https://danesh24.um.ac.ir

**Research Article**

**Vol. 29, No. 2, 2023, p. 327-358**

# Detection of Money Laundering Activities in Financial Transactions by Using Data Mining Methods, Benford's Law and GANs algorithm

**A. Kazemi [1]*, A. Moghadamfalahi [2], A. Abdali [3], S. Aryaee [4]**

1- Professor of Operations Management and Decision Sciences, Faculty of Management, University of Tehran, Tehran, Iran
2- MSc in Industrial Management, Faculty of Management, University of Tehran, Tehran, Iran
3- Assistant Professor, Amin Police University, Tehran, Iran
4- Ph. D Student of Industrial Management, Faculty of Management, University of Tehran, Tehran, Iran

(*- Corresponding Author Email: aliyehkazemi@ut.ac.ir)
https://doi.org/10.22067/mfe.2023.72676.1117
ORCID ID: 0000-0002-0755-7800

| | |
|---|---|
| Received:2021/09/27<br>Revised:2023/02/19<br>Accepted:2023/02/27 | **How to cite this article:**<br>Kazemi, A.; Moghadamfalahi, A., Abdali, A., & Aryaee S. (2023). Detection of Money Laundering Activities in Financial Transactions by Using Data Mining Methods, Benford's Law and GANs algorithm. *Quarterly Monetary & Financial Economics*, *29(2):* 327-358. (in Persian with English abstract).<br>https://doi.org/10.22067/mfe.2023.72676.1117 |
| Available Online: 2023/02/27 | |

## 1- INTRODUCTION

Nowadays, as a result of advancements in technology, criminals use a broad range of sophisticated fraudulent activities to gain remarkable financial profits. Therefore, the detection and identification of these activities are becoming a complex problem. One of the fraudulent activities that experienced significant developments during recent years is money laundering. Money laundering is the process of making cash that is earned

from illicit activities appear to be earned from legal activities. Since money laundering has several negative effects on the economy than, anti-money laundering activities could seriously bolster the situation of economic and encourage legitimate investment opportunities. As traditional approaches for anti-money laundering activities need enormous person-hour work, these approaches need to be supported by automated tools. Machine learning and data mining techniques could be efficiently and appropriately used for the detection of money laundering. This research aims to investigate the effectiveness of these tools in detecting money laundering activities by applying data mining methods, Benford's law, and GANs algorithm.

## 2- THEORETICAL FRAMEWORK

Money laundering is a process of disguising the origin and ownership of illegal proceeds by integration them into the legitimate financial system. It involves three stages: placement, layering, and integration. The placement stage involves the introduction of illegal proceeds into the financial system, while the layering stage involves creating multiple transactions to conceal the origin of the funds. The integration stage involves using the laundered funds for legitimate purposes. Data mining methods, Benford's law, and GANs algorithm have been proposed as effective techniques for detecting money laundering activities in each of these stages.

## 3- METHODOLOGY

The methodology used in this study was CRISP-DM, and the statistical population was banking transactions. First, the system and the

database were identified. Then, in the data preparation stage, the data required for the implementation of the algorithms were selected, the required new variables were added and the redundant variables were removed, the missing data and the frequency distribution of the data were examined, and the initial visualization of the data set was done. Next, in the modelling phase, two approaches were used to discover suspected money laundering cases that require further investigation by bank inspectors and audits.

In the first approach, according to the conventional methods in outlier data discovery, bank transactions were first clustered, and then clusters that contained a small percentage of data, along with examples of dense clusters that seemed unusual compared to other examples of their cluster, were introduced as outlier transactions that may be suspected of money laundering and need further investigation by banking experts. K-means clustering algorithm was used for data clustering, and the number of clusters was identified using the elbow method. Also, Silhote's score was applied to measure the reliability of the clustering.

In the second approach, by using Benford's law and GANs algorithm, accounts that may have fake numbers in their transaction were investigated. To do this, a data set that is suitable for the problem conditions was simulated. A data set containing the transaction of 50,000 bank accounts was simulated, half of which did not have fake numbers in the transaction and the other half had fake numbers. Then, using an artificial neural network, the model was trained to distinguish natural data from fake data. After that, using GANs algorithm, synthetic data for fake numbers

were created to detect suspicious cases in which an attempt was made to hide the fake numbers.

## 4- RESULTS & DISCUSSION

The results of the study showed that data mining methods, Benford's law, and GANs algorithm were effective in detecting money laundering activities in financial transactions. The first approach that uses clustering and anomaly detection had an accuracy of about 93%, while the second approach that uses Benford's law and GANs algorithm had an accuracy of about 60%.

## 5- CONCLUSIONS & SUGGESTIONS

This paper presented a significant contribution to the field of money laundering detection by using data mining techniques, Benford's law and GANs algorithm. The study demonstrated the potential of these techniques to detect money laundering activities, and it could be beneficial for financial institutions to utilize these techniques to prevent financial crimes. However, the study has some limitations, such as the small sample size of transactions from one bank branch. Future research could replicate the study with larger database from multiple banking institutions. Additionally, future research could also explore the apply of other advanced data mining algorithms to improve the accuracy of detecting suspicious transactions.

**Keywords**: Money Laundering, Data mining, K-means algorithm, GANs algorithm, Benford’s law.

اقتصاد پولی مالی، دوره جدید، سال بیست و نهم، شمارهٔ ۲، پیاپی ۲۴، پاییز و زمستان ۱۴۰۱
https://danesh24.um.ac.ir

# شناسایی فعالیت‌های پول‌شویی در تراکنش‌های مالی با استفاده از روش‌های داده‌کاوی، قانون بنفورد و الگوریتم GANs

**عالیه کاظمی***
استاد مدیریت عملیات و علوم تصمیم، دانشکده مدیریت دانشگاه تهران، تهران، ایران
**امیر مقدم‌فلاحی**
کارشناس ارشد مدیریت صنعتی، دانشکده مدیریت، دانشگاه تهران، تهران، ایران
**علی ابدالی**
استادیار دانشگاه علوم انتظامی امین، تهران، ایران
**سارا آریایی**
دانشجوی دکترای مدیریت صنعتی، دانشکده مدیریت، دانشگاه تهران، تهران، ایران
دستیار پژوهشی، کالج کسب‌وکار کاگین، دانشگاه فلوریدا شمالی، جکسون‌ویل، فلوریدا، ایالات‌متحده آمریکا

نوع مقاله: پژوهشی
https://doi.org/10.22067/mfe.2023.72676.1117

## چکیده
امروزه پدیده پول‌شویی به تهدیدی جدی برای اقتصاد جهانی تبدیل شده است. روش‌های سنتی مقابله با پول‌شویی هزینه‌بر و ناکارآمد هستند. اخیراً تکنیک‌های داده‌کاوی گسترش پیدا کرده‌اند و به‌عنوان روش‌های مناسب برای کشف فعالیت‌های پول‌شویی موردتوجه قرار گرفته‌اند. هدف این تحقیق، استفاده از الگوریتم‌های داده‌کاوی در کشف موارد مشکوک به پول‌شویی با استفاده از داده‌های واقعی تراکنش‌های بانکی است که ممکن است نیاز به بررسی‌های بیشتر داشته باشند. تحلیل داده‌ها با استفاده از فرآیند CRISP-DM انجام شده است. جامعه آماری تراکنش‌های بانک و نمونه آماری تراکنش‌های مربوط به یکی از شعب بانک است. داده‌ها از بانک اطلاعاتی بانک موردمطالعه جمع‌آوری شده است. برای انجام این کار از دو رویکرد استفاده شده است. در رویکرد اول با استفاده از الگوریتم k-میانگین ابتدا تراکنش‌های بانکی افراد خوشه‌بندی شده‌اند، سپس با استفاده از به‌کارگیری الگوریتم‌های کشف موارد مشکوک، تراکنش‌هایی که ممکن است مشکوک به پول‌شویی باشند مشخص گردیده‌اند. در رویکرد دوم، روشی نوین با به‌کارگیری قانون بنفورد و روش GANs برای کشف حساب‌هایی که در تراکنش‌های آن‌ها از ارقام ساختگی استفاده شده است و ممکن است مشکوک به پول‌شویی باشند، معرفی شده است. رویکرد اول می‌تواند حساب‌هایی را که در تراکنش‌های آن‌ها موارد پرت وجود دارد، با دقتی حدود ۹۳٪ درصد و رویکرد دوم می‌تواند حساب‌های مشکوکی را که در پنهان نمودن ارقام ساختگی در تراکنش‌های آن‌ها، از روش‌های حرفه‌ای استفاده نشده است، با دقتی حدود ۶۰٪ به‌درستی تشخیص دهد.

**کلیدواژه‌ها**: پول‌شویی، داده‌کاوی، الگوریتم k-میانگین، الگوریتم GANs، قانون بنفورد.
**طبقه‌بندی JEL:** C02, C44, C53, G21, P44

---

* نویسنده مسئول: aliyehkazemi@ut.ac.ir
تاریخ دریافت: ۱۴۰۰/۰۷/۰۵ تاریخ پذیرش: ۱۴۰۱/۱۲/۰۸
صفحات: ۳۵۸-۳۲۷

**مقدمه**

در دنیای امروزی استفاده از فناوری اطلاعات در تمامی ابعاد زندگی بشر امکان‌پذیر شده و درنتیجه تحولی شگرف در نظام پولی و بانکی جهانی به‌واسطه انقلاب فناوری اطلاعات و ارتباطات ایجاد گردیده است. استفاده از ابزارهای مختلف مبتنی بر علوم رایانه و ارتباطات و فناوری‌های بر پایه سرویس‌های شبکه، امکانات و فرصت‌های بسیار متنوع را در اختیار مسئولین و مشتریان مؤسسات مالی قرار داده است. وجود انواع ابزارهایی همچون بانکداری الکترونیک، پول الکترونیکی و انواع کارت پرداخت به‌طورکلی بانکداری سنتی را دستخوش تغییرات جدی و بسیار زیادی نموده است. در این میان و همراه با فواید مختلفی همچون سرعت، افزایش کارایی، امنیت، تنوع کاربرد و گمنامی که در این ابزارها وجود دارد، کاربرد آن‌ها نیز با چالش‌هایی مواجه است؛ چراکه این خصوصیات باعث گردیده است که این ابزار جدید، به‌عنوان راهکاری جذاب و جدید برای فعالیت‌های مجرمانه تبدیل گردد (Masjidi, 2015).

یکی از معایب این تحول اطلاعاتی، ایجاد بستری مناسب برای فعالیت‌های غیرقانونی ازجمله پول‌شویی است. پول‌شویی فعالیتی است که در طی انجام آن، عواید و درآمدهای ناشی از اعمال خلاف قانون، مشروعیت می‌یابد؛ به‌عبارت‌دیگر پول‌های کثیف ناشی از اعمال خلاف به پول‌های به‌ظاهر تمیز تبدیل می‌شوند و در بدنه اقتصاد جایگزین می‌شوند (Jantani, 2017). پول‌شویی یکی از شایع‌ترین انواع جرائم مالی است. بسیاری از دولت‌ها و ارگان‌ها، واحدهایی با عنوان واحد اطلاعات مالی برای تحقیق در این مورد در نظر گرفته‌اند. مؤسسات مالی مانند بانک‌ها، سازمان‌های بیمه و کسب‌وکارهای پولی باید سوابق معاملات مشتریان خود را در آرشیو الکترونیکی ذخیره کنند و گزارش‌های معاملات مشکوک را به واحد اطلاعات مالی گزارش دهند (Didimo, Liotta & Montecchiani, 2014).

شناسایی پول‌شویی با توجه به تعداد زیاد تراکنش درگیر، کار بسیار دشواری است. در فرایند ضد پول‌شویی باید فعالیت‌های پول‌شویی در سیستم مالی واقعی به‌درستی تشخیص داده شوند. به‌طور سنتی روش‌های آماری و روش‌های مبتنی بر قانون برای کشف فعالیت‌های پول‌شویی استفاده می‌شود؛ اما تحلیل تراکنش مبتنی بر قانون برای کشف الگوهای تراکنشی پیچیده ناکافی است و معاملات بزرگ و غیریکنواخت زمان و انرژی زیادی را به خود اختصاص می‌دهند.

یکی از بزرگ‌ترین بنگاه‌های اقتصادی هر کشوری که مورد سوءاستفاده مجرمان به‌منظور پول‌شویی قرار می‌گیرد بانک‌ها هستند که بزرگ‌ترین ضربه را به اقتصاد یک کشور وارد می‌کنند (Asadi, 2015) و تشخیص تراکنش‌های مشکوک به پول‌شویی بزرگ‌ترین چالش مبارزه با پول‌شویی در حوزه بانک است. عدم مبارزه با پول‌شویی موجب شیوع بیشتر جرائمی مانند فرار مالیاتی، اختلاس، قاچاق کالا و همچنین

خریدوفروش مواد مخدر و موارد دیگری از این قبیل می‌شود و تمایل سرمایه‌گذاری در فعالیت‌های مولد را کاهش داده و موجب تضعیف بنیان‌های اقتصادی کشور می‌گردد (Sarraf & Heidari, 2015).

بر اساس گزارش‌های موسسه بازل که به‌صورت انحصاری در زمینه مبارزه با پول‌شویی کار می‌کند، کشور ایران از سال ۲۰۱۴ تاکنون در صدر جدول پر ریسک‌ترین کشورها از نظر پول‌شویی قرار دارد و بر اساس شاخص‌های اندازه‌گیری این موسسه، کشور ایران آسیب‌پذیرترین کشور در این زمینه محسوب می‌شود.

ازجمله علل احتمالی به وجود آمدن این مسئله می‌توان به، به‌کارگیری روش‌های قدیمی در تشخیص موارد مشکوک به پول‌شویی در سیستم‌های حال حاضر بانکی کشور اشاره کرد که این امر موجب تحمیل هزینه‌های بسیار هنگفت به بانک‌های کشور شده و همچنین از دقت کافی نیز برخوردار نیستند. درحالی‌که با پیشرفت روزافزون سیستم‌های اطلاعاتی و استفاده مناسب از آن‌ها می‌توان این هزینه‌ها را به‌شدت کاهش و دقت بررسی آن‌ها را به‌مراتب افزایش داد.

استفاده از فناوری‌های جدید در فرآیند این جرم، استفاده از راهکارهای جدید و نرم‌افزاری را ضروری می‌نماید. امروزه با ایجاد بسترهای ارتباطی بین مؤسسات مختلف، دیگر مانند گذشته ردیابی تراکنش‌های مختلف به‌راحتی و آسانی و تنها با تکیه بر تجربیات نیروی انسانی قابل‌اجرا نیست. این امر وجود سیستم‌های نرم‌افزاری هوشمند در اجرای سیاست‌های مبارزه با پول‌شویی در مؤسسات مالی را اجتناب‌ناپذیر می‌سازد.

امروزه همه کسانی که به‌نوعی با پیشگیری و کشف این فرایند نامشروع در ارتباط‌اند به دنبال راه‌حل‌هایی هستند که آن‌ها را در این مسیر سخت و دشوار یاری رسانند و یکی از این راه‌حل‌ها استفاده از تکنیک‌های داده‌کاوی است (Sarraf & Heidari, 2015). لازم به ذکر است که تحقیقات انجام شده در این زمینه بسیار محدود بوده‌اند و تاکنون تحقیقی در ارتباط با به‌کارگیری تکنیک‌های داده‌کاوی در تشخیص موارد مشکوک به پول‌شویی با استفاده از داده‌های واقعی برای بانک‌های کشور صورت نگرفته است. هدف از انجام این تحقیق به‌کارگیری روش‌های داده‌کاوی در تشخیص و کشف موارد مشکوک به پول‌شویی در انبار داده بانک است. در ادامه به بررسی پیشینه پژوهش پرداخته شده است. سپس روش‌شناسی پژوهش توضیح داده شده است. پس‌ازآن یافته‌های پژوهش ارائه و نهایتاً نتایج ذکر شده است.

## پیشینه پژوهش

در این بخش به بررسی تحقیقات انجام شده در حوزه مبارزه با پول‌شویی با استفاده از الگوریتم‌های داده‌کاوی پرداخته می‌شود.

الکساندر و بالسا یک سیستم چند عاملی (MAS[1]) برای مبارزه با پول‌شویی ارائه کردند که یادگیری ماشینی و یک استراتژی مبتنی بر ریسک را در بر می‌گیرد. سیستم پیشنهادی از سه عامل جمع‌آوری داده، تجزیه‌وتحلیل تراکنش و ارزیابی ریسک تشکیل شده است. نویسندگان سیستم پیشنهادی را با استفاده از مجموعه داده‌های دنیای واقعی ارزیابی کرده و عملکرد آن را با سایر سیستم‌های مبارزه با پول‌شویی مقایسه می‌کنند. نتایج نشان داد که سیستم پیشنهادی به دقت بالاتری در تشخیص تراکنش‌های مشکوک دست می‌یابد و تعداد موارد مثبت کاذب را کاهش می‌دهد و درنتیجه کارایی فرآیند مبارزه با پول‌شویی را بهبود می‌بخشد (Alexandre & Balsa, 2023). لوکانن مطالعه‌ای در مورد استفاده از یادگیری ماشین و الگوریتم‌های شبکه عصبی مصنوعی برای پیش‌بینی فعالیت‌های پول‌شویی در بانک‌ها ارائه داد. این مطالعه از نمونه‌ای از تراکنش‌های بانکی در آفریقای جنوبی استفاده می‌کند. نتایج این مطالعه نشان داد مدل‌های شبکه عصبی مصنوعی از سایر مدل‌های یادگیری ماشین در شناسایی فعالیت‌های پول‌شویی بهتر عمل می‌کنند (Lokanan, 2022). کومار و همکاران با استفاده از دسته‌بند بیز ساده[2] به شناسایی فعالیت‌های پول‌شویی پرداختند. آن‌ها مجموعه داده‌ای از تراکنش‌های بانکی را جمع‌آوری کردند و از روش‌های مختلف پیش‌پردازش برای استخراج ویژگی‌های مرتبط استفاده کردند و عملکرد الگوریتم بیز ساده را ارزیابی کردند. آن‌ها همچنین رویکرد پیشنهادی را با سایر روش‌های یادگیری ماشینی مانند درخت‌های تصمیم‌گیری و ماشین‌های بردار پشتیبان مقایسه کردند (Kumar, Das & Tyagi, 2020).

مارتینز سانچز و همکاران با استفاده از درخت رگرسیون به شناسایی فعالیت‌های پول‌شویی در مکزیک پرداختند. نویسندگان یک رویکرد مدیریت ریسک را پیشنهاد دادند که از داده‌های تراکنش‌های تاریخی برای شناسایی مناطق پرخطر برای پول‌شویی و توسعه یک مدل پیش‌بینی‌کننده استفاده می‌کند و می‌تواند به مؤسسات مالی کمک کند تا پول‌شویی را شناسایی و از آن جلوگیری کنند. نتایج مطالعه نشان داد رویکرد پیشنهادی می‌تواند در شناسایی فعالیت‌های بالقوه پول‌شویی مؤثر باشد (Martínez-Sánchez, Cruz-García & Venegas-Martínez, 2020). ژانگ و تروبی به بررسی عملکرد پنج الگوریتم یادگیری ماشین برای تشخیص پول‌شویی با استفاده از تراکنش‌های واقعی افراد که متعلق به بانک اطلاعاتی یک موسسه مالی در آمریکا بود، پرداختند (Zhang & Trubey, 2019). فیور و همکاران برای افزایش دقت

---

[1] Multi-Agent System
[2] Naïve Bayes Classifier

الگوریتم‌های خوشه‌بندی در تشخیص کلاه‌برداری‌های مالی در کارت‌های اعتباری و همچنین برای مقابله با مشکل متوازن نبودن کلاس‌ها در تشخیص موارد غیرعادی[1]، از الگوریتم GANs[2] استفاده نمودند (Fiore, De Santis, Perla, Zanetti & Palmieri, 2019). بدل ولرو و همکاران با ترکیب قانون بنفورد و الگوریتم‌های یادگیری ماشین به کشف الگوهای پول‌شویی برای یک مورد قضایی در دادگاه اسپانیا پرداختند (Badal-Valero, Alvarez-Jareño & Pavía, 2018).

سورش و همکاران یک رویکرد ترکیبی ارائه کرده‌اند که چندین روش داده‌کاوی ازجمله خوشه‌بندی، کاوش قواعد وابستگی و درخت تصمیم را برای شناسایی حساب‌های مشکوک ترکیب می‌کند. این مقاله توضیح می‌دهد که چگونه می‌توان از هر یک از این روش‌ها برای استخراج الگوها و قوانین مفید از داده‌ها استفاده کرد و چگونه می‌توان نتایج را برای بهبود دقت فرآیند تشخیص ترکیب کرد (Suresh, Reddy, & Sweta, 2016). احمد و همکاران به بررسی انواع روش‌های خوشه‌بندی در تشخیص عوامل غیرعادی و مقایسه آن‌ها از مناظر مختلف با یکدیگر پرداختند. آن‌ها بیان کردند که به دلیل پیشرفت تقلب‌های مالی و در دسترس نبودن داده‌های کافی، باید یک روش جهانی در حوزه تشخیص تقلب‌های مالی کشف شود (Ahmed, Mahmood & Islam, 2016). صدیقی و سجادی‌نژاد برای طراحی سیستمی هوشمند که بتواند با استفاده از ویژگی‌های مختلف یک تراکنش مالی، قانونی یا غیرقانونی بودن آن را تشخیص دهد از الگوریتمی مبتنی بر یادگیری عمیق بهره گرفتند (Seddighi, & Sajedinejad, 2009). فرشادی‌نیا و بصیری قائمی‌پسند با استفاده از روش‌های درخت تصمیم، شبکه‌های عصبی مصنوعی و ماشین بردار پشتیبان به کشف موارد مشکوک به پول‌شویی پرداختند (Farshadinia, & Basiri Ghaemi Pasand, 2016). تقوا و همکاران برای کشف ناهنجاری در تراکنش کارت‌های بانکی با رویکرد پردازش موازی و راه‌حل نگاشت کاهش، از شبکه عصبی مدل کوهونن[3] استفاده کردند (Taghva, Mansouri, Feizi, Akhgar, 2016).

در بررسی موارد مشکوک به پول‌شویی در حوزه بانک، یکی از اصلی‌ترین چالش‌ها دسترسی به انبار داده تراکنش‌ها و حساب‌های واقعی اشخاص است. با توافقاتی که با یکی از بانک‌های کشور صورت گرفت، شرایط موردنیاز برای استفاده از انبار داده بانک مذکور جهت انجام این پژوهش فراهم گردید. یکی دیگر

---

[1] Anomaly Detection
[2] Generative Adversarial Networks
[3] Kohonen

از چالش‌های اساسی در این حوزه نبود اطلاعات موردنیاز برای استفاده از مدل‌های داده‌کاوی با ناظر است؛ زیرا اولاً بانک‌ها و مؤسسات مالی یا خود به این سطح از اطلاعات دسترسی ندارند یا اگر دسترسی داشته باشند به علت تصمیم‌گیری‌های سیاسی و کسب‌وکاری تمایلی به اشتراک آن‌ها ندارند، ثانیاً روش‌های انجام پول‌شویی روز به روز در حال پیشرفت هستند، لذا روش‌های کشف شده با استفاده از روش‌های داده‌کاوی و یادگیری ماشین با ناظر با استفاده از داده‌های حال حاضر قادر به تشخیص موارد مشکوک به پول‌شویی در آینده که با استفاده از روش‌های پول‌شویی جدیدتر انجام شده باشند را ندارند. بنابراین تصمیم گرفته شد که پژوهش در دو بخش اجرا شود. در بخش اول، با استفاده از الگوریتم‌های خوشه‌بندی، خوشه‌بندی تراکنش‌های مالی بانک مذکور انجام می‌شود و با استفاده از روش‌های کشف موارد پرت به کشف موارد مشکوک به پول‌شویی پرداخته می‌شود. در بخش دوم، ابتدا با استفاده از قانون بنفورد، تعداد زیادی تراکنش مالی در دو دسته شبیه‌سازی می‌شوند. دسته اول کاملاً از قانون بنفورد پیروی می‌کنند و دسته دوم به‌گونه‌ای شبیه‌سازی می‌شوند که بیشترین انحراف را از قانون بنفورد داشته باشند. سپس با استفاده از روش GANs به تولید داده‌هایی که در حالت میانی دو دسته قبلی هستند پرداخته می‌شود و به‌طور همزمان مدل یادگیری الگوهای پول‌شویی در هر تکرار این مدل قوی‌تر می‌شود. سپس از مدل قوی شده جهت تشخیص موارد مشکوک به پول‌شویی در پایگاه داده تراکنش‌های واقعی استفاده می‌شود. قابل ذکر است روش GANs به‌تازگی و توسط (Goodfellow et al., 2014) معرفی گردیده است.

## روش‌شناسی پژوهش

پژوهش حاضر به دنبال پاسخ به سؤالات زیر است:

- کدام تراکنش‌های مالی با حجم زیاد توجیه قابل قبولی ندارند؟
- کدام حساب‌های بانکی به‌یک‌باره تغییر اساسی داشته‌اند؟
- در کدام حساب‌های بانکی تراکنش‌های پرتکرار با حجم و میزان معین وجود دارد؟

قلمرو مکانی این پژوهش یکی از بانک‌های کشور است. قلمرو زمانی داده‌های مربوط به اسفندماه سال ۱۳۹۶ است که از طرف بانک در اختیار محققین قرار گرفته است. مطابق با شکل ۱ که چهارچوب مفهومی پژوهش را نشان می‌دهد، پایگاه داده در نظر گرفته شده متشکل از داده‌های تاریخی حساب‌های مشتریان و تراکنش‌های آن‌ها است و درنهایت با گذر از مراحل تحقیق به ارائه دانش (موارد مشکوک به پول‌شویی) منجر می‌شود. جامعه آماری این تحقیق شامل کلیه مشتریان و تراکنش‌های بانک موردمطالعه

است. اطلاعات مشتریان به‌صورت کد شده و با حذف نام آن‌ها استفاده شده است. نمونه آماری انتخاب شده شامل مشتریان و تراکنش‌های بانکی ۳ شعبه بانک است که شامل ۳۱۱۲۳ تراکنش از ۳۰۲۹ حساب بانکی است. نمونه آماری به‌گونه‌ای انتخاب شده است که شامل تمامی انواع مشتری و تراکنش‌های بانکی باشد و همچنین به اندازه کافی بزرگ باشد که بتوان به کمک آن خوشه‌بندی را انجام داد و تراکنش‌های غیرعادی را که نیاز به بررسی‌های بیشتر دارند به‌عنوان تراکنش‌های مشکوک به پول‌شویی توسط روش‌های تشخیص عوامل غیرعادی کشف و گزارش داد. تحلیل داده‌ها در این تحقیق با به‌کارگیری روش‌های داده‌کاوی انجام شده است. پس از جمع‌آوری داده‌ها، پیش‌پردازش و آماده‌سازی داده‌ها انجام می‌شود. شاخص‌های مناسب برای تشخیص موارد مشکوک به پول‌شویی (همچون سپرده نقدی با حجم زیاد که توجیه قابل قبولی ندارد، نحوه سابقه حساب به‌گونه‌ای که به‌یک‌باره تغییر اساسی داشته باشد، تکرار تراکنش‌های با حجم و میزان معین) با مرور ادبیات تحقیق و دریافت نظر خبرگان بانک مشخص شده است. سپس از الگوریتم‌های مختلف داده‌کاوی برای تعیین نقاط پرت استفاده می‌شود. برای بررسی عملکرد الگوریتم‌های انتخابی، هنگامی‌که برچسب داده‌ها موجود نیست (رویکرد اول در مدل‌سازی) از روش‌های تعیین دقت داخلی همچون روش سیلهوته[1] استفاده می‌شود و هنگامی‌که برچسب داده‌ها موجود است (رویکرد دوم در مدل‌سازی) به بررسی دقت مدل در تشخیص درست موارد مشکوک و تشخیص نادرست موارد غیر مشکوک پرداخته می‌شود.

در ادامه فرایند اجرایی پژوهش با جزئیات تشریح گردیده است.

**فرایند اجرایی پژوهش**

فرآیند اجرائی به‌صورت چندین مرحله مجزا و پیوسته انجام می‌گردد که عبارتند از:

**۱. تعیین حدود مسئله و تشکیل پایگاه داده:** با بررسی پایگاه اطلاعاتی بانک موردنظر، تصحیح اطلاعات پرت، استانداردسازی داده‌ها، کاهش ابعاد و خارج‌کردن ویژگی‌های مؤثر به ایجاد پایگاه داده‌ای معتبر برای داده‌کاوی پرداخته می‌شود. سپس با استفاده از الگوریتم‌های معتبر، پایگاه داده به دو بخش یادگیری و تست تقسیم می‌شود. از داده‌های یادگیری برای آموزش سیستم و از داده‌های تست بعد برای بررسی اعتبار مدل استفاده می‌شود.

**۲. پیاده‌سازی الگوریتم‌های داده‌کاوی:** با استفاده از الگوریتم‌های متفاوت داده‌کاوی همچون k میانگین،

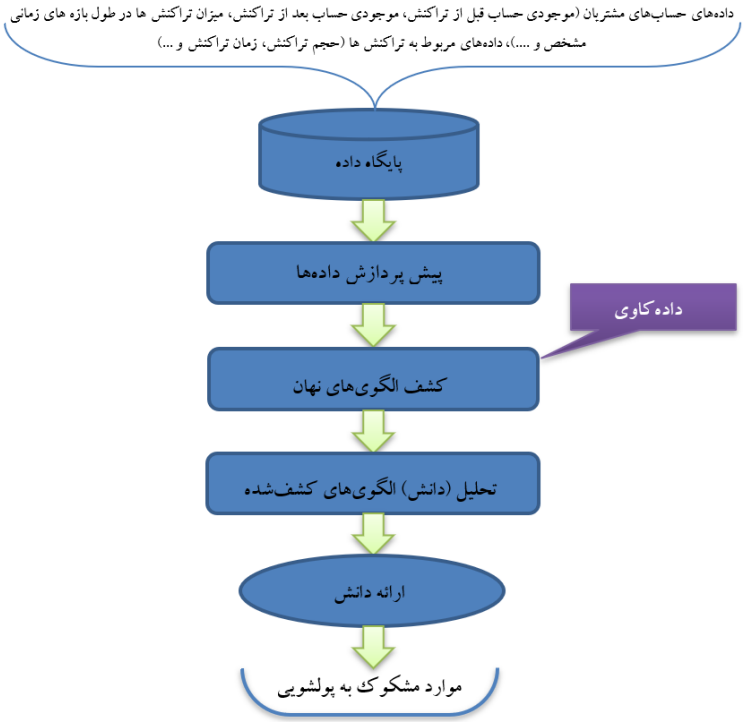
---

[1] Silhouette Score

ماشین بردار پشتیبان و شبکه‌های عصبی مصنوعی و همچنین با توجه به الگوریتم‌های مختلف تشخیص داده‌های غیرعادی[1] و تشخیص نقاط پرت[2] و مفروضات آن‌ها، به تشخیص نقاط پرت از منظرهای گوناگون پرداخته می‌شود.

**۳. بررسی نتایج حاصل از داده‌کاوی:** نتایج حاصل از الگوریتم‌های داده‌کاوی از نظر میزان عملکرد و دقت، حجم محاسبات، هزینه و سرعت محاسبات و کاربردی بودن آن‌ها با یکدیگر مقایسه می‌گردند، سپس الگوریتم بهینه و همچنین پارامترهای آن مشخص می‌گردند.

**٤. مشخص‌سازی تراکنش‌ها و مشتریان مشکوک به پول‌شویی:** بعد از انجام داده‌کاوی در پایگاه داده بانک، تراکنش‌ها و مشتریانی که بیشترین احتمال پول‌شویی در آن‌ها وجود دارد مشخص خواهند شد تا بانک مورد نظر بتواند تحقیقات کامل‌تری در خصوص این تراکنش‌ها انجام دهد.




**شکل ۱: چهارچوب مفهومی پژوهش**

[1] Anomaly Detection
[2] Outlier Detection

برای انجام مراحل ذکر شده، روش‌شناسی CRISP-DM مورد استفاده قرار گرفته است. پس از آماده‌سازی و پیش‌پردازش داده‌ها، مجموعه داده‌ها با به‌کارگیری روش‌های خوشه‌بندی و تشخیص موارد پرت مورد بررسی قرار می‌گیرند و تراکنش‌های بانکی مشکوک به پول‌شویی مشخص می‌شوند. همچنین روشی بر پایه قانون بنفورد و با به‌کارگیری الگوریتم GANs معرفی می‌شود. برای پیاده‌سازی مدل‌های در این پژوهش از زبان برنامه‌نویسی Python استفاده شده است. در ادامه روش‌های مورد استفاده در این پژوهش به‌اختصار تشریح می‌شوند.

### ❖ روش آرنج

یکی از اصلی‌ترین مشکلات استفاده از الگوریتم k-میانگین این است که تعداد خوشه‌ها (k) باید به‌عنوان ورودی به مدل داده شود. درحالی‌که در خیلی از مواقع، محقق دانش و اطلاعات دقیقی از تعداد دقیق خوشه‌ها ندارد بنابراین تعیین تعداد خوشه‌ها امری مشکل است. یکی از روش‌های پرکاربرد در زمینه تعیین مقدار k، رسم نمودار آرنج[1] است. در این نمودار، خطای به‌دست‌آمده به ازای مقادیر متفاوت k در نموداری رسم می‌شود. عدد k در قسمت بازوی نمودار، بهترین مقدار k است (James, Witten, Hastie & Tibshirani, 2013).

### ❖ روش ارزیابی کیفیت خوشه‌بندی سیلهوته

برای محاسبه ضریب سیلهوته یک نمونه در یک مجموعه داده، می‌توان سه مرحله را طی کرد: ۱. محاسبه پیوستگی خوشه، $a(i)$، با استفاده از محاسبه میانگین فاصله نمونه، $x(i)$، با تمامی نمونه‌های دیگر در خوشه خودش، ۲. محاسبه گسستگی خوشه، $b(i)$، از نزدیک‌ترین خوشه مجاور با محاسبه میانگین فاصله بین نمونه مورد نظر، $x(i)$ و تمامی نمونه‌ها در نزدیک‌ترین خوشه و ۳. محاسبه ضریب سیلهوته برای نمونه مورد نظر، $s(i)$، با استفاده از محاسبه تفاوت بین پیوستگی و گسستگی خوشه تقسیم‌بر بزرگ‌ترین آن دو طبق فرمول $s(i) = \frac{b(i)-a(i)}{max\{b(I),a(I)\}}$ به دست می‌آید. ضریب سیلهوته محدود به بازه ۱ تا ۱- است. هرچه $b(i)$ از $a(i)$ بزرگ‌تر باشد، نمونه به ضریب سیلهوته ایده‌آل یک نزدیک‌تر می‌شود (Raschka, & Mirjalili, 2017). امتیاز کیفیت خوشه‌بندی سیلهوته برای مدل مورد استفاده را می‌توان با استفاده از

---

[1] Elbow

محاسبه میانگین ضرایب سیلهوته تمامی نمونه‌ها در مجموعه داده به دست آورد.

**❖ الگوریتم GANs**

الگوریتم GANs توسط (Goodfellow et al., 2014) مطرح شدند. در این الگوریتم شبکه‌ها بر اساس رویکرد تئوری بازی‌ها بنا شدند. یک شبکه یادگیری عمیق که مولد[1] نامیده می‌شود به رقابت می‌پردازد. شبکه عمیق دیگری که متمایزکننده[2] نامیده می‌شود نمونه‌های تولیدشده از شبکه مولد را از داده‌های اصلی متمایز می‌کند. رقابت بین این دو شبکه درنهایت باعث یادگیری بهتر و بهبود عملکرد هر دو می‌شود. برای یادگیری توزیع شبکه مولد بر روی داده‌های ورودی، ابتدا از یک توزیع نویز به‌عنوان ورودی استفاده می‌شود. هدف بهبود همزمان تابع مولد و تابع متمایزکننده است. درنهایت برای بهینه‌کردن شبکه مولد و متمایزکننده از فرمول زیر استفاده می‌شود.

$$\min_G \max_D V(G,D) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z_{p_z(z)}}\left[\log\left(1 - D(G(z))\right)\right] \quad (۱)$$

D و G یک بازی حداقل حداکثر دو بازیکنه با تابع ارزش $V(G,D)$ را ادامه می‌دهند. $P_g$ توزیع شبکه مولد و $p_z(z)$ متغیر نویز ورودی است. فرمول بالا متمایزکننده D را به‌گونه‌ای استخراج می‌کند که بتواند به‌درستی داده‌های واقعی و مصنوعی را از هم تفکیک کند. فرمول داده‌شده به‌صورت فرم بسته قابل حل نیست و لذا از روش‌های تکراری و عددی به‌منظور حل آن استفاده می‌شود. روند کلی مراحل یادگیری الگوریتم GANs به‌صورت زیر است (Goodfellow et al., 2014).

آموزش شبکه‌های GANs با روش کمترین گرادیان خطا: تعداد گام‌ها برای آموزش متمایزکننده برابر k و به‌عنوان پارامتر اولیه در نظر گرفته شده است.

برای تعداد k تکرار شود:

- به تعداد m از فضای نویز اولیه $p_g(z)$ نمونه‌برداری می‌شود: $z = \{z^{(1)}, \dots, z^{(m)}\}$
- به تعداد m از توزیع اولیه داده‌ها $p_{data}$ نمونه‌برداری می‌شود: $x = \{x^{(1)}, \dots, x^{(m)}\}$
- با محاسبه گرادیان مطابق با فرمول $\left[\log D(x^{(i)}) + \log\left(1 - D\left(G(z^{(I)})\right)\right)\right]$ مقدار تابع $\nabla\theta_d \frac{1}{m}\sum_{i=1}^{m}$

---

[1] Generator
[2] Discriminator

متمایز کننده محاسبه می‌شود.

پایان حلقه دوم.

- به تعداد m از فضای نویز اولیه $p_g(z)$ نمونه‌برداری می‌شود: $z = \{z^{(1)}, \dots, z^{(m)}\}$

- تابع مولد با روش کاهش گرادیان به‌صورت مقابل $\nabla\theta_g \frac{1}{m}\sum_{i=1}^{m}\log\left(1 - D\left(G(z^{(I)})\right)\right)$ به‌روزرسانی می‌شود.

**یافته‌ها**

در این بخش مراحل مختلف روش‌شناسی CRISP-DM[1] پیاده‌سازی می‌شود و پس از انجام مراحل مختلف مربوط به آماده‌سازی و پیش‌پردازش داده‌ها، مجموعه داده‌ها با به‌کارگیری روش‌های خوشه‌بندی و تشخیص موارد پرت مورد بررسی قرار می‌گیرند و تراکنش‌های بانکی مشکوک به پول‌شویی معرفی می‌شوند. در آخر الگوریتم‌ها مورد بررسی قرار خواهند گرفت و نتایج حاصل از آن‌ها شرح داده می‌شود.

➢ **مراحل روش‌شناسی CRISP-DM**

روش‌شناسی CRISP-DM به‌عنوان روش مرجع فرآیند داده‌کاوی به کار می‌رود. این روش‌شناسی از گام‌های شناخت سیستم، شناخت داده‌ها، آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی و توسعه سیستم تشکیل شده است (Ghazanfari, Alizadeh & Teymour Pour, 2008).

✓ **شناخت سیستم**

یکی از زمینه‌هایی که در سال‌های اخیر موردتوجه بانکداران بوده است، ایجاد سیستم‌های جامع و نوین بر پایه استفاده از علوم کامپیوتر برای مبارزه با پول‌شویی بوده است. در این پژوهش با استفاده از داده‌های یکی از بانک‌های کشور، سعی بر این است که سیستمی جامع برای کشف موارد مشکوک به پول‌شویی ایجاد گردد.

✓ **شناخت داده‌ها**

هدف این گام شناخت مجموعه داده‌ای است که باید مورد کاوش قرار گیرد و شامل به دست آوردن و

[1] Cross Industry Standard Process for Data Mining

شناخت داده‌ها است. بانک اطلاعاتی شامل ۳۱۱۲۳ رکورد از تراکنش‌های ۳۰۲۹ حساب بانکی مربوط به یک دوره یک ماهه و همچنین در مجموعه‌ای دیگر شامل مانده حساب تمامی این ۳۰۲۹ حساب بانکی در ابتدای دوره است. مجموعه داده‌ها جمعاً شامل ۱۵ متغیر است. بررسی جامعه به‌منظور شناخت بیشتر داده‌های موجود در بانک اطلاعاتی انجام شد و نتایج در جدول ۱ مشاهده می‌شود.

**جدول (۱): شناسایی متغیرهای پژوهش**

| نام متغیر | شرح | نام متغیر | شرح |
|---|---|---|---|
| BRNCD | کد شعبه بانکی که حساب در نزد آن شعبه افتتاح گردیده است. | SEX | جنسیت مشتری:<br>مرد = M<br>زن = F<br>نامشخص = None |
| MODCD | مشخص‌کننده نوع حساب بانکی:<br>قرض‌الحسنه = ۱<br>سپرده = ۵ | TRNCD | کد تراکنش:<br>واریز نقدی و واریز نقدی با دفترچه = ۱۰۱<br>واریز انتقالی و واریز انتقالی با دفترچه = ۱۰۳<br>واگذاری چک عهده سایر بانک‌ها = ۱۰۴<br>واریز چک بانک مورد بررسی به حساب = ۱۰۵<br>واریز نقدی بدون دفترچه = ۱۰۷<br>واریز انتقالی بدون دفترچه = ۱۰۹<br>واریز چک بانک مورد بررسی به حساب بدون دفترچه = ۱۱۵<br>واریز چک عهده سایر بانک‌ها = ۱۸۴<br>واریز سود = ۱۹۳<br>پایا/شاپرک و کارت‌خوان = ۱۹۸<br>برداشت نقدی و برداشت نقدی با دفترچه = ۲۰۱<br>برداشت انتقالی و برداشت انتقالی با دفترچه = ۲۰۳<br>برداشت نقدی بدون دفترچه = ۲۰۷<br>برداشت انتقالی بدون دفترچه = ۲۰۹<br>برداشت طی چک = ۲۱۲<br>هزینه‌های متفرقه (شامل برداشت‌های مربوط به کارمزد بانکی) = ۲۹۹<br>برداشت چک عهده سایر بانک‌ها = ۳۰۶ |

| نام متغیر | شرح | نام متغیر | شرح |
|---|---|---|---|
| | | | برداشت چک بانک مورد بررسی = ۳۲۲<br>واریز به کارت یا حساب قوامین = ۷۰۳<br>برداشت وجه از خودپرداز = ۷۰۸<br>انتقال از کارت یا حساب بانک مورد بررسی = ۷۰۹<br>خرید شارژ یا پرداخت قبض = ۷۱۹<br>خرید از پایانه فروشگاهی = ۷۲۶ |
| ACNO | شماره حساب | TRANSACTIONDATE | تاریخ تراکنش |
| CHKDGT | شماره چک‌پوینت حساب[1] | TRANSACTIONTIME | تاریخ و ساعت تراکنش |
| PRLINE | مشخص‌کننده نوع حساب:<br>جاری = PROD<br>کوتاه‌مدت = SHORT | TRNAMT | مبلغ تراکنش (ریال):<br>مبالغ منفی نشانه برداشت از حساب و مبالغ مثبت نشانه واریز به حساب |
| PRTYP | مشخص‌کننده نوع حساب:<br>متمرکز = QJCA<br>فاقد دسته‌چک = QJNC<br>زرین = ZARIN<br>خاتم = QHSA<br>عادی = S91SA<br>کارت هدیه = CIFT<br>۹۶ = SJNC<br>پشتیبان جاری = POSA | TRNBRN | کد درگاهی که تراکنش در آن انجام شده است. |
| CIFKEY | کد مشتری | MANDEH | مانده حساب مشتریان در ابتدای بازه زمانی (ریال) |
| BIRTHDATE | تاریخ تولد مشتری | | |

### ✓ آماده‌سازی داده‌ها

این مرحله شامل گام‌های انتخاب داده، افزودن متغیرهای جدید، حذف متغیرهای زائد، شناسایی داده‌های مفقوده، توزیع فراوانی متغیرهای مسئله و تصویرسازی متغیرها است.

---

[1] چهار متغیر BRNCD، MODCD، ACNO و CHKDGT در کنار هم شماره حساب‌های بانکی را شکل می‌دهند.

**انتخاب داده‌ها**

در این مرحله به بررسی مجموعه داده‌ها پرداخته شده و با مشورت با خبرگان، اساتید، مدیریت بانک و کارشناسان بانک داده‌هایی که دارای خطا بودند از مجموعه داده‌ها حذف گردیدند. این خطاها عمدتاً ناشی از اشتباه اپراتور بانک در هنگام ورود اطلاعات یا اشتباهات سیستمی در محاسبات هستند. تراکنش‌های غیرعادی که نیاز به بررسی‌های بیشتر دارند به‌عنوان تراکنش‌های مشکوک به پول‌شویی توسط روش‌های تشخیص عوامل غیرعادی کشف و گزارش شد. درنهایت، تعداد تراکنش‌های باقی‌مانده برابر ۲۹۸۵۷ تراکنش و تعداد حساب‌های باقی‌مانده برابر ۳۰۲۶ حساب بانکی شد.

**افزودن متغیرهای جدید**

متغیر مانده حساب قبل و بعد از انجام تراکنش محاسبه و به مجموعه داده‌ها اضافه گردیدند. سپس متغیری با عنوان trn_failure به مجموعه داده‌ها اضافه شد که در مورد تراکنش‌هایی که ناموفق بودند برابر با یک و برای تراکنش‌های موفق برابر با صفر است. با ترکیب ۷ متغیر BRNCD، MODCD، ACNO، CHKDGT، PRLINE، PRTYP و CRLINE متغیر جدید ’acc_num‘ به‌عنوان شماره حساب دارای مقادیر ۱ تا ۳۰۲۶ تعریف شد که هر یک از این شماره‌ها متعلق به یک حساب بود. در مورد تاریخ و زمان تراکنش، از متغیر TRANSACTIONTIME استفاده شد. این متغیر به دو متغیر ساعت تراکنش و تاریخ تراکنش تفکیک گردید. متغیر کد تراکنش TRNCD دارای ۲۳ حالت است که با توجه به نظر کارشناسان بانکی به ۷ حالت کلی، تبدیل شد و سپس ۷ حالت به وجود آمده به ۷ متغیر دودویی تبدیل گردید.

**حذف متغیرهای زائد**

در این مرحله پس از بررسی مقادیر موجود در خصوص هر یک از متغیرها، برخی از متغیرها از مجموعه داده‌های موردمطالعه حذف شده‌اند. به‌عنوان‌مثال متغیر TRNCD که مشخص‌کننده نوع تراکنش است به ۷ متغیر دسته‌ای تبدیل شده بود؛ بنابراین وجود این متغیر در مجموعه داده‌ها زائد بوده و حذف گردید.

**شناسایی داده‌های مفقود**

با توجه به نوع داده‌ها و تکمیل آن‌ها با استفاده از سیستم‌های خودکار بانکی، هیچ‌یک از متغیرها دارای داده مفقوده نبودند و نیاز به حذف هیچ سطر یا ستونی از مجموعه داده نبود.

**توزیع فراوانی متغیرهای مورد بررسی**

در جدول ۲ توزیع فراوانی‌های متغیرهای کمّی مورد بررسی در این پژوهش قبل و بعد از استانداردسازی آورده شده است.

**جدول (۲): توزیع آماری متغیرهای مورد بررسی**

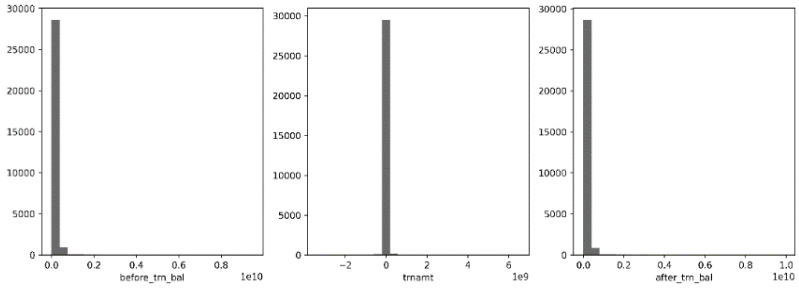
| نام متغیر | توزیع فراوانی | مقدار p-value | پارامترهای توزیع قبل از استاندارد شدن | پارامترهای توزیع بعد از استاندارد شدن |
|---|---|---|---|---|
| مانده حساب قبل از انجام تراکنش | نرمال لگاریتمی | 0 | Shape = 2.535<br>Location = -2927.554<br>Scale = 6107673.548<br>Mean = 69876435.936<br>Standard Deviation = 268117128.179 | Shape = 2.535<br>Location = - 0.261<br>Scale = 0.023<br>Mean = 0<br>Standard Deviation = 1 |
| مانده حساب بعد از انجام تراکنش | نرمال لگاریتمی | 0 | Shape = 2.516<br>Location = -2990.3443<br>Scale = 6251533.035<br>Mean = 70932428.253<br>Standard Deviation = 276197030.385 | Shape = 2.516<br>Location = - 0.257<br>Scale = 0.023<br>Mean = 0<br>Standard Deviation = 1 |
| مبلغ تراکنش | نرمال | 0 | 952137.174 = Mean<br>= Standard Deviation<br>88659771.464 | 0.000 Mean =<br>1 = Standard Deviation |

قابل ذکر است متغیرهای کمی این مسئله به دو علت استانداردسازی شده‌اند. دلیل اولی این است که مدل‌های داده‌کاوی و یادگیری ماشینی در شرایطی که متغیرهای ورودی آن‌ها استاندارد باشند (دارای میانگین صفر و انحراف معیار یک باشند) عملکرد بهتری دارند. دلیل دوم این است که با توجه به مقادیر بالای تراکنش‌های مالی و مانده‌حساب‌ها مقادیر مجموع مربعات خطای نمونه که در مراحل مدل‌سازی محاسبه می‌شوند مقادیر زیادی خواهند بود و تصمیم‌گیری در ارتباط با میزان عملکرد مدل را دشوار خواهند کرد.

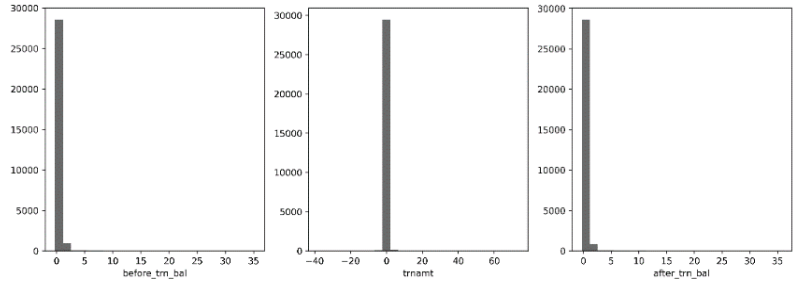
### ✓ تصویرسازی داده‌ها

نمودارهای هیستوگرام مربوط به سه متغیر کمّی پژوهش (مبلغ تراکنش 'trnamt'، مانده حساب قبل از انجام تراکنش 'before_trn_bal' و مانده حساب بعد از انجام تراکنش 'after_trn_bal') قبل و بعد از استانداردسازی در شکل‌های ۲ و ۳ نشان داده شده‌اند. در نمودارهای زیر منظور از 'before_trn_bal' مانده حساب قبل از انجام تراکنش، 'after_trn_bal' مانده حساب بعد از انجام تراکنش و 'trnamt' مبلغ

تراکنش است.




**شکل ۲: نمودار هیستوگرام متغیرهای کمّی قبل از استانداردسازی**




**شکل ۳: نمودار هیستوگرام متغیرهای کمّی بعد از استانداردسازی**

✓ **مدل‌سازی داده‌ها**

در این گام در مرحله اول با بهره‌گیری از الگوریتم‌های خوشه‌بندی و کشف موارد پرت در مجموعه داده‌ها، تراکنش‌های مالی که ممکن است مشکوک به پول‌شویی باشند شناسایی می‌شوند. سپس در مرحله دوم با توجه به شرایط مسئله و عدم وجود راه‌حلی جامع برای کشف این موارد راه‌حلی با استفاده از قانون بنفورد و الگوریتم GANs برای کشف موارد مشکوک به پول‌شویی در تراکنش‌های مالی افراد ارائه می‌شود و نتایج حاصل از این روش بررسی می‌شوند.

○ **مدل‌سازی با استفاده از خوشه‌بندی و کشف موارد پرت**

▪ **کاهش ابعاد مسئله توسط الگوریتم تحلیل مؤلفه‌های اصلی**

یکی از عمومی‌ترین روش‌های آماری به‌منظور کاهش ابعاد داده‌ها روش تحلیل مؤلفه‌های اصلی (PCA)[1] است (Ghazanfari et al., 2008). روش PCA در سال ۱۹۰۱ برای اولین بار توسط کارل پیرسن[2] ارائه شد. هدف PCA کاهش تعداد زیاد متغیرهای اصلی به تعداد کمی از مؤلفه‌های اصلی است (Azar & Khadivar, 2014). در این پژوهش، الگوریتم PCA با استفاده از زبان برنامه‌نویسی Python و با به‌کارگیری کتابخانه sklearn در این زبان، پیاده‌سازی شده است. در ابتدا ورودی‌های الگوریتم ۱۱ متغیر شامل ۴ متغیر trnamt، before_trn_bal، after_trn_brn، failure_trn و ۷ متغیر مربوط به نوع تراکنش هستند. از ۳ مؤلفه اصلی اول برای نمایش داده‌ها استفاده شده است زیرا آن‌ها خود حدود ۸۷ درصد از اطلاعات مسئله را پوشش داده و نماینده نسبتاً خوبی برای تصویرسازی داده‌ها خواهند بود.

**▪ خوشه‌بندی داده‌ها**

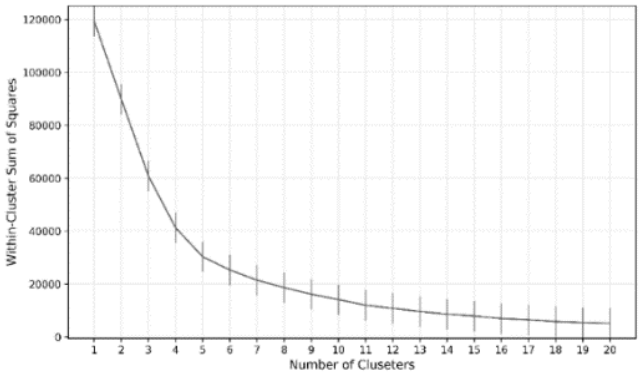
خوشه‌بندی نوعی عملیات داده‌کاوی غیرمستقیم است. در این روش هیچ دسته‌ای از قبل وجود ندارد و درواقع متغیرها به دوطبقه مستقل و وابسته تقسیم نمی‌شوند (Ghazanfari et al., 2008). روش‌های مبتنی بر خوشه‌بندی، خوشه‌های کوچک را به‌عنوان داده‌های پرت در نظر می‌گیرند. منظور از خوشه‌های کوچک، خوشه‌هایی هستند که میزان قابل‌توجهی نقاط داده کمتری نسبت به سایر خوشه‌ها دارند. این روش بدون ناظر است و می‌تواند پس از خوشه‌بندی، نقاط جدید را وارد و پرت بودن آن‌ها را مورد آزمایش قرار دهد (Kiani & Montazeri, 2015).

در این بخش با استفاده از الگوریتم خوشه‌بندی k-میانگین تراکنش‌های مالی افراد خوشه‌بندی می‌شوند. در روش خوشه‌بندی k-میانگین، مجموعه n شی‌ای را به k خوشه افراز می‌کند. پارامتر k به‌عنوان ورودی گرفته می‌شود (Berkhin, 2006). در این پژوهش از روش آرنج برای تشخیص تعداد خوشه‌ها (k) استفاده شده است. در شکل ۴ مقدار مجذور مربعات در خوشه‌ها (WCSS)[3] به ازای مقادیر k از ۱ تا ۲۰ نشان داده شده است. میزان کاهش در مقدار WCSS بعد از k=5 به‌شدت کاهش می‌یابد. در ابتدا شیب نمودار بسیار زیاد بوده و WCSS به‌شدت در حال کاهش است، اما بعد از k=5 این تغییر به نسبت نامحسوس می‌شود و شیب نمودار کاهش می‌یابد؛ بنابراین برای خوشه‌بندی داده‌ها از ۵ خوشه استفاده می‌شود. سپس

---

[1] Principal Component Analysis (PCA)
[2] Pearson
[3] Within-Cluster Sum of Squares (WCSS)

خوشه‌بندی بر مجموعه داده‌ها اعمال گردید. برای سنجش میزان عملکرد خوشه‌بندی از امتیاز کیفیت خوشه‌بندی سیلهوته استفاده شد. امتیاز کیفیت خوشه‌بندی سیلهوته در این خوشه‌بندی به عدد ۰/۷۹ می‌رسد که بیانگر آن است که در خوشه‌بندی نمونه‌ها با سایر اعضای خوشه خودشان به‌خوبی مطابقت دارند و از نمونه‌های سایر خوشه‌ها به‌خوبی جدا هستند.




**شکل ٤: نمودار حاصل از پیاده‌سازی روش آرنج**

**▪ تشخیص موارد پرت**

اعضای خوشه‌های ۴ و ۵ جمعاً ۲۴۲ نمونه بوده و ترکیب آن دو ۰/۸ درصد از تعداد کل نمونه‌ها را تشکیل می‌دهد. درنتیجه این دو خوشه به‌عنوان موارد پرت دسته‌ای (ناهنجاری‌های دسته‌ای) در نظر گرفته شد.

برای نمونه‌هایی که متعلق به خوشه‌های ۱، ۲ و ۳ هستند، آن نمونه‌هایی که دارای احتمال پایینی برای عضویت در خوشه هستند به‌عنوان موارد پرت نقطه‌ای (ناهنجاری‌های نقطه‌ای) در نظر گرفته شدند. برای یافتن این نمونه‌ها، ازآنجایی‌که در الگوریتم خوشه‌بندی k-means از معیار فاصله اقلیدسی استفاده می‌شود، برای هر خوشه ۰/۵ درصد از نمونه‌هایی که بیشترین فاصله اقلیدسی را تا مرکز خوشه خود دارند، به‌عنوان موارد پرت نقطه‌ای شناسایی شدند.

خوشه‌های ۴ و ۵ و ۰/۵ درصد نمونه‌های خوشه‌های ۱، ۲ و ۳ که بیش‌ترین فاصله اقلیدسی را با مرکز خوشه خود داشتند به‌عنوان موارد پرت مشخص شدند که ممکن است مشکوک به پول‌شویی باشند و نیاز به بررسی بیشتر دارند.

با توجه به پارامترهای در نظر گرفته شده، موارد پرت کشف شده تنها ۱/۳ درصد از کل تراکنش‌ها را تشکیل می‌دهند که این عدد مقدار مناسبی برای بررسی بیشتر این تراکنش‌ها خواهد بود. همچنین الزامی

وجود ندارد که این تراکنش‌ها حتماً مربوط به پول‌شویی باشند، در این پژوهش تنها به کشف موارد مشکوک به پول‌شویی پرداخته شده که ممکن است نیاز به بررسی بیشتر داشته باشند. نهایتاً تعداد تراکنش‌های مشکوک به پول‌شویی با استفاده از این الگوریتم برابر با ۳۸۹ تراکنش شد که توسط ۶۹ حساب بانکی انجام گرفته بودند. لازم به ذکر است که ازاین‌بین، تعداد ۲۶۴ تراکنش بانکی (۶۸٪) تنها توسط ۱۲ حساب بانکی انجام گردیده‌اند که بیش از دیگر حساب‌ها، مشکوک به پول‌شویی بوده و نیاز به بررسی‌های بیشتر توسط بانک مربوطه دارند.

**○ مدل‌سازی داده‌ها با استفاده از قانون بنفورد و با به‌کارگیری الگوریتم GANs**

یک راه‌حل برای کشف اعداد و ارقام ساختگی در مجموعه اعداد استفاده از قانون بنفورد است و بسیاری از محققان از این قانون برای تشخیص اعداد ساختگی از اعدادی که به شکل طبیعی تولید شده‌اند، در زمینه‌های مختلف استفاده نموده و جواب‌های مورد قبولی گرفته‌اند (Busta & Weinberg, 1998; Nigrini, 1996; Durtschi, Hillison & Pacini, 2004; Diekmann, 2007; Huang, Yen, Yang & Hua, 2008; Álvarez-Jareño, Badal-Valero & Pavía, 2017). در این بخش به استفاده از قانون بنفورد برای کشف تراکنش‌های مالی که ممکن است دارای اعداد ساختگی باشند و مشکوک به پول‌شویی باشند پرداخته می‌شود. داده‌های موردنیاز برای پیاده‌سازی مدل آماده‌سازی می‌شوند، سپس به ایجاد مدلی بر پایه هوش مصنوعی و با استفاده از قانون بنفورد و به‌کارگیری الگوریتم GANs پرداخته می‌شود و این روش به‌عنوان روشی نوین و بدون ناظر برای کشف موارد مشکوک به پول‌شویی در انبار داده بانک‌ها معرفی می‌شود.

**• محدودیت‌های به‌کارگیری قانون بنفورد**

از بین ۲۸ نوع تراکنش، تنها تراکنش‌های شماره ۱۹۸ و ۷۲۶ که مربوط به تراکنش‌های انجام گرفته توسط دستگاه‌های کارت‌خوان فروشگاهی هستند قابلیت بررسی توسط قانون بنفورد را دارند؛ بنابراین این دسته از تراکنش‌ها از مجموعه داده‌های بانکی برای بررسی بیشتر جدا شدند. فقط برای حساب‌هایی که حداقل ۵۰ مورد تراکنش مالی با شماره تراکنش ۱۹۸ و ۷۲۶ را دارا بودند تبعیت از قانون بنفورد بررسی شد.

**• آماده‌سازی داده‌های موردنیاز جهت پیاده‌سازی مدل**

برای بررسی وجود اعداد ساختگی در متغیر مبلغ تراکنش، تنها کافی است مبالغ تراکنش‌ها برای

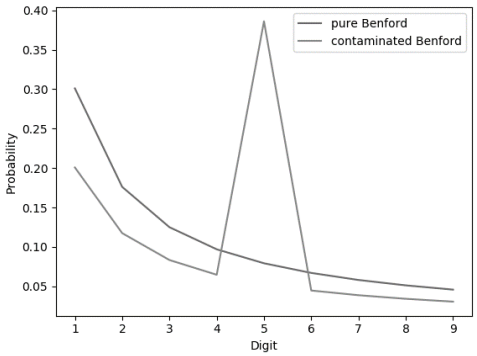
حساب‌های متفاوت جمع‌آوری شوند و مورد بررسی قرار گیرند. ساده‌ترین راه ممکن برای انجام این کار، جمع‌آوری تراکنش‌های انجام گرفته توسط هر یک از حساب‌ها و بررسی آماری میزان تطابق آن‌ها با قانون بنفورد است. با توجه به اینکه قانون بنفورد تنها با ارقام اول اعداد سر و کار دارد، برای بررسی تراکنش‌های مالی افراد توسط قانون بنفورد، تنها رقم اول تراکنش‌ها شبیه‌سازی شدند. تعداد تراکنش‌های حساب‌ها با توجه به تعداد واقعی تراکنش‌ها در مجموعه داده بانکی مورد بررسی، شبیه‌سازی شدند. از ورود حساب‌هایی که کمتر از ۵۰ عدد تراکنش انجام داده بودند به مدل جلوگیری شد. به‌عنوان داده‌های ورودی مسئله، دو دسته داده شبیه‌سازی شدند. داده‌های دسته اول تراکنش‌های مالی بودند که به‌صورت تصادفی انجام شده بودند اما توزیع رقم اول آن‌ها از قانون بنفورد پیروی می‌کرد. دسته دوم داده‌هایی بودند که آن‌ها نیز به‌صورت تصادفی انجام شده بودند اما به میزان M=0.5 و I=0.5 از قانون بنفورد پیروی نمی‌کردند؛ یعنی ۵۰ درصد از این تراکنش‌ها به میزان $\gamma = 0.5$ از قانون بنفورد انحراف داشتند. انتخاب مقادیر ۵۰ درصد و ۰/۵ با توجه به تحقیقات انجام شده توسط (Bhattacharya et al., 2011) انتخاب گردیدند. همچنین برای شبیه‌سازی این دسته از تراکنش‌ها، ابتدا ۵۰ درصد از تراکنش‌ها به‌گونه‌ای شبیه‌سازی شدند که از قانون بنفورد پیروی کنند، سپس برای ۵۰ درصد مابقی، یکی از ارقام ۱ تا ۹ به‌صورت کاملاً تصادفی انتخاب گردیدند و به میزان $\gamma = 0.5$ به مقدار توزیع بنفورد آن‌ها اضافه گردید و سپس به‌احتمال تکرار ارقام به‌گونه‌ای مقیاس داده شد که مجموع آن‌ها برابر با یک شود. در شکل ۵، نمودار احتمال رخداد ارقام با استفاده از قانون بنفورد و با داده‌های ساختگی برای رقم ۵ و $\gamma = 0.5$ نشان داده شده‌اند.

دسته دوم داده‌ها یا همان داده‌هایی که با داده‌های ساختگی ترکیب شده‌اند، تفاوت کاملاً زیادی با داده‌هایی که به شکل طبیعی به وجود آمده‌اند، دارند. انتخاب اعداد ۵۰ درصد و ۰/۵ برای متغیرهای I و M بدین منظور بود که این تفاوت بین داده‌های طبیعی و ساختگی وجود داشته باشد تا مدل هوش مصنوعی به‌خوبی بتواند تفاوت آن‌ها را آموزش ببیند. پس از آموزش اولیه مدل شبکه‌های عصبی مصنوعی از الگوریتم تولیدکننده در روش GANs جهت تولید داده‌های مصنوعی به‌گونه‌ای که الگوریتم متمایزکننده گمراه شود استفاده می‌کنند. برای هر سطر از داده‌ها (برای هر حساب) آزمون‌های برازش آماری خی‌دو و کولموگروف–اسمیرنف انجام شد تا میزان تشابه این داده‌ها با توزیع آماری بنفورد مقایسه گردد.

نتایج این آزمون‌ها به‌عنوان دو متغیر دودویی به مجموعه داده‌های جدید اضافه گردید به‌گونه‌ای که درصورتی‌که داده‌ها در هر یک از این دو آزمون شکست می‌خوردند، متغیر متناظر با آن آزمون در

مجموعه داده‌ها برابر ۱ خواهد بود و در غیر این صورت برابر صفر خواهد بود. همچنین تعداد تراکنش‌های هر حساب نیز به مجموعه داده‌ها اضافه گردید و ازآنجایی‌که این متغیر نسبت به سایر متغیرهای از واریانس خیلی بالایی برخوردار خواهد بود که باعث می‌شود در آموزش مدل تأثیر بسیار زیادی داشته باشد، بعدازآنکه مقادیر ذکر شده برای تمامی حساب‌ها محاسبه شدند، ستون مربوط به تعداد تراکنش‌ها استانداردسازی (نرمال‌سازی) شد بدین گونه که میانگین این ستون از تمامی مقادیر آن کم شد و سپس تمامی مقادیر به انحراف‌معیار آن ستون تقسیم شدند.

درنهایت مجموعه داده به ازای هر حساب دارای ۱۳ متغیر شد. ۹ متغیر نشان‌دهنده نسبت تکرار ارقام ۱ تا ۹، ۲ متغیر نشان‌دهنده وضعیت داده‌های آن سطر با توجه به آزمون‌های آماری خی‌دو و کولموگروف-اسمیرنف، یک متغیر مشخص‌کننده تعداد تراکنش‌های حساب و متغیر نهایی مشخص‌کننده طبیعی بودن آن سطر و یا وجود داده‌های ساختگی در آن است. شبیه‌سازی داده‌ها برای ۵۰۰۰۰ حساب انجام شد که ازاین‌بین نیمی از آن‌ها کاملاً از توزیع قانون بنفورد تبعیت می‌کردند و نیمی دیگر از آن‌ها به ازای متغیرهای $I = 0.5$ و $M = 0.5$ دارای داده‌های ساختگی بودند.




**شکل ۵: مقایسه داده‌های ساختگی و داده‌های طبیعی که از قانون بنفورد پیروی می‌کنند.**

**• مدل‌سازی با استفاده از الگوریتم GANs**

برای آموزش مدل GANs برای تشخیص تراکنش‌های مالی مشکوک به پول‌شویی و دارای ارقام ساختگی در مبلغ تراکنش‌ها، ابتدا یک مدل شبکه عصبی مصنوعی با استفاده از مجموعه داده‌ای که در قسمت قبل شبیه‌سازی شد آموزش داده می‌شود. برای این کار از یک شبکه عصبی مصنوعی ساده با یک لایه مخفی و نرخ یادگیری ۰/۰۰۱ با ۱۵۰ تکرار استفاده شد. مجموعه داده‌ها به دو قسمت آموزش و تست تقسیم شدند،

بدین گونه که ۸۰ درصد از داده‌های به‌صورت تصادفی با رعایت نرخ ۵۰ درصدی داده‌های مشکوک و ۵۰ درصدی داده‌های طبیعی در مجموعه آموزش و ۲۰ درصد مابقی در مجموعه تست قرار گرفتند. دقت این مدل در هنگام استفاده از داده‌های یادگیری برابر ۹۱/۷۳ درصد و با به‌کارگیری داده‌های تست برای آزمایش نتایج دارای دقت ۹۲/۹۱ درصد بود. از بین ۵۰۰۰ حساب دارای تراکنش‌های طبیعی و ۵۰۰۰ حساب دارای تراکنش‌های مشکوک، این مدل موفق به تشخیص درست ۴۷۳۵ مورد از حساب‌هایی که تراکنش‌های آن‌ها مشکوک نبودند و ۴۵۵۶ مورد از حساب‌هایی که دارای اعداد دست‌کاری شده بودند شد. سپس با به‌کارگیری الگوریتم GANs در هر تکرار این الگوریتم، بخش مولد تلاش کرد مجموعه داده‌هایی با توجه به داده‌های اولیه مسئله تولید کند که متمایزکننده قادر به تشخیص این امر نباشد که آیا در این مجموعه داده، اعداد به‌صورت طبیعی به وجود آمدند یا مجموعه دارای اعداد ساختگی است و متمایزکننده تلاش می‌کند در هر تکرار این موارد را بهتر شناسایی کند. انجام این کار توسط دو شبکه عصبی مصنوعی انجام شد که یک بازی حداکثر حداقل را بهینه می‌کنند. سپس این امکان وجود دارد که از مدل آموزش داده شده برای کشف موارد مشکوک به پول‌شویی در انبار داده واقعی تراکنش‌های بانکی استفاده شود. برای انجام این کار یک مدل GANs با ۱۵۰ تکرار و شبکه‌های عصبی با پارامترهای مشابه مدل قبلی پیاده‌سازی شد. این شبکه قادر نبود با دقت بالایی که شبکه عصبی در تشخیص موارد طبیعی و ساختگی با متغیرهای M=0.5 و I=0.5 داشت، موارد مشکوک را از موارد طبیعی تشخیص دهد، زیرا در هر تکرار بخش مولد احتمالاً اعداد مختلفی برای این دو متغیر انتخاب می‌کند که این امر باعث پیچیدگی مسئله و سخت شدن کار برای شبکه متمایزکننده می‌شود. درنهایت دقت مدل در داده‌های یادگیری به‌طور میانگین در تکرارهای مدل برابر ۶۰/۵ و در داده‌های تست به‌طور میانگین برابر ۶۱/۶ بود. احتمالاً درصد بالایی از مجرمانی که تلاش در انجام پول‌شویی و پنهان نمودن اعداد و ارقام ساختگی در تراکنش‌های مالی خود دارند، آن‌قدر در این کار پیشرفت نکرده‌اند که توسط قسمت اول شبکه عصبی با دقت بالای ۹۰ درصد به دام نیفتند.

**✓ارزیابی**

با ارائه مستندات لازم به بانک مورد بررسی، ارزیابی نتایج پژوهش به عهده کارشناسان بانکی قرار داده شد تا در صورت رضایت از نتایج از روش‌های بحث شده جهت جلوگیری از اعمال مجرمانه و پول‌شویی استفاده نمایند.

**✓توسعه**

بر اساس نتایج به‌دست‌آمده می‌توان یک طرح توسعه اولیه برای مدل پیشنهاد نمود؛ به این صورت که اطلاعات تراکنش‌های تمام مشتریان بانک از زمانی که افتتاح حساب نموده‌اند تا زمانی که واحد مبارزه با پول‌شویی در بانک بخواهد حساب‌ها را بررسی کند در پایگاه داده بانک ذخیره گردد.

## بحث و نتیجه‌گیری

پول‌شویی یکی از شریان‌های تجارت مجرمانه جهانی تلقی می‌شود؛ زیرا ناشی از فعالیت‌های اقتصادی ناسالم بوده و نقش اساسی آن ترغیب یا تسهیل فعالیت بزهکاران یا تقویت جرائم سازمان‌یافته است (Salehi & Ghazanfari, 2014). امروزه با ایجاد بسترهای ارتباطی بین مؤسسات مختلف، دیگر مانند گذشته ردیابی تراکنش‌های مختلف به‌راحتی و آسانی و تنها با تکیه بر تجربیات نیروی انسانی قابل‌اجرا نیست. این امر وجود سیستم‌های نرم‌افزاری هوشمند در اجرای سیاست‌های مبارزه با پول‌شویی در مؤسسات مالی را اجتناب‌ناپذیر می‌سازد (Masjidi, 2015). یکی از روش‌های مبارزه با پول‌شویی که در سال‌ها اخیر موردتوجه محققان و افراد و سازمان‌های ذینفع قرار گرفته است، استفاده از روش‌های داده-کاوی در کشف فعالیت‌های مشکوک به پول‌شویی است. در این پژوهش با به‌کارگیری روش‌های داده‌کاوی به کشف موارد مشکوک به پول‌شویی در مجموعه داده‌های یکی از بانک‌های کشور پرداخته شد. با به‌کارگیری روش تحقیق CRISP-DM ابتدا به شناخت سیستم و شناخت داده‌های موجود در پایگاه داده پرداخته شد. سپس در مرحله آماده‌سازی داده‌ها، داده‌های موردنیاز برای پیاده‌سازی الگوریتم‌های انتخاب گردیدند، متغیرهای جدید موردنیاز اضافه و متغیرهای زائد حذف گردیدند، داده‌های مفقوده و توزیع فراوانی داده‌ها مورد بررسی قرار گرفتند و تصویرسازی ابتدایی از مجموعه داده انجام گرفت. سپس در مرحله مدل‌سازی، برای ساخت مدل‌های داده‌کاوی و یادگیری ماشین برای کشف موارد مشکوک به پول‌شویی که نیاز به بررسی بیشتر توسط بازرسان و ممیزی‌های بانکی دارند، از دو رویکرد استفاده شد.

در رویکرد اول، با توجه به روش‌های مرسوم در کشف داده‌های پرت، ابتدا تراکنش‌های بانکی خوشه‌بندی شدند و سپس خوشه‌هایی که درصد کمی از داده‌ها را در خود جای می‌دادند، به همراه نمونه‌هایی از خوشه‌های چگال که نسبت به سایر نمونه‌های خوشه خود غیرعادی به نظر می‌رسیدند، به‌عنوان تراکنش‌های پرت که ممکن است مشکوک به پول‌شویی باشند و نیاز به بررسی بیشتر توسط کارشناسان بانکی دارند، معرفی گردیدند. برای خوشه‌بندی داده‌ها از الگوریتم خوشه‌بندی k میانگین استفاده گردید و تعداد خوشه‌ها به‌عنوان پارامتر ورودی الگوریتم با استفاده از روش آرنج برابر با ۵ خوشه

به دست آمد. همچنین از امتیاز سیلهوته برای سنجش میزان اعتبار این خوشه‌بندی استفاده شد که مقدار نتیجه این اعتبارسنجی برابر با امتیاز ۰/۷۹ سیلهوته بوده و بدین معناست که خوشه‌بندی به‌خوبی انجام گرفته است.

در رویکرد دوم کشف موارد مشکوک، با استفاده از قانون بنفورد و به‌کارگیری الگوریتم GANs به کشف حساب‌هایی که ممکن است در مبالغ تراکنش‌های آن‌ها اعداد ساختگی وجود داشته باشد، پرداخته شد. برای انجام این کار، پس از بررسی مختصر تحقیقات سایر محققینی که در این زمینه فعالیت کرده بودند و بررسی نقاط ضعف و قوت پژوهش‌های آن‌ها، همچنین بررسی محدودیت‌های به‌کارگیری قانون بنفورد، به شبیه‌سازی مجموعه داده‌ای که مناسب شرایط مسئله باشد پرداخته شد. مجموعه داده‌ای حاوی مبلغ تراکنش‌های ۵۰۰۰۰ حساب بانکی شبیه‌سازی شد که نیمی از آن‌ها فاقد اعداد ساختگی در مبلغ تراکنش و نیمی دیگر دارای اعداد ساختگی، با توجه به مقادیر متغیرهای میزان دست‌کاری و میزان درگیری بودند. سپس با استفاده از یک شبکه عصبی مصنوعی، به آموزش مدل برای تشخیص داده‌های طبیعی از دادهای ساختگی پرداخته شد. دقت این مدل حدود ۹۳ درصد بود. سپس با استفاده از یک شبکه GANs، به ایجاد داده‌های مصنوعی برای اعداد ساختگی و درعین‌حال افزایش قدرت مدل برای تشخیص موارد مشکوکی که در آن‌ها تلاش شده بود ساختگی بودن اعداد پنهان شود، پرداخته شد. این مدل نیز از قدرت بالایی برخوردار بود و قادر بود بیش از ۶۰ درصد حساب‌هایی که به‌صورت حرفه‌ای تلاش در پنهان نمودن داده‌های ساختگی داشتند را شناسایی کند. درنهایت نیز به گام‌های ارزیابی مدل و توسعه آن پرداخته شد.

این روش در مقایسه با پژوهش‌های قبلی که با استفاده از سایر روش‌های یادگیری ماشین توسط (Bhattacharya et al., 2011) انجام شده بود، در کشف این دسته از تراکنش‌ها از دقت بالاتری برخوردار است. همچنین در کشف موارد مشکوکی که در پنهان نمودن ارقام ساختگی در مبلغ تراکنش‌های آن‌ها از روش‌های پیشرفته استفاده شده باشد، این روش دقتی در حدود ۶۰ درصد دارد.

در استفاده از قانون بنفورد و الگوریتم یادگیری عمیق GANs برای کشف موارد مشکوک به پول‌شویی پیشنهاد می‌شود تا دقت این روش در تشخیص مجرمان افزایش یابد. الگوریتم GANs به‌تازگی معرفی گردیده است و محققین بیشتر در حال بهبود این الگوریتم در زمینه پردازش تصویر هستند؛ بنابراین پیشنهاد

می‌شود محققان علوم کامپیوتر به بهبود این الگوریتم به‌طور خاص برای مبارزه با پول‌شویی بپردازند. همچنین توصیه می‌شود برای مبارزه با پول‌شویی از روش‌های سری‌های زمانی[1] برای تشخیص موارد پرت زمینه‌ای[2]، روش‌های کشف الگو[3] و تحلیل شبکه[4] استفاده شود.

**References**

Ahmed, M.; Mahmood, A. N. & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, *55*, 278-288.

Alexandre, C. R., & Balsa, J. (2023). Incorporating machine learning and a risk-based strategy in an anti-money laundering multiagent system. *Expert Systems with Applications*, 217, 119500.

Álvarez-Jareño, J. A.; Badal-Valero, E., & Pavía, J. M. (2017). Using machine learning for financial fraud detection in the accounts of companies investigated for money laundering. Economics Department, Universitat Jaume I, Castellón (Spain).

Asadi, M. (2015). Detection of money laundering in the banking system using genetic algorithms and neural networks. *International Conference on New Research in Engineering Sciences*. Tehran. (in Persian)

Azar, A., & Khadivar, A. (2014). *Application of multivariate statistical analysis in management*. Tehran: Negahe Danesh. (in Persian)

Badal-Valero, E.; Alvarez-Jareño, J. A., & Pavía, J. M. (2018). Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case. *Forensic science international*, *282*, 24-34.

[1] Time Series
[2] Contextual Anomalies
[3] Pattern Recognition
[4] Network Analysis

Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Springer, Berlin, Heidelberg.

Bhattacharya, S.; Xu, D. & Kumar, K. (2011). An ANN-based auditor decision support system using Benford's law. *Decision support systems, 50*(3), 576-584.

Busta, B., & Weinberg, R. (1998). Using Benford's Law and neural networks as a review procedure. *Managerial Auditing Journal*. *13*(6), 356-366.

Diekmann, A. (2007). Not the first digit! using benford's law to detect fraudulent scientif ic data. *Journal of Applied Statistics, 34*(3), 321-329.

Didimo, W.; Liotta, G., & Montecchiani, F. (2014). Network visualization for financial crime detection. *Journal of Visual Languages & Computing, 25*(4), 433-451.

Durtschi, C.; Hillison, W. & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of forensic accounting*, *5*(1), 17-34.

Farshadinia, M. & Basiri Ghaemi Pasand, A. (2016). Proposing a novel method to detect money laundering using data mining, *11th International Conference on Accounting and Management and 7th Conference on Entrepreneurship and Open Innovation*. Tehran, Mehr Ishraq. (in Persian)

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.

Ghazanfari, M.; Alizadeh, S. & Teymour Pour, B. (2008). *Data mining and knowledge discovery*. Tehran: Iran University of Science and Technology. (in Persian)

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*. 2672-2680.

Huang, S. M.; Yen, D. C.; Yang, L. W. & Hua, J. S. (2008). An investigation of Zipf's Law for fraud detection. *Decision Support Systems*, *46*(1), 70-83.

James, G.; Witten, D.; Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning*, 112, New York: springer.

Jantani, M. (2017). Identifying the role of money control systems in preventing electronic money laundering on the website of the Court of Audit. *Studies of Economy, Financial Management and Accounting, 3*(2/2), 69-61. (in Persian)

Kiani, R. & Montazeri, M. (2015). An overview on anomaly detection methods. *International Conference on Research in Science and Technology*. Tehran, Karin Conference. (in Persian)

Kumar, A.; Das, S. & Tyagi, V. (2020). Anti money laundering detection using Naïve Bayes classifier. *IEEE International Conference on Computing, Power and Communication* Technologies, 568-572.

Lokanan, M. E. (2022). Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks. *Journal of Applied Security Research*, 1-25.

Martínez-Sánchez, J. F.; Cruz-García, S. & Venegas-Martínez, F. (2020). Money laundering control in Mexico: a risk management approach through regression trees (data mining). *Journal of Money Laundering Control*. *23*(2), 427-439.

Masjidi, A. (2015). Electronic money laundering and a case study of data mining methods in its prevention of that. *International Conference on Applied Research in Information Technology, Computer and Telecommunications*. Torbat-e Heydarieh, Khorasan Razavi Telecommunication Company. (in Persian)

Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law. *The Journal of the American Taxation Association*, *18*(1), 72.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559-572.

Raschka, S. & Mirjalili, V. (2017). *Python machine learning*. Packt Publishing Ltd.

Salehi, A. & Ghazanfari, M. (2014). Introducing and reviewing the data mining methods to identify money laundering in electronic banking. *The Second National Conference on Applied Research in Computer Science and Information Technology*. Tehran. University of Applied Science and Technology. (in Persian)

Sarraf, F. & Heidari, B. (1394). The need for proper implementation of internal control, auditing and training. Fourth National Conference and Second International Conference on Accounting and Management. Tehran. (in Persian)

Seddighi, A. & Sajedinejad, A. (2009). A Deep Learning Approach to Fraud Detection in Financial Payment Services. *Journal of Information Management*, *5*(1). 166-182. (in Persian)

Suresh, C.; Reddy, K. T. & Sweta, N. (2016). A hybrid approach for detecting suspicious accounts in money laundering using data mining techniques. *International Journal of Information Technology and Computer Science*, 8(5), 37-43.

Taghva, M.; Mansouri, T.; Feizi, K. & Akhgar, B. (2016). Fraud Detection in Credit Card Transactions; Using Parallel Processing of Anomalies in Big Data. *Journal of Information Technology Management*, 8(3), 477-498. (in Persian)

Zhang, Y. & Trubey, P. (2019). Machine learning and sampling scheme: An empirical study of money laundering detection. *Computational Economics*, *54*(3), 1043-1063.