Journal of Information Technology Management
ISSN: 2008-5893
EISSN: 2423-5059
Vol. 9, No. 1; PP. 1 - 20
Spring 2017

University of Tehran
Faculty of Management

An Intelligent System for Fraud Detection in Coin Futures Market's Transactions of Iran Mercantile Exchange Based on Bayesian Network

Amir-Reza Abtahi [1], Fatemeh Elahi [2], Reza Yousefi-Zenouz[3]

Abstract: In order to gain more illicit profit, some traders in the stock market try to make a targeted impact on prices by placing fake orders and false advertising. Due to the high customer population, it is not possible to discover these frauds using traditional methods. The present study seeks to provide a system for preventing the frauds in future market-trading coins based on Bayesian classifier model for Iran Mercantile Exchange. The proposed model has polynomial time complexity and high accuracy because of considering important dependencies among different features of data. The primary labeling of data has been done by Kmeans clustering. The test of model shows 94.55 percent similarity between model's output and labeled data. Using this system can helps to identify the fraudulent from non-fraudulent traders.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

Key words: *Bayesian network, Fraud detection, Futures contract, Induction behavior, Mercantile exchange.*

1. Assistant Prof., Dep. of IT., Faculty of Management, Kharazmi University, Tehran, Iran
2. MSc. Student in Decision Sciences and Knowledge Engineering, Faculty of Management Kharazmi University, Tehran, Iran
3. Assistant Prof., Dep. of IT., Faculty of Management, Kharazmi University, Tehran, Iran

Submitted: 14 / June / 2016
Accepted: 25 / February / 2017
Corresponding Author: Amir-Reza Abtahi
Email: abtahi@khu.ac.ir

مدیریت فناوری اطلاعات
Journal of Information Technology Management
دانشکدهٔ مدیریت دانشگاه تهران
دورهٔ ۹، شمارهٔ ۱
بهار ۱۳۹۶
صص. ۲۰- ۱

طراحی نوعی سیستم هوشمند برای کشف تقلب حاصل از حرکت القایی در معاملات بازار قراردادهای آتی سکهٔ بورس کالای ایران مبتنی بر شبکهٔ بیزی

سید امیررضا ابطحی[۱]، فاطمه الهی[۲]، رضا یوسفی زنوز[۳]

چکیده: در بازار سهام، برخی سوداگران به‌منظور کسب منفعت بیشتر سعی می‌کنند به‌وسیلهٔ سفارش‌های صوری، سبب تأثیر هدفمند بر نوسان قیمت‌ها شوند و از وضعیت پیش‌آمده در قیمت‌ها سوءاستفاده کنند که به این عمل، حرکت القایی گفته می‌شود. به‌دلیل جمعیت آماری زیاد مشتریان، امکان نظارت چشمی یا نظارت سیستمی از طریق روش‌های سنتی برای کشف حرکت القایی وجود ندارد. پژوهش حاضر به‌دنبال ارائهٔ سیستمی برای پیشگیری از وقوع حرکت القایی در معاملات بازار قراردادهای آتی سکهٔ بورس کالای ایران مبتنی بر مدل‌سازی طبقه‌بند بیزی است. سیستم ارائه‌شده ضمن برخورداری از پیچیدگی زمانیِ چندجمله‌ای، به‌دلیل لحاظ‌کردن وابستگی‌های مهم میان ویژگی‌های مختلف داده، دقت زیادی دارد. برچسب‌گذاری اولیهٔ داده‌ها با استفاده از خوشه‌بندی کا ـ میانگین انجام گرفته است و آزمون مدل، تشابه ۹۴/۵۵ درصدی را میان نتیجهٔ روش بیزی پیشنهادی و داده‌های برچسب‌گذاری‌شده نشان می‌دهد. استفاده از سیستم نوین ارائه‌شده به شناسایی افراد فریبکار در معاملات بازار از افراد سالم کمک به‌سزایی می‌کند.

واژه‌های کلیدی: *رفتار القایی، سیستم هوشمند، شبکهٔ بیزی، قرارداد آتی، کشف تقلب.*

۱. استادیار گروه مدیریت فناوری اطلاعات، دانشکدهٔ مدیریت دانشگاه خوارزمی، تهران، ایران
۲. دانشجوی کارشناسی ارشد مهندسی دانش و علوم تصمیم، دانشکدهٔ مدیریت، دانشگاه خوارزمی، تهران، ایران
۳. استادیار گروه مدیریت فناوری اطلاعات، دانشکدهٔ مدیریت دانشگاه خوارزمی، تهران، ایران

تاریخ دریافت مقاله: ۱۳۹۵/۰۳/۲۵
تاریخ پذیرش نهایی مقاله: ۱۳۹۵/۱۲/۰۷
نویسندهٔ مسئول مقاله: سید امیر رضا ابطحی
E-mail: abtahi@khu.ac.ir

مقدمه

بحث فریبکاری به‌منظور منافع فردی در بازارهای سرمایه، همواره مطرح است و شدت و ضعف آن در بازارهای مختلف با توجه به بهره‌مندی آنها از قوانین مناسب ضد دستکاری و ابزارهای نظارتی، تفاوت دارد (مدیریت مطالعات اقتصادی و برنامه‌ریزی و آموزش شرکت بورس کالای ایران، ۱۳۹۱)؛ این موضوع در بازار قراردادهای آتی نیز مشاهده می‌شود. اگر بازار بورس کارایی لازم در انجام کارکردهای خود را نداشته نباشد، اختلالات عمده‌ای در سطوح اقتصاد ملی و فراملی ایجاد خواهد کرد. در این میان، افرادی بدون رعایت اخلاق در معاملات، جریان واقعی عرضه و تقاضا را منحرف می‌کنند و با فریب دیگران زمینهٔ کسب سود خود را فراهم می‌آورند. مقابله با این اعمال، به شناسایی و پیشگیری کارشناسان نظارت بر بازار بورس نیاز دارد.

آمار معاملات قراردادهای آتی در ایران، نشان می‌دهد حجم و ارزش این معاملات هر سال نسبت به سال قبل افزایش بسیار چشمگیری را تجربه کرده است. ناظران نیز برای بررسی، افزون بر روش‌هایی چون بایگانی و کنترل اسناد مربوط به قراردادها، نظارت بر اخبار منتشرشده و سیستم نظارت، از برنامه‌های نرم‌افزاری نیز استفاده می‌کنند (مدیریت مطالعات اقتصادی و برنامه‌ریزی و آموزش شرکت بورس کالای ایران، ۱۳۹۱). ناظران پس از آگاهی از معاملات مشکوک، برای اطمینان، سابقهٔ داد و ستدها را بررسی می‌کنند. در حال حاضر نرم‌افزارهایی که برای این کار استفاده می‌شوند، با ضعف‌هایی مانند نادیده‌انگاشتن وابستگی بین متغیرها و روابط علّی بین ویژگی‌های تراکنش‌های روزانه و استدلال‌های احتمالی مواجه‌اند. البته با توجه به سرعت روبه‌رشد بازار بورس، فریبکاری در رفتار معاملاتی متنوع است و با توجه به تخصصی‌بودن بحث، در این پژوهش صرفاً به حرکت القایی در معاملات قراردادهای آتی سکه در بورس کالای ایران پرداخته می‌شود. از آنجا که پراکندگی داده‌های بورس متنوع و همچنین توزیع داده‌های مثبت و منفی بین مجموعهٔ داده‌ها نامتوازن است، باید به دنبال آن بود تا راهکار مناسبی برای کشف الگوهای پنهان و ضمنی پیدا کرد. انجام این کار در بستر کشف دانش از روی داده‌ها امکان‌پذیر است. پس از مطالعه در زمینهٔ رفتار القایی، مدلی ارائه شد که شرایط زیر را محقق ساخته است:

- روابط بین متغیرها و میزان تأثیر هریک از آنها را بر یکدیگر مشخص می‌کند؛
- توانایی استدلال‌ها و استنباط‌های احتمالی را دارد؛
- می‌تواند پس از آموزش و یادگیری، به‌عنوان تصمیم‌گیرنده‌ای مطمئن افراد مشکوک را از سالم تفکیک کند.

بنابراین، پژوهش حاضر در پی پاسخ به این سؤال است که چگونه می‌توان سیستمی طراحی کرد که به‌کمک آن از وقوع تقلب در رفتارهای القاییِ معاملات بورس کالا جلوگیری کرد؟

پیشینهٔ نظری پژوهش

طبقه‌بندی عبارت است از تخصیص رکوردها یا هر مجموعه‌ای از اشیا به مجموعهٔ مشخصی از طبقه‌ها (دسته‌ها) (میتشل، ۱۹۹۷: ۱۹۸). همان‌گونه که گفته شد، در این مقاله به‌دنبال طراحی سیستمی برای شناسایی احتمال وقوع رفتارهای القایی و طبقه‌بندی مشتریان روزانهٔ بورس به دو دستهٔ افراد سالم و متخلّف هستیم. قبل از پرداختن به روش اجرای پژوهش، ابتدا لازم است دربارهٔ مفاهیم کلیدی تحقیق همچون، شبکه‌های بیزی، خوشه‌بندی و قرارداد آتی شرح مختصری بیان شود.

شبکه‌های بیزی

شبکهٔ بیزی، نوعی ساختار گرافیکی (GMs)[1] است که امکان بیان و استدلال دربارهٔ حوزه‌ای غیرقطعی را میسر می‌کند. اساس روش یادشده بر این اصل استوار است که برای هر کمیت، یک توزیع احتمال وجود دارد و با مشاهدهٔ دادهٔ جدید و استدلال دربارهٔ توزیع احتمال آن، می‌توان تصمیمات بهینه‌ای اتخاذ کرد (میتشل، ۱۹۹۷). از نظر گرافیکی، این شبکه‌ها گراف‌های جهت‌دار بدون دور هستند، مجموعهٔ متناهی $(X_1, \dots, X_n)$ نشان‌دهندهٔ رأس‌های گراف است و در واقع همان متغیرهای فضای مسئله‌اند. ساختار شبکهٔ بیزی باید رابطهٔ کیفی بین متغیرها را به‌درستی نشان دهد. به‌دلیل برقراری حالت مارکوف در شبکه‌های بیزی، توزیع احتمال توأم این شبکه‌ها به‌صورت زیر محاسبه می‌شود:

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i | \Pi X_i) \qquad \text{رابطهٔ ۱)}$$

در رابطهٔ ۱، ΠX_i نشان‌دهندهٔ گره‌های والدگرهٔ X_i در شبکه است (هکرمن، گیگر و چیکرینگ، ۱۹۹۵).

در شبکه‌های بیزی مفهومی با عنوان یادگیری وجود دارد. یادگیری، فرایندها و تکنیک‌هایی را دربرمی‌گیرد که یک شبکه با استفاده از آنها، به‌صورت خودکار می‌تواند ایجاد شود و عملکرد

1. Graphical Models

خود را بهبود بخشد. در شبکه‌های بیزی یادگیری به دو صورت کلی ساختاری و پارامتری انجام می‌پذیرد:

۱. یادگیری ساختاری: این شیوهٔ یادگیری به‌دنبال پیداکردن ساختار بهینه برای شبکهٔ بیزی مد نظر، از نظر موقعیت گره‌ها و روابط کیفی آنها با یکدیگر توسط یال‌های جهت‌دار است. این کار در دو مرحلهٔ کلی جست‌وجو در فضای نمونهٔ مدل‌های علّی و جست‌وجو برای یافتن معیارهای قابل قبول برای ارزشیابی شبکه‌ها انجام می‌پذیرد.

۲. یادگیری پارامتری: در این شیوه، از نمونه‌ها برای برآورد پارامترها (توزیع احتمال شرطی)[1] استفاده می‌شود (نیلسن و جنسن، ۲۰۰۹: ۵۷). به‌طور کلی، داده‌ها را می‌توان به دو دسته داده‌های کامل[2] و داده‌های ناکامل[3] دسته‌بندی کرد و برای هریک به برآورد پارامترها پرداخت (جدول ۱)

جدول ۱. انواع داده‌ها و ساختارهای شبکه

مدل	ساختار معلوم	ساختار نامعلوم
داده‌های کامل	برآورد پارامترهای آماری	بهینه‌سازی روی ساختار[4]
داده‌های ناکامل	بهینه‌سازی پارامترها	روش‌های ترکیبی[5]

با توجه به مسئلهٔ بررسی‌شده، برای یادگیری پارامتری شبکه با حالت ساختار معلوم و داده‌های کامل مواجه‌ایم.

استنتاج در شبکه‌های بیزی

در اغلب سیستم‌های هوشمند نیاز است که براساس تعدادی از مشاهدات، احتمال وقوع رویداد برآورد شود. در سیستم‌های دسته‌بندی، احتمال عضویت یک شیء در هریک از دسته‌ها، براساس ویژگی‌های شیء مد نظر مشخص می‌شود. به فرایند پاسخ به این درخواست‌ها، استنتاج[6] گفته می‌شود، هر فرایند استنتاج به داده‌های قلمرویی[7] نیاز دارد که قصد کنترل عدم قطعیت آن را

1. Conditional Probabilities Distribution (CPD)
2. Complete Data
3. Incomplete Data
4. Optimization over structure
5. Combined
6. Inference
7. Domain

داریم. استنتاج در شبکه‌های بیزی به دو نوع استنتاج در ساختار و استنتاج در پارامترهای احتمالاتی دسته‌بندی می‌شود (نیلسن و جنسن، ۲۰۰۹: ۱۲۰).

الگوی طبقه‌بندی بیزی ساده

یکی از روش‌های بسیار کاربردی در یادگیری بیزی، روش یادگیرندهٔ سادهٔ بیزی است که عموماً روش طبقه‌بندی سادهٔ بیزی نامیده می‌شود.

طبقه‌بندی سادهٔ بیزی برای مسائلی کاربرد دارد که هر نمونهٔ x در آن توسط مجموعه‌ای از مقادیر صفات و تابع هدف $f(x)$ از مجموعه‌ای مانند V انتخاب می‌شود. روش بیزی برای طبقه‌بندی نمونهٔ جدید، محتمل‌ترین طبقه یا مقدار هدف V_{MAP} را با داشتن مقادیر صفات $< a_1, a_2, \dots, a_n >$ که توصیف‌کنندهٔ نمونهٔ جدید است، شناسایی می‌کند (باربر، ۲۰۱۰: ۱۰۲).

رابطهٔ ۲) $$V_{MAP} = \arg max \ P(V_j | a_1, a_2, \dots, a_n)$$

خوشه‌بندی کا – میانگین[1]

در سال ۱۹۷۵ هارتیگان برای اولین‌بار الگوریتم کا ــ میانگین را ارائه داد و در سال ۱۹۷۹ به‌کمک ونگ تغییراتی در آن ایجاد کرد و اکنون متداول‌ترین ابزار خوشه‌بندی استفاده‌شده در کاربردهای صنعتی و علمی است. در این روش، خوشه‌ها با مراکزشان که معمولاً میانگین نقاط درون یک خوشه است، بیان می‌شوند. در این روش فاصلهٔ هر نقطه تا مرکز آن خوشه، به‌عنوان تابع هدف در نظر گرفته می‌شود. فاصله می‌تواند تعاریف مختلف و گسترده‌ای را شامل شود. هر نقطه به خوشه‌ای تعلق دارد که به مرکز جرم آن نزدیک‌تر است. در این روش تعداد خوشه‌ها (K) باید مشخص باشد. کلیت الگوریتم به‌صورت زیر است:

۱. انتخاب K نقطه به‌عنوان مراکز اولیهٔ خوشه‌ها؛
۲. تخصیص هر نقطه به خوشه‌ای که به مرکز آن خوشه نزدیک‌تر است؛
۳. محاسبهٔ مجدد مراکز خوشه‌ها تا جایی که تکرار مراحل یادشده، تغییری در خوشه‌ها و مراکز آنها ایجاد نکند.

1. K-Means

قرارداد آتی

قراردادی است که می‌توان به‌عنوان شرط ضمن عقد لازم معاملهٔ سلف آورد و به موجب آن، عرضه‌کنندهٔ کالا (فروشنده) هنگام معاملهٔ سلف به مقدار وجه دریافتی، تعهد می‌کند که مقدار معینی از همان کالا با مشخصات منطبق بر استاندارد بورس را در زمان معین و با قیمت مشخص به طرف قرارداد تحویل دهد و خریدار نیز تعهد خرید می‌دهد. هر طرفی که از انجام تعهد تخلف کند، باید مبلغ معینی به طرف دیگر بپردازد (شرکت بورس کالای ایران، ۱۳۸۶).

پیشینهٔ تجربی

مطالعات مختلفی در زمینهٔ پیش‌بینی دستکاری قیمت، معاملات کلاه‌برداری[1] و لایه‌بندی[2] انجام شده است. برخی از این مطالعات به شرح زیر است:

مادا و فردوسی (۲۰۰۶) برای تشخیص رفتار نامعمول معاملاتی، الگوریتم تشخیص تقلب بدون ناظر برای داده‌های سری زمانی را ارائه کردند و با اعمال روش بدون ناظر تحلیل گروه همسان، عملکرد زیادی را از روش پیشنهادی خود گزارش کردند.

فرانک، هوزر و اسکرودر (۲۰۰۸) با هدف تحلیل رفتار معاملاتی نامنظم در بازار سهام، روشی را طراحی و اجرا کردند که در آن سه گام اصلی تشکیل گراف نمایش‌دهندهٔ تجارت بین سرمایه‌گذاران و تاجران، تشکیل ماتریس همجواری هرمیتی و تحلیل نهایی براساس روش تحلیل سیستم ویژه[3] وجود داشت. مزیت مهم این پژوهش، تشخیص نوع خاصی از تقلب در معاملات با دقت زیاد است.

چن و همکارانش در مقاله‌ای با عنوان «تشخیص داده‌های پرت مجاور»، الگوریتمی را برمبنای ترکیب روش‌های محاسبات گرانول و الگوریتم تشخیص داده‌های پرت، پیشنهاد کردند. این الگوریتم عملکرد بهتری نسبت به روش‌های نزدیک‌ترین همسایه کا- ام[4]، الگوریتم تشخیص داده‌های پرت مبتنی بر فاصله[5] و شبکه‌های عصبی تکرارشونده،[6] برای مجموعه داده‌های مختلط داشت (چن، میائو و ژانگ، ۲۰۱۰).

فلاح شمس و کردلوئی (۱۳۹۰) پژوهشی با عنوان «آزمون مدل‌های لاجیت و شبکهٔ عصبی مصنوعی برای پیش‌بینی دستکاری قیمت در بورس اوراق بهادار تهران» انجام دادند. آنها در این

---

1. Spoof Trading
2. Layering
3. Eigen System
4. K- Nearest Neighbor
5. Distance Based Outlier Detection
6. Replicator Neural Network

پژوهش به بسط و توسعهٔ روش‌های مبتنی بر ساختارهای شبکه‌ای و اقتصادسنجی پرداختند و به تشخیص قیمت‌های دستکاری‌شده در بورس اوراق بهادار تهران دست یافتند.

در تحقیقی با عنوان «طراحی الگوی پیش‌بینی دستکاری قیمت در بورس اوراق بهادار تهران» فلاح شمس، کردلوئی و رشنو (۱۳۹۱) به طبقه‌بندی و تفکیک گروه‌ها به‌منظور پیش‌بینی دستکاری قیمت‌ها در بورس اوراق بهادار پرداختند. روش پیشنهادی آنها برمبنای تحلیل داده‌های غیرخطی با استفاده از روش آماری تجزیه و تحلیل مؤلفه‌های اصلی[1] و طراحی مدلی با استفاده از مدل ماشین بردار پشتیبان در انجام پیش‌بینی بوده است.

صفری، حشمتی‌پور، مهرابی و نصابی (۱۳۹۱) در پژوهشی به ارائهٔ مدلی با استفاده از روش‌های ترکیبی نگاشت علّی و شبکه‌های بیزی برای تعیین عوامل مؤثر بر به اشتراک‌گذاری اطلاعات در زنجیرهٔ تأمین شرکت ایران خودرو پرداختند.

کیم و سان (۲۰۱۲) با استفاده از روش‌های یادگیری بدون ناظر- تحلیل گروه همسان[2]، برای تشخیص الگوهای مشکوک دستکاری قیمت سهام از سیستم‌های خبره بهره بردند. پژوهشگران توانستند به بهبود کارایی تحلیل گروه همسان از طریق به‌کارگیری اوزان و به‌هنگام‌آوری آنها و تشخیص محلی[3] معاملهٔ غیرمعمول انجام‌گرفته از طریق الگوهای یادگرفته به‌جای مقایسه با کل جمعیت دست یابند.

وثوق، تقوی‌فرد و البرزی (۱۳۹۳) در مدلی برمبنای شبکهٔ عصبی مصنوعی، تراکنش‌ها را به انواع سالم و متقلب طبقه‌بندی کردند که عملکرد آنها روایی و پایایی زیادی داشت.

پانیامورتی و توپان (۲۰۱۳) پژوهشی را به‌منظور نظارت و بررسی بازار سهام و مروری بر برنامه‌های کاربردی اجرا کردند. آنان پیاده‌سازی شبکهٔ عصبی که تعیین و به‌روزرسانی اوزان آن مبتنی بر الگوریتم ژنتیک بود را برای روش پژوهش انتخاب کردند. محققان توانستند به نظارت بر بازار بورس و طبقه‌بندی شواهد از نوع تراکنشی برای تشخیص فعالیت‌های بالقوهٔ فریبکارانه با استفاده از مدل هیبریدی دست یابند.

لی، اوم و پارک (۲۰۱۳) در مقاله‌ای با عنوان «دستکاری مبتنی بر زیرساختار: رفتار استراتژیک و معامله‌گران فریبکار» به تعیین الگوهای عام و خاص معامله‌گران فریبکار با استفاده از روش‌های تحلیلی و مبتنی بر زیرساختار پرداختند.

---

1. Principal Component Analysis
2. Peer Group Analysis
3. Local

گل‌محمدی، دایاز و زایان (۲۰۱۴) دستکاری بازار سهام را با استفاده از الگوریتم‌های یادگیری با ناظری مانند ماشین‌های بردار پشتیبان تشخیص دادند. آنها تراکنش‌های مشکوک به دستکاری در بازار و عملکرد بهتر شبکهٔ بیزی ساده نسبت به سایر روش‌ها را در مقایسهٔ چند روش تشخیصی با یکدیگر گزارش کردند.

اولزوسکی روشی طراحی کرد که در آن ماتریس به‌دست‌آمده از فعالیت‌های حساب کاربری که چندبعدی است، به یک بردار تبدیل می‌شود؛ سپس از نگاشت خودسازمان‌ده[1] برای تصویرسازی استفاده می‌کند و بعد از تصویرسازی با الگوریتم کلاس‌بندی مبتنی بر مقدار حد آستانه[2]، مورد مشکوک و کشف تقلب را تشخیص می‌دهد. روش ارائه‌شده به‌دلیل ترکیب روش تصویرسازی داده‌ها از طریقی مثل نگاشت خودسازمان‌ده با الگوریتم کلاسه‌بندی، به نتایج بهتر و رضایت‌بخشی نسبت به سایر روش‌ها دست یافته است (اولزوسکی، ۲۰۱۴).

یکی از روش‌های کشف تقلب متکی بر پایگاه دانش مستخرج از دانش خبرگان، استفاده از سیستم‌های خبرهٔ فازی است؛ تقوی‌فرد و جعفری (۱۳۹۴) در پژوهش خود از این مدل برای کشف تقلب در بیمهٔ بدنهٔ خودرو استفاده کردند.

تقوا، منصوری، فیضی (۱۳۹۵) در پژوهشی با رویکرد پردازش موازی و راه حل نگاشت کاهش، از نوعی شبکهٔ عصبی مدل کوهنن برای کشف ناهنجاری در تراکنش‌های کارت‌های بانکی استفاده کردند.

یائو، ژای، کائو و دینگ (۲۰۱۵) چارچوبی برای پیش‌بینی دستکاری قیمت ارائه کردند. در این چارچوب به‌کمک مدل‌های ایستا و پویا، به تعیین الگوهای غیرمعمول تراکنش‌های معاملاتی برمبنای روابط بین تراکنش‌ها و بدون داشتن هیچ فرض اولیه، پرداخته می‌شود.

مدل مفهومی

فرایند تعیین مشتریان سالم یا متخلف بازار در هر روز، بسیار پیچیده و حساس است. بنابراین، برای مشخص‌کردن قانونی یا غیرقانونی بودن عملیات مشتریان، علاوه‌بر اجرای سیستم پیشنهادی و تعیین خروجی آن، تأیید نهایی کارشناسان بازار ضروری است. وضعیت خاص بازار و تجربهٔ کارشناسان در تعیین نتیجهٔ دقیق‌تر کمک بسزایی می‌کند.

هر روزی که سفارش و معامله‌ای صورت می‌گیرد، به‌صورت لحظه‌ای در سیستم ثبت می‌شود و در بازه‌های زمانی ثابت، کوتاه‌مدت و متوالی، احتمال متخلف‌بودن هر یک از مشتریان در سیستم طراحی‌شده، براساس چارچوب بیزی تعیین می‌شود. با وجود آنکه طبقه‌بند بیزی براساس

1. Self-Organizing Map
2. Threshold Type

پارامترهای بسیاری تصمیم‌گیری می‌کند، برای اطمینان از نتیجهٔ به‌دست‌آمده، موارد متعددی همچون تاریخچهٔ فعالیت‌های اقتصادی فرد در بورس کالا ضروری به‌نظر می‌رسد. می‌توان روزانه حد آستانه‌ای را براساس سیاست‌های مختلف اتخاذ کرد تا در هر یک از بازه‌های زمانی، چنانچه به تشخیص سیستم، میزان تخلف فرد حداقل برابر با مقدار حد آستانه باشد، متخلف شناخته‌شده و پیگیری‌های بیشتر به بخش‌های ویژه واگذار شود.

همان‌طور که از مطالعهٔ آثار پژوهشی پیشین نمایان است، روش‌های گوناگونی در حل مسائل با موضوع مشابه، ارائه شده است. با بررسی ویژگی‌های مثبت و کاستی‌های پژوهش‌های پیشین، در پژوهش حاضر نوعی سیستم هوشمند مبتنی بر طبقه‌بند شبکهٔ بیزی طراحی و پیاده‌سازی شده است. خصوصیات این روش هوشمند و نوآوری‌های این پژوهش عبارت‌اند از:

- قابلیت ترکیب با تکنیک‌های آمار بیزی در ارائهٔ دانش؛
- تعیین روابط علّی بین متغیرهای فضای مسئله؛
- قابلیت استنتاج و تصمیم‌سازی در فضاهای غیرقطعی با استفاده از احتمالات؛
- قابلیت به‌روزرسانی همهٔ بخش‌ها (ساختاری و پارامتری) مبتنی بر المان یادگیری؛
- قابل استفاده بودن برای داده‌های ناقص؛
- قابلیت کاربرد در حوزه‌های متغیری کمّی (گسسته و پیوسته) و کیفی؛
- بیان نتایج طبقه‌بندی به‌صورت احتمالی و پرهیز از خطای تصمیم‌گیری.

روش‌شناسی پژوهش

این پژوهش از نظر گردآوری اطلاعات و مدارک، کتابخانه‌ای است که به‌صورت توصیفی ـ تحلیلی بررسی شده است و با توجه به هدف پژوهش (بررسی و استفاده از عملکرد تکنیک‌های شبکهٔ بیزی برای پیشگیری از وقوع حرکت القایی در معاملات بازار قراردادهای آتی سکهٔ بورس کالای ایران)، تحقیق جاری، پژوهش توسعه‌ای ـ کاربردی به‌شمار می‌رود. در ادامه، مراحل اجرای پژوهش شرح داده می‌شود.

برچسب‌گذاری اولیهٔ داده‌ها

برای برچسب‌گذاری داده‌ها، از روش خوشه‌بندی کا- میانگین برای شناسایی تراکنش‌های سالم یا متخلف مشتریان استفاده شده است. پس از تعیین تراکنش سالم یا متخلف و برچسب‌گذاری روی آنها، مقادیر به‌دست‌آمده به‌صورت متغیر جدید به مجموعهٔ داده‌ها اضافه می‌شود. شکل ۱ الگوریتم استفاده‌شده برای برچسب‌گذاری را نمایش می‌دهد.

Algorithm: K means clustering

1- Minimize distance between every data point and the correspondent centroid, by this function:

$$J=\sum_{i=1}^{k}\sum_{j=1}^{n}\|x_j-c_i\|^2$$

2- K←choose randomly *k* of our points as partition centers.

3- **For** m=1 to *number of centers* **do**

For n=1 **to** *number of data* **do**

Distance [m, n]←compute the distance between every data point[n] on the set with center[m]

End For

End For

4-Assign each point to the nearest cluster center.

5- **For** i=1 to *number of centers* **do**

Update the cluster center positions by using the following formula:

$$c_i=\frac{1}{|ki|}\sum xj$$

End For

6- **If** the cluster centers change, **Then** repeat the process from **3 Else** finish the process of k means clustering algorithm and get the **partition's members** and **centroids**.

شکل ۱. الگوریتم برچسب‌گذاری اولیهٔ داده‌ها

آماده‌سازی داده‌ها (گسسته‌سازی)

پس از برچسب‌گذاری داده‌ها، در گام بعد از طبقه‌بندی بیزی سادهٔ درخت افزوده[1] بهره برده می‌شود. این طبقه‌بند شکل توسعه‌یافتهٔ بیزی ساده است که اجازه می‌دهد ویژگی‌ها به‌صورت درخت با هم ارتباط داشته باشند. برای استفاده از این الگوریتم، ابتدا باید گسسته‌سازی داده‌ها صورت گیرد. برای گسسته‌سازی داده‌ها از تابع گسسته‌ساز[2] در نرم‌افزار R استفاده شده است که از توزیع خی‌دو کمک می‌گیرد. الگوریتم خی‌دو به‌صورت خودکار مقادیر حد آستانه‌ای را تعیین می‌کند و به‌طور صحیح و دقیق مجموعهٔ داده‌های عددی را گسسته‌سازی می‌کند. پارامترهای تابع عبارت‌اند از مجموعهٔ داده‌ها، آلفا و دلتا که به‌ترتیب داده‌ها، سطح معناداری و حداکثر میزان ناسازگاری در مجموعهٔ داده را نشان می‌دهند.

الگوریتم خی‌دو برمبنای توزیع خی‌دو رفتار می‌کند. ابتدا برای همهٔ ویژگی‌های عددی که گسسته‌سازی شده‌اند، با یک سطح معنادار بالا آغاز می‌گردد و تمام متغیرها براساس مقدارشان ذخیره می‌شوند. در مرحلهٔ نخست، مقدار خی‌دو برای هر زوج بازهٔ مجاور محاسبه می‌شود. گفتنی است که قبل از انجام گسسته‌سازی، هر متغیر را به‌عنوان یک بازه در نظر می‌گیرد که فقط همان متغیر به آن بازه تعلق دارد. در پیاده‌سازی مرحلهٔ دو، بازه‌های مجاور با کمترین مقدار خی‌دو را

1 .Tree Augmented Naïve Bayesian (TAN)
2. Discretization

ادغام می‌کند و فرایند ادغام ادامه می‌یابد تا اینکه مقدار خی‌دو همهٔ زوج بازه‌های مجاور، بزرگ‌تر از سطح معناداری‌ای شود که ابتدا مشخص شده است. این فرایند با سطح معنادار کاهش‌یافته‌ای ادامه می‌یابد تا با افزایش تدریجی میزان ناسازگاری به یک مقدار مطابق با شرط حداکثری دلتا برسد (لیو و سشنو، ۱۹۹۵). شکل ۲ الگوریتم استفاده‌شده برای گسسته‌سازی داده‌ها را در این پژوهش نمایش می‌دهد.

Algorithm: Discretization

```
Chi2 Function (att - attribute)
Phase 1:
    α ←.5
    While (InConCheck (data) <δ) do /Function InConCheck() returns an inconsistency rate found in the discretized data/
        For each numeric att do
            Sort (att, data)
            Chi-sq-init (att, data)
            While (Merge (data)) do/Function Merge () returns true or false depending on whether the concerned attribute is merged or not/
                Chi-sq-calculation (att, data)
            End While
        End For
        α0 ← α
        α ← decreSigLevel (α)
    End While
Phase 2:
    For i=1 to number of tt do
        sigLvl[i] ← α0
    End For
    While (no-att-can-be-merged) do
        For each mergeable att do
            Sort (att, data)
            chi-sq-init(att, data)
            While (Merge (data)) do
                Chi-sq-calculation (att, data)
            End While
            If (InConCheck(data) < δ)
                sigLvl[i] ←decreSigLevel(sigLvl[i]);
            Else
                att i is not mergeable;
        End For
    End While
End Function
```

شکل ۲. الگوریتم گسسته‌سازی

پیاده‌سازی الگوریتم TAN و تعیین پارامترهای احتمالی شبکهٔ حاصل

فرض اساسی روش طبقه‌بندی سادهٔ بیزی این است که مقادیر صفات با داشتن مقادیر تابع هدف از یکدیگر مستقل شرطی هستند. به‌بیان دیگر، این فرض بیان می‌کند که به شرط مشاهدهٔ خروجی تابع هدف، احتمال مشاهدهٔ صفات $a_1, a_2, \dots, a_n$ برابر ضرب احتمالات هر صفت به‌طور جداگانه است. یعنی:

$$V_{NB} = \arg max\ P(v_j) \prod_i P(a_i|v_j)\ V_{MAP} = \arg max\ P(V_j|a_1, \dots, a_n) \qquad \text{رابطهٔ ۳)}$$

V_{NB} خروجی طبقه‌بندی سادهٔ بیزی برای تابع هدف است. توجه کنید که تعداد جملات $P(a_i|v_j)$ که در این روش باید محاسبه شوند با تعداد صفات ضرب در تعداد دسته‌های خروجی برای تابع هدف برابر است که این مقدار از تعداد جملات $P(a_i, \dots, a_n|V_j)$ بسیار کمتر است.

نتیجه اینکه یادگیری سادهٔ بیزی سعی در تخمین مقادیر مختلف $P(v_j)$ و $P(a_i|v_j)$ با استفاده از میزان تکرار آنها در داده‌های آموزشی دارد. این مجموعه تخمین‌ها متناظر با فرض یاد گرفته شده است. در گام بعد، از فرض یادشده برای طبقه‌بندی نمونه‌های جدید استفاده می‌شود که این کار با استفاده از رابطهٔ ۳ صورت می‌گیرد (پرل، ۱۹۸۲). الگوریتم TAN طراحی‌شده برای پژوهش حاضر در شکل ۳ نمایش داده شده است.

Algorithm: TAN classification

1- Compute the mutual information between each pair of attributes: A_i, A_j

$$I(A_i ; A_j | C) = \sum_{x \in Ai} \sum_{y \in Aj} \sum_{c \in C} P(x, y, z) \log \left(\frac{P(x,y | z)}{P(x | z)P(y | z)}\right)$$

2- Build a complete undirected graph in which the vertices are the attributes *n* variables. The edges are weighted according to the pairwise mutual information

3- Build a maximum weighted spanning tree (MWST) using **Algorithm: Maximum weighted spanning tree**

4- Transform the resulting undirected graph to a directed graph by selecting the class variable as the root node and setting the direction of all edges outward from it

5- Construct a TAN model by adding an arc from the class variable to all other variables

6- Determine each customer is guilty or not and its probability based on parameters are gained in classifier

شکل ۳. الگوریتم TAN

درخت شبکهٔ افزوده

رابطهٔ ۴ میزان اطلاعاتی را نشان می‌دهد که متغیر Y، در حالی که متغیر Z مشخص شده باشد، می‌تواند از متغیر X ارائه دهد.

$$I\left(A_i ; A_j \middle| C\right) = \sum_{x \in A_i} \sum_{y \in B_i} \sum_{c \in C} P(x, y, z) \log\left(\frac{P(x, y)|z}{P(x|z)P(y|z)} \right)$$ رابطهٔ ۴)

در این روش با استفاده از رابطهٔ ۴ به‌جای چند درخت، یک درخت به‌دست می‌آید. مزیت این روش نسبت به روش‌های پیشین، ساده‌بودن از نظر پیچیدگی زمانی، درنظرگرفتن رابطهٔ وابستگی‌های مهم بین ویژگی‌های داده‌ها و در نتیجه، دقت بیشتر است (فریدمن، ایفتاک و دادنا، ۱۹۹۹). زمانی تفاوت عملکرد روش یادشده با سایر روش‌ها شایان توجه است که در مجموعهٔ داده‌های تحت پردازش، روابط بین ویژگی‌های در نظر گرفته‌شده، به حد کافی معنادار باشد؛ در غیر این‌صورت، نتایج مشابه نتایج سایر روش‌ها خواهد بود (چنگ، بل و ویرو، ۱۹۹۸). الگوریتم درخت شبکهٔ افزودهٔ طراحی‌شده در این پژوهش در شکل ۴ نمایش داده شده است.

Algorithm: Maximum weighted spanning tree
MWST-KRUSKAL Function (G,w)
1- A ← Ø
2 - For each vertex v ϵ G.V
3 - MAKE-SET (v)
4- Sort the edges of G.E into decreasing order by weight w
5- For each edge (u,v) ϵ G.E, taken in decreasing order by weight
6- If FIND-SET (u) ≠ FIND-SET (v)
7- A ← A ∪ {(u,v)}
8- UNION (u,v)
9- Return A
End Function

شکل ٤. الگوریتم درخت شبکهٔ افزوده

تعیین احتمال تخلف مشتری

در این سیستم برای بررسی تراکنش‌های هر فرد، بازه‌های زمانی ثابت در نظر گرفته می‌شود و نتایج میزان سالم یا متخلف‌بودن هریک از تراکنش‌های وی در هر دوره درج می‌گردد. میانگین وزنی نتایج به‌دست‌آمده از هر دوره، محاسبه‌شده و نتیجهٔ آن دوره اعلام می‌شود؛ نتیجهٔ نهایی، بیشینهٔ همهٔ دوره‌ها است. در این محاسبه، عدد صفر سلامت کامل تراکنش‌ها در نظر گرفته می‌شود و عدد صد یعنی معامله‌گر به احتمال نزدیک به یقین فریبکار است. الگوریتم اصلی و طراحی مدل این پژوهش در شکل ۵ ارائه شده است.

Algorithm: Main
1- For i=1 **to** *number of data* **do** Assign the Label of each data using **Algorithm: Kmeans clustering** /non guilty or guilty/ **End For** **2- For** j=1 **to** *number of customers* **do** Determine customer [j] is not guilty or guilty using **Algorithm: Model Designing** **End For** **3- Validate proposed method** Compute the amount of percentage of compliance between two labeling.
Algorithm: Model Designing
1-**For** j=1 to *number of continuous feature* **do** Discrete the amount of data in feature [j] using **Algorithm: Discretization** **End For** 2-Make TAN classification using **Algorithm: TAN classification** and labeling data

شکل ۵. الگوریتم اصلی و الگوریتم طراحی مدل

جمع‌آوری داده‌ها

داده‌های بررسی‌شده از شرکت بورس تهیه شده است. از آنجا که برای هر دوره‌ای که اعلامیهٔ بازار قراردادهای آتی سکه اعلام می‌شود، مشتریان می‌توانند طی چندین ماه برای این دورهٔ بازار و حتی به موازات آن برای سایر بازارها نیز فعالیت کنند، در این پژوهش بر یک دورهٔ خاص تمرکز شده است.

جدول ۲. ویژگی‌های داده‌ها

داده‌های سفارش (خرید/ فروش)	داده‌های معاملاتی
شناسهٔ بازاری که برای آن سفارش می‌گذارند	شناسهٔ بازاری که برای آن معامله انجام می‌دهند
نام سفارش‌دهنده (به‌صورت کدشده)	شناسهٔ معامله
حجم سفارش	نام فروشنده (به‌صورت کدشده)
قیمت پیشنهادی برای یک قرارداد	نام کارگزاری مرتبط با فروشنده (به‌صورت کدشده)
نام کارگزاری (به‌صورت کدشده)	نام خریدار
آخرین وضعیت مشتری در هر لحظه (جدید، ویرایش، حذف، انطباق)	نام کارگزاری مرتبط با خریدار (به‌صورت کدشده)
شناسهٔ سفارش	قیمت معامله
وضعیت مشتری در هر روز کاری (فعال، غیرفعال)	حجم
نوع سفارش (خرید، فروش)	زمان و تاریخ
زمان و تاریخ	

محیط پیاده‌سازی

برای خوشه‌بندی کا- میانگین، از نرم‌افزار متلب و به‌منظور گسسته‌سازی و طبقه‌بندی از نرم‌افزار R نسخهٔ ۲۰۱۵ و توابع bnlearn و discretization استفاده شده است.

اعتبارسنجی

با توجه به اینکه برچسب‌گذاری واقعی داده‌ها در دسترس نبود، برچسب‌گذاری داده‌ها به‌کمک روش خوشه‌بندی کا- میانگین انجام گرفت و از این طریق، سالم یا متخلف‌بودن هر یک از تراکنش‌ها مشخص شد. پس از آن، نتایج روش بیزی با نتایج خوشه‌بندی مقایسه شدند تا دقت روش بیزی مشخص شود.

یافته‌های پژوهش

با توجه به حجم داده‌های دریافتی، برای ایجاد ساختار طبقه‌بند بیزی مطمئن و از سوی دیگر دسته‌بندی داده‌ها به دو گروه داده‌های آموزشی و آزمایشی، از 10-fold cross validation استفاده شده است. با انجام این کار، داده‌های موجود به ۱۰ زیرمجموعهٔ[1] مساوی دسته‌بندی شدند. در ۱۰ اجرای مجزا، هر بار یک داده به‌عنوان دادهٔ تست و ۹ قسمت باقی دادهٔ آموزشی در نظر گرفته می‌شود و ساختار شبکه متناسب با آن شکل می‌گیرد؛ به این‌ترتیب حتی نمونه‌برداری‌های گوناگون برای داده‌های آموزشی و آزمایشی صورت می‌گیرد.

برای تعیین میزان خطای طبقه‌بند حاصل، توابع مختلفی وجود دارند که در اینجا از تابع logl[2] استفاده شده است. در هر یک از ۱۰ اجرا، مقدار خطا محاسبه شده و خطای کلی برابر میانگین خطای ۱۰ اجرا در نظر گرفته می‌شود. برای افزایش دقت، روال 10-fold cross validation پنج مرتبه اجرا شد و خطای میانگین اجرای این پنج مرحله ۰/۰۹۰۴۵ به‌دست آمد.

در جدول ۳ فقط نمونه‌ای از آزمون و نوع خروجی مدل طراحی‌شده مشاهده می‌شود که در آن دو نمونهٔ ۵۰ نفری جداگانه از مشتریان خرید و مشتریان فروش در یک روز مشخص را به نمایش گذاشته است. همان‌طور که در جدول ۳ مشاهده می‌شود، برای هر یک از مشتریان درصد احتمالی تخلف درج شده است. به‌طور مثال، برای مشتری با کد ۳ در بین سفارش‌دهندگان فروش عدد ۶۵/۴۷۷ درصد درج شده است؛ یعنی با این احتمال فرد متخلف شناخته می‌شود و تا زمانی که مقدار احتمالی ۱۰۰ درصد برای فردی تعیین نشود، نمی‌توان به یقین آن را متخلف شمرد.

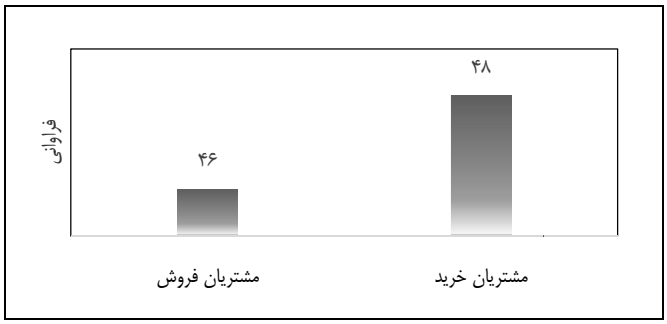
---

1. Subset
2. Log-Likelihood Loss

جدول ۳. نتایج احتمال تخلف مشتریان فروش و خرید

مشتریان فروش		مشتریان خرید					
کد مشتری	احتمال تخلف (٪)	کد مشتری	احتمال تخلف (٪)	کد مشتری	احتمال تخلف (٪)	کد مشتری	احتمال تخلف (٪)
۱	۰/۰۰	۲۶	۰۰۱/۰	۱	۹/۹۹	۲۶	۰/۰۰
۲	۷۳۱/۰	۲۷	۶۹/۲۱	۲	۹۹/۱۴۵	۲۷	۷۸/۲۵۱
۳	۴۷۷/۶۵	۲۸	۰/۶۹۴	۳	۶۳/۲۹۹	۲۸	۰/۰۴۴۷
۴	۱۱۳/۷۷	۲۹	۰/۰۱۶	۴	۰/۰۰۲۶۵	۲۹	۰/۲۱۳
۵	۷۷۹/۴۹	۳۰	۶/۵۴۵	۵	۵/۰۹۷	۳۰	۵۷/۱۲
۶	۰/۰۰	۳۱	۰/۰۰۲	۶	۸۴/۱۸۱	۳۱	۲/۶۵۵
۷	۵/۱۰	۳۲	۰/۰۰	۷	۰/۰۹۱	۳۲	۲۲/۲۲۴
۸	۱۰۰	۳۳	۳۴/۹۵	۸	۲۵/۶۶۳	۳۳	۰/۰۱۵۸
۹	۰۰۶/۰	۳۴	۶۴/۲۵۱	۹	۱/۳۱۴۸	۳۴	۸/۰۹۶
۱۰	۶۲۹/۲۱	۳۵	۳۲/۱۰۸	۱۰	۰/۰۸۶۳	۳۵	۲۶/۳۴۴
۱۱	۷۹۶/۰	۳۶	۶۹/۵۰۲	۱۱	۱/۰۳۸۶	۳۶	۱۳/۸۳۴
۱۲	۵۴/۳۵۳	۳۷	۵۰/۰۹۲	۱۲	۰/۰۸۵	۳۷	۰/۰۱۵
۱۳	۰/۲۲۱	۳۸	۲۹/۹۷۴	۱۳	۴۸/۲۶	۳۸	۹۹/۹
۱۴	۹۹/۹۰۵	۳۹	۱۶/۸۰۹	۱۴	۰/۰۰	۳۹	۶۱/۲۹۰
۱۵	۰/۳۶	۴۰	۴۹/۸۹۳	۱۵	۰/۰۵	۴۰	۷/۰۲۶
۱۶	۴۵/۸۱۶	۴۱	۳۰/۷۷۵	۱۶	۵۱/۷۹۶	۴۱	۱۶/۲۲
۱۷	۰/۰۰	۴۲	۱۱/۹۵۲	۱۷	۴۶/۸۲۵	۴۲	۵۰/۰۴
۱۸	۰/۰۰۴	۴۳	۰/۵۷	۱۸	۳۰/۷۲	۴۳	۰/۰۰
۱۹	۰/۰۰۳	۴۴	۳۳/۱۸۱	۱۹	۰/۰۰	۴۴	۶۸/۱۷
۲۰	۰/۰۰۳	۴۵	۱۲/۸۴۸	۲۰	۰/۰۲۸	۴۵	۱/۱۵۳
۲۱	۰/۰۰	۴۶	۰/۰۰	۲۱	۴۶/۵۲۱	۴۶	۴/۶۹
۲۲	۵۹/۹۹۴	۴۷	۰/۰۵۲	۲۲	۲/۰۷۶	۴۷	۸۷/۷۳
۲۳	۸۷/۰۰۳	۴۸	۰/۳۸۲	۲۳	۰/۰۰	۴۸	۹۹/۹
۲۴	۱۱/۶۸۸	۴۹	۰/۲۸۲	۲۴	۳/۵۴۹	۴۹	۰/۰۰
۲۵	۰/۰۰	۵۰	۱۰۰	۲۵	۰/۰۰	۵۰	۰/۰۰

برای طبقه‌بندی کلی مشتریان در دو دستهٔ افراد سالم و متخلف، با توجه به مقدار حد آستانهٔ مشخص‌شده در هر روز و نیز با در نظرگرفتن وضعیت بازار، هر مشتری که درصد احتمالی تخلف وی بیشتر از مقدار آستانه باشد، در دستهٔ افراد متخلف قرار می‌گیرد. به‌طور مثال، مقدار حد آستانه در اینجا ۵۰ درصد لحاظ شده است. بدین ترتیب، بنا بر آنچه در اعتبارسنجی بیان شد، فراوانی میزان انطباق نتیجهٔ آزمون انجام‌شده با برچسب‌گذاری حاصل از کا ـ میانگین، به‌صورت شکل ۶ است.



شکل ٦. نتایج اعتبارسنجی

همان‌طور که در شکل ۶ مشاهده می‌شود، بین خروجی مدل و برچسب‌گذاری اولیه برای مشتریان فروش ۹۲ درصد و برای مشتریان خرید ۹۶ درصد انطباق وجود دارد. بنابراین، میانگین تشابه نتایج ۹۴ درصد است.

نتیجه‌گیری و پیشنهادها

به‌کارگیری ابزارهای متداول مدل‌سازی در سیستم‌های اقتصادی اجتماعی و مدیریت فناوری اطلاعات، علاوه‌بر تبیین صریح صورت مسئله، به کشف راه‌حل‌های سازنده منجر می‌شود. این ابزارها در صورتی که به‌خوبی استفاده شوند، درک و شناخت مسائل دنیای واقعی را تسهیل می‌کنند و برای سرمایه‌گذاران به‌منظور شرکت در معاملات، رضایت و اعتماد بیشتری فراهم می‌آورند.

با توجه به کاستی‌های موجود در ابزار نظارتی بازار بورس و راهبردهای متنوع سوداگران برای کسب منافع بیشتر از طریق تراکنش‌های به‌ظاهر قانونی، استفاده از مدل‌های احتمالی برای تشخیص افرادی که تراکنش‌های بالقوهٔ فریبکارانه دارند، مفید به‌نظر می‌رسید؛ از این رو، مدلی برمبنای شبکهٔ بیزی، در چهار گام اساسی طراحی شده است که نتیجهٔ آن درصد احتمال متخلف‌بودن مشتری را در هر لحظه و در هر روز از بازار نشان می‌دهد. انتخاب طبقه‌بند به‌گونه‌ای است که همهٔ تراکنش‌های فرد را در یک روز خاص و حتی در زمینهٔ سایر فعالیت‌های وی بررسی می‌کند. طبقه‌بند TAN رقیب بسیاری از طبقه‌بندهای معروف بوده و عملکرد بسیار خوبی از خود نشان داده است. همان‌طور که در جدول ۳ مشاهده می‌شود، خروجی مدل روی نمونهٔ آماری ۱۰۰ نفره، متشکل از ۵۰ مشتری فروش و ۵۰ مشتری خرید آورده شده است و برای هر یک از مشتریان احتمال متخلف‌بودن وی در آن روز خاص بیان شده است؛ با انتخاب

مقدار حد آستانهٔ ۵۰ درصد، ۱۲ نفر از مشتریان فروش و ۱۲ نفر از مشتریان خرید به‌عنوان متخلف شناسایی شده‌اند.

از آنجا که هیچ برچسب‌گذاری اولیه‌ای برای راستی‌آزمایی وجود ندارد، از روش خوشه‌بندی کا ـ میانگین برای ایجاد برچسب‌های اولیه استفاده شد که مقایسهٔ نتایج کا ـ میانگین و مدل بیزی، ۹۴ درصد تطابق را نشان داد.

با توجه به مزایای استفاده از مدل شبکهٔ بیزی ارائه‌شده در این پژوهش، از جمله تشخیص هوشمندانه و خودکار، شفافیت در ارائهٔ دلایل قابل پذیرش، تعیین روابط علّی بین متغیرها، مدیریت مجموعهٔ داده‌های آموزشی در حجم زیاد، سهولت در مدیریت مجموعه داده‌های قاعده‌مند و پراکنده، کوتاه‌مدت‌بودن زمان آموزش در این شبکه‌ها و نیز یادگیری خودآموز در این نوع از شبکه‌ها (کورب و نیکولسن، ۲۰۱۰: ۱۵۰–۱۴۵)، استفاده از سیستم هوشمند پیشنهادی مطابق با شرایط، می‌تواند به شرکت‌های بورس کالا، بورس اوراق بهادار و فرابورس در تشخیص تقلب کمک بسیاری نماید.

پژوهش حاضر را می‌توان به‌صورت‌های مختلف توسعه داد؛ برای نمونه می‌توان با استفاده از سایر سیستم‌های تصمیم‌گیری و ترکیب آنها با شبکهٔ بیزی، سیستم دقیق‌تری ایجاد کرد. همچنین برای افزایش قابلیت شبکه، می‌توان به بررسی تأثیر و شیوهٔ ترکیب شبکه‌های بیزی به‌لحاظ پارامتری و ساختاری پرداخت. از سوی دیگر، با توجه به اینکه بازار قراردادهای آتی سکهٔ بورس کالای ایران هر روز دچار نوسان می‌شود، می‌توان برای بهبود عملکرد سیستم هوشمند، روشی برای درنظرگرفتن این نوسان‌ها ارائه کرد.

فهرست منابع

تقوا، م. ر.؛ منصوری، ط.؛ فیضی، ک.؛ اخگر، ب. (۱۳۹۵). کشف تقلب در تراکنش‌های کارت‌های بانکی با استفاده از پردازش موازی ناهنجاری در بزرگ‌داده. *نشریهٔ مدیریت فناوری اطلاعات*، ۸(۳)، ۴۷۷–۴۹۸.

تقوی‌فرد، م.ت.؛ جعفری، ز. (۱۳۹۴). کشف تقلب در بیمهٔ بدنهٔ خودرو با بهره‌مندی از سیستم خبرهٔ فازی. *نشریهٔ مدیریت فناوری اطلاعات*، ۷(۲)، ۲۵۸–۲۳۹.

صفری، ح.؛ حشمتی‌پور، ف.؛ مهرابی، ع.؛ نصابی، و. (۱۳۹۱). مدل‌سازی عوامل مؤثر بر به اشتراک‌گذاری اطلاعات در زنجیرهٔ تأمین شرکت ایران‌خودرو خراسان با استفاده از روش ترکیبی نگاشت علّی و شبکه‌های بیزین. *نشریهٔ مدیریت فناوری اطلاعات*، ۴(۱۰)، ۹۲–۶۵.

فلاح شمس، م.؛ کردلوئی، ح.ر. (۱۳۹۰). آزمون مدل‌های لاجیت و شبکهٔ عصبی مصنوعی جهت پیش‌بینی دستکاری قیمت در بورس اوراق بهادار تهران. *مجلهٔ مهندسی مالی و مدیریت اوراق بهادار*، ۲(۷)، ۶۹–۳۷.

فلاح شمس، م.؛ کردلوئی، ح. ر.؛ رشنو، م. (۱۳۹۱). بررسی دستکاری قیمت‌ها در بورس اوراق بهادار تهران با استفاده از مدل ماشین بردار پشتیبان. *مجلهٔ تحقیقات مالی*، ۱(۱۴)، ۸۴- ۶۹.

مدیریت مطالعات اقتصادی. (۱۳۹۱). دستکاری بازار، ترفندها و ابزارهای مطالعه. شرکت بورس کالای ایران.

وثوق، م.؛ تقوی‌فرد، م. ت.؛ البرزی، م. (۱۳۹۳). شناسایی تقلب در کارت‌های بانکی با استفاده از شبکه‌های عصبی مصنوعی. *نشریهٔ مدیریت فناوری اطلاعات*، ۶(۴)، ۷۴۶-۷۲۱.

Barber, D. (2010). *Bayesian Reasoning and Machine Learning;* first edition. Cambridge: University Press.

Chen, Y., Miao, D. & Zhang, H. (2010). Neighborhood outlier detection. *Expert Systems with Applications*, 37(12), 8745-8749.

Cheng, J., Bell, D. & Weiru, L. (1998). *Learning Bayesian networks from data: An efficient approach based on information theory*. On World Wide Web at http://www. Cs. ualberta. ca/~ jcheng/bnpc. Htm.

Fallah Shams, M. & Kordlouei, H. (2011). Logit Model Test and Neural Network to Predict Price Manipulation in Tehran Stock Exchange. *Financial Engineering and Portfolio Management*, 2(7), 37-69. (*in Persian*)

Fallah Shams, M., Kordlouei, H. & Rashnow, M. (2013). Investigating the Price Manipulation in Tehran Stock Exchange by Using the SVM Model. *Financial Research, University of Tehran*, 1(14), 69-84. (*in Persian*)

Ferdousi, Z. & Maeda, A. (2006). Unsupervised Fraud Detection in Time Series data. In Proceedings. *22nd International Conference on Data Engineering Workshops*, Atlanta, Georgia, 3-7 April 2006.

Franke, M., Hoser, B. & Schröder, J. (2008). *On the analysis of irregular stock market trading behavior*. In Data Analysis, Machine Learning and Applications (pp. 355-362). Springer Berlin Heidelberg.

Friedman, N., Iftach, N. & Dana, P. (1999). Learning Bayesian network structure from massive datasets: the sparse candidate algorithm. *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc.

Golmohammadi, K., Zaiane, O. R. & Díaz, D. (2014, October). Detecting stock market manipulation using supervised learning algorithms. In Data Science and Advanced Analytics (DSAA). *2014 International Conference on* (pp. 435-441). IEEE.

Heckerman, D, Geiger, D., &Chickering, D M. (1995). *Learning Bayesian networks: The combination of knowledge and statistical data*. Machine learning20.3: 197-243.

Kim, Y. & Sohn, S.Y. (2012). Stock fraud detection using peer group analysis. *Expert Systems with Applications*, 39(10), 8986-8992.

Korb, K. B. & Nicholson, A.E. (2010). *Bayesian artificial intelligence*. London: CRC press.

Lee, E.J., Eom, K.S. & Park, K.S. (2013). Microstructure-based manipulation: Strategic behavior and performance of spoofing traders. *Journal of Financial Markets*, 16(2), 227-252.

Liu, H. & Setiono, R. (1995, November). Chi2: Feature selection and discretization of numeric attributes. In *Tools with artificial intelligence, 1995. proceedings., seventh international conference on* (pp. 388-391). IEEE.

The management of Economic Studies. (2012). Market Manipulation, Tricks & Prevention Tools. Iran Mercantile Exchange.

Mitchell, T. M. (1997). *Machine learning*. Burr Ridge, IL: McGraw Hill.

Nielsen, T. D. & Jensen, F. V. (2009). *Bayesian networks and decision graphs*. Springer Science & Business Media.

Olszewski, D. (2014). Fraud detection using self-organizing map visualizing the user profiles. *Knowledge-Based Systems*, 70, 324-334.

Pearl, J. (1982). Reverend Bayes on inference engines: A distributed hierarchical approach. *AAAI'82 Proceedings of the Second AAAI Conference on Artificial Intelligence*. Pennsylvania, August 18 – 20.

Safari, H., Heshmatipour, F., Mehrabi, A. & Nesabi, V.R. (2012). Modeling of Factors Affecting Information Sharing in Supply Chain of IKKCO Using the Integrated Cognitive Mapping Method and Bayesian Networks. *The Journal of Information Technology Management*, 4(10), 65-92 (*in Persian*)

Taghavifard, S. M. & Jafari, Z. (2015). Fraud Detection Using a Fuzzy Expert System in Motor Insurance. *The Journal of Information Technology Management*, 7(2), 239- 258. (*in Persian*)

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

Thoppan, J.J. & Punniyamoorthy, M. (2013). Market manipulation and surveillance–a survey of literature and some practical implications. *International Journal of Value Chain Management*, 7(1), 55-75.

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

Yao, Y., Zhai, J., Cao, Y. & Ding, X. (2015). WITHDRAWN: Static and dynamic models: A framework for price manipulation detection. *Expert Systems with Applications.* Available in: http://www.elsevier.com/locate/withdrawalpolicy.