

الگوی یادگیری جمعی بهبودیافته با هوش

ازدحامی جهت پیش‌بینی ریزش مشترکان

تلفن همراه

بیژن مرادی^۱، مهران خلج^{۱*} و علی تقی‌زاده هرات^۱

^۱ گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد پرند و رباط کریم، تهران، ایران

چکیده

از آنجاکه در شرکت‌های مخابرات همراه، هزینه حفظ مشتریان فعلی بسیار کمتر از هزینه جذب مشتریان جدید است، پیش‌بینی دقیق امکان ریزش هریک از مشتریان و جلوگیری از آن، امری ضروری است. بنابراین، پژوهشگران روش‌های کارآمدی را با استفاده از ابزارهای داده‌کاوی و هوش مصنوعی برای شناسایی مشتریانی که قصد روی‌گردانی دارند، ارائه کرده‌اند. در این مقاله، ما به‌منظور بهبود فرایند پیش‌بینی ریزش مشتری، یک راهکار مؤثر مبتنی بر یادگیری جمعی پیشنهاد می‌کنیم که در آن از الگوریتم بهینه‌سازی گرگ خاکستری، به‌منظور انتخاب ویژگی‌های مؤثر و همچنین تنظیم شاخص‌های آزاد الگوی پیشنهادی، استفاده شده‌است. سپس، به‌منظور ارزیابی عملکرد روش پیشنهادی، آن را با استفاده از دو مجموعه داده ریزش مشتری شبیه‌سازی کرده و نتایج حاصل را به کمک معیارهای ارزیابی شامل صحت، دقت، یادآوری، امتیاز F1 و AUC با سایر روش‌های مشابه مقایسه کرده‌ایم. نتایج به‌دست‌آمده برتری روش پیشنهادی بر سایر راهکارهای ارزیابی‌شده را نشان می‌دهد.

واژگان کلیدی: پیش‌بینی ریزش مشتری، مخابرات همراه، یادگیری جمعی، هوش ازدحامی، الگوریتم بهینه‌سازی گرگ خاکستری

Improved Ensemble Learning Model by Swarm Intelligence for Mobile Subscribers' Churn Prediction

Bijan Moradi¹, Mehran Khalaj^{1*} & Ali Taghizadeh Herat¹

¹ Department of Industrial Engineering, Faculty of Engineering and Technology, Islamic Azad University, Parand and Robat Karim Branch, Tehran, Iran

Abstract

In today's competitive world, companies need to analyze, identify and predict the behavior of their customers and respond to their demands earlier than their competitors. Moreover, in many industries such as mobile telecommunications, the cost of maintaining existing customers (customer retention) is much lower than the cost of attracting a new customer. Therefore, the problem of identifying customers who are going to leave the company, so-called Customer Churn Prediction (CCP), and preventing them by offering Incentives is essential in these industries. In this direction, researchers have presented competent methods using data mining and artificial intelligence tools to identify potential churners. Machine learning (ML) methods are one of the most powerful and widely used techniques to deal with the CCP problem, since they can properly extract and learn complex relationships between the customers' attributes and their churn intention. Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Decision Trees (DTs), Logistic Regression (LR), and Naïve Bayes (NB) are among the well-known ML models utilized in numerous studies to tackle the CCP problem. Also, ensemble learning techniques such as Adaboost, Gradient Boost, and Extreme Gradient Boost (XG_boost) have been widely used to solve the CCP problem since they can aggregate the capabilities of multiple ML models. Hence, in order to improve the process of predicting customer churn, in this paper we propose a novel ensemble learning based approach, which is designed based on the two-level stacking technique. We employ six prominent ML models in each level of our proposed ensemble model including MLP, SVM-

* Corresponding author

* نویسنده عهده‌دار مکاتبات

سال ۱۴۰۲ شماره ۴ پیاپی ۵۸

• تاریخ ارسال مقاله: ۱۳۹۸/۹/۲۸ • تاریخ پذیرش: ۱۴۰۰/۹/۱۵ • تاریخ انتشار: ۱۴۰۱/۷/۷ • نوع مطالعه: پژوهشی



RBF, DT, NB, KNN, and LR. We also benefit from the Gray Wolf Optimization (GWO) algorithm as an efficient swarm intelligence based search algorithm to select the most effective features and also adjust the hyper-parameters in the proposed model. We have implemented our proposed model using Python and simulated it on two well-known customer churn datasets in the telecom market (IBM_Telco and Duke_Cell2Cell) to evaluate its performance. In this direction, we first demonstrated the optimal features' subset and the parameter values obtained from applying the GWO algorithm on each dataset. Next, we compared the performance of the proposed ensemble model with each of the base learners using common evaluation criteria including accuracy, precision, recall, F1 score and AUC. The results show that the proposed ensemble model can collect the capabilities of all the base learners and it works better than each of the basic ML models. Afterward, we compared the obtained results from the suggested model with the common ensemble models (Adaboost, Gradient_boost, XG_boost, and Cat_boost) The experimental results show the superiority of the proposed method over other evaluated ensemble models in all the evaluation criteria. Eventually, our method is compared with two recent CCP approaches introduced in the literature. This analysis reveals that, except for the recall criterion in the Duke_Cell2Cell dataset, our introduced method achieves superior results compared to the considered approaches in both datasets.

Keywords: Customer Churn Prediction, Mobile Telecommunication, Ensemble Learning, Swarm Intelligence, Gray Wolf Optimization Algorithm

۱- مقدمه

در دنیای رقابتی امروز، بیشتر شرکت‌ها و سازمان‌ها در بخش‌های مختلف تلاش می‌کنند با استفاده از راهکارهای مدیریت ارتباط با مشتری (CRM^۱) برای دستیابی به روابط طولانی‌مدت با مشتریان و حفظ رضایت آن‌ها تلاش کنند. این عمل، به‌طور عمده به این دلیل است که نتایج پژوهش‌های علمی نشان داده‌اند جذب مشتریان جدید در بیشتر موارد بسیار پرهزینه‌تر از حفظ مشتریان موجود است [۱]. در این راستا، توسعه الگوهای دقیقی که رفتار ریزش مشتریان را در صنایع مختلف (به‌عنوان مثال، بانک‌داری [۲-۴]، بخش خرده‌فروشی [۵، ۶]، مخابرات و ...) پیش‌بینی می‌کنند، برای ابداع برنامه‌های تجاری اثربخش و جلوگیری از خروج مشتریان از شرکت‌ها از اهمیت بالایی برخوردار است.

رقابت جدی در بازار مخابرات همراه و سهولت جابجایی بین ارائه‌دهندگان خدمات گوناگون، مشتریان را تشویق می‌کند تا به‌طور مداوم شرکت‌های مخابراتی مختلف را مقایسه کنند و به شرکتی بپیوندند که خدمات با کیفیت بالاتر و فناوری‌های جدیدتر را با هزینه کمتری ارائه می‌دهد [۷]؛ بنابراین، توسعه الگوهای پیش‌بینی ریزش مشتری (CCP^۲) کارآمد برای شناسایی علائم اولیه فرسایش مشتریان^۳ و کاهش آن با به‌کارگیری راهبردهای بهبود رضایت مشتری برای حفظ پایداری در بازار مخابرات همراه امری ضروری است [۸]؛ با این حال، پرداختن به مسئله پیش‌بینی ریزش مشتری در بخش مخابرات، یعنی تعیین مشتریانی که به احتمال زیاد یک شرکت مخابراتی

را ترک می‌کنند، یک مسئله چالش‌برانگیز و دلایل آن به‌طور عمده دو چیز است:

۱. اندازه بزرگ داده‌های مرتبط با هریک از مشتریان شامل داده‌های مربوط به اطلاعات جمعیت‌شناسی مشتریان، گزارش خدمات ارائه‌شده، اطلاعات صورت‌حساب، شکایات مطرح شده از سوی مشتری و ... [۹]
۲. ماهیت نامتعادل مجموعه داده مربوط به ریزش مشتریان به دلیل تعداد کم مشتریان روی‌گردانده در مقایسه با سایر مشتریان [۱۰]

فنون یادگیری ماشین (ML^۴) یکی از قوی‌ترین راهکارها برای حل مسئله پیش‌بینی ریزش مشتریان است. این فنون شامل مجموعه‌ای از الگوریتم‌ها هستند که می‌توانند روابط بین داده‌های ورودی و خروجی‌های متناظر آن‌ها را بیاموزند و به‌طور خودکار پیش‌بینی‌هایی با کیفیت بالا ارائه دهند [۱۱]. این روش‌ها که به‌طور گسترده برای رسیدگی به مسئله پیش‌بینی ریزش مشتری مطالعه شده‌اند، شامل رگرسیون لجستیک^۵ [۱۲]، درختان تصمیم^۶ [۱۳]، الگوی بیز ساده^۷ [۱۴]، نظریه مجموعه خشن^۸ [۱۵]، ماشین‌های بردار پشتیبان^۹ [۱۶] و شبکه‌های عصبی مصنوعی^{۱۰} [۱۷] است. علاوه بر فنون یادشده، روش‌های یادگیری جمعی^{۱۱} که در ساختار خود از روش‌های مختلف یادگیری ماشین (موسوم به یادگیرندگان پایه) بهره می‌برند، عمل‌کرد بسیار مناسبی

⁴ Machine Learning

⁵ Logistic Regression

⁶ Decision Tree

⁷ Naïve Bayes model

⁸ Rough Set Theory

⁹ Support Vector Machine

¹⁰ Artificial Neural Networks

¹¹ Ensemble Learning Methods

¹ Customer Relationship Management

² Customer Churn Perdition

³ Customer Attrition

در حل مسئله^۱ پیش‌بینی ریزش مشتری داشته‌اند [۷, ۱۸-۲۰]. این به‌طور عمده، به دلیل توانایی رویکردهای یادگیری جمعی برای جبران کاستی‌های هریک از یادگیرندگان پایه با تجمیع قابلیت‌های چند الگوی یادگیری ماشین است. این برتری پژوهشگران را بر آن داشته‌است تا از راهبردهای یادگیری جمعی مختلف مانند بسته‌بندی^۱، تقویت^۲ و انباشتگی^۳ در حل مسئله^۱ پیش‌بینی ریزش مشتری استفاده کنند. همچنین، به‌منظور کاهش اثرات مخرب ناشی از تعداد زیاد ویژگی‌ها در مجموعه داده‌های مربوط به ریزش مشتری که منجر به کاهش دقت پیش‌بینی الگوهای یادگیری ماشین، پیچیدگی محاسباتی بالا، و مشکلات مربوط به برازش بیش‌ازحد^۴ می‌شود، از روش‌های انتخاب ویژگی^۵ در راهکارهای حل مسئله^۱ پیش‌بینی ریزش مشتری استفاده شده‌است. روش‌های انتخاب ویژگی، همان‌طور که از نامشان پیداست، سعی می‌کنند با استفاده از الگوریتم‌های جستجو و معیارهای ارزیابی گوناگون، مرتبط‌ترین و مؤثرترین زیرمجموعه^۶ ویژگی‌ها را انتخاب کنند. الگوریتم‌های فراابتکاری^۶ مانند الگوریتم‌های بهینه‌سازی تکاملی^۷ و الگوریتم‌های مبتنی بر هوش ازدحامی^۸ نامزدهای عالی برای جستجوی ویژگی‌های موردنظر هستند، زیرا می‌توانند مجموعه ویژگی‌های بهینه را در تعداد محدودی از تکرارها پیدا کنند.

با الهام از مطالعات پیشین، در این مقاله ما یک الگوریتم جدید پیش‌بینی ریزش مشتری به کمک یک الگوی یادگیری جمعی مبتنی بر هوش ازدحامی پیشنهاد می‌کنیم تا مشتریانی را که می‌خواهند یک شرکت مخابرات همراه را ترک کنند، شناسایی کنیم. در این راهکار، نخست یک الگوی یادگیری جمعی دوسطحی بر اساس روش انباشتگی طراحی می‌کنیم. در ساختار الگوی یادگیری جمعی پیشنهادی، ما شش الگوریتم یادگیری ماشین متداول برای رسیدگی به مسئله^۱ پیش‌بینی ریزش مشتری شامل یک الگوی شبکه^۹ عصبی پرسپترون چندلایه (MLP^۹)، یک ماشین بردار پشتیبان با تابع پایه^{۱۰} شعاعی (RBF^{۱۰})، یک درخت تصمیم، یک الگوی ساده^{۱۱} بیز، یک

الگوی K-نزدیک‌ترین همسایگان (KNN^{۱۱}) و یک الگوی رگرسیون لجستیک، را به‌عنوان یادگیرندگان پایه به‌کار می‌گیریم. پس از آن، از الگوریتم بهینه‌سازی گرگ خاکستری (GWO^{۱۲}) [۲۱] به‌عنوان یک الگوریتم مبتنی بر هوش ازدحامی برای انتخاب مؤثرترین ویژگی‌های ورودی و همچنین برای بهینه‌سازی فراشاخص‌های موجود در الگوی یادگیری جمعی، برای دستیابی به بهترین نتایج ممکن استفاده می‌کنیم. تابع برازش^{۱۳} تعریف‌شده در الگوریتم GWO شامل پنج شاخص آزاد است که کاربر می‌تواند به کمک آن‌ها وزن هریک از معیارهای صحت^{۱۴}، دقت^{۱۵}، یادآوری^{۱۶}، امتیاز^{۱۷} F1 و مساحت زیرمنحنی (AUC^{۱۸}) را با توجه به سیاست‌های مورد نظر شرکت تنظیم کند. نوآوری‌های اصلی راهکار پیشنهادی در این مقاله را می‌توان به شرح زیر فهرست کرد:

- (۱) ما در این پژوهش یک روش یادگیری جمعی مبتنی بر هوش ازدحامی پیشنهاد می‌کنیم که می‌تواند رفتار ریزش مشتریان را بر اساس ویژگی‌های موجود در مجموعه داده‌های مشتریان شرکت‌های مخابرات همراه آموخته و سپس مشتریانی را که قصد ترک یک شرکت مخابراتی را دارند، به‌دقت پیش‌بینی کند.
- (۲) راهکار پیشنهادی ما در این مقاله، یک الگوی یادگیری جمعی دوسطحی است که بر اساس روش انباشتگی طراحی شده‌است و از الگوهای یادگیری ماشین پایه شامل MLP، SVM-RBF، DT، NB، KNN، و LR در هر سطح استفاده می‌کند. بنابراین، برای دستیابی به نتایجی با کیفیت بالا الگوی پیشنهادی از شش یادگیرنده^۱ پایه در دو سطح بهره می‌برد و نتایج حاصل از آن‌ها را بر اساس روش تجمیع آرای وزن‌دار ترکیب می‌کند.
- (۳) در روش پیشنهادی از الگوریتم بهینه‌سازی گرگ خاکستری برای انتخاب ویژگی و تنظیم فراشاخص‌های الگوی یادگیری جمعی استفاده می‌شود. به عبارت دیگر، الگوریتم GWO به‌طور هم‌زمان بهترین ویژگی‌ها را در مجموعه داده‌های ورودی جستجو، و شش وزن مربوط به فرایند رأی‌گیری وزن‌دار را در الگوی یادگیری جمعی دوسطحی برای تصمیم‌گیری نهایی تنظیم می‌کند.

¹¹ K Nearest Neighbors

¹² Gray Wolf Optimization

¹³ Fitness Function

¹⁴ Accuracy

¹⁵ Precision

¹⁶ Recall

¹⁷ F1-Score

¹⁸ Area Under the Curve

¹ Bagging

² Boosting

³ Stacking

⁴ Over-Fitting

⁵ Feature Selection

⁶ Metaheuristic Algorithms

⁷ Evolutionary Optimization Algorithms

⁸ Swarm Intelligence based Algorithms

⁹ Multi-Layer Perceptron

¹⁰ Radial Basis Function

۴) تابع برازش در الگوریتم GWO به گونه‌ای تعریف شده‌است که کاربر می‌تواند قدرت هریک از معیارهای صحت، یادآوری، دقت و امتیاز FI را بر اساس سیاست‌های مدیریتی شرکت مخابراتی موردنظر تنظیم کند. به‌عنوان مثال، اگر بر اساس سیاست‌های حفظ مشتری یک شرکت مخابراتی، شناسایی همه مشتریانی که احتمال دارد شرکت را ترک کنند، مهم‌تر از تعیین مشتریانی باشد که بااطمینان شرکت را ترک خواهند کرد، باید وزن معیار یادآوری را افزایش داد. برعکس، اگر یک شرکت مخابراتی نیاز به پیش‌بینی خاص مشتریانی داشته‌باشد که احتمال خروج آن‌ها بسیار زیاد است (پیش‌بینی با حداقل تعداد مثبت کاذب)، وزن معیار دقت در تابع برازش الگوریتم بهینه‌سازی گرگ خاکستری باید افزایش یابد.

سایر بخش‌های این مقاله به شرح زیر سازمان‌دهی شده‌است: در بخش دوم تعدادی از روش‌های برجسته معرفی شده در مقالات پژوهشی برای مقابله با مسئله پیش‌بینی ریزش مشتری بررسی شده‌است. در بخش سوم، به معرفی مجموعه‌داده‌های ریزش مشتری در صنعت مخابرات، شامل دو مجموعه‌داده به نام‌های IBM_Telco [9] و Duke_Cell2Cell [29] می‌پردازیم. در بخش چهارم، الگوریتم بهینه‌سازی گرگ خاکستری را معرفی می‌کنیم. در این پژوهش از این الگوریتم برای انجام انتخاب ویژگی و تنظیم فراشاخص‌های الگوی یادگیری جمعی استفاده می‌شود. در بخش پنجم مقاله جزئیات راهکار پیشنهادی را به تفصیل شرح داده و هریک از بخش‌های ساختاری آن را به دقت بررسی می‌کنیم. پس از آن، در بخش ششم مقاله، نتایج آزمایشی به دست آمده از به‌کارگیری راهکار پیشنهادی برای پیش‌بینی ریزش مشتریان در مجموعه‌داده‌های IBM_Telco و Duke_Cell2Cell گزارش و بررسی می‌شوند. همچنین، در این بخش نتایج به دست آمده از شبیه‌سازی الگوی پیشنهادی را با نتایج حاصل از سایر راهکارهای شناخته‌شده در حوزه پیش‌بینی ریزش مشتری مقایسه می‌کنیم. در نهایت، در بخش هفتم مقاله به جمع‌بندی نتایج حاصل از الگوریتم پیشنهادی پرداخته و در پایان، چند پیشنهاد برای انجام مطالعات آینده مطرح می‌شود.

۲- مروری بر پژوهش‌های پیشین

استفاده از روش‌های داده‌کاوی و الگوهای یادگیری ماشین سابقه‌ای به نسبت طولانی در پژوهش‌های مربوط به حل مسئله پیش‌بینی ریزش مشتری در صنعت مخابرات دارد.

پژوهش موزر و همکاران [22] شامل الگوهای مختلف یادگیری ماشین شامل رگرسیون لاجیستیک، درخت‌های تصمیم، شبکه‌های عصبی و همچنین یادگیری جمعی مبتنی بر روش تقویت برای پیش‌بینی ریزش مشتریان استفاده شده‌است. در این پژوهش از یک مجموعه‌داده شامل 47 هزار مشترکان محلی تلفن همراه استفاده شده‌است که مشتمل بر اطلاعات مربوط به سابقه مصرف، صورت حساب، اعتبار، کاربرد و شکایات است. نتایج به دست آمده نشان می‌دهد استفاده از روش‌های یادگیری ماشین برای شناسایی مشتریانی که قصد ترک شرکت را دارند و ارائه مشوق‌های مناسب به آن‌ها می‌تواند صرفه‌جویی قابل توجهی را برای یک شرکت سرویس‌دهنده به همراه داشته‌باشد. همچنین، در این مقاله نشان داده شده که الگوی شبکه عصبی مصنوعی عملکرد بهتری نسبت به رگرسیون لاجیستیک و درختان تصمیم جهت پیش‌بینی ریزش مشترکان در مجموعه‌داده استفاده شده دارد.

هادن و همکاران [23]، تحلیل عمیقی از روش‌های انتخاب ویژگی و فواید استفاده از آن‌ها در حل مسئله ریزش مشتری ارائه کرده و مرور جامعی بر الگوهای یادگیری ماشینی متداول تا آن زمان شامل تحلیل رگرسیون، شبکه‌های عصبی، درختان تصمیم و الگوهای مارکوف انجام دادند. نتایج پژوهش‌های آن‌ها نشان داد که استفاده از روش‌های انتخاب ویژگی سه مزیت عمده دارد. نخست این که، کاهش تعداد ویژگی‌های اطلاعاتی می‌تواند باعث ایجاد بهبود قابل توجهی در عملکرد یک الگوی طبقه‌بند شود؛ دوم این که، با کاهش تعداد ویژگی‌های مورد استفاده در الگوی یادگیری ماشین، زمان پردازش مورد نیاز بسیار کاهش می‌یابد؛ و سوم این که، در بعضی موارد، مانند هنگامی که مقدار محدودی از داده‌های آموزشی در دسترس است، استفاده از مجموعه‌داده‌های با ابعاد پایین‌تر مناسب‌تر است. همچنین، آن‌ها دریافتند در فرایند انتخاب ویژگی، به‌کارگیری روش‌های بهینه‌سازی (مانند الگوریتم ژنتیک) می‌تواند نقش بزرگی در بهبود عملکرد فنون پیش‌بینی ریزش مشتری ایفا کند. به‌علاوه، در خصوص الگوهای پیش‌بینی، بررسی انجام شده نشان داد تا آن زمان درختان تصمیم، محبوب‌ترین انتخاب در مسئله پیش‌بینی ریزش مشتری بودند و به شبکه‌های عصبی اندکی بیشتر از تحلیل‌های مبتنی بر رگرسیون توجه شده بود. بنابراین، با توجه به محدود بودن انواع الگوهای استفاده شده، نویسندگان پیشنهاد استفاده از

به‌کارگیری روش‌های نمونه‌برداری کاهش یافته را نشان می‌دهند.

پندهارکار [۲۵]، دو الگوی شبکه عصبی مبتنی بر الگوریتم ژنتیک را برای پیش‌بینی ریزش مشتری در بین مشترکان مخابرات بی‌سیم پیشنهاد کرده‌است. اولین الگوی پیشنهادی در این مطالعه از یک معیار مبتنی بر آنتروپی متقاطع^۷ برای پیش‌بینی ریزش مشتری استفاده کرده و دومین الگو تلاش می‌کند تا به‌طور مستقیم دقت پیش‌بینی ریزش مشتری را به کمک الگوریتم ژنتیک به حد بیشتری برساند. نتایج حاصل از اعمال الگوهای مبتنی بر الگوریتم ژنتیک پیشنهادی بر روی مجموعه داده‌های مربوط به خدمات بی‌سیم سلولی دنیای واقعی و مقایسه آنها با یک الگو Z-Score آماری بر اساس چندین معیار ارزیابی شامل صحت، Lift دهک بالای دهرصد و AUC نشان می‌دهد که هر دو الگوی مبتنی بر الگوریتم ژنتیک از لحاظ تمام معیارهای یادشده عملکرد بهتری دارند. علاوه بر این، نتایج به‌دست‌آمده مشخص می‌کند که شبکه‌های عصبی با اندازه متوسط بهترین عملکرد را دارند و معیار مبتنی بر آنتروپی متقاطع می‌تواند مقاومت بیشتری در برابر بروز بیش‌برازش در فرایند آموزش داشته‌باشد.

در مطالعه دیگری که توسط ادیس و همکاران [۱۸] انجام شده‌است، پژوهشگران نشان داده‌اند اگرچه استفاده از الگوریتم‌های یادگیری جمعی به‌ویژه در راهکارهای مبتنی بر روش‌های تقویتی بر پایه الگوهای درختان تصمیم برای حل مسئله پیش‌بینی ریزش مشتری به نتایج خوبی دست یافته‌اند، اما اندازه بزرگ مجموعه داده‌ها، ماهیت نامتعادل این مجموعه‌ها، و همچنین تعداد زیاد ویژگی‌های موجود در آن‌ها، به‌ویژه در مجموعه داده‌های مربوط به شرکت‌های مخابراتی، به‌طور عمده باعث می‌شود که الگوریتم‌های طبقه‌بندی در پیش‌بینی دقیق ریزش‌ها دچار مشکل شوند. بنابراین، در این مقاله جهت مقابله با مشکلات یادشده، از یک راهکار مبتنی بر ترکیب برنامه‌نویسی ژنتیک^۸ و الگوی تقویتی Adaboost برای حل مسئله پیش‌بینی ریزش مشتری در صنعت مخابرات استفاده شده‌است. در نهایت، دقت پیش‌بینی روش پیشنهادی با استفاده از اعتبارسنجی متقابل ده‌لایه در مجموعه داده‌های مخابراتی ارزیابی و با روش‌های KNN و جنگل تصادفی مقایسه شده‌است که برتری آن را نشان می‌دهد.

الگوهای دیگر مانند الگوی بیز ساده و سامانه استنتاج فازی را مطرح کرده‌اند.

کوزمنت و همکاران [۲۴]، مطالعه جامعی را در خصوص استفاده از ماشین‌های بردار پشتیبان در زمینه پیش‌بینی ریزش مشتریان انجام دادند تا به الگویی با عملکرد پیش‌بینی بهتر دست یابند. علاوه بر این، در این پژوهش مقایسه‌ای بین دو روش انتخاب شاخص موردنیاز برای پیاده‌سازی ماشین‌های بردار پشتیبان انجام شده و سپس، عملکرد هر دو نوع الگوی ماشین بردار پشتیبان با روش‌های رگرسیون لجستیک و جنگل‌های تصادفی از لحاظ دقت پیش‌بینی‌کنندگی مقایسه شده‌است. نتایج به‌دست‌آمده در این مطالعه نشان می‌دهد که ماشین‌های بردار پشتیبان عملکرد کلی خوبی را هنگام استفاده از داده‌های نوفه‌ای از خود نشان می‌دهند. با این وجود، روند بهینه‌سازی شاخص نقش مهمی را در عملکرد این الگوها ایفا می‌کند. به عبارت دیگر، زمانی که روش بهینه انتخاب شاخص اعمال می‌شود، ماشین‌های بردار پشتیبان عملکرد بهتری نسبت به روش رگرسیون لجستیک دارند، درحالی‌که جنگل‌های تصادفی از هر دو نوع ماشین بردار پشتیبان عملکرد بهتری دارند.

بورز و وزدن [۱۰] بر روی مسئله عدم تعادل^۱ در مجموعه داده‌های مربوط به پیش‌بینی ریزش مشتری تمرکز کردند. منظور از عدم تعادل در یک مجموعه داده اختلاف زیاد در تعداد نمونه‌های یک طبقه نسبت به سایر طبقه‌ها در آن مجموعه داده است. در مجموعه داده مربوط به پیش‌بینی ریزش مشتری نیز با توجه به این که تعداد مشتریانی که شرکت را ترک کرده‌اند، بسیار کمتر از سایر مشتریان است، این مجموعه‌ها نامتعادل هستند و این امر می‌تواند باعث کاهش کارایی فرایند آموزش در الگوی یادگیری ماشین شود. از این‌رو در پژوهش یادشده، نتایج به‌دست‌آمده از اعمال راهکارهای متداول، برای متعادل‌سازی مجموعه داده شامل نمونه‌برداری کاهش تصادفی^۲ و نمونه‌برداری کاهش پیشرفته^۳ مانند روش مکعب^۴ بر روی الگوهای پیش‌بینی مبتنی بر جنگل‌های تصادفی وزن‌دار^۵ و الگوی تقویت گرادیان^۶، بررسی و مقایسه شده‌اند. برای ارزیابی نتایج به‌دست‌آمده معیارهای AUC و Lift اندازه‌گیری شده‌اند که تأثیرات مثبت

¹ Class Imbalance

² Random Under-Sampling

³ Advanced Under-Sampling

⁴ CUBE

⁵ Weighted Random Forests

⁶ Gradient Boosting

⁷ Cross Entropy

⁸ Genetic Programming

و/فیدایس و همکاران [۸]. یک مطالعه تطبیقی جامع در مورد محبوب‌ترین روش‌های یادگیری ماشین به‌کاررفته در مسئله پیش‌بینی ریزش مشتری در صنعت مخابرات ارائه کرده‌اند، که در آن، نخست، تمام الگوها، شامل شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان، درختان تصمیم، الگوی بیز ساده و رگرسیون لاجیستیک با استفاده از روش اعتبارسنجی متقابل بر روی یک مجموعه داده عمومی اعمال و ارزیابی شده‌اند. در گام دوم، دربارهٔ بهبود عملکرد حاصل از به‌کارگیری روش یادگیری جمعی مبتنی بر تقویت مطالعه، و به‌منظور تعیین کارآمدترین ترکیب‌های شاخصی از شبیه‌سازی مونت کارلو برای هر روش و برای طیف گسترده‌ای از شاخص‌ها استفاده شده‌است. نتایج به‌دست‌آمده به‌وضوح برتری نسخه‌های تقویت‌شدهٔ الگوها را در برابر نسخه‌های ساده آن‌ها نشان می‌دهد.

ادریس و خان [۷]، جهت رسیدگی به مسئله چالش برانگیز پیش‌بینی ریزش مشتری در شرکت‌های مخابراتی، یک سامانهٔ پیش‌بینی هوشمند به نام FW-ECP پیشنهاد کرده‌اند که در آن توانایی ترکیب روش‌های انتخاب ویژگی مبتنی بر فیلتر و بسته‌بندی^۱، و همچنین بهرهٔ قابلیت یادگیری جمعی با استفاده از یادگیرندگان پایهٔ گوناگون لحاظ شده‌است. در گام فیلتر روش پیشنهادی، از روش بهینه‌سازی ازدحام ذرات برای انجام نمونه‌برداری کاهشی جهت متعادل‌سازی مجموعه داده‌ها، سپس از روش کمینهٔ افزونگی و بیشینهٔ ارتباط ($mRMR^2$) برای انتخاب ویژگی استفاده شده‌است. سپس، در گام بسته‌بندی، برای حذف هرچه بیشتر ویژگی‌های نامربوط و اضافی، الگوریتم ژنتیک به‌کارگرفته شده‌است. در نهایت، دو پیش‌بینی‌کنندهٔ مبتنی بر یادگیری جمعی، با استفاده از فنون رأی اکثریت و انباشتی ساخته شده و عملکرد سامانهٔ پیشنهادی FW-ECP بر روی دو مجموعه داده مخابراتی عمومی آزمون و مقایسه شده‌اند. از آنجاکه سامانهٔ پیشنهادی در ساختار خود روش‌هایی را هم برای رسیدگی به ماهیت نامتعادل و هم ابعاد بزرگ مجموعه‌های آموزشی در نظر گرفته‌است، نتایج به‌دست‌آمده عملکرد پیش‌بینی بهتری را در قیاس با سایر روش‌های مقایسه‌شده نشان می‌دهد.

در پژوهش انجام‌شده توسط یو و همکاران [۲۶]، یک شبکهٔ عصبی پس‌انتشار^۳ مبتنی بر بهینه‌سازی

طبقه‌بندی ذرات (PCO^4) برای مسئلهٔ پیش‌بینی ریزش مشتریان مخابراتی (به نام $PBCCP^5$) پیشنهاد شده‌است که به‌طور مکرر بهینه‌سازی طبقه‌بندی ذرات و محاسبهٔ میزان برازش ذرات را اجرا می‌کند. الگوریتم PCO که برگرفته از الگوریتم بهینه‌سازی ازدحام ذرات است، ذرات را با توجه به میزان برازش آن‌ها به سه دسته طبقه‌بندی و سرعت ذرات دسته‌های مختلف را با استفاده از روابط متمایزی به‌روز می‌کند. در نهایت، مقادیر وزن‌ها و آستانه‌های اولیهٔ شبکهٔ عصبی BP در فرایند آموزش روبه‌جلوی شبکه بهینه شده و بهبود قابل‌توجهی را در دقت پیش‌بینی ریزش مشتری به ارمغان می‌آورد.

اگرچه پژوهش‌های متعددی در خصوص عملکرد انواع روش‌های یادگیری ماشین برای حل مسئلهٔ پیش‌بینی ریزش مشتریان در صنعت مخابرات همراه انجام شده‌است، هنوز فقدان پژوهش جامعی که عملکرد طیف گسترده‌ای از روش‌های طبقه‌بندی، آشکارسازی هدف و کاهش تعداد ویژگی‌ها بر اساس روش‌های استخراج ویژگی‌ها و انتخاب ویژگی‌ها را در کنار یکدیگر ارزیابی کند، حس می‌شد. از این‌رو در مطالعهٔ انجام‌شده توسط ایمانی [۲۷]، روش‌های مختلف یادگیری ماشین شامل هفت طبقه‌بندی‌کنندهٔ (درخت تصمیم، رگرسیون لاجستیک، جنگل تصادفی، شبکهٔ عصبی پیش‌خور، حافظهٔ کوتاه‌مدت طولانی^۶، ماشین بردار پشتیبان و K نزدیک‌ترین همسایگی)، هفت آشکارساز هدف (آشکارساز زیرفضای همسان^۷، آشکارساز زیرفضای تطبیقی^۸، طرح‌ریزی زیرفضای متعامد^۹، نقشه‌بردار زاویه طیفی^{۱۰}، نقشه‌بردار زاویهٔ طیفی هسته^{۱۱}، کمینه‌سازی انرژی مقید^{۱۲} و آشکارساز هدف مبتنی بر پراکندگی^{۱۳})، ده روش کاهش ویژگی شامل چهار الگوریتم استخراج ویژگی (تجزیه و تحلیل مؤلفه‌های اصلی^{۱۴}، تجزیه و تحلیل تفکیک خطی^{۱۵}، استخراج ویژگی مبتنی بر خوشه‌بندی^{۱۶} و تعبیهٔ خط میانه- میانگین و ویژگی^{۱۷}) و شش مورد انتخاب

⁴ Particle Classification Optimization

⁵ Back-Propagation Customer Churn Prediction

⁶ Long Short Term Memory

⁷ Matched Subspace Detector

⁸ Adaptive Subspace Detector

⁹ Orthogonal Subspace Projection

¹⁰ Spectral Angle Mapper

¹¹ Kernel Spectral Angle Mapper

¹² Constrained Energy Minimization

¹³ Sparsity-based Target Detector

¹⁴ Principal Component Analysis

¹⁵ Linear Discriminant Analysis

¹⁶ Clustering-Based Feature Extraction

¹⁷ Median-Mean and Feature Line Embedding

¹ Wrapper

² minimum redundancy and maximum relevance

³ Back-Propagation Neural Network

مشتریان ریزشی استفاده می‌شود. در نهایت، تقسیم‌بندی مشتریان ریزش کرده با استفاده از الگوریتم خوشه‌بندی K-means انجام شده و مشتریان به گروه‌های مختلفی تقسیم می‌شوند که به گروه بازاریابی و تصمیم‌گیرندگان اجازه می‌دهد تا راهبردهای حفظ مشتریان را با دقت بیشتری انتخاب کنند.

با الهام از عملکرد خوب الگوهای پیش‌بینی‌کننده مبتنی بر یادگیری جمعی، در مطالعه انجام شده توسط بیهاری و فوکونه [۱۹]، یک الگوی مجموعه رأی‌گیری انعطاف‌پذیر دولایه برای پیش‌بینی نرخ ریزش مشتری در صنایع مخابراتی پیشنهاد شده است. مجموعه داده‌های مورد استفاده در این مطالعه شامل دو مجموعه داده عمومی مربوط به اطلاعات مشترکان تلفن همراه هستند که پس از مرحله پیش‌پردازش، به دو دسته نامتعادل (حالت عادی) و متعادل تبدیل شده‌اند. مجموعه‌های متعادل شامل تعداد مساوی نمونه برای هر دو طبقه است («Churn» و «Not-Churn») که از طریق اعمال روش نمونه‌برداری کاهشی تصادفی به دست آمده‌اند. همچنین، بررسی‌های گسترده‌ای برای تعیین شرایطی که الگو بهترین عملکرد را ارائه می‌دهد، انجام شده است. نتایج آزمایش‌های انجام شده نشان می‌دهد که الگوی پیش‌بینی پیشنهادی به طور قابل توجهی معیار امتیاز-F1 را هنگام در نظر گرفتن یک مجموعه داده متعادل افزایش داده است.

در پژوهش دیگری که توسط لالوانی و همکاران [۲۰] انجام شده، یک راهکار که شامل شش مرحله است، برای پیش‌بینی ریزش مشتری در صنعت مخابرات همراه ارائه شده است. در دو مرحله اول، پیش‌پردازش داده‌ها و تحلیل ویژگی‌ها انجام می‌شود. در مرحله سوم، فرایند انتخاب ویژگی با استفاده از الگوریتم جستجوی گرانشی^۸ اجرا شده و در مرحله بعد، داده‌ها به دو بخش آموزش و آزمون به ترتیب با نسبت هشتاد و بیست درصد تقسیم شده‌اند. در فرایند پیش‌بینی این پژوهش، از رایج‌ترین الگوهای پیش‌بینی‌کننده شامل رگرسیون لجستیک، بیز ساده، ماشین بردار پشتیبان، جنگل تصادفی، درخت‌های تصمیم‌گیری و k نزدیک‌ترین همسایه استفاده شده است. همچنین، روش‌های یادگیری جمعی مبتنی بر تقویت برای مشاهده اثر روی دقت الگوهای پیش‌بینی بررسی شده‌اند. به علاوه، از روش اعتبارسنجی متقابل K-fold روی مجموعه آموزش برای تنظیم ابرشاخص‌ها و جلوگیری از برازش بیش از حد الگوها استفاده شده است. در نهایت، نتایج

ویژگی (بهینه‌سازی پیشرفته کلونی مورچه‌های دودویی^۱، Relief-F، انتخاب ویژگی با یادگیری ساختار تطبیقی^۲، عملگر کمترین قدر مطلق گزینش و انقباض^۳، الگوریتم ژنتیک و انتخاب متوالی رو به عقب^۴) بررسی شده‌اند. عملکرد این روش‌ها بر روی سه مجموعه داده مخابراتی و با شش معیار ارزیابی شده‌اند.

از آنجا که مطالعات محدودی برای ترکیب عملیات پیش‌بینی ریزش مشتری و تقسیم‌بندی مشتریان به گروه‌های مختلف انجام شده است، در پژوهش ارائه شده توسط وئو و همکاران [۲۸]، هدف، ارائه یک چارچوب تجزیه و تحلیل مشتری یک پارچه برای مدیریت ریزش است. چارچوب پیشنهاد شده در این پژوهش شش بخش دارد که شامل پیش‌پردازش داده‌ها، تجزیه و تحلیل داده‌های اکتشافی (EDA^۵)، پیش‌بینی ریزش، تجزیه و تحلیل عامل، تقسیم‌بندی مشتریان و تجزیه و تحلیل رفتار مشتری است. این چارچوب، پیش‌بینی ریزش و فرایند تقسیم‌بندی مشتری را ادغام می‌کند تا به اپراتورهای مخابراتی امکان یک تجزیه و تحلیل کامل برای مدیریت بهتر و جلوگیری مؤثر از ریزش مشتری ارائه دهد. در آزمایش‌های انجام شده در این مقاله از سه مجموعه داده مخابراتی عمومی و شش الگوی یادگیری ماشین (رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، بیز ساده، Adaboost و شبکه عصبی مصنوعی) استفاده شده است. همچنین، برای مقابله با مشکلات ناشی از عدم تعادل در مجموعه داده‌های ریزش مشتری، در این پژوهش از روش نمونه‌برداری افزایشی اقلیت ساختگی (SMOTE^۶) استفاده شده است. برای ارزیابی الگوها نیز، معیارهای صحت و امتیاز-F1 محاسبه شده‌اند. نتایج به دست آمده نشان می‌دهد در مجموعه داده نخست، الگوی AdaBoost و در مجموعه داده دوم، الگوی جنگل تصادفی بهترین عملکرد را داشته‌اند. همچنین، در مجموعه داده سوم، اگرچه الگوی جنگل تصادفی بهترین عملکرد را از نظر صحت داشت، ولی از نظر امتیاز-F1، شبکه عصبی مصنوعی دارای بهترین عملکرد است. پس از انجام پیش‌بینی، از رگرسیون لجستیک بیزی^۷ برای بررسی عوامل و کشف برخی ویژگی‌های مهم برای تقسیم‌بندی

¹ Advanced Binary Ant Colony Optimization

² Feature Selection with Adaptive Structure Learning

³ Least Absolute Shrinkage and Selection Operator

⁴ Sequential Backward Selection

⁵ Exploratory Data Analysis

⁶ Synthetic Minority Oversampling Technique

⁷ Bayesian Logistic Regression

⁸ Gravitational Search Algorithm

به دست آمده در مجموعه آزمون با استفاده از ماتریس درهم‌ریختگی^۱ و معیار AUC ارزیابی شده‌اند. بر اساس نتایج به دست آمده مشخص شده‌است که الگوهای Adaboost و XGboost دارای بالاترین میزان صحت و امتیاز AUC هستند و نسبت به سایر الگوها عملکرد بهتری دارند.

۳- مجموعه داده‌های استفاده شده

به منظور مطالعه عملکرد روش پیشنهادی در پیش‌بینی ریزش مشتری در صنعت مخابرات، در این مقاله از دو مجموعه داده شناخته شده به نام‌های IBM_Telco [۹] و Duke_Cell2Cell [۲۹] استفاده می‌کنیم. این مجموعه داده‌ها به طور گسترده در مطالعات مختلف در خصوص مسئله پیش‌بینی ریزش مشتری استفاده شده و به صورت عمومی در دسترس هستند. مشخصات این مجموعه داده‌ها در جدول (۱) نشان داده شده‌است.

(جدول ۱-): مشخصات مجموعه داده‌های مورد مطالعه

(Table-1): Specifications of the studied datasets

Duke_Cell2Cell	IBM_Telco	مجموعه داده
71047	7043	تعداد کل نمونه‌ها
51047	7043	تعداد نمونه‌های قابل استفاده
58	21	تعداد ویژگی‌ها
14711	1869	تعداد مشتریان ریزشی
36336	5174	تعداد مشتریان غیر ریزشی
36	4	تعداد ویژگی‌های عددی
22	17	تعداد ویژگی‌های رسته‌ای

مجموعه داده IBM_Telco، یک مجموعه داده شناخته شده در حوزه پیش‌بینی ریزش مشتریان مخابراتی و همان‌طور که در جدول (۱) مشاهده می‌شود، شامل ۷۰۴۳ رکورد و ۲۱ ویژگی است. این مجموعه داده که «انجمن تحلیل تجاری IBM» آن را منتشر کرده، در مطالعات متعدد مربوط به ارزیابی عملکرد روش‌های پیش‌بینی ریزش مشتری در صنعت مخابرات همراه استفاده شده‌است [۱۹، ۲۰، ۲۷، ۲۸، ۳۰-۳۷]. همان‌طور که در جدول (۱) مشاهده می‌شود، مجموعه داده

¹ Confusion Matrix

IBM شامل ۱۸۶۹ رکورد ریزش مثبت است که ۲۶/۵ درصد از کل نمونه‌های موجود را تشکیل می‌دهد و در نتیجه، یک مجموعه داده نامتعادل است. در این مجموعه داده، تعداد ویژگی‌های عددی و رسته‌ای به ترتیب چهار و هفده ویژگی است.

دومین مجموعه داده مورد استفاده در این مقاله مجموعه داده Duke_Cell2Cell است که «مرکز مدیریت ارتباط با مشتری دانشگاه دوک» آن را ارائه کرده و در مطالعات متعددی مانند [۷، ۱۸، ۱۹، ۲۸، ۳۶، ۳۸-۴۱] استفاده شده‌است. این مجموعه داده شامل ۷۱۰۴۷ رکورد با ۳۶ ویژگی عددی و ۲۲ ویژگی رسته‌ای است. با این حال، ۵۱،۰۴۷ رکورد از ۷۱،۰۴۷ رکورد دارای یک برچسب مشخص در زمینه ریزش مشتری هستند و می‌توانند برای تجزیه و تحلیل عملکرد الگوهای معرفی شده استفاده شوند. این مجموعه داده نیز یک مجموعه نامتعادل است و ۱۴۷۱۱ نمونه دارای برچسب «بله» در زمینه ریزش مشتری هستند که ۲۸/۸ درصد از کل نمونه‌ها را تشکیل می‌دهد.

۴- روش بهینه‌سازی گرگ خاکستری

الگوریتم گرگ خاکستری یک روش فراابتکاری مبتنی بر هوش ازدحامی است که از رفتار گروهی گرگ‌ها در هنگام شکار الهام گرفته‌است. در طبیعت، چهار دسته گرگ خاکستری به نام‌های آلفا (α)، بتا (β)، دلتا (δ) و امگا (ω) که سلسله‌مراتب رهبری را مشخص می‌کنند، در هنگام شکار به کار گرفته می‌شوند. در ساختار الگوریتم بهینه‌سازی گرگ خاکستری نیز این سلسله‌مراتب چهارگانه به همراه سه مرحله اصلی شکار، شامل جستجوی طعمه، احاطه طعمه و حمله به آن برای انجام فرایند بهینه‌سازی شبیه‌سازی می‌شوند. به منظور الگوسازی ریاضی سلسله‌مراتب اجتماعی گرگ‌ها هنگام پیاده‌سازی الگوریتم بهینه‌سازی گرگ خاکستری، مناسب‌ترین راه‌حل را به‌عنوان آلفا (α) و دومین و سومین راه‌حل برتر را به ترتیب بتا (β) و دلتا (δ) در نظر می‌گیریم و بقیه راه‌حل‌ها امگا (ω) فرض می‌شوند. در الگوریتم GWO فرایند بهینه‌سازی به وسیله α ، β و δ هدایت می‌شود و سایر گرگ‌های ω به دنبال این سه گرگ می‌آیند.

همان‌طور که در بالا گفته شد، گرگ‌های خاکستری در طول شکار، نخست طعمه را محاصره می‌کنند. به منظور الگوسازی ریاضی رفتار محاصره‌ای گرگ‌ها روابط (۱) و (۲) به کار گرفته می‌شوند:

باید توجه کرد که در الگوریتم بهینه‌سازی گرگ خاکستری، به منظور الگوسازی ریاضی فرایند جستجوی طعمه (اکتشاف جواب بهینه^۱) و حمله به طعمه (استخراج جواب بهینه^۲)، مقدار \vec{a} را به تدریج در طول تکرارهای مختلف از دو به صفر کاهش می‌دهیم که منجر به کاهش \vec{A} در بازه $[-2a, 2a]$ می‌شود. هنگامی که $|A| > 1$ است، گرگ‌های خاکستری را مجبور می‌کند تا از طعمه جدا شوند و شکار مناسب‌تری را در محیط جستجو اکتشاف کنند. برعکس، هنگامی که مقادیر تصادفی \vec{A} در بازه $[-1, 1]$ باشد، موقعیت بعدی یک عامل جستجو می‌تواند در هر موقعیتی بین موقعیت فعلی آن و موقعیت طعمه باشد که منجر به فرایند استخراج جواب بهینه می‌شود. یکی دیگر از مؤلفه‌های GWO که به فرایند اکتشاف جواب بهینه کمک می‌کند، \vec{C} است که مقادیر تصادفی در بازه $[0, 2]$ دارد. این مؤلفه وزن‌های تصادفی را برای طعمه فراهم می‌کند تا هنگامی که مقدار آن بزرگ‌تر از یک است، باعث افزایش تأکید بر تأثیر طعمه ($C > 1$) و کاهش تأکید بر تأثیر طعمه در تعریف فاصله در معادله (۲) شود. به عبارت دیگر، این مؤلفه به GWO کمک می‌کند تا رفتار تصادفی بیشتری را در طول فرایند بهینه‌سازی از خود نشان دهد، که این امر به نفع افزایش قدرت اکتشاف الگوریتم و اجتناب از گیرافتادن در نقاط بهینه محلی است. گفتنی است \vec{C} برخلاف \vec{A} به صورت خطی کاهش نمی‌یابد و باعث افزایش قدرت اکتشاف نه تنها در طول تکرارهای اولیه، بلکه در تکرارهای نهایی می‌شود. بنابراین، این مؤلفه در جلوگیری از به دام افتادن در نقاط بهینه محلی، به‌ویژه در تکرارهای نهایی، بسیار مفید است.

۵- الگوی پیشنهادی پیش‌بینی ریزش

در این مقاله یک رویکرد یادگیری جمعی مبتنی بر هوش ازدحامی را برای حل مسئله پیش‌بینی ریزش مشتری در صنعت مخابرات همراه ارائه می‌کنیم. مراحل کلی رسیدگی به این مسئله با استفاده از روش‌های یادگیری ماشین در شکل (۱) نشان داده شده‌است. این روش شامل چهار مرحله اصلی شامل پیش‌پردازش داده‌ها، انتخاب ویژگی،

¹ Exploration
² Exploitation
³ Local Optima

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (2)$$

در روابط بالا، t نشان‌دهنده تکرار فعلی، \vec{X}_p بردار موقعیت طعمه، \vec{X} بردار موقعیت گرگ و \vec{C} و \vec{A} بردارهای ضرایب هستند که مطابق روابط (۳) و (۴) محاسبه می‌شوند:

$$\vec{A} = 2 \vec{a} \cdot r_1 - a \quad (3)$$

$$\vec{C} = 2 \cdot r_2 \quad (4)$$

در این روابط مقادیر \vec{a} در طول تکرارهای مختلف به صورت خطی از دو تا صفر کاهش می‌یابند و r_1 و r_2 بردارهای تصادفی بین صفر و یک هستند.

به منظور شبیه‌سازی ریاضی رفتار شکار گرگ‌های خاکستری، فرض می‌کنیم آلفا (بهترین راه‌حل)، بتا (دومین راه‌حل بهتر) و دلتا (سومین راه‌حل بهتر) دانش بهتری در مورد موقعیت بالقوه طعمه (راه‌حل بهینه) دارند. بنابراین، سه راه‌حل برتر به دست آمده در هر تکرار را ذخیره، و سایر عوامل جستجو (گرگ‌های امگا) را موظف می‌کنیم موقعیت‌های خود را مطابق با موقعیت بهترین عوامل جستجو بر اساس روابط (۵) تا (۱۱) به‌روزرسانی کنند. با این معادلات، یک عامل جستجو موقعیت خود را بر اساس موقعیت آلفا، بتا و دلتا در یک فضای جستجوی Ω بعدی به‌روز می‌کند. به عبارت دیگر، آلفا، بتا و دلتا موقعیت طعمه را تخمین می‌زنند و سایر گرگ‌ها موقعیت خود را به‌طور تصادفی در اطراف طعمه به‌روز می‌کنند.

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right| \quad (5)$$

$$\vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \quad (6)$$

$$\vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \quad (7)$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha) \quad (8)$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta) \quad (9)$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta) \quad (10)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (11)$$



آموزش و اعتبارسنجی الگوی یادگیری ماشین و در نهایت، ارزیابی عملکرد الگوی آموزش دیده با استفاده از زیرمجموعه آزمون است. جزئیات هر مرحله را در بخش‌های زیر شرح می‌دهیم.

۵-۱- پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها مجموعه‌ای از عملیات است که بر روی داده‌های خام اعمال می‌شود تا آن را برای فرایند یادگیری ماشین آماده کند. این مرحله از اهمیت بالایی برخوردار است، زیرا تأثیر مستقیمی بر عملکرد الگوریتم‌های پیش‌بینی ریزش مشتری دارد و مرحله پیش‌پردازش با کیفیت پایین، دقت پیش‌بینی‌ها را کاهش می‌دهد. مرحله پیش‌پردازش شامل مراحل مختلفی است که شامل حذف ویژگی‌های نامربوط، مدیریت داده‌های ناموجود، تبدیل ویژگی‌های رسته‌ای به عددی، مقیاس‌بندی ویژگی‌ها در یک محدوده مشخص و رسیدگی به عدم تعادل داده‌هاست. در ادامه، به توضیح هر یک از این مراحل می‌پردازیم.

۵-۱-۱- حذف ویژگی‌های نامربوط

در این مرحله داده‌های مربوط به ویژگی‌های نامرتبب مانند شناسه یکتای هر مشتری که دارای مقادیر منحصر به فرد برای هر مشترک است، حذف می‌شوند. این داده‌ها هویت مشتریان را نشان می‌دهند و هیچ ارتباطی با خروجی مورد نظر الگوریتم یادگیری ماشین، یعنی پیش‌بینی ریزش مشتری ندارند.

۵-۱-۲- مدیریت داده‌های ناموجود

هر دو مجموعه داده IBM_Telco و Duke_Cell2Cell مقادیری از دست رفته در برخی از ویژگی‌ها دارند که باید پیش از آموزش الگوی یادگیری ماشین به آنها رسیدگی شود. به طور مشخص، مجموعه داده IBM_Telco دارای یازده نمونه با مقادیر ناموجود در ویژگی «Total Charges» است که با توجه به این که تعدادشان کمتر از ۵٪ از کل نمونه‌های این مجموعه داده است، آن‌ها را حذف می‌کنیم [۴۲]. همچنین، در مورد مجموعه داده Duke_Cell2Cell، پانزده ویژگی با مقادیر ناموجود وجود دارد. از آنجاکه درصد نمونه‌هایی با مقادیر ناموجود در چهارده ویژگی کمتر از پنج درصد است، نمونه‌های مربوط به آن‌ها را حذف می‌کنیم و برای ویژگی «HandsetPrice»، که شامل ۲۸۹۸۲ نمونه با مقادیر

ناموجود است، این مقادیر را با مقدار متوسط این ویژگی (یعنی ۵۶/۳۴) جایگزین می‌کنیم. به این ترتیب، ۴۹۷۷۶ نمونه در مجموعه داده Duke_Cell2Cell باقی خواهد ماند که برای آموزش الگوی پیش‌بینی ریزش استفاده خواهند شد.

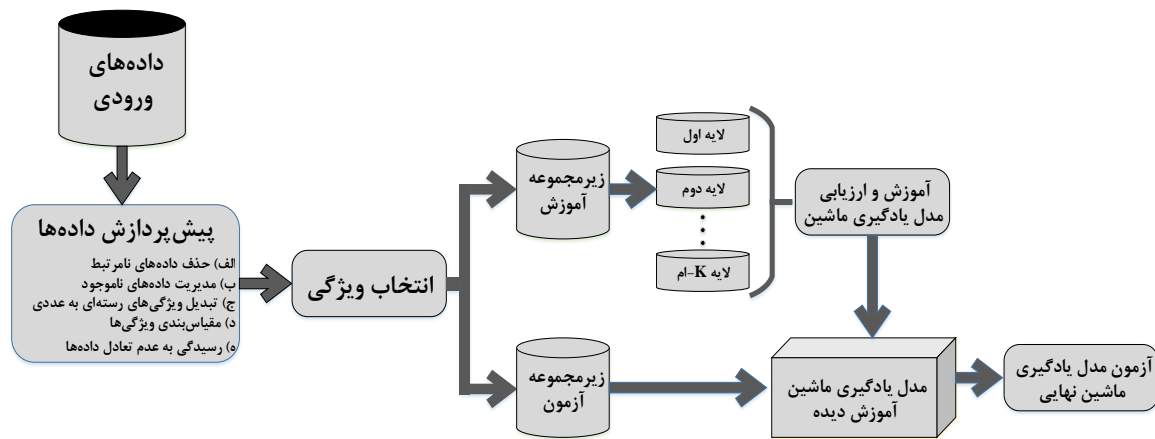
۵-۱-۳- تبدیل ویژگی‌های رسته‌ای به عددی

در این مرحله داده‌های ویژگی‌های مختلف به گونه‌ای تبدیل می‌شوند که بتوان از آنها برای آموزش الگوی پیش‌بینی ریزش مشتری به درستی استفاده کرد. در این راستا، نخست، ویژگی‌های رسته‌ای با دو مقدار (یعنی بله/خیر، مذکر/مونث و شناخته شده/ناشناس) را با استفاده از روش رمزگذاری برچسب به مقادیر عددی تبدیل می‌کنیم. به طور مشخص، در مجموعه داده IBM_Telco مقادیر بله و خیر در ویژگی‌های «Partner»، «Dependents»، «PhoneService»، «PaperlessBilling» و «Churn» به ترتیب با مقادیر یک و صفر جایگزین می‌شوند؛ همچنین، در ویژگی «جنسیت»، مقادیر مرد و زن را به یک و صفر تبدیل می‌کنیم؛ علاوه بر این، در مورد مجموعه داده Duke_Cell2Cell، نیز همین فرایند برای ویژگی‌های رسته‌ای اعمال می‌شود؛ سپس، با استفاده از روش One Hot Encoding (OHE)، ویژگی‌های باقی‌مانده با بیش از دو مقدار رسته‌ای را به بردارهای دودویی با اندازه N تبدیل می‌کنیم. گفتنی است در این مرحله ویژگی «ServiceArea» در مجموعه داده Duke_Cell2Cell را به دلیل این که دارای ۷۴۴ مقدار منحصر به فرد است، حذف می‌کنیم، زیرا تبدیل آن با روش OHE منجر به بردارهای دودویی با ابعاد بالا، و در نتیجه، افزایش پیچیدگی و کاهش کارایی فرایند آموزش الگوی پیش‌بینی ریزش می‌شود.

۵-۱-۴- مقیاس‌بندی ویژگی‌ها

در این مرحله، ویژگی‌های هر دو مجموعه داده IBM_Telco و Duke_Cell2Cell با استفاده از رابطه (۱) به محدوده [۰ ۱] نرمال می‌شوند. به این ترتیب، اهمیت همه ویژگی‌ها در الگوی پیش‌بینی به طور تقریبی برابر می‌شود و فرایند آموزش به نسبت ساده‌تر و با کیفیت‌تر خواهد بود.

$$x_{normal} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$



(شکل-۱): مراحل کلی پیش‌بینی ریزش مشتری با استفاده از الگوی یادگیری ماشین
(Figure-1): Overall steps of customer churn prediction using machine learning model

۵-۱-۵- متعادل‌سازی مجموعه‌های داده

همان‌طور که در جدول (۱) مشاهده شد، تعداد کم نمونه‌های ریزش در مقایسه با نمونه‌های بدون ریزش در هر دو مجموعه‌داده IBM_Telco و Duke_Cell2Cell، آن‌ها را به شدت نامتعادل می‌کند؛ که می‌تواند بر مرحله آموزش الگوی پیش‌بینی ریزش مشتری تأثیر منفی بگذارد [۴۳، ۱۰]. نمونه‌گیری تصادفی کاهشی و نمونه‌برداری تصادفی افزایشی، دو روش اصلی به منظور کاهش اثرات عدم تعادل در مجموعه‌داده‌های نامتعادل هستند [۲۰، ۲۸، ۳۷، ۴۴]. روش‌های نمونه‌برداری تصادفی کاهشی، همان‌طور که از نامشان پیداست، تلاش می‌کنند با حذف تصادفی برخی از نمونه‌های طبقه اکثریت (نمونه‌هایی از مشتریان ریزش نکرده) مجموعه‌داده‌ها را متعادل کنند. با این حال، این روش می‌تواند کارایی الگوی پیش‌بینی نهایی را کاهش دهد، زیرا برخی از داده‌های مفید نیز ممکن است در طول این فرایند حذف شوند. از سوی دیگر، هدف روش‌های نمونه‌برداری تصادفی افزایشی، کاهش مشکل عدم تعادل مجموعه‌داده‌ها با افزودن نمونه‌های طبقه اقلیت (نمونه‌های طبقه ریزش کرده) به مجموعه‌داده است. نتایج پژوهش‌های انجام شده در خصوص استفاده از روش‌های نمونه‌برداری تصادفی به منظور متعادل‌سازی مجموعه‌داده‌های ریزش مشتریان، نشان داده‌اند که روش‌های نمونه‌برداری تصادفی افزایشی به طور معمول به نتایج بهتری در خصوص دقت الگوی پیش‌بینی نهایی نسبت به نوع نمونه‌برداری تصادفی کاهشی دست می‌یابند، اما باید توجه کرد که این روش‌ها می‌توانند خطر برآزش بیش‌از حد در مرحله آموزش الگوی پیش‌بینی را افزایش دهند [۴۵]. با توجه به مطالب بالا، در این پژوهش از روش نمونه‌برداری تصادفی افزایشی به منظور متعادل‌سازی مجموعه‌داده‌های IBM_Telco و

Duke_Cell2Cell استفاده می‌کنیم. همچنین، به منظور دفع خطر امکان رخداد بیش‌برآزش در مرحله آموزش الگوی یادگیری ماشین از روش اعتبارسنجی متقابل ده‌لایه در روش پیشنهادی بهره می‌بریم.

۵-۲- انتخاب ویژگی و بهینه‌سازی ابرشاخص‌ها

اگرچه پس از مرحله پیش‌پردازش، داده‌ها برای استفاده در فرایند آموزش الگوی یادگیری ماشین آماده می‌شوند، اما وجود ویژگی‌های نوفه‌ای و زائد در مجموعه‌داده‌ها، همچنان می‌تواند بر عملکرد الگوی یادگیری ماشین تأثیر منفی بگذارد. بنابراین، مرحله انتخاب ویژگی که فرایند یافتن زیرمجموعه‌ای از ویژگی‌های مؤثر است، به طور معمول روی داده‌های از پیش‌پردازش شده انجام می‌شود. انتخاب ویژگی مزایای متعددی را برای افزایش کارایی الگوی پیش‌بینی فراهم می‌کند، زیرا فرایند آموزش الگوی یادگیری ماشین بر روی یک زیرمجموعه مؤثر از ویژگی‌های غنی از اطلاعات منجر به یادگیری با کیفیت بالا می‌شود و از برآزش بیش‌از حد آن بر روی مجموعه گسترده‌ای از ویژگی‌های نامربوط اضافی جلوگیری می‌کند [۴۱]. علاوه بر این، کاهش تعداد ویژگی‌ها، تفسیرپذیری الگوی توسعه‌یافته را افزایش می‌دهد و شرکت‌های مخابراتی را قادر می‌سازد ریشه‌های اصلی مشکلات را بهتر درک کنند و سیاست‌های مقابله مؤثری طراحی کنند.

فنون انتخاب ویژگی به طور کلی به دو دسته روش‌های صافی (فیلتر)^۱ و بسته‌بندی^۲ تقسیم می‌شوند. روش‌های صافی (فیلتر)، زیرمجموعه‌ای از ویژگی‌های مناسب را با توجه به نتایج به دست آمده از برخی معیارهای

¹ Filter Methods

² Wrapper Methods

آماري مانند امتياز فيشر^۱ [۴۶] انتخاب مي‌کنند. به عبارت ديگر، در روش‌هاي فيلتر، ويژگي‌ها بدون در نظر گرفتن تأثير واقعي آنها بر عملکرد الگوي پيش‌بيني انتخاب مي‌شوند. در مقابل، در روش‌هاي بسته‌بندی بازخوردی از كيفيت نتايج پيش‌بيني الگوي يادگيري ماشين به زيرمجموعه انتخابی ويژگي‌هاي ورودی وجود دارد. در روش‌هاي بسته‌بندی به‌طور معمول از يك الگوريتم بهينه‌سازي فراابتكاري به‌منظور يافتن زير مجموعه‌اي از ويژگي‌ها كه منجر به بالاترين عملکرد در الگوي يادگيري ماشين مي‌شوند، استفاده مي‌شود.

در اين پژوهش ما از روش بسته‌بندی با استفاده از الگوريتم بهينه‌سازي گرگ خاكستري به‌عنوان يك روش فراابتكاري مبتني بر هوش ازدحامی به‌منظور يافتن مؤثرترين ويژگي‌ها و همچنين يافتن مقادير بهينه ابرشاخص‌هاي الگو شامل ضرايب وزني هريك از يادگيرندگان پايه در مرحله نهايي تجميع آرا استفاده مي‌کنيم. شكل (۲) نحوه بازنمايي راه‌حل‌ها براي انتخاب ويژگي‌هاي بهينه به‌وسيله الگوريتم گرگ خاكستري را نشان مي‌دهد.

F_1	F_2	...	F_n	W_1	W_2	...	W_6
-------	-------	-----	-------	-------	-------	-----	-------

(شكل-۲): بازنمايي راه‌حل‌ها در الگوريتم بهينه‌سازي خاكستري گرگ

(Figure-2): Solution representation in GWO algorithm

در اين شكل F_1 تا F_n مقادير دودویی هستند كه وجود يا عدم‌وجود هر ويژگي در فرايند آموزش الگوي يادگيري ماشين را تعيين مي‌کنند. به عبارت ديگر، چنانچه F_i برابر با يك باشد، يعني ويژگي i -ام در فرايند آموزش حضور دارد و چنانچه مقدار F_i برابر با صفر باشد، ويژگي i -ام از فرايند آموزش حذف مي‌شود. همچنين، مقادير W_1 تا W_6 مقادير وزني مربوط به خروجی هريك از يادگيرندگان پايه هستند كه در بخش بعد به توضيح عملکرد آنها مي‌پردازيم.

۵-۳- آموزش الگوي يادگيري ماشين

در اين مرحله، نخست ساختار الگوي يادگيري ماشين موردنظر جهت پيش‌بيني ريزش مشتريان را طراحي کرده، سپس به آموزش و اعتبارسنجی اين الگو مي‌پردازيم. همان‌طور كه در بالا گفته شد، با توجه به عملکرد مناسب الگوريتم‌هاي يادگيري جمعی در حل مسئله پيش‌بيني

ريزش مشتريان، در اين پژوهش ما يك الگوي يادگيري جمعی دوسطحي مبتني بر روش انباشتگی را پيشنهاده مي‌کنيم كه از شش الگوي يادگيري پايه شامل MLP، SVM-RBF، DT، NB، KNN، و LR در هر سطح استفاده مي‌کند. ساختار اين الگوي پيشنهادهی در شكل (۳) نشان داده شده است. همان‌طور كه در اين شكل مشاهده مي‌شود، براي دست‌يابی به نتايج با كيفيت بالا در الگوي پيشنهادهی، در مجموع، از نظرات تركيب‌هاي مختلف از يادگيرندگان پايه استفاده شده است و در نهايت، نتايج حاصل از آنها به كمك روش ميانگين وزن‌دار آرا تجميع مي‌شوند.

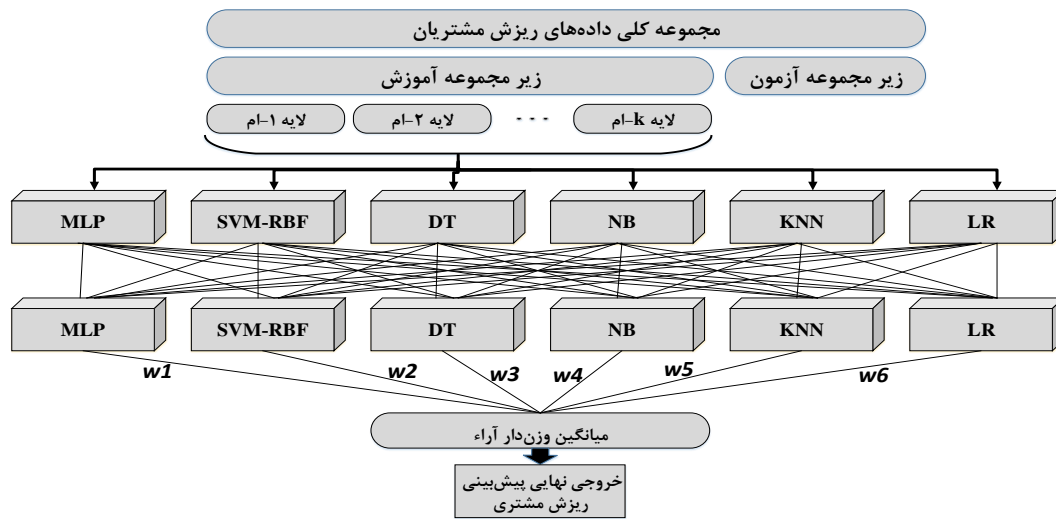
به‌منظور آموزش الگوي يادگيري جمعی پيشنهادهی، نخست هريك از مجموعه‌داده‌هاي ريزش مشتريان را با استفاده از روش برون‌گذاري^۲ به دو زيرمجموعه آموزش و آزمون با نسبت هفتاد به سي درصد تقسيم مي‌کنيم. سپس، به‌منظور بهره‌گيري از روش اعتبارسنجی متقابل k -لايه، زيرمجموعه آموزش را به ده زيربخش مساوی تقسيم مي‌کنيم. جدول (۲) شبه‌رمز مربوط به مراحل آموزش الگوي يادگيري جمعی پيشنهادهی را نشان مي‌دهد كه به كمك الگوريتم بهينه‌سازي گرگ خاكستري عمليات انتخاب ويژگي و يافتن مقادير بهينه براي وزن‌هاي يادگيرندگان پايه را به‌طور هم‌زمان انجام مي‌دهد. مطابق شبه‌رمز نشان داده شده، نخست، شاخص‌هاي اولیة الگوريتم بهينه‌سازي گرگ خاكستري شامل اندازه جمعيت و بشينه تعداد تکرارها و همچنين، مقادير شاخص‌هاي تابع برازش شامل وزن‌هاي W_A ، W_R ، W_P و W_{F1} را مشخص مي‌کنيم.

سپس، در دور آغازين هريك از عامل‌هاي جستجوی جمعيت اوليه را به‌صورت تصادفی مقداردهی اوليه کرده و اعداد تصادفی a ، r_1 و r_2 را تعيين و از روی آنها مقادير \vec{A} و \vec{C} را بر اساس روابط (۴) و (۵) محاسبه مي‌کنيم. سپس، براي هريك از عامل‌هاي جستجو مقدار تابع برازش را مطابق با رابطه (۱۲) محاسبه کرده و پس از مرتب‌سازي مقادير به‌دست‌آمده، اولين، دومين و سومين عامل جستجو با بهترين مقادير برازش را به‌ترتيب \vec{X}_α ، \vec{X}_β و \vec{X}_δ مي‌ناميم.

$$\text{Fitness} = W_A \times \text{Accuracy} + W_R \times \text{Recall} + W_P \times \text{Precision} + W_{F1} \times F1_score \quad (12)$$

² Hold-Out

¹ Fisher Score



(شکل-۳): ساختار الگوی یادگیری جمعی دوسطحی مبتنی بر روش انباشتگی پیشنهادی

(Figure-3): Structure of the proposed two level ensemble learning model based on the stacking method

آزمون که شامل داده‌های تاکنون دیده نشده‌است، با استفاده از معیارهای مختلف صحت، یادآوری، دقت و امتیاز F_1 بررسی می‌کنیم. چهار معیار بالا که در روابط (۱۴) تا (۱۷) تعریف شده‌اند، با استفاده از اطلاعات جمع‌آوری شده به وسیله ماتریس درهم‌ریختگی محاسبه می‌شوند؛ که در جدول (۳) نشان داده شده‌است.

(جدول-۲): شبه‌رمز مربوط به مراحل آموزش الگوی یادگیری

جمعی پیشنهادی به همراه الگوریتم بهینه‌سازی گرگ خاکستری

(Table-2): Pseudo code of the proposed ensemble learning model along with the GWO algorithm

Inputs:

et GWO and Fitness function parameters:

PopSize, MaxIter, W_A , W_R , W_P , and W_{F1}

Output:

Optimized features and weights of each base learner

GWO algorithm:

1. $t = 0$ (initial population)
2. Initialize a population of wolves (search agents) \vec{X}_p ($p=1,2,\dots,PopSize$)
3. Initialize \vec{a} , \vec{r}_1 , \vec{r}_2 , \vec{A} , and \vec{C}
4. Calculate the fitness function for each search agent according to Eq. 13
5. \vec{X}_α =the best search agent
6. \vec{X}_β =the second-best search agent
7. \vec{X}_δ =the third best search agent
8. **while** ($t \leq MaxIter$)
9. **for** each search agent
10. Update the position of each search

همان‌طور که در رابطه (۱۳) مشاهده می‌شود، تابع برازش در الگوریتم GWO به گونه‌ای تعریف شده‌است که کاربر می‌تواند قدرت هریک از معیارهای صحت، یادآوری، دقت و امتیاز F_1 را با تغییر مقادیر وزن‌های W_A ، W_P ، W_R ، W_{F1} بر اساس اهداف و سیاست‌های مورد نظر شرکت مخابراتی تنظیم کند. در ادامه الگوریتم بهینه‌سازی گرگ خاکستری، در یک حلقه تکرار تا زمانی که به بیشینه تعداد تکرارهای تعیین شده برسد، موقعیت هریک از عامل‌های جستجو را بر اساس روابط (۵) تا (۱۱) به‌روزرسانی می‌کند؛ سپس مطابق با مجموعه ویژگی‌ها و وزن‌های انتخاب‌شده به‌وسیله هریک از عوامل جستجو، هریک از یادگیرندگان پایه را بر روی $k-1$ لایه آموزش داده و نتایج حاصل از آن را بر روی لایه باقی‌مانده ارزیابی می‌کنیم. پس از این مرحله، دوباره مقادیر تصادفی \vec{r}_1 ، \vec{a} ، \vec{r}_2 ، \vec{A} و \vec{C} تعیین شده و مقدار تابع برازش برای هریک از عوامل جستجو مطابق با رابطه (۱۳) محاسبه و سه عامل برتر $(\vec{X}_\alpha, \vec{X}_\beta, \vec{X}_\delta)$ مشخص می‌شوند. در نهایت با اتمام تعداد دورهای تعیین شده، بهترین راه‌حل (\vec{X}_δ) به‌عنوان جواب بهینه که زیرمجموعه‌ای از ویژگی‌های برتر و همچنین وزن‌های بهینه را در عملیات میانگین‌گیری نهایی مشخص می‌کند، تعیین می‌شود.

۴-۵- ارزیابی عملکرد الگوی آموزش دیده بر

روی زیرمجموعه آزمون

برای ارزیابی الگوی یادگیری ماشین پیشنهادی، عملکرد آن را در پیش‌بینی ریزش مشتری بر روی زیرمجموعه



به درستی آن‌ها را در طبقه ریزش نکرده قرار داده‌است؛ و منفی کاذب (FN) نمونه‌هایی را در بر می‌گیرد که طبقه واقعی آن‌ها به طبقه ریزش کرده تعلق دارد، ولی الگوی پیش‌بینی طبقه آن‌ها را اشتباه تشخیص داده و آن‌ها را در طبقه ریزش نکرده قرار داده‌است. علاوه بر چهار معیار بالا، به منظور ارزیابی عملکرد الگوی پیشنهادی در تشخیص طبقه‌های مثبت و منفی، معیار مساحت زیرمنحنی (AUC) را نیز بر روی نتایج به دست آمده از زیرمجموعه آزمون محاسبه می‌کنیم. معیار AUC مساحت زیر نمودار مشخصه عملکرد (ROC¹) را تعیین می‌کند. منحنی ROC، یک نمودار گرافیکی است که توانایی تشخیص یک سامانه طبقه‌بند دودویی را به‌ازای مقادیر مختلف سطح آستانه نشان می‌دهد. محور عمودی در منحنی ROC، معیار نرخ مثبت صحیح (TPR²) و محور افقی آن، معیار نرخ مثبت کاذب (FPR³) برای سامانه یادگیری تحت بررسی را نشان می‌دهد که تعاریف آن‌ها در روابط (۱۷) و (۱۸) نشان داده شده‌است.

$$TPR = \frac{TP}{TP + FN} \quad (17)$$

$$FPR = \frac{FP}{FP + TN} \quad (18)$$

هرچه مقدار مساحت زیر نمودار ROC (یعنی مقدار AUC) یک الگوی یادگیری ماشین بالاتر باشد، نشان‌دهنده عملکرد بهتر آن الگو در تشخیص هر دو طبقه ریزش کرده و ریزش نکرده است.

۶- نتایج پیاده‌سازی الگوی پیشنهادی

در این بخش به منظور بررسی و مقایسه عملکرد الگوی پیش‌بینی ریزش مشتری پیشنهادی، نتایج حاصل از پیاده‌سازی آن را با استفاده از زبان برنامه‌نویسی پایتون ۳.۱۰.۸ ارزیابی می‌کنیم.

۶-۱- تعیین مقادیر شاخص‌ها

جدول (۴) مقادیر شاخص‌های استفاده‌شده در فرایند شبیه‌سازی، الگوی ارائه‌شده را نشان می‌دهد که به صورت تجربی تنظیم شده‌اند. همان‌طور که در این جدول مشاهده می‌شود، تعداد بیشینه تکرارها و اندازه جمعیت اولیه در الگوریتم بهینه‌سازی گرگ خاکستری به ترتیب برابر با صد تکرار و پنجاه راه‌حل تصادفی اولیه انتخاب شده‌اند.

¹ Receiver Operating Characteristic

² True Positive Rate

³ False Positive Rate

agent using Eq. 6 to Eq. 12

11. **end**
12. **for** each search agent
13. **for** each base-learner
14. **for** $k=1: K$ -fold
the classifier on $K-1$ folds, and evaluate on the remaining fold
16. **end**
17. **end**
19. **end**
20. Update \vec{a} , \vec{r}_1 , \vec{r}_2 , \vec{A} , and \vec{C}
Calculate the fitness function for each search agent according to Eq. 13
22. Update \vec{X}_α , \vec{X}_β , and \vec{X}_δ
23. $t = t+1$
24. **end while**
25. Return the global best search agent (\vec{X}_α)

(جدول ۳-): ماتریس درهم‌ریختگی برای ارزیابی الگوی

پیش‌بینی ریزش مشتری

(Table-3): Confusion matrix for evaluation of customer churn prediction model

طبقه پیش‌بینی / طبقه واقعی	ریزش کرده	ریزش نکرده
ریزش کرده	مثبت صحیح (TP)	منفی کاذب (FN)
ریزش نکرده	مثبت کاذب (FP)	منفی صحیح (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$F1_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (16)$$

همان‌طور که در جدول (۳) مشاهده می‌شود، مثبت صحیح (TP) شامل نمونه‌هایی از زیرمجموعه آزمون است که طبقه واقعی آنها متعلق به طبقه ریزش کرده و الگوی پیش‌بینی ریزش مشتری نیز طبقه آن‌ها را به درستی پیش‌بینی کرده‌است؛ و مثبت کاذب (FP) نیز شامل نمونه‌هایی که طبقه واقعی آن‌ها متعلق به طبقه ریزش نکرده است، ولی الگوی پیش‌بینی به اشتباه آن‌ها را در طبقه ریزش کرده قرار داده‌است. همچنین منفی صحیح (TN) شامل نمونه‌هایی از زیرمجموعه آزمون است که در طبقه واقعی ریزش نکرده قرار دارند و الگوی پیش‌بینی نیز

(Table-4): Parameters values for proposed model simulation

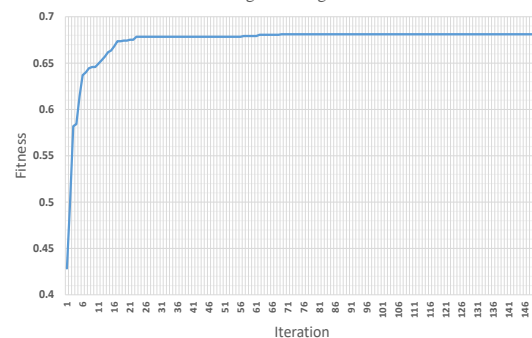
مقدار	شاخص
۱۰۰	بیشینه تکرارها
۵۰	اندازه جمعیت اولیه
۳	تعداد لایه‌های MLP
Sigmoid	تابع فعال‌سازی MLP
RBF	تابع هسته SVM
۱۰	تعداد لایه‌ها در روش اعتبارسنجی (K)
۰.۲۵	وزن معیار صحت در تابع برازش (W_A)
۰.۲۵	وزن معیار یادآوری در تابع برازش (W_R)
۰.۲۵	وزن معیار دقت در تابع برازش (W_P)
۰.۲۵	وزن معیار F_1 در تابع برازش (W_{F1})

همچنین، برای الگوی MLP در الگوی یادگیری جمعی پیشنهادی، یک شبکه عصبی سه‌لایه با تابع فعال‌سازی سیگموئید و تابع هسته در الگوی SVM از نوع RBF در نظر گرفته شده است. به علاوه، تعداد لایه‌های در نظر گرفته شده برای روش اعتبارسنجی متقابل برابر با ده لایه است و مقادیر وزن‌های معیارهای صحت، یادآوری، دقت و امتیاز F_1 در تابع برازش به صورت برابر با ۰.۲۵ انتخاب شده‌اند.

۶-۲- آموزش الگوی یادگیری جمعی پیشنهادی

پس از تعیین مقادیر شاخص‌ها، نسبت به آموزش الگوی پیشنهادی مطابق با الگوریتم نشان‌داده شده در جدول (۲) اقدام می‌کنیم. شکل (۴) نمودار همگرایی الگوریتم بهینه‌سازی گرگ خاکستری را نشان می‌دهد.

Convergence Diagram



(شکل-۴): نمودار همگرایی الگوریتم گرگ خاکستری

(Figure-4): Convergence diagram of GWO algorithm

جداول (۵) تا (۸) به ترتیب ویژگی‌های انتخاب شده و وزن‌های بهینه را برای هر یک از یادگیرندگان پایه در هر یک از مجموعه‌داده‌های IBM_Telco و Duke_Cell2Cell که به وسیله الگوریتم GWO به دست آمده‌اند، نمایش می‌دهند.

همان‌طور که در جدول (۵) دیده می‌شود، ۱۱ ویژگی از ۲۱ ویژگی موجود در مجموعه‌داده IBM_Telco به وسیله الگوریتم بهینه‌سازی گرگ خاکستری به عنوان ویژگی‌های مؤثر در پیش‌بینی ریزش مشتری انتخاب شده‌اند. همچنین، جدول (۶) ضرایب وزنی تعیین شده برای هر یک از یادگیرندگان پایه را نشان می‌دهد که در آن شبکه عصبی پرسپترون چندلایه دارای بیشترین ضریب و طبقه‌بند بیز ساده دارای کمترین ضریب است.

(جدول-۵): ویژگی‌های انتخاب شده به وسیله الگوریتم

بهینه‌سازی گرگ خاکستری برای مجموعه‌داده IBM_Telco

(Table-5): Selected features by grey wolf optimization algorithm for IBM_Telco dataset

ردیف	ویژگی
۱	Contract
۲	Internet Service
۳	Monthly Charges
۴	Online Backup
۵	Online Security
۶	Paperless Billing
۷	Phone Service
۸	Senior Citizen
۹	Tech Support
۱۰	Tenure
۱۱	Total Charges

(جدول-۶): مقادیر وزن‌های یادگیرندگان پایه تعیین شده

به وسیله الگوریتم بهینه‌سازی گرگ خاکستری برای

IBM_Telco

(Table-6): Optimum weights of base learners determined by gray wolf optimization algorithm for IBM_Telco

مقدار	وزن یادگیرندگان پایه
۰.۷۱	Weight of MLP (w_1)
۰.۶۷	Weight of SVM (w_2)
۰.۵۹	Weight of DT (w_3)
۰.۵۲	Weight of NB (w_4)
۰.۵۶	Weight of KNN (w_5)
۰.۶۴	Weight of LR (w_6)

همچنین، در جدول (۷) ویژگی‌های انتخاب شده برای پیش‌بینی ریزش مشتریان در مجموعه‌داده Duke_Cell2Cell نشان می‌دهد که پانزده ویژگی مؤثر از مجموع ۵۸ ویژگی موجود در این مجموعه‌داده به وسیله الگوریتم بهینه‌سازی گرگ خاکستری انتخاب شده‌اند. تعداد کم ویژگی‌های انتخاب شده به وسیله الگوریتم گرگ خاکستری نشان‌دهنده وجود تعداد زیادی ویژگی در مجموعه‌داده Duke_Cell2Cell است، که ارتباط ضعیفی با خروجی مورد نظر (ریزش یا عدم ریزش مشتری) دارند. البته باید توجه کرد که تعداد کم ویژگی‌های ورودی در الگوی پیش‌بینی، خود یک مزیت برای شرکت مخابراتی

۳-۶- بررسی نتایج حاصل از یادگیرندگان پایه و الگوی پیشنهادی بر روی زیرمجموعه آزمون

در این بخش به ارائه و تحلیل نتایج به دست آمده از یادگیرندگان پایه و همچنین، الگوی یادگیری جمعی پیشنهادی، بر روی زیرمجموعه آزمون که تاکنون به وسیله هیچ یک از الگوهای یادگیری ماشین به کار گرفته دیده نشده است، می پردازیم. در این راستا، مقادیر معیارهای متعارف برای اندازه گیری نحوه عملکرد الگوهای یادگیری ماشین، شامل معیارهای صحت، دقت، یادآوری، امتیاز F1 و AUC، برای مجموعه داده های IBM_Telco و Duke_Cell2Cell به ترتیب در جداول (۹) و (۱۰) گزارش شده اند.

(جدول-۹): مقادیر معیارهای ارزیابی یادگیرندگان پایه و الگوی

پیشنهادی در مجموعه داده IBM_Telco

(Table-9): Values of the evaluation metrics for base learners and proposed model in IBM_Telco dataset

الگوی پیش بینی	ACC	PRE	REC	F1	AUC
MLP	۷۶/۳	۵۳/۹	۷۸/۵	۶۳/۹	۸۳/۹
SVM	۷۶/۵	۵۴/۸	۷۶/۲	۶۳/۷	۸۳/۴
DT	۷۷/۱	۵۴/۵	۷۴/۸	۶۳/۱	۸۲/۷
NB	۷۴/۳	۵۱/۶	۷۶/۷	۶۱/۷	۸۲/۲
KNN	۷۵/۷	۵۴/۴	۷۱/۶	۶۱/۸	۸۲
LR	۷۵/۴	۵۳/۱	۷۹/۱	۶۳/۵	۸۴/۱
الگوی پیشنهادی	۸۳/۷	۷۶/۴	۸۱/۲	۷۸/۷	۸۴/۸

همان طور که در جدول (۹) مشاهده می شود، الگوی پیشنهادی در تمام معیارها دارای برتری است؛ که این امر عملکرد مطلوب آن را در پیش بینی ریزش مشتریان بر روی مجموعه داده IBM_Telco نشان می دهد. علت اصلی برتری الگوی پیشنهادی به دلیل تجمع نقاط قوت تمامی الگوهای پایه در قالب یک الگوی یادگیری جمعی دوسطحی مبتنی بر روش انباشتگی است.

مقادیر گزارش شده در جدول (۱۰) به وضوح نشان می دهند که الگوی یادگیری جمعی پیشنهادی در مجموعه داده Duke_Cell2Cell نیز از لحاظ تمامی معیارهای بررسی شده به نتایج بهتری نسبت به یادگیرندگان پایه دست می یابد.

محسوب می شود، زیرا این امکان را فراهم می کند تا کارشناسان بازاریابی شرکت بتوانند به شکل دقیق تری تأثیر هر یک از ویژگی ها را تفسیر کرده و سیاست های کارآمدتری برای جلوگیری از ریزش مشتریان پیش گیرند. به علاوه، جدول (۸) نشان می دهد که از میان یادگیرندگان پایه که برای پیش بینی ریزش مشتری در مجموعه داده Duke_Cell2Cell به کار گرفته شده اند، الگوی رگرسیون لجستیک بالاترین و الگوی درخت تصمیم پایین ترین ضریب وزنی را دارند.

(جدول-۷): ویژگی های انتخاب شده به وسیله الگوریتم

بهینه سازی گرگ خاکستری برای مجموعه داده

Duke_Cell2Cell

(Table-7): Selected features by grey wolf optimization algorithm for Duke_Cell2Cell dataset

ردیف	ویژگی
۱	AgeHH1
۲	Credit Rating
۳	Current Equipment Days
۴	Handset Models
۵	Handset Refurbished
۶	Handsets
۷	Made Call To Retention Team
۸	Monthly Minutes
۹	Off Peak Calls In Out
۱۰	Outbound Calls
۱۱	Received Calls
۱۲	Responds To Mail Offers
۱۳	Retention Calls
۱۴	Retention Offers Accepted
۱۵	Total Recurring Charge

(جدول-۸): مقادیر وزن های یادگیرندگان پایه تعیین شده

به وسیله الگوریتم بهینه سازی گرگ خاکستری برای

Duke_Cell2Cell

(Table-8): Optimum weights of base learners determined by gray wolf optimization algorithm for Duke_Cell2Cell

مقدار	وزن یادگیرندگان پایه
۰/۷۲	Weight of MLP (w_1)
۰/۶۶	Weight of SVM (w_2)
۰/۵۱	Weight of DT (w_3)
۰/۶۱	Weight of NB (w_4)
۰/۵۷	Weight of KNN (w_5)
۰/۷۶	Weight of LR (w_6)

(جدول-۱۰): مقادیر معیارهای ارزیابی یادگیرندگان پایه و

الگوی پیشنهادی در مجموعه داده Duke_Cell2Cell

(Table-10): Values of the evaluation metrics for base learners and proposed model in Duke_Cell2Cell dataset

الگوی پیش‌بینی	ACC	PRE	REC	F1	AUC
MLP	۵۴/۵	۳۹/۸	۵۷/۴	۴۷	۶۰/۷
SVM	۵۴/۸	۴۱/۹	۵۴/۵	۴۷/۴	۵۸/۳
DT	۵۹/۱	۳۸/۲	۴۴/۱	۴۰/۹	۵۷/۶
NB	۵۴/۲	۳۹/۳	۵۲/۲	۴۴/۸	۵۶/۱
KNN	۵۵/۳	۳۷/۴	۴۹/۸	۴۲/۷	۵۵/۸
LR	۵۷/۹	۴۱/۶	۶۲/۴	۴۹/۹	۶۱/۲
الگوی پیشنهادی	۷۴/۳	۶۴/۱	۶۷/۳	۶۵/۷	۶۳/۴

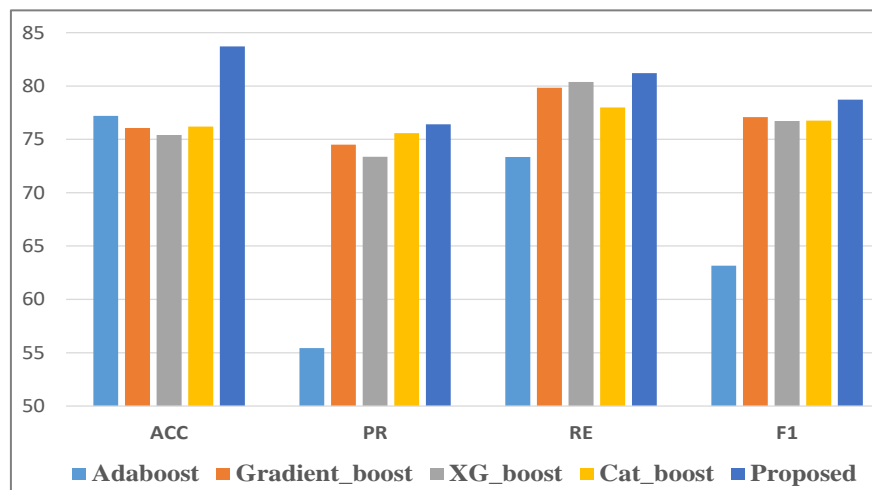
۴-۶- مقایسه نتایج الگوی پیشنهادی با سایر الگوهای یادگیری جمعی

در این بخش، به منظور انجام یک بررسی منصفانه در خصوص نحوه عملکرد الگوی یادگیری جمعی پیشنهادی در این پژوهش، نتایج حاصل از آن را با سایر روش‌های یادگیری جمعی متداول شامل الگوهای Adaboost، Gradient_boost، XG_boost و Cat_boost مقایسه می‌کنیم. به این منظور مقادیر معیارهای ارزیابی عملکرد سامانه‌های یادگیری ماشین را برای هر یک از الگوهای یادگیری جمعی ذکر شده محاسبه کرده و نتایج به دست آمده را برای مجموعه داده‌های IBM_Telco و

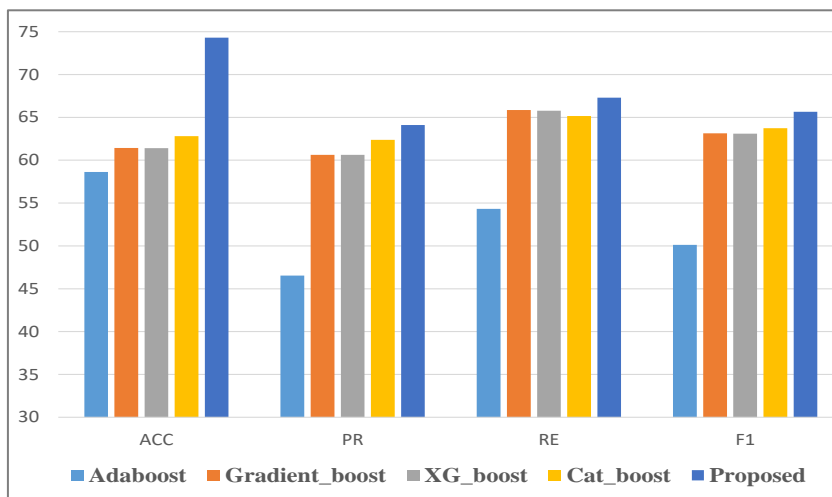
Duke_Cell2Cell به صورت نمودار میله‌ای به ترتیب در شکل‌های (۵) و (۶) نشان داده‌ایم. همان‌طور که در این اشکال مشاهده می‌شود، الگوی یادگیری جمعی پیشنهادی در این پژوهش برتری محسوسی را نسبت به سایر الگوهای یادگیری جمعی نشان داده شده، به ویژه از لحاظ معیار صحت از خود نشان می‌دهد. دلیل اصلی این امر استفاده از یک الگوی یادگیری جمعی دوسطحی بر اساس روش انباشتگی است که شاخص‌های آن با استفاده از هوش ازدحامی بهینه‌سازی شده‌اند. همچنین، نتایج به دست آمده در این بخش مبین این نکته است که استفاده از الگوریتم بهینه‌سازی گرگ خاکستری در مرحله انتخاب ویژگی‌ها منجر به یافتن مجموعه ویژگی‌های مؤثری می‌شود که عملکرد الگوی پیش‌بینی ریزش مشتری را به شدت ارتقا می‌بخشد.

۵-۶- مقایسه نتایج الگوی پیشنهادی با الگوهای ارائه شده در مطالعات پیشین

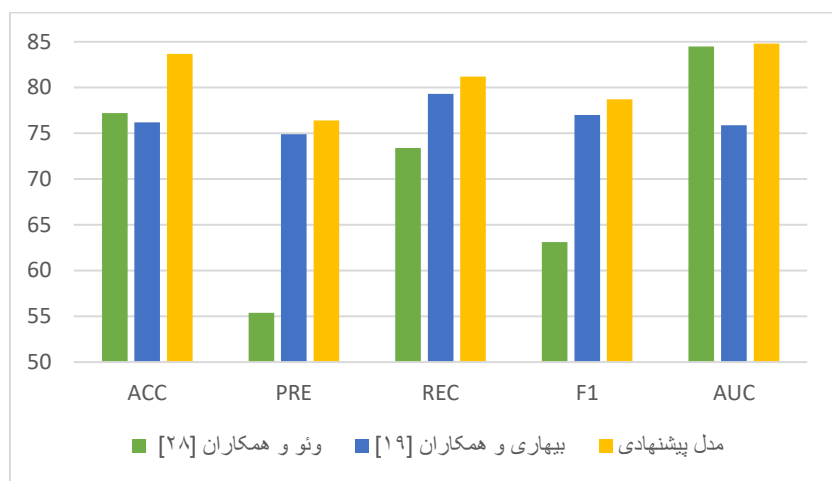
در قسمت پایانی از این بخش، نتایج به دست آمده را از الگوی پیشنهادی با نتایج حاصل از الگوهای ارائه شده در مقالات [۲۸] و [۱۹] مقایسه می‌کنیم. در این راستا، نخست، الگوهای ارائه شده در این مقالات را پیاده‌سازی کرده و سپس نتایج حاصل از اجرای آن‌ها را بر روی مجموعه داده‌های IBM_Telco و Duke_Cell2Cell، به کمک محاسبه معیارهای ارزیابی بررسی می‌کنیم.



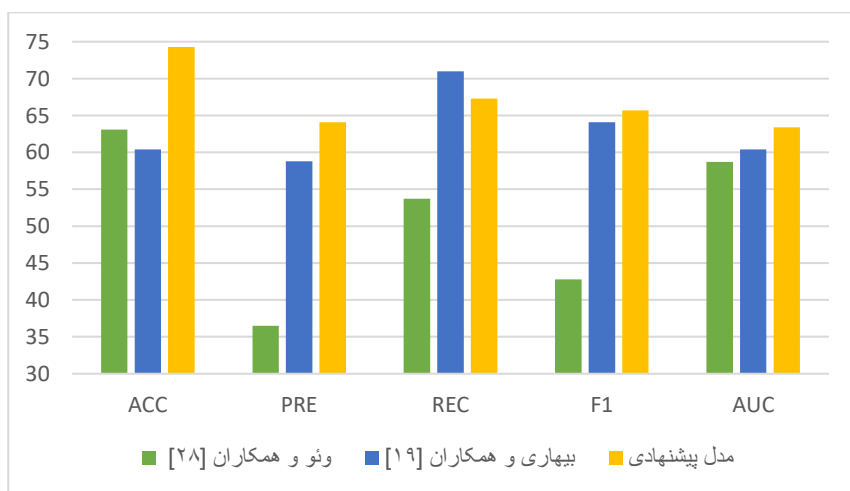
(شکل-۵): مقایسه نتایج الگوی پیشنهادی با سایر الگوهای یادگیری جمعی در مجموعه داده IBM_Telco (Figure-5): Comparing the results of the proposed model with other ensemble learning models in the IBM_Telco dataset



(شکل-۶): مقایسه نتایج الگوی پیشنهادی با سایر الگوهای یادگیری جمعی در مجموعه داده Duke_Cell2Cell
 (Figure-6): Comparing the results of the proposed model with other ensemble learning models in the Duke_Cell2Cell dataset



(شکل-۷): مقایسه الگوی پیشنهادی با الگوهای ارائه شده در مطالعات پیشین بر روی مجموعه داده IBM_Telco
 (Figure-7): Comparing the proposed model with the models presented in previous studies on the IBM_Telco dataset



(شکل-۸): مقایسه الگوی پیشنهادی با الگوهای ارائه شده در مطالعات پیشین بر روی مجموعه داده Duke_Cell2Cell
 (Figure-8): Comparing the proposed model with the models presented in previous studies on the Duke_Cell2Cell dataset

ارائه شده به وسیله زبان برنامه‌نویسی پایتون، سه دسته آزمایش جهت ارزیابی عملکرد این الگو انجام شده و خروجی‌های به دست آمده از لحاظ معیارهای متعارف ارزیابی عملکرد سامانه‌های یادگیری ماشین شامل معیارهای صحت، دقت، یادآوری، امتیاز F1 و AUC بررسی شده‌اند. نتایج به دست آمده از آزمایش‌های انجام شده، به وضوح برتری الگوی ارائه شده نسبت به عملکرد هریک از یادگیرندگان پایه، الگوهای شناخته شده یادگیری جمعی شامل الگوهای Adaboost، Gradient_boost، XG_boost و Cat_boost و همچنین، دو الگوی ارائه شده در مطالعات پیشین بررسی شده را نشان می‌دهد. به عنوان پیشنهادهایی جهت انجام پژوهش‌های آینده در حوزه ارائه راه حل برای مسئله پیش‌بینی ریزش مشتریان، می‌توان از سایر الگوریتم‌های فراابتکاری به ویژه الگوریتم‌های بهینه‌سازی الهام گرفته شده از انسان^۱ برای انجام فرایند بهینه‌سازی استفاده کرد. یکی از این الگوریتم‌های بهینه‌سازی جدید که می‌تواند تناسب خوبی با مسئله پیش‌بینی ریزش مشتریان داشته باشد، الگوریتم بهینه‌سازی معاملات بورس^۲ [۴۷] است که در مطالعات آینده به آن پرداخته خواهد شد. همچنین به کارگیری انواع دیگر ساختارهای یادگیری جمعی و یادگیرندگان پایه از مواردی است که می‌توان در پژوهش‌های پیش‌رو به آن‌ها پرداخت.

همان‌طور که در شکل (۷) مشاهده می‌شود، الگوی پیشنهادی در این مقاله عملکرد بهتری را از لحاظ تمامی معیارهای ارزیابی شده نسبت به الگوهای ارائه شده در مطالعات [۲۸] و [۱۹] بر روی مجموعه داده IBM_Telco از خود نشان می‌دهد. دلیل اصلی این امر، علاوه بر ساختار یادگیری جمعی دوسطحی استفاده شده، عملکرد مناسب الگوریتم بهینه‌سازی گرگ خاکستری در انتخاب ویژگی‌های مؤثر و همچنین یافتن شاخص‌های بهینه برای وزن‌های هریک از یادگیرندگان پایه در الگوی پیشنهادی است. همچنین، شکل (۸) نشان می‌دهد که الگوی پیشنهادی ما در تمامی معیارهای ارزیابی شده به جز معیار یادآوری نسبت به دو الگوی دیگر دارای برتری است. البته باید توجه کرد همان‌طور که پیشتر گفته شد، الگوی پیشنهادی در این پژوهش از قابلیت تنظیم وزن‌های هریک از معیارهای صحت، دقت، یادآوری و امتیاز F1 در تابع برازش الگوریتم بهینه‌سازی برخوردار است. بنابراین، چنانچه کاربری بسته به کاربرد مورد نظر نیاز به افزایش قدرت معیار یادآوری را داشته باشد، می‌تواند با افزایش ضریب وزنی به صورتی تنظیم کند که منجر به افزایش این معیار در نتایج خروجی آن شود.

۷- جمع بندی

در این پژوهش به مسئله پیش‌بینی ریزش مشتریان در شرکت‌های ارائه‌دهنده خدمات مخابرات همراه پرداخته شده است. به این منظور، یک الگوی یادگیری جمعی دوسطحی مبتنی بر روش انباشتگی ارائه شده است که از یادگیرندگان پایه شامل MLP، SVM-RBF، DT، NB، KNN و LR در هر سطح بهره می‌برد. در این مقاله، همچنین، به منظور انتخاب ویژگی‌های مناسب برای پیش‌بینی ریزش مشتریان، از الگوریتم بهینه‌سازی گرگ خاکستری استفاده شده است. این الگوریتم علاوه بر انجام فرایند انتخاب ویژگی، نسبت به تعیین ضرایب وزنی بهینه در خروجی هریک از یادگیرندگان پایه اقدام می‌کند. تابع برازش استفاده شده در این الگوریتم به نحوی تعریف شده است که کاربر می‌تواند بسته به سیاست‌های شرکت و کاربرد مورد نظر، ضرایب هریک از معیارهای ارزیابی را تعیین کند. در این پژوهش به منظور ارزیابی عملکرد الگوی پیشنهادی، نتایج خروجی آن بر روی مجموعه داده‌های IBM_Telco و Duke_Cell2Cell بررسی شده‌اند. به این منظور پس از پیاده‌سازی الگوی

8-Refrence

۸- مراجع

- [1] W. Jianxun, "A study on customer acquisition cost and customer retention cost: Review and outlook," *INNOVATION AND MANAGEMENT*, 2012.
- [2] A. Bilal Zorić, "Predicting customer churn in banking industry using neural networks," *Interdisciplinary Description of Complex Systems: INDECS*, vol. 14, no. 2, pp. 116-124, 2016.
- [3] K. G. M. Karvana, S. Yazid, A. Syalim, and P. Mursanto, "Customer churn analysis and prediction using data mining models in banking industry," in *2019 International Workshop on Big Data and Information Security (IWBIIS)*, 2019, pp. 33-38: IEEE.
- [4] A. Keramati, H. Ghaneei, and S. M. Mirmohammadi, "Developing a prediction model for customer churn from electronic banking services using data mining," *Financial Innovation*, vol. 2, no. 1, pp. 1-13, 2016.

¹ Human-inspired Optimization Algorithms

² Stock Exchange Trading Optimization Algorithm



- machine learning approach," *Computing*, vol. 104, no. 2, pp. 271-294, 2022.
- [21] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46-61, 2014.
- [22] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," *IEEE Transactions on neural networks*, vol. 11, no. 3, pp. 690-696, 2000.
- [23] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends," *Computers & Operations Research*, vol. 34, no. 10, pp. 2902-2917, 2007.
- [24] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert systems with applications*, vol. 34, no. 1, pp. 313-327, 2008.
- [25] P. C. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6714-6720, 2009.
- [26] R. Yu, X. An, B. Jin, J. Shi, O. A. Move, and Y. Liu, "Particle classification optimization-based BP network for telecommunication customer churn prediction," *Neural Computing and Applications*, vol. 29, no. 3, pp. 707-720, 2018.
- [27] M. Imani, "Customer Churn Prediction in Telecommunication Using Machine Learning: A Comparison Study," *AUT Journal of Modeling and Simulation*, vol. 52, no. 2, pp. 8-8, 2020.
- [28] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated churn prediction and customer segmentation framework for telco business," *IEEE Access*, vol. 9, pp. 62118-62136, 2021.
- [29] "telecom churn (cell2cell)," <https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom>, 2018.
- [30] E. Hanif, "Applications of data mining techniques for churn prediction and cross-selling in the telecommunications industry," Dublin Business School, 2019.
- [31] J. Pamina, B. Raja, S. SathyaBama, M. Sruthi, and A. VJ, "An effective classifier for predicting churn in telecommunication," *Jour of Adv Research in Dynamical & Control Systems*, vol. 11, 2019.
- [32] N. I. Mohammad, S. A. Ismail, M. N. Kama, O. M. Yusop, and A. Azmi, "Customer churn prediction in telecommunication industry using machine learning classifiers," in *Proceedings of the 3rd international conference on vision, image and signal processing*, 2019, pp. 1-7.
- [5] J. Kaur, V. Arora, and S. Bali, "Influence of technological advances and change in marketing strategies using analytics in retail industry," *International journal of system assurance engineering and management*, vol. 11, no. 5, pp. 953-961, 2020.
- [6] A. Dingli, V. Marmara, and N. S. Fournier, "Comparison of deep learning algorithms to predict customer churn within a local retail industry," *International journal of machine learning and computing*, vol. 7, no. 5, pp. 128-132, 2017.
- [7] A. Idris and A. Khan, "Churn prediction system for telecom using filter-wrapper and ensemble classification," *The Computer Journal*, vol. 60, no. 3, pp. 410-430, 2017.
- [8] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1-9, 2015.
- [9] "IBM Telco customer churn," <https://www.kaggle.com/datasets/blastchar/telco-customer-churn>, 2018.
- [10] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4626-4636, 2009.
- [11] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.
- [12] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.
- [13] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81-106, 1986.
- [14] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, vol. 3, no. 22, pp. 41-46.
- [15] Z. Pawlak, "Rough sets," *International journal of computer & information sciences*, vol. 11, no. 5, pp. 341-356, 1982.
- [16] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273-297, 1995.
- [17] M. H. Hassoun, *Fundamentals of artificial neural networks*. MIT press, 1995.
- [18] A. Idris, A. Khan, and Y. S. Lee, "Genetic programming and adaboosting based churn prediction for telecom," in *2012 IEEE international conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 1328-1332: IEEE.
- [19] Y. Beeharry and R. Tsokizep Fokone, "Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, p. e6627, 2022.
- [20] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a

"Telecommunication subscribers' churn prediction model using machine learning," in *Eighth international conference on digital information management (ICDIM 2013)*, 2013, pp. 131-136: IEEE.

- [45] N. V. Chawla, "Data mining for imbalanced datasets: An overview," *Data mining and knowledge discovery handbook*, pp. 875-886, 2009.
- [46] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," *arXiv preprint arXiv:1202.3725*, 2012.
- [47] H. Emami, "Stock exchange trading optimization algorithm: a human-inspired method for global optimization," *The Journal of Supercomputing*, vol. 78, no. 2, pp. 2125-2174, 2022.



بیژن مرادی مدرک کارشناسی خود در رشته برق مخابرات را از دانشگاه امام حسین و مدرک کارشناسی ارشد خود در رشته صنایع گرایش مدیریت سامانه (سیستم) و بهره‌وری از دانشگاه

علم و صنعت ایران گرفته‌است. ایشان هم‌اکنون دانشجوی دکتری رشته صنایع در دانشگاه آزاد اسلامی واحد پرند و رباط کریم هستند. از حوزه‌های موردعلاقه ایشان می‌توان به مدیریت راهبردی (استراتژیک)، هوش مصنوعی و مسائل بهینه‌سازی اشاره کرد.

نشانی رایانامه ایشان عبارت است از:

bijanmoradi53@yahoo.com



مهران خلیج مدرک دکتری خود را در رشته مدیریت حرفه‌ای کسب و کار (DBA) با گرایش مدیریت استراتژی و همچنین دکتری مهندسی صنایع با گرایش تحقیق در عملیات از دانشگاه آزاد اسلامی گرفته و از سال ۱۳۷۸ تا

کنون عضو هیأت علمی تمام وقت دانشگاه آزاد اسلامی بوده‌است. ایشان هم‌اکنون با مرتبه علمی استادیار در دانشگاه آزاد اسلامی واحد پرند و رباط کریم مشغول به تدریس است. از حوزه‌های موردعلاقه ایشان می‌توان به مدیریت خطر راهبردی (ریسک استراتژیک)، مدیریت دارایی‌ها و سرمایه‌گذاری، نظریه تصمیم‌گیری و کاربرد تئوری فازی در تصمیم‌گیری اشاره کرد.

نشانی رایانامه ایشان عبارت است از:

m Khalaj@rkiau.ac.ir

- [33] S. Agrawal, A. Das, A. Gaikwad, and S. Dhage, "Customer churn prediction modelling based on behavioural patterns analysis using deep learning," in *2018 International conference on smart computing and electronic enterprise (ICSCEE)*, 2018, pp. 1-6: IEEE.
- [34] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo, and S. Anwar, "Customer churn prediction in telecommunication industry using data certainty," *Journal of Business Research*, vol. 94, pp. 290-301, 2019.
- [35] S. Momin, T. Bohra, and P. Raut, "Prediction of customer churn using machine learning," in *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*, 2020, pp. 203-212: Springer.
- [36] S. Wael Fujo, S. Subramanian, and M. Ahmad Khder, "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," *Information Sciences Letters*, vol. 11, no. 1, p. 24, 2022.
- [37] I. V. Pustokhina, D. A. Pustokhin, P. T. Nguyen, M. Elhoseny, and K. Shankar, "Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector," *Complex & Intelligent Systems*, pp. 1-13, 2021.
- [38] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *European Journal of Operational Research*, vol. 269, no. 2, pp. 760-772, 2018.
- [39] V. Umayaparvathi and K. Iyakutti, "Automated feature selection and churn prediction using deep learning models," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 3, pp. 1846-1854, 2017.
- [40] U. Ahmed, A. Khan, S. H. Khan, A. Basit, I. U. Haq, and Y. S. Lee, "Transfer learning and meta classification based deep churn prediction system for telecom industry," *arXiv preprint arXiv:1901.06091*, 2019.
- [41] A. Idris, A. Khan, and Y. S. Lee, "Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification," *Applied intelligence*, vol. 39, no. 3, pp. 659-672, 2013.
- [42] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *European journal of operational research*, vol. 218, no. 1, pp. 211-229, 2012.
- [43] Y. Xie, X. Li, E. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445-5449, 2009.
- [44] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman,



علی تقی‌زاده هرات مدرک

دکترای تخصصی مهندسی صنایع را

با گرایش تحقیق در عملیات از

دانشگاه آزاد اسلامی واحد علوم و

تحقیقات گرفته‌است. او از سال

۱۳۷۸ تدریس خود را آغاز کرده و از

سال ۱۳۸۶ عضو هیئت علمی تمام وقت دانشگاه آزاد

اسلامی واحد پرند و رباط کریم است. ایشان هم‌اکنون با

رتبه علمی استادیار فعالیت می‌کند. حوزه‌های موردعلاقه

ایشان مدیریت کیفیت و تعالی سازمانی است.

نشانی رایانامه ایشان عبارت است از:

taghizadeh@piau.ac.ir