. Popescu, Lonea, Zmaranda, Vancea & Tiurbe
6. He, Williard, Osterman & Pecht
7. Tabassian, Ghaderi & Ebrahimpour
8. Khatibi & Montazer

استنتاج فازی در تشخیص کیفیت خدمت سرویس‌های صدا روی اینترنت را بررسی کردند (منصوری، نبوی، زارع رواسان و آهنگربهان[1]، ۲۰۱۵).

با توجه به بررسی مطالعات پیشین، به این نتیجه رسیدیم که تنها یک مقاله در حوزهٔ ترکیب الگوریتم‌های مختلف، به‌روش نظریهٔ گواه و هم‌جوشی احتمالی برای مجموعه داده‌های بانکی انجام شده که در آن نیز از داده‌هایی که به‌طور مصنوعی تولید شده‌اند، به‌جای داده‌های واقعی استفاده شده است (پانیگراهی و همکاران، ۲۰۰۹). مجموعه دادهٔ بانک برزیلی، به‌صورت یک مجموعه دادهٔ جهانی مورد تأیید است. در تحقیق سال ۲۰۰۸، گادی و همکارانش الگوریتم‌های مختلف شبکهٔ عصبی، شبکهٔ بیزی، بیز ساده، سیستم ایمنی مصنوعی، درخت تصمیم و الگوریتم ژنتیک را روی این مجموعه داده پیاده‌سازی کردند؛ اما نه در آن تحقیق و نه در تحقیقات پس از آن، روی این مجموعه دادهٔ جهانی، از روش ترکیب الگوریتم‌ها به‌منظور کاهش هزینه استفاده نشده است. تلاش ما در این مطالعه، پوشش این شکاف تحقیقاتی و مقایسهٔ آن با مقالهٔ گادی و همکاران (۲۰۰۸) است.

روش‌شناسی پژوهش

روش این پژوهش، از نوع روش‌های تحقیق تحلیلی ریاضی در طبقه‌بندی واکر است و هدف آن، کاهش هزینهٔ کشف تقلب در کارت‌های اعتباری با رویکرد هم‌جوشی است. در این پژوهش از تراکنش‌های یک بانک بزرگ برزیلی، در بازهٔ زمانی دوماهه، از۱۴ جولای تا ۱۲ سپتامبر سال ۲۰۰۴ استفاده شده که دلیل این امر، در دسترس نبودن مجموعه داده‌های سایر بانک‌هاست. همچنین، از داده‌های این بانک در مقاله‌های معتبر کشف تقلب استفاده شده و قابل دسترسی و ارجاع است. برای تهیهٔ این مجموعه داده‌ها، در دو مرحله نمونه‌گیری از تراکنش‌ها انجام شد و در نهایت اطلاعاتی شامل ۴۱۶۴۷ رکورد به‌دست آمد که ۳/۷۴ درصد آن‌ها تقلب‌آمیز بود. همچنین پایگاه داده در سه مرحله پیش پردازش شد:

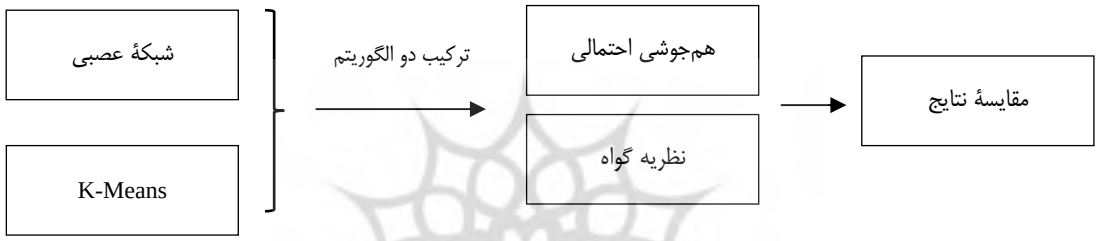
۱. برای تجزیه‌وتحلیل، متغیرهایی که برای مدل مهم نبودند حذف شدند که پس از این مرحله، از ۳۳ متغیر اولیه، ۱۷ متغیر مستقل و یک متغیر وابسته (برچسب تقلب) به‌دست آمد.
۲. همهٔ متغیرها به‌جز کد دستهٔ بازرگانی در حداکثر ۱۰ گروه یک رقمی مانند جدول ۳ طبقه‌بندی شدند.
۳. در آخر ۹ تقسیم‌بندی از پایگاه داده تولید شده که هر بخش شامل یه جفت داده است: ۷۰ درصد از تراکنش‌ها برای توسعه و ۳۰ درصد برای اعتبارسنجی مدل (گادی و همکاران، ۲۰۰۸).

برای این تحقیق ابتدا داده‌ها به دو روش شبکهٔ عصبی که رویکرد با سرپرستی است و الگوریتم K-means که رویکردی بدون سرپرستی است، به دو دسته تراکنش‌های قانونی و تقلب‌آمیز طبقه‌بندی شدند و تابع هزینه برای هر الگوریتم محاسبه شد. این تابع اولین‌بار در مقاله گادی و همکارانش در سال ۲۰۰۸ برای همین مجموعه داده‌ها معرفی و استفاده شد. پس از آن، در سطح تصمیم با استفاده از دو روش هم‌جوشی احتمالی و نظریهٔ گواه دمستر ـ شفر دو الگوریتم با هم ترکیب و نتایج مقایسه خواهد شد. گام‌های اجرای پژوهش در شکل ۲ مشاهده می‌شود.

1. Mansouri, Nabavi, Zare Ravasan & Ahangarbahan

جدول ۳. تعداد دسته‌ها برای هر متغیر، کلمه قبلی به‌معنای آخرین ارزش ساخته‌شده برای مشتری

نام متغیر	کد دسته بازرگانی[1]	کد دسته بازرگانی قبلی[2]	کد پستی[3]	کد پستی قبلی[4]	ارزش تراکنش[5]	ارزش تراکنش قبلی[6]
تعداد	۳۳	۳۳	۱۰	۱۰	۱۰	۱۰
نام متغیر	حالت ورودی پوز[7]	محدودیت اعتبار[8]	برند[9]	نوع[10]	امتیاز[11]	نوع شخص[12]
تعداد	۱۰	۱۰	۶	۶	۱۰	۲
نام متغیر	نوع تراکنش[13]	تعداد اعلان[14]	سرعت[15]	امتیاز تفاوت[16]	خط اعتبار[17]	برچسب تقلب[18]
تعداد	۲	۴	۸	۶	۹	۲



شکل ۲. گام‌های اجرای پژوهش

شبکهٔ عصبی

در این پژوهش از شبکهٔ عصبی پرسپترون سه‌لایه با یادگیری پس انتشار خطا استفاده کرده‌ایم. لایه‌های مخفی اول و دوم، هر یک از ۱۰۰ نورون ورودی تشکیل شده و برای فعال‌سازی، از تابع ReLU استفاده شده است؛ اما لایهٔ خروجی تک نورون و تابع فعال‌سازی آن تابع سیگموئید است. در اجرای شبکهٔ عصبی ۳۰۰ تکرار داریم و مراحل کلی اجرای این الگوریتم به‌صورت جدول ۴ است.

1. Merchant Category Code
2. Merchant Category Code previous
3. zip code
4. zip code previous
5. Value transaction
6. Value transaction previous
7. Pos entry mode
8. Credit limit
9. Brand
10. Variant
11. Score
12. Type person
13. Type of transaction
14. Number of statements
15. Speed
16. Diff score
17. Credit line
18. Flag fraud

جدول ٤. نحوهٔ اجرای الگوریتم شبکهٔ عصبی

نرمال‌سازی داده‌ها به‌صورت زیر:

رابطهٔ ۱)
$$a_{ij} = \frac{x_{ij} - \bar{x}}{\sigma}$$

که در آن x_{ij} عناصر هر ستون و $\bar{x}$ میانگین عناصر هر ستون و σ انحراف معیار عناصر هر ستون است.

۱. اضافه‌کردن لایهٔ ورودی و ساختن لایهٔ مخفی اول با ۱۰۰ نورون، توزیع یکنواخت و تابع فعال‌سازی ReLU

رابطهٔ ۲)
$$ReLU(x) = \begin{cases} 0 \; x \leq 0 \\ x \; x > 0 \end{cases}$$

۲. ساختن لایهٔ مخفی دوم با ۱۰۰ نورون، توزیع یکنواخت و تابع فعال‌سازی ReLU

۳. ساختن لایهٔ مخفی سوم با ۱۰۰ نورون، توزیع یکنواخت و تابع فعال‌سازی ReLU

۴. ساختن لایهٔ خروجی با ۱ نورون، توزیع یکنواخت و تابع سیگموئید

رابطهٔ ۳)
$$S(x) = \frac{1}{1+e^{-x}}$$

الگوریتم K-means

الگوریتم K-Means یک الگوریتم ساده، کارا و پیش‌رونده برای حل مدل خوشه‌بندی بر اساس نمایش استفاده شده توسط ویند است (مک‌کوئین[1]، ۱۹۶۷). این الگوریتم مجموعه‌های داده‌ای را به K زیرمجموعهٔ از پیش تعریف‌شده افراز می‌کند. در هر تکرار، همهٔ مشاهدات با همهٔ خوشه‌ها مقایسه شده و تخصیص بر اساس سنجش یک معیار فاصله صورت می‌گیرد (کتو و دشپاند[2]، ۲۰۱۴). ورودی‌های این الگوریتم، هیچ محدودیتی ندارند. هرچند ارزش‌های عددی در آن بهتر کار می‌کنند، لازم است تا این ارزش‌ها قبل از به‌کارگیری نرمال شوند (کتو و دشپاند، ۲۰۱۴). این الگوریتم به داده‌های پرت حساس است و به‌دلیل سنجش فاصله‌ای در فضای چندبُعدی، به‌سمت نقطهٔ بهینهٔ محلی حرکت می‌کند. بنابراین به نقطهٔ شروع حساس است و تضمینی برای پیدا کردن نقطهٔ بهینه جهانی در آن وجود ندارد (منصوری، زارع رواسان و غلامیان[3]، ۲۰۱۴). برای اجرای این الگوریتم نیز از داده‌های نرمال شده استفاده شده و از خروجی این الگوریتم، به‌عنوان ورودی الگوریتم K نزدیک‌ترین همسایه استفاده می‌شود. شبه‌کد این الگوریتم به صورت جدول ۵ است.

جدول ۵. مراحل اجرای الگوریتم K-means

۱.	ایجاد هسته‌های اولیهٔ تصادفی به تعداد K = ۲۵ خوشهٔ از پیش تعیین شده برای داده‌های تقلب‌آمیز و K = ۲۵ خوشه برای داده‌های قانونی و تشکیل ماتریس $C_{50\times18}$.
۲.	سنجش فاصلهٔ همهٔ بردارهای ماتریس آموزش با همهٔ هسته‌ها و اختصاص به نزدیک‌ترین خوشه بر اساس کمترین فاصله.
۳.	محاسبهٔ دوبارهٔ هسته‌های جدید با گرفتن میانگین از مشاهداتی که عضو خوشه هستند.
۴.	محاسبهٔ تابع هدف که همان جمع فواصل همهٔ خوشه‌هاست و مقایسهٔ آن با تکرار قبل در صورت کاهش، دوباره گام دو تکرار می‌شود و در صورت عدم تغییر، پایان الگوریتم خواهد بود.

1. MacQueen
2. Kotu & Deshpande
3. Mansouri, Ravasan & Gholamian

هم‌جوشی احتمالی

در این رویکرد هم‌جوشی داده‌ها بر اساس موجودیت احتمالی و جایگزین‌های ادغام چندگانه پردازش می‌شود (جاردات و همکاران، ۲۰۲۲). هم‌جوشی احتمالی به‌طورکلی بر قانون بیز تکیه دارد. مهم است توجه داشته باشیم که برای ترکیب دو توزیع احتمالی، آن‌ها باید روی یک چارچوب تشخیص قرار گیرند. دانش ناکامل در مورد کلاس $\omega \in \Omega$ بعد از رؤیت $x \in X$، به‌وسیلهٔ یک توزیع احتمالی روی Ω مدل می‌شود. اگر $p(\omega_i|x_1)$ و $p(\omega_i|x_2)$ توزیع‌های احتمالی روی Ω باشند که بعد از مشاهده داده‌های $x_1 \in X$ و $x_2 \in X$ اندازه‌گیری شده‌اند، با فرض استقلال شرطی داریم:

رابطهٔ ۴) $$p(x_1.x_2|\omega_i) = p(x_1|\omega_i)p(x_2|\omega_i) \;\; \forall i\epsilon\{1.2.\ldots.k\}$$

رابطهٔ ۴ احتمال وقوع هم‌زمان دو داده x_1 و x_2 را محاسبه می‌کند و طبق قانون بیز برای هر $i\epsilon\{1.2.\ldots.k\}$ خواهیم داشت:

رابطهٔ ۵) $$p(\omega_i|x_1.x_2) = \frac{p(\omega_i)p(x_1.x_2|\omega_i)}{p(x_1.x_2)}$$

همچنین با توجه به رابطهٔ ۴ خواهیم داشت:

رابطهٔ ۶) $$p(\omega_i|x_1.x_2) = \frac{p(\omega_i)}{p(x_1.x_2)}p(x_1|\omega_i)p(x_2|\omega_i)$$

و در نهایت رابطهٔ ۶ با توجه به قانون احتمال شرطی به‌صورت رابطهٔ ۷ قابل تعمیم است:

رابطهٔ ۷) $$p(\omega_i|x_1.x_2) = \frac{p(\omega_i)}{p(x_1.x_2)}\frac{p(x_1)p(x_1|\omega_i)}{p(\omega_i)}\frac{p(x_2)p(x_2|\omega_i)}{p(\omega_i)} = \frac{p(x_1)p(x_2)}{p(x_1.x_2)}\frac{p(\omega_i|x_1)p(\omega_i|x_2)}{p(\omega_i)}$$

با توجه به اینکه مقادیر $\frac{p(x_1)p(x_2)}{p(x_1.x_2)}$ مشخص و قابل محاسبه هستند، به‌عنوان ضریب ثابت در معادله محاسبه می‌شوند و رابطهٔ ۷ به رابطهٔ ۸ تبدیل خواهد شد.

رابطهٔ ۸) $$p(\omega_i|x_1.x_2) \propto \frac{p(\omega_i|x_1)p(\omega_i|x_2)}{p(\omega_i)}$$

در عمل ارزیابی توزیع کلاس اولیه $p(\omega_i)$ مشکل است و اغلب با یک توزیع یکنواخت جایگزین می‌شود (ژو و همکاران، ۲۰۱۶). روش اجرای الگوریتم هم‌جوشی احتمالی در این مقاله در جدول ۶ ارائه شده است.

جدول ٦. نحوهٔ اجرای هم‌جوشی احتمالی

۱. برای هر نمونه اعتبارسنجی، احتمال قانونی بودن و احتمال تقلب‌آمیز بودن را محاسبه می‌کنیم.

۲. با استفاده از دستور زیر ترکیب احتمالات تقلب‌آمیز و قانونی بودن را جداگانه محاسبه می‌کنیم:

رابطهٔ ۹) $$p(\omega_i|x_1.x_2)=\frac{p(x_1)p(x_2)}{p(x_1.x_2)}\frac{p(\omega_i|x_1)p(\omega_i|x_2)}{p(\omega_i)}$$

که با توجه به مستقل بودن دو پیشامد تقلب‌آمیز بودن و قانونی بودن تراکنش، حاصل کسر اول یک می‌شود. در کسر دوم $p(\omega_i)$ همان مقدار احتمال پیشین است که در گام اول محاسبه شده است.

۳. اگر احتمال قانونی بودن به‌دست‌آمده از رابطهٔ فوق بیشتر بود، برچسب قانونی و در غیر این صورت برچسب تقلب‌آمیز می‌زنیم و هزینه را محاسبه می‌کنیم.

نظریهٔ گواه

این نظریه به‌عنوان توسعه‌ای از نظریهٔ بیز در نظر گرفته می‌شود و امکان قضاوت بر اساس اطلاعات غیردقیق را می‌دهد و کمک می‌کند تا درجهٔ احتمال به‌جای نقطه‌ای با بازه سنجیده شود. این نظریه از مفهوم حد بالایی و پایینی احتمال با به‌کارگیری یک نگاشت چندگانه سرچشمه می‌گیرد (ین، ۱۹۹۰). اجزای این نظریه عبارت‌اند از: چارچوب تشخیص[1]، توابع انتساب باور اولیه[2]، تابع باور[3]، تابع مقبولیت[4]، قاعدهٔ ترکیب دمستر که در ادامه به توضیح هر یک می‌پردازیم (آهنگربهان و منتظر، ۱۳۹۵):

چارچوب تشخیص مجموعه‌ای ثابت از گزاره‌های کامل و منحصربه‌فرد در بررسی مسئله است؛ به‌طوری‌که این گزاره‌ها، چارچوبی را برای پوشش همه گزاره‌های درگیر در مسئله ارائه می‌کند. مجموعه چارچوب دید را با نماد Θ نمایش می‌دهند (گوآن و بل[5]، ۱۹۹۱):

رابطهٔ ۱۰) $$\Theta=\{\theta_1,\theta_2,\dots,\theta_N\}$$

θ_i ها حالت‌های ممکن رخداد در مسئله هستند و منظور از تشخیص این است که امکان تشخیص یک پاسخ صحیح از میان همهٔ پاسخ‌های ممکن وجود دارد. با توجه به موضوع تحقیق، چارچوب تشخیص این مسئله، شامل چهار حالت تراکنش «قانونی»، «تقلب‌آمیز»، «قانونی یا تقلب‌آمیز» و «هیچ‌کدام» است.

انتساب باور پایه را با استفاده از نگاشتی مانند m برای بیان باور خود دربارهٔ یک گزاره، به شکل عددی در بازهٔ [۰،۱] تعریف می‌کنیم (خطیبی و منتظر، ۲۰۱۰):

رابطهٔ ۱۱) $$\mathrm{m}: 2^{\Theta}\rightarrow[0,1]$$

این تابع دارای خواص زیر است:

1. Frame of Discernment
2. Basic Probability Assignment Function
3. Belief Function
4. Plausibility Function
5. Guan & Bell

۱. برای هر یک از گزاره‌ها، باور می‌تواند مقداری بزرگ‌تر یا مساوی صفر داشته باشد؛ یعنی:

رابطهٔ ۱۲) $\forall A \in 2^{\theta}: \mathrm{m}(\mathrm{A}) \geq 0$

۲. هیچ درجه‌ای از باور برای گزارهٔ تهی در نظر گرفته نمی‌شود؛ یعنی:

رابطهٔ ۱۳) $\mathrm{m}(\emptyset) = 0$

۳. مجموع کل باور برابر با یک است؛ یعنی:

رابطهٔ ۱۴) $\sum_{A \in 2^{\theta}} \mathrm{m}(\mathrm{A}) = 1$

درواقع انتساب باور اولیه فقط برای تعداد محدودی از مجموعه‌ها به نام «عنصر مرکزی» غیر صفر است (الآنی و دریچه[1]، ۲۰۰۲).

تابع باور نشان‌دهندهٔ درجه باور کامل به گزاره‌ای است که نسبت به آن اطمینان وجود دارد. در این تابع، تنها گزاره‌هایی وارد می‌شود که به شکل کامل مؤید گزاره مدنظر است و به‌شکل زیر تعریف می‌شود (بائه و همکاران، ۲۰۰۴):

رابطهٔ ۱۵) $\mathrm{bel}(\mathrm{A}) = \sum_{B \subseteq A} m(B)$

$\mathrm{bel}: 2^{\theta} \rightarrow [0,1]$

که در آن m(B) بخشی از باور کامل منتسب به گزاره A است. بنابراین، در این مسئله برای محاسبهٔ تابع باور تراکنش «قانونی» و «تقلب‌آمیز»، لازم است انتساب باور اولیهٔ هر یک به تنهایی در نظر گرفته شود؛ اما در خصوص تابع باور «قانونی یا تقلب‌آمیز» باید مقادیر تابع انتساب اولیه تراکنش «قانونی»، «تقلب‌آمیز» و «قانونی یا تقلب‌آمیز» با هم جمع شوند.

تابع مقبولیت نیز بیانگر درجهٔ مقبولیت یک گزاره است. درواقع درجه مقبولیت یک گزاره، به‌معنای حد بیشینه وقوع آن است. از آنجا که هر گزاره‌ای که با گزارهٔ مدنظر اشتراک دارد، حداقل بخشی از آن گزاره را پوشش می‌دهد و بدین ترتیب، درجه باور گزاره دارای اشتراک، بر امکان‌پذیری و مقبولیت گزاره مدنظر دلالت می‌کند. این تابع به‌صورت زیر تعریف می‌شود (گوآن و بل، ۱۹۹۱):

رابطهٔ ۱۶) $\mathrm{pls}(\mathrm{A}) = \sum_{B \cap A \neq \emptyset} m(B)$

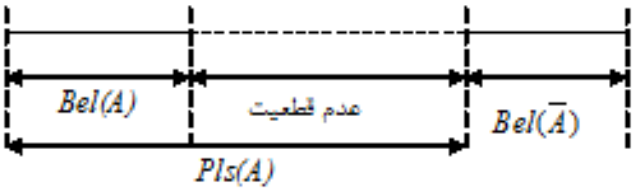
$\mathrm{pls}: 2^{\theta} \rightarrow [0,1]$

در صورت فقدان اطلاعات و دانش کافی دربارهٔ مسئله، بسیار معقول است که به‌جای ارائه یک مقدار معین برای متغیر نامطمئن (همانند آنچه در نظریه احتمال انجام می‌شود) بازه‌ای برای محدودهٔ عدم قطعیت ارائه شود. با توجه به تعریف تابع باور، این تابع حداقل درجه گواه کامل گزاره را که موجود است و دربارهٔ آن اطمینان وجود دارد، ارائه می‌کند.

1. Al-Ani & Deriche

اما تابع مقبولیت، حداکثر درجه باور گزاره را بیان می‌کند که به‌صورت بالقوه می‌تواند داشته باشد. بدین ترتیب، مقدار متغیر عدم قطعی (درجه باور) بین دو کران پایین و بالا که به‌ترتیب تابع باور و تابع مقبولیت هستند، محصور می‌شود. همان‌طور که در شکل ۳ ملاحظه می‌شود، درجه باور گزاره فرضی A با بازه زیر ارائه می‌شود. با توجه به اینکه مقادیر توابع باور و مقبولیت در بازه [۰،۱] هستند، درجه باور گزاره نیز در همین بازه خواهد بود (خطیبی و منتظر، ۲۰۱۰).

رابطهٔ ۱۷) $$I(A) = [bel(A), pls(A)]$$



شکل ۳. توابع باور و مقبولیت یک گزاره

اگرچند تابع انتساب گواه پایه برای چارچوب دید مسئله ارائه شود، باید آن‌ها را باهم ترکیب کرد. این ترکیب به ایجاد تابع انتساب گواه پایهٔ جدیدی منجر می‌شود که حاوی معتبرترین و کامل‌ترین اطلاعات است (گوآن و بل، ۱۹۹۱). دو گواه m1 و m2 را که از طریق دو منبع اطلاعاتی مستقل برای مسئله حاصل‌شده است، می‌توان طبق قاعدهٔ دمستر به‌صورت زیر با یکدیگر ترکیب کرد (ین، ۱۹۹۰):

رابطهٔ ۱۸) $$m_{1,2}(A) = m_1(A) \oplus m_2(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{\sum_{A_i\cap B_j\neq\emptyset} m_1(A_i)m_2(B_j)} = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-\sum_{A_i\cap B_j=\emptyset} m_1(A_i)m_2(B_j)}$$

که در آن، هر یک از B و C به گزاره‌هایی از یک منبع مستقل اشاره دارد. در مسئله حاضر، m1 از الگوریتم شبکهٔ عصبی و m2 از الگوریتم k-means به‌دست آمده‌اند. با توجه به مسئلهٔ تحقیق با در نظر گرفتن L برای تراکنش قانونی، F برای تراکنش تقلب‌آمیز و $L \cup F$ برای تراکنش قانونی یا تقلب‌آمیز، محاسبهٔ ترکیب دو منبع مستقل m1 و m2 به‌روش دمستر ـ شفر به‌صورت رابطهٔ ۱۹ است.

رابطهٔ ۱۹) $$m_{1,2}(L) = \frac{\sum_{i\cap j=L} m_1(i)m_2(j)}{1-\sum_{i\cap j=\emptyset} m_1(i)m_2(j)} = \frac{m_1(L)m_2(L)+m_1(L)m_2(L\cup F)+m_1(L\cup F)m_2(L)}{1-(m_1(L)m_2(F)+m_1(F)m_2(L))}$$

اعتبارسنجی مدل

با توجه به مسئله حاضر، یعنی کشف تقلب، ماتریس درهم‌ریختگی آن به‌صورت جدول ۷ است.

جدول۷. ماتریس درهم‌ریختگی

	واقعی		
	کل جامعه	شرایط درست	شرایط نادرست
پیش‌بینی	پیش‌بینی مثبت	اعلان درست[1] (TP)	اعلان نادرست[2](FP)
	پیش‌بینی منفی	عدم اعلان نادرست[3](FN)	عدم اعلان درست[4](TN)

گادی در مصاحبه با خبرگان جلوگیری از تقلب و همچنین مشاهدهٔ نتایج اولیهٔ منحنی ROC متوجه شد که در صورت استفاده از تابع هزینهٔ فوق، نتایج کاربردی‌تری به‌دست خواهد آورد. در تابع فوق، هزینهٔ متوسط ۱ دلار برای هر تأیید و متوسط ۱۰۰ دلار برای تقلب کشف نشده در نظر گرفته شده است:

رابطهٔ ۲۰) $$cost = \$100 \times FN + \$1 \times (FP + TP)$$

این تابع هزینه، شاخص‌های دقت و صحت کشف تقلب را هم‌زمان در نظر می‌گیرد. با توجه به اینکه در دیتابیس اولیه ۱۰۰ درصد از داده‌های تقلب‌آمیز و ۱۰ درصد از داده‌های قانونی در نظر گرفته شده‌اند، تابع هزینه نهایی به صورت زیر تعدیل شده است (گادی و همکاران، ۲۰۰۸):

رابطهٔ ۲۱) $$cost = 100 \times FN + 10 \times FP + TP$$

یافته‌های پژوهش

در این بخش نتایج الگوریتم‌های به‌کار رفته به تفصیل بررسی خواهند شد. در گام اول، الگوریتم شبکهٔ عصبی در نرم‌افزار پایتون اجرا و هزینهٔ هر بخش از مجموعه دیتا، به‌صورت جدول ۸ محاسبه شد.

جدول ۸. نتایج نهایی اجرای الگوریتم شبکهٔ عصبی در هر بخش

مجموعه داده‌ها	مجموعه داده ۱	مجموعه داده ۲	مجموعه داده ۳	مجموعه داده ٤	مجموعه داده ٥	مجموعه داده ٦	مجموعه داده ۷	مجموعه داده ۸	مجموعه داده ۹
هزینه	۲۶۲۳۶	۲۵۲۶۴	۲۵۷۶۵	۲۸۰۵۴	۲۵۷۶۰	۲۵۲۲۳	۲۳۳۹۸	۲۴۳۹۴	۲۳۹۵۳

1. True Positive
2. False Positive
3. False Negative
4. True Negative

در گام بعدی، الگوریتم خوشه‌بندی K-means اجرا شده است که در آن، ۵۰ خوشه به‌دست آمده را به‌عنوان داده آموزش ورودی به الگوریتم K نزدیک‌ترین همسایه می‌دهیم تا طبقه‌بندی بر اساس این الگوریتم انجام شود و نتایج نهایی را در جدول ۹ ذخیره می‌کنیم.

جدول ۹. نتایج نهایی اجرای الگوریتم K-means در هر بخش

مجموعه داده‌ها	مجموعه داده ۱	مجموعه داده ۲	مجموعه داده ۳	مجموعه داده ٤	مجموعه داده ٥	مجموعه داده ٦	مجموعه داده ۷	مجموعه داده ۸	مجموعه داده ۹
هزینه	۳۱۶۵۰	۲۹۶۱۳	۳۰۰۵۲	۳۲۲۱۰	۳۲۱۶۶	۳۰۴۱۴	۳۱۲۳۵	۳۱۷۹۷	۳۰۴۰۶

در مرحلهٔ آخر دو منبع داده‌ای ورودی متفاوت طبقه‌بندی و خوشه‌بندی برای الگوریتم‌های هم‌جوشی در سطح تصمیم‌گیری با دو روش هم‌جوشی احتمالی و هم‌جوشی به‌روش دمستر ـ شفر ترکیب می‌شود.

برای اجرای هم‌جوشی احتمالی، در هر بخش که الگوریتم‌ها اجرا شده‌اند، احتمالات هر برچسب‌گذاری را نیز ذخیره کرده‌ایم. از طرفی، برای محاسبهٔ احتمالات پیشین، در هر نمونه آموزش احتمال صفر بودن و احتمال یک بودن را نیز محاسبه کرده‌ایم. نتایج حاصل از اجرای این الگوریتم، در جدول ۱۰ گزارش شده است.

جدول ۱۰. نتایج نهایی اجرای الگوریتم هم‌جوشی احتمالی در هر بخش

مجموعه داده‌ها	مجموعه داده ۱	مجموعه داده ۲	مجموعه داده ۳	مجموعه داده ٤	مجموعه داده ٥	مجموعه داده ٦	مجموعه داده ۷	مجموعه داده ۸	مجموعه داده ۹
هزینه	۱۹۸۰۱	۲۰۰۳۹	۲۰۵۰۹	۲۰۹۴۸	۱۸۷۳۱	۱۹۸۳۹	۲۰۰۸۷	۱۹۳۳۲	۱۹۹۳۷

سپس هم‌جوشی به‌روش نظریه گواه انجام شد. در این بخش با توجه به اینکه مجموعه چارچوب تصمیم این مسئله از چهار عضو تراکنش قانونی، تراکنش تقلب‌آمیز، تراکنش قانونی یا تقلب‌آمیز و هیچ‌کدام تشکیل شده است و پیش از این ذکر کردیم که احتمال انتساب باور اولیه «هیچ‌کدام»، صفر در نظر گرفته می‌شود و با توجه به فرض استقلال دو پیشامد «تراکنش تقلب‌آمیز» و «تراکنش قانونی»، برای پیشامد «تقلب‌آمیز یا قانونی» حاصل‌ضرب آن دو را قرار دادیم. همچنین با قاعدهٔ ترکیب دمستر، تابع انتساب باورهای اولیه را ترکیب کردیم و با توجه به بزرگ‌ترین عدد به‌دست آمده، به تراکنش‌ها برچسب قانونی یا تقلب‌آمیز زدیم و هزینه را محاسبه کردیم. هزینهٔ محاسبه شده در هر مجموعه داده در جدول ۱۱ گزارش شده است.

جدول ۱۱. نتایج نهایی اجرای نظریه گواه در هر بخش

مجموعه داده‌ها	مجموعه داده ۱	مجموعه داده ۲	مجموعه داده ۳	مجموعه داده ٤	مجموعه داده ٥	مجموعه داده ٦	مجموعه داده ۷	مجموعه داده ۸	مجموعه داده ۹
هزینه	۲۵۲۳۸	۲۵۷۰۳	۲۴۷۱۱	۲۷۴۱۳	۲۵۳۴۲	۲۴۹۵۹	۲۳۴۵۹	۲۳۹۶۱	۲۳۷۶۸

با توجه به جدول‌های ۱۰ و ۱۱ که جمع‌بندی آن‌ها در جدول ۱۲ آمده است، می‌توان نتیجه گرفت که در مقایسه با نظریهٔ گواه دمستر ـ شفر، هم‌جوشی احتمالی هزینهٔ کمتر و درنتیجه صحت بیشتری در تشخیص نشان می‌دهد.

جدول ۱۲. مقایسهٔ نتایج نظریهٔ گواه و هم‌جوشی احتمالی در هر بخش

شماره مجموعه	۱	۲	۳	٤	٥	٦	۷	۸	۹
هم‌جوشی احتمالی	۱۹۸۰۱	۲۰۰۳۹	۲۰۵۰۹	۲۰۹۴۸	۱۸۷۳۱	۱۹۸۳۹	۲۰۰۸۷	۱۹۳۳۲	۱۹۹۳۷
نظریه گواه	۲۵۲۳۸	۲۵۷۰۳	۲۴۷۱۱	۲۷۴۱۳	۲۵۳۴۲	۲۴۹۵۹	۲۳۴۵۹	۲۳۹۶۱	۲۳۷۶۸

در نهایت، میانگین هزینهٔ محاسبه شده ۹ مجموعه برای الگوریتم شبکهٔ عصبی معادل ۲۵۳۳۸ و برای الگوریتم K-means برابر با ۳۱۰۶۰ است. برای کاهش این هزینه، از دو الگوریتم هم‌جوشی احتمالی و نظریهٔ گواه دمستر ـ شفر استفاده شد که نتیجه آن در جدول ۱۳ نشان داده شده است.

جدول ۱۳. مقایسهٔ نتایج الگوریتم‌های مختلف

الگوریتم‌ها	شبکهٔ عصبی	K-means	هم‌جوشی احتمالی	نظریه گواه
میانگین هزینه‌ها	۲۵۳۳۸	۳۱۰۶۰	۱۹۹۱۴	۲۴۹۵۰

همان طور که در این جدول مشاهده می‌شود، هر دو روش هم‌جوشی به کاهش هزینه منجر شده‌اند؛ اما بهترین هزینه با روش هم‌جوشی احتمالی به‌دست آمده است. دلیل آن نیز کاهش شاخص TN (عدم اعلان درست) در هم‌جوشی احتمالی است، در حالی‌که در نظریهٔ گواه این شاخص نسبت به شبکهٔ عصبی افزایش داشته است. از طرفی شاخص TP (اعلان درست) در هم‌جوشی احتمالی، افزایش چشمگیری داشته است که بهبود عملکرد سیستم تشخیص تقلب را نشان می‌دهد. از طرفی کاهش شاخص FN (عدم اعلان نادرست) در هم‌جوشی احتمالی، به این معناست که سیستم کشف تقلب تعداد کمتری از تراکنش‌های تقلب‌آمیز را به اشتباه قانونی اعلام می‌کند که در کاهش هزینهٔ خطا به‌شدت تأثیرگذار است؛ اما افزایش مقدار شاخص FP در هم‌جوشی احتمالی به نشانه اعلان تقلب نادرست است که این افزایش برای سیستم مناسب نیست. در نهایت، با توجه به هزینه‌های محاسبه شده، هم‌جوشی احتمالی در سطح تصمیم‌گیری، در مقایسه با نظریهٔ گواه و مقاله مرجع عملکرد بهتری دارد. مقایسهٔ نتایج در جدول ۱۴ مشاهده می‌شود.

جدول ۱٤. مقایسهٔ نتایج هم‌جوشی در سطح تصمیم‌گیری

روش‌ها	AIS	هم‌جوشی احتمالی	نظریه گواه
میانگین هزینه‌ها	۲۳۳۰۳	۱۹۹۱۴	۲۴۹۵۰

نتیجه‌گیری و پیشنهادها

هدف این پژوهش کاهش هزینهٔ تشخیص کشف تقلب در کارت‌های اعتباری با رویکرد هم‌جوشی اطلاعات بود. در این

راستا، مدل‌ها و الگوریتم‌های مختلفی ارائه شده در این حوزه مرور شدند و در خصوص ضعف‌ها و قوت‌های هر یک بحث کردیم. در این تحقیق از الگوریتم طبقه‌بندی شبکهٔ عصبی پرسپترون سه‌لایه با یادگیری پس انتشار خطا و الگوریتم خوشه‌بندی K-means به‌عنوان منابع داده‌ای مستقل برای هم‌جوشی در سطح تصمیم‌گیری به دو روش هم‌جوشی احتمالی و نظریهٔ گواه دمستر ـ شفر استفاده کردیم و نتایج این هم‌جوشی‌ها را در جدول ۱۲ نشان دادیم. به‌دلیل عدم انجام مقایسه در روش‌های مختلف هم‌جوشی در سطح تصمیم‌گیری در ادبیات تحقیق، در این مطالعه به این مبحث پرداختیم.

با توجه به پیشینهٔ تحقیق، تنها در نظر گرفتن منبع تاریخی مالک تراکنش برای استنتاج نهایی، هزینهٔ پایینی در طبقه‌بندی ارائه نمی‌دهد. این هزینهٔ بالا، از ذات داده‌های تراکنش‌های کارت اعتباری نشئت می‌گیرد که باعث عدم اطمینان در تصمیم نهایی می‌شود. بنابراین مدل باید از دو یا چند منبع مستقل و رویکردی برای هم‌جوشی نتایج استفاده کند. برای این منظور، یک الگوریتم خوشه‌بندی K-means با الگوریتم طبقه‌بندی شبکهٔ عصبی به‌عنوان منابع مستقل برای هم‌جوشی در نظر گرفته شدند.

در مقالهٔ گادی و همکاران (۲۰۰۸) بهترین روش، روش سیستم ایمنی مصنوعی، در بین سایر الگوریتم‌های به‌کار رفته، کمترین هزینه را داشته است. با توجه به مقایسهٔ نتایج دو الگوریتم هم‌جوشی احتمالی و نظریهٔ گواه با روش سیستم ایمنی مصنوعی مقاله مرجع که در جدول ۱۴ گزارش شده، مشاهده شد که هم‌جوشی احتمالی، کاهش هزینهٔ چشمگیری نسبت به مقاله مرجع داشته است. بنابراین برای به‌کارگیری این الگوریتم‌ها به‌منظور کاهش هزینه کشف تقلب کارت‌های اعتباری، هم‌جوشی احتمالی در سطح تصمیم‌گیری توصیه می‌شود.

اصلی‌ترین محدودیت این پژوهش، فقدان داده‌های در دسترس بانک‌های داخل کشور برای پژوهش در حوزهٔ کشف تقلب است که مانع اجرای الگوریتم این پژوهش در فضای واقعی ایران شد و مقایسه را با مطالعات مشابه بین‌المللی غیرممکن ساخت. دلایل امنیتی و کسب‌وکاری، به همراه رقابتی بودن راه‌حل‌های کشف تقلب بر این محدودیت بسیار تأثیر گذاشته است. از طرفی نبود یک مجموعه داده‌ای استاندارد برای مقایسهٔ کارایی الگوریتم‌ها و مدل‌ها نیز، چالش اساسی در حوزهٔ کشف تقلب است که باعث جلوگیری از اعتبارسنجی الگوریتم‌ها و مدل‌ها می‌شود. در سطح کاربردی، این تحقیق می‌تواند مبنای استفاده در بانک‌ها و مؤسسه‌های اعتباری در داخل کشور شود. پیشنهاد می‌شود که به‌منظور رفع محدودیت در استفاده از این پژوهش، مجموعه داده‌ای از بانک‌ها طراحی، ایجاد و بازآزمایی شود که آیا متغیرهای رفتاری و فرهنگی، اثر تعدیل‌کننده‌ای روی نتایج الگوریتم‌ها دارند یا خیر. همچنین می‌توان تابع هزینهٔ جدیدی متناسب با مجموعه داده‌های بانک‌های داخلی تعریف کرد.

محدودیت تأثیرگذار دیگر در این حوزه، کمبود منابع آکادمیک و مطالعات پیشین در کشف تقلب با رویکرد هم‌جوشی اطلاعات در هنگام مطالعه است.

در مطالعات آتی، می‌توان بر روش‌های دیگر هم‌جوشی منابع داده‌ای مستقل با استنتاج‌های متعارض تمرکز کرد و به مقایسهٔ نتایج آن با پژوهش حاضر پرداخت. استفاده از سایر الگوریتم‌های طبقه‌بندی و خوشه‌بندی، به‌منظور دستیابی

به نتایج بهتر، پیشنهاد می‌شود. همچنین هم‌جوشی الگوریتم‌ها در سطح داده‌ای و مقایسهٔ آن با پـژوهش حاضـر توصـیه می‌شود.

منابع

آهنگربهان، حمید و منتظر، غلامعلی (۱۳۹۵). طراحی سامانه تشخیص دستبرد ادبی جمله بنیاد در متون فارسی به کمک هم‌جوشی گواه‌ها. *پردازش علایم و داده‌ها*، ۱ (۲۷)، ۷۱-۸۵.

وثوق، ملیحه؛ تقوی‌فرد، محمدتقی و البرزی، محمود (۱۳۹۳). شناسایی تقلب در کارت های بانکی با استفاده از شبکه های عصـبی مصنوعی. *مدیریت فناوری اطلاعات*، ۶(۴)، ۷۲۱- ۷۴۶.

References

Abdallah, A., Maarof, M. A. & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, *68*, 90-113.

Adibi, M. A. & Shahrabi, J. (2015). Online anomaly detection based on support vector clustering. *International Journal of Computational Intelligence Systems*, *8*(4), 735-746.

Ahangarbahan, H. & Montazer, Gh.A. (2016). Design A Sentence Based Plagiarism Detection System by Evidences Fusion in Persian Text. *Signal and Data Processing*, 1(27), 71-85. (*in Persian*)

Akila, S. & Reddy, U. S. (2018). Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection. *Journal of computational science*, *27*, 247-254.

Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science*, *14*(3), 553-569.

Al-Ani, A. & Deriche, M. (2002). A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence. *Journal of Artificial Intelligence Research*, *17*, 333-361.

Aral, K. D., Güvenir, H. A., Sabuncuoğlu, İ. & Akar, A. R. (2012). A prescription fraud detection model. *Computer methods and programs in biomedicine*, *106*(1), 37-46.

Awasthi, A. & Chauhan, S. S. (2011). Using AHP and Dempster–Shafer theory for evaluating sustainable transport solutions. *Environmental Modelling & Software*, *26*(6), 787-796.

Awoyemi, J. O., Adetunmbi, A. O. & Oluwadare, S. A. (2017, October). Credit card fraud detection using machine learning techniques: A comparative analysis. In *2017 international conference on computing networking and informatics (ICCNI)* (pp. 1-9). IEEE.

Bae, H. R., Grandhi, R. V. & Canfield, R. A. (2004). An approximation approach for uncertainty quantification using evidence theory. *Reliability Engineering & System Safety*, *86*(3), 215-225.

Bahnsen, A. C., Aouada, D., Stojanovic, A. & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications*, *51*, 134-142.

Bhatia, S., Bajaj, R. & Hazari, S. (2016). Analysis of credit card fraud detection techniques. *International Journal of Science and Research*, *5*(3), 1302-1307.

Boström, H., Andler, S. F., Brohede, M., Johansson, R., Karlsson, A., Van Laere, J., ... & Ziemke, T. (2007). On the definition of information fusion as a field of research.

Brause, R., Langsdorf, T. & Hepp, M. (1999, November). Neural data mining for credit card fraud detection. In *Proceedings 11th international conference on tools with artificial intelligence* (pp. 103-106). IEEE.

Carneiro, N., Figueira, G. & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. *Decision Support Systems*, *95*, 91-101.

Chang, R. I., Lai, L. B., Su, W. D., Wang, J. C. & Kouh, J. S. (2007). Intrusion detection by backpropagation neural networks with sample-query and attribute-query. *International Journal of Computational Intelligence Research*, *3*(1), 6-10.

Duman, E. & Ozcelik, M. H. (2011). Detecting credit card fraud by genetic algorithm and scatter search. *Expert Systems with Applications*, *38*(10), 13057-13063.

Fiore, U., De Santis, A., Perla, F., Zanetti, P. & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, *479*, 448-455.

Gadi, M. F. A., Wang, X. & do Lago, A. P. (2008, December). Comparison with parametric optimization in credit card fraud detection. In *2008 Seventh International Conference on Machine Learning and Applications* (pp. 279-285). IEEE.

Ghobadi, F. & Rohani, M. (2016, December). Cost sensitive modeling of credit card fraud using neural network strategy. In *2016 2nd international conference of signal processing and intelligent systems (ICSPIS)* (pp. 1-5). IEEE.

Gravina, R., Alinia, P., Ghasemzadeh, H. & Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, *35*, 68-80.

Guan, J. W. & Bell, D. A. (1991). Evidential reasoning and its applications. North-Holand.

Gupta, S., Malsa, N. & Gupta, M. V. (2017). Credit card fraud detection and prevention—a survey. *International Journal for Innovative Research in Science & Technology*, *4*, 1-7.

Hadavandi, E., Shahrabi, J. & Hayashi, Y. (2016). SPMoE: a novel subspace-projected mixture of experts model for multi-target regression problems. *Soft Computing*, *20*, 2047-2065.

Halvaiee, N. S. & Akbari, M. K. (2014). A novel model for credit card fraud detection using Artificial Immune Systems. *Applied soft computing*, *24*, 40-49.

He, W., Williard, N., Osterman, M. & Pecht, M. (2011). Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method. *Journal of Power Sources*, *196*(23), 10314-10321.

Hormozi, H., Akbari, M. K., Hormozi, E. & Javan, M. S. (2013, May). Credit cards fraud detection by negative selection algorithm on hadoop (To reduce the training time). In *The 5th Conference on Information and Knowledge Technology* (pp. 40-43). IEEE.

Jaradat, A., Safieddine, F., Deraman, A., Ali, O., Al-Ahmad, A. & Alzoubi, Y. I. (2022). A probabilistic data fusion modeling approach for extracting true values from uncertain and conflicting attributes. *Big Data and Cognitive Computing*, *6*(4), 114.

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L. & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert systems with applications*, *100*, 234-245.

Khatibi, V. & Montazer, G. A. (2010). A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. *Expert Systems with Applications*, *37*(12), 8536-8542.

Kotu, V. & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann.

Kumar, K. & Rao, P. (2013). A Valuable Progress on the Way to Credit Card Deception Revelation System. *International Journal of Computer and Electronic research*.

Lu, X. Y., Chu, X. Q., Chen, M. H. & Chang, P. C. (2015). Data Analytics for Bank Term Deposit by Combining Artificial Immune Network and Collaborative Filtering. In *Proceedings of the ASE BigData & SocialInformatics 2015* (pp. 1-6).

MacQueen, J. (1967, January). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (Vol. 5, pp. 281-298). University of California press.

Maes, S., Tuyls, K., Vanschoenwinkel, B. & Manderick, B. (2002, January). Credit card fraud detection using Bayesian and neural networks. In *Proceedings of the 1st international naiso congress on neuro fuzzy technologies* (Vol. 261, p. 270).

Malini, N. & Pushpa, M. (2017). Analysis on credit card fraud detection techniques by data mining and big data approach. *International journal of research in computer applications and robotics*, *5*(5), 38-45.

Mansouri, T., Nabavi, A., Zare Ravasan, A. & Ahangarbahan, H. (2016). A practical model for ensemble estimation of QoS and QoE in VoIP services via fuzzy inference systems and fuzzy evidence theory. *Telecommunication Systems*, *61*, 861-873.

Mansouri, T., Ravasan, A. Z. & Gholamian, M. R. (2014). A novel hybrid algorithm based on k-means and evolutionary computations for real time clustering. *International Journal of Data Warehousing and Mining (IJDWM)*, *10*(3), 1-14.

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y. & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, *50*(3), 559-569.

Panigrahi, S., Kundu, A., Sural, S. & Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning. *Information Fusion*, *10*(4), 354-363.

Phua, C., Lee, V., Smith, K. & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

Popescu, D. E., Lonea, M., Zmaranda, D., Vancea, C. & Tiurbe, C. (2010). Some aspects about vagueness & imprecision in computer network fault-tree analysis. *International Journal of Computers Communications & Control*, *5*(4), 558-566.

Sherly, K. K. (2012). A comparative assessment of supervised data mining techniques for fraud prevention. *International Journal of Science and Technology*, *1*(16).

SamanehSorournejad, Z. Z., Atani, R. E. & Monadjemi, A. H. (2016). A survey of credit card fraud detection techniques: Data and technique oriented perspective. *arXiv preprint ArXiv:1611.06439 [Cs]*.

Stolfo, S., Fan, D. W., Lee, W., Prodromidis, A. & Chan, P. (1997, July). Credit card fraud detection using meta-learning: Issues and initial results. In *AAAI-97 Workshop on Fraud Detection and Risk Management* (Vol. 83).

Tabassian, M., Ghaderi, R. & Ebrahimpour, R. (2012). Combination of multiple diverse classifiers using belief functions for handling data with imperfect labels. *Expert systems with applications*, *39*(2), 1698-1707.

Tripathi, K. K. & Pavaskar, M. A. (2012). Survey on credit card fraud detection methods. *International Journal of Emerging Technology and Advanced Engineering*, *2*(11), 721-726.

Vosough, M., Taghavi Fard, M.T. & Alborzi, M. (2015). Bank Card Fraud Detection Using Artificial Neural Network. *Journal of Information Technology Management*, 6(4), 721-746. (*in Persian*)

Xu, P., Davoine, F., Bordes, J. B., Zhao, H. & Denœux, T. (2016). Multimodal information fusion for urban scene understanding. *Machine Vision and Applications*, *27*, 331-349.

Yen, J. (2002). Generalizing the Dempster-Schafer theory to fuzzy sets. *IEEE Transactions on Systems, man, and Cybernetics*, *20*(3), 559-570.

Zareapoor, M. & Shamsolmoali, P. (2015). Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia computer science*, *48*(2015), 679-685.

Zareapoor, M., Seeja, K. R. & Alam, M. A. (2012). Analysis on credit card fraud detection techniques: based on certain design criteria. *International journal of computer applications*, *52*(3).