

An Online Group Feature Selection Algorithm using Mutual Information

Maryam Rahmaninia^{1*}, Sondos Bahadori²

1. Assistant Professor, Department of Computer Engineering, Qasreshirin Branch, Islamic Azad University, Qasreshirin, Iran. * *Corresponding Author*, ma.rahmaninia@gmail.com
2. Assistant Professor, Department of Computer Engineering, Ilam Branch, Islamic Azad University, Ilam, Iran. sondos.bahadori@iau.ac.ir

Abstract

Introduction: In the area of big data, the dimension of data in many fields are increasing dramatically. To deal with the high dimensions of training data, online feature selection algorithms are considered as very important issue in data mining. Recently, online feature selection methods have attracted a lot of attention from researchers. These algorithms deal with the process of selecting important and efficient features and removing redundant features without any pre-knowledge of the set of features. Despite all the progress in this field, there are still many challenges related to these algorithms. Among these challenges, we can mention scalability, minimum size of selected features, sufficient accuracy and execution time. On the other hand, in many real-world applications, features are entered into the dataset in groups and sequentially. Although many online feature selection algorithms have been presented so far, but none of them have been able to find trade of between these criteria.

Method: In this paper, we propose a group online feature selection method with feature stream using two new measures of redundancy and relevancy using mutual information theory. Mutual information can compute linear and non-linear dependency between the variables. With the proposed method, we try to create a better tradeoff between all the challenges.

Results: In order to show the effectiveness of the proposed online group feature selection method, a number of experiments have been conducted on six large multi-label training data sets named ALLAML, colon, SMK-CAN-187, credit-g, sonar and breast-cancer in different applications and 3 online group feature selection algorithms named FNE_OGSFS, Group-SAOLA and OGSFS which are presented recently. Also, 3 evaluation criteria including average accuracy using KNN (k - nearest neighborhood), SVM (Support Vector Machine) and NB (Naïve Bayesian) classifiers, number of selected features and executing time were used as criteria for comparing the proposed method. According to the obtained results, the proposed algorithm has obtained better results in almost of cases compared to other algorithms which it shows the efficiency of the proposed method.

Discussion: In this paper, we will show that proposed online group feature selection method will achieve better performance by considering label group dependency between the new arrival features.

Keywords: Online Group Feature selection algorithm, feature stream datasets, mutual information.



یک الگوریتم انتخاب ویژگی برخط گروهی با استفاده از نظریه اطلاعات متقابل

دوره پنجم، بهار ۱۴۰۳
شماره اول، صص: ۳۷-۴۵

تاریخ دریافت: ۱۶/۱۰/۱۴۰۲
تاریخ پذیرش: ۲۵/۱۱/۱۴۰۲

مریم رحمانی نیا^{۱*}، سندس بهادری^۲

۱. استادیار، گروه کامپیوتر، مرکز قصرشیرین، دانشگاه آزاد اسلامی، قصرشیرین، ایران. (نویسنده مسئول) ma.rahmaninia@gmail.com
۲. استادیار، گروه کامپیوتر، واحد ایلام، دانشگاه آزاد اسلامی، ایلام، ایران.
Sondos.bahadori@iau.ac.ir

چکیده

امروزه ویژگی‌ها به صورت پیوسته و یکی یکی یا گروهی به مجموعه داده‌ها اضافه می‌شوند. این امر باعث افزایش بی‌رویه ابعاد داده‌های آموزشی می‌شود. روش‌های انتخاب ویژگی برخط، با انتخاب یک زیر مجموعه مهم از میان مجموعه ویژگی‌های دیده شده تا آن لحظه اندازه داده‌های آموزشی را کاهش می‌دهند و در نتیجه عملکرد الگوریتم‌های یادگیری ماشین را بهبود می‌بخشند. اما به دلیل عدم قطعیت و ابهام در داده‌های آموزشی، روش‌های انتخاب ویژگی برخط موجود معمولاً از مشکلاتی مانند مقیاس پذیری، دقت پایین و زمان اجرای بالا رنج می‌برند. همچنین بسیاری از این روش‌ها توانایی انتخاب ویژگی‌های مهم هنگامی که ویژگی‌ها به صورت گروهی به مجموعه داده‌ها اضافه می‌شوند را ندارند. بنابراین در این مقاله، یک روش انتخاب ویژگی برخط گروهی بر اساس نظریه اطلاعات متقابل ارائه داده‌ایم. در روش پیشنهادی در ابتدا مقدار ارتباط هر یک از ویژگی‌های موجود در گروه تازه‌وارد شناسایی و سپس مقدار افزونگی ویژگی‌های انتخابی در گروه تازه‌وارد و مجموعه ویژگی‌هایی که قبلاً انتخاب شده محاسبه و ویژگی‌های اضافی حذف می‌شوند. نتایج به دست آمده توسط الگوریتم پیشنهادی در مقایسه با ۳ الگوریتم انتخاب ویژگی برخط گروهی دیگر ارزیابی شده است. طبق نتایج حاصل، الگوریتم پیشنهادی کارایی بالاتری نسبت به سایر الگوریتم‌ها به دست آورده است.

واژه‌های کلیدی: الگوریتم‌های انتخاب ویژگی گروهی برخط، داده‌های با جریان برخط، اطلاعات متقابل.

۱. مقدمه

امروزه با حجم وسیعی از اطلاعات سروکار داریم. این اطلاعات از منابع مختلفی مانند آزمایشگاه، تصاویر ماهواره‌ای، بیولوژیک و بینایی ماشین به دست آمده است [۱-۳]. از طرف دیگر، اندازه و حجم بالای این داده‌ها کارایی بسیاری از الگوریتم‌های یادگیری ماشین را تحت شعاع قرار داده و باعث شده بسیاری از این روش‌ها عملکرد خود را از دست بدهند [۴، ۵]. روش‌های انتخاب ویژگی، یک رویکرد مهم به منظور کاهش تعداد ویژگی‌ها و در نتیجه کاهش اندازه داده‌های آموزشی است [۶]. در این روش‌ها معنای اولیه ویژگی‌ها بدون تغییر مانده و فقط تعدادی از ویژگی‌های مهم انتخاب شده و ویژگی‌های اضافی و نامرتب از مجموعه ویژگی‌ها حذف می‌شود. تاکنون روش‌های انتخاب ویژگی مختلفی ارائه شده است که می‌توان آن‌ها را در سه دسته فیلتر، پروانه‌ای و جاسازی شده دسته‌بندی کرد. در روش‌های فیلتر با استفاده از معیارهای مختلفی چون بهره اطلاعاتی، نرخ فیشر [۷]، همبستگی [۸]، فاصله اقلیدسی [۹]، اطلاعات متقابل [۱۰] و غیره مقدار کارایی ویژگی‌ها اندازه‌گیری شده و بعد از مرتب‌سازی آن‌ها، k تا از بهترین ویژگی‌ها که در ابتدای صف به وجود آمده قرار گرفته‌اند انتخاب می‌شوند. در مقابل، روش‌های پوششی با استفاده از سیاست‌های مختلف زیر مجموعه‌های مختلف از ویژگی‌ها را انتخاب می‌کنند و بر اساس معیارهای مختلف آن‌ها را ارزیابی و رتبه‌دهی می‌کند [۷، ۱۱]. سپس هر کدام از زیر مجموعه‌ها که دارای مطلوبیت بالاتری بود انتخاب می‌شوند. در نهایت روش‌های انتخاب ویژگی از نوع جاسازی شده، با ادغام دو رویکرد فیلتر و پوششی به‌طور همزمان از مزایای هر دو روش استفاده کرده است [۱۲، ۱۳].

اگرچه تاکنون روش‌های انتخاب ویژگی معرفی شده در هر کدام از این سه دسته توانسته است مسائل زیادی را برطرف کند اما هنوز چالش‌های زیادی وجود دارد که مرتفع نشده است. یکی از این چالش‌ها کامل شدن تدریجی داده‌های آموزشی است. امروزه به دلیل رشد تکنولوژی داده‌های آموزشی به تدریج و با گذر زمان اضافه می‌شوند. به عنوان مثال شبکه‌های اجتماعی چون فیسبوک و توئیتر روزانه با هزاران کلمه جدید روبرو و مخزن داده‌های مربوط به این شبکه‌ها کامل تر می‌شوند [۱۴]. از طرف دیگر با اضافه شدن هر کدام از این کلمات، اضافی بودن هر یک از کلمات قبلی و یا حتی کلمه‌های جدید ممکن است رخ دهد. گاهی اوقات نیز، ممکن است کلمات جدید هیچ ربطی به هیچ یک از دامنه‌های کاربردی نداشته باشد. به همین دلیل، در اینگونه داده‌های آموزشی که ویژگی‌ها به صورت تدریجی به داده‌ها اضافه می‌شوند نیز ممکن است تعداد ویژگی‌ها آنقدر زیاد شود که باعث افزایش بی‌رویه حجم داده‌های آموزشی شود. در این مواقع به کار بردن روش‌های انتخاب ویژگی برخط می‌تواند باعث کاهش حجم داده‌ها و بهبود کیفیت الگوریتم‌های یادگیری ماشین شود [۱۵، ۱۶]. این الگوریتم‌ها از آنجاکه در محیط‌های برخط استفاده می‌شوند بهتر است در حداقل‌ترین زمان ممکن کارایی یک ویژگی را از لحاظ افزونگی و یا ارتباط با برچسب کلاس بررسی کنند. همچنین در هر لحظه زمانی تعداد ویژگی‌های انتخابی باید حداقل باشد.

از طرف دیگر چون هیچ‌گونه اطلاعاتی از داده‌هایی که در آینده قرار است به مجموعه آموزشی وارد شوند در دسترس نیست، پس بهتر است بدون هیچ دانش پیش زمینه‌ای به فرایند انتخاب ویژگی‌ها بپردازد. تاکنون الگوریتم‌های انتخاب ویژگی برخط متعددی ارائه شده است که هر کدام سعی در برطرف کردن یک یا چند چالش دارد [۲۰-۱۶]. اگر چه همه این روش‌ها تا حدود زیادی توانسته است شرایط یک الگوریتم انتخاب ویژگی برخط با ویژگی جریان را برآورده کند. اما در همه آن‌ها فرض بر این است که ویژگی‌ها به صورت منفرد و یکی یکی به مجموعه داده‌ها وارد می‌شوند. در دنیای واقعی گاهی اوقات ویژگی‌ها به صورت گروهی به مجموعه داده اضافه می‌شوند. به عنوان مثال، برای یک شخص، آیتم‌های مختلف آزمایش گروه‌های مختلفی از ویژگی‌ها مانند معاینه معمول خون، معاینه معمول ادرار و غیره را تولید می‌کند [۲۱]. اخیراً چندین روش انتخاب ویژگی برخط گروهی در جریان داده‌ها ارائه شده است [۲۲-۲۴]. اما به دلیل عدم قطعیت و وجود ابهام در داده‌های آموزشی هیچ کدام از این روش‌ها نتوانسته‌اند تاکنون حد واسط مطلوبی میان چالش‌هایی چون مقیاس پذیری، زمان اجرا، دقت و تعداد ویژگی‌های انتخاب شده، ایجاد کنند.

در این مقاله قصد داریم یک الگوریتم انتخاب ویژگی برخط در داده‌های آموزشی جریانی که در آن ویژگی‌ها به صورت گروهی و برخط به مجموعه داده‌ها اضافه می‌شوند، ارائه دهیم که به منظور ارزیابی ویژگی‌های درون یک گروه و نیز، گروه‌های مختلف از ویژگی‌ها از معیار اطلاعات متقابل استفاده شده است. اطلاعات متقابل در نظریه اعداد یک معیار کارا و مؤثر در جهت کشف وابستگی‌های خطی و غیرخطی میان متغیرها به حساب می‌آید.

ساختار کلی مقاله پس از مقدمه‌ای که گذشت، شامل پیشینه تحقیق و ذکر کارهایی است که تاکنون در این زمینه انجام شده است (بخش دوم)، بخش سوم و چهارم به ترتیب شامل روش پیشنهادی و نتایج آزمایشگاهی به همراه تجزیه و تحلیل نتایج است. بخش پنجم نیز شامل نتیجه‌گیری و کارهای آینده است.

۲. پیشینه تحقیق

۲.۱. پیشینه تحقیق

به طور کلی می‌توان براساس نوع داده‌های آموزشی، الگوریتم‌های انتخاب ویژگی مختلف را در دو دسته ایستا و برخط دسته‌بندی کرد. الگوریتم‌های انتخاب ویژگی ایستا فرض می‌کنند که در لحظه شروع الگوریتم همه داده‌ها در دسترس است [۲۵-۲۹]. در صورتی که در دنیای واقعی این گونه نیست و ویژگی‌ها به صورت پی‌درپی و تدریجی به مجموعه داده‌ها اضافه می‌شوند [۱۴، ۲۱]. در چنین داده‌هایی انتظار کامل شدن داده‌های آموزشی قبل از انجام فرایند انتخاب ویژگی، عملاً بی‌فایده است. به همین دلیل، الگوریتم‌های انتخاب ویژگی برخط مطرح شد. در این الگوریتم‌ها هر لحظه که داده جدیدی به مجموعه داده‌ها اضافه می‌شود فرایند انتخاب ویژگی انجام می‌شود. بر این اساس که ویژگی‌ها به صورت فردی

گسترش یافته روش SAOLA است. در این روش در ابتدا براساس اطلاعات متقابل در فاز درون گروهی ویژگی‌های مرتبط و افزونه در گروه تازه‌وارد شنا سایی می‌شود. سپس در فاز برون گروهی میزان افزونگی و ارتباط گروه تازه‌وارد در مقایسه با سایر گروه‌ها محاسبه و گروه‌های اضافی شناسایی و از مجموعه حذف می‌شوند. در [۲۲]، یک روش انتخاب ویژگی به نام OGSFS ارائه شده است که ویژگی‌ها را طی دو فاز درون گروهی و میان گروهی به‌طور همزمان با یکدیگر انتخاب می‌کند. در فاز درون گروهی ویژگی‌هایی از یک گروه که با هم تعامل دارند، شنا سایی شده و سپس در فاز انتخاب میان گروه، با استفاده از روش تنظیم‌سازی و انتخاب متغیرهای الاستیک نت گروه‌های تأثیرگذارتر شنا سایی می‌شوند. در [۳۵] یک روش جدید دیگر به منظور انتخاب ویژگی جریان گروهی برخط به نام FNE-OGSFS با استفاده از آنتروپی همسایگی فازی و مجموعه‌های خشن همسایگی ارائه شده است که ویژگی‌های مهم در گروه‌ها را شنا سایی و انتخاب می‌کند. سپس، تمام ویژگی‌های انتخابی توسط مدل Lasso مورد ارزیابی مجدد قرار می‌گیرند تا ویژگی‌های اضافی حذف شوند.

۲.۲ پیش زمینه

در نظریه احتمالات، اطلاعات متقابل بین دو متغیر تصادفی معیاری برای نشان دادن میزان وابستگی متقابل آن دو متغیر می‌باشد. مفهوم اطلاعات متقابل ذاتاً مرتبط با آنتروپی یک متغیر تصادفی که میزان اطلاعات موجود در یک متغیر تصادفی را نشان می‌دهد، است. آنتروپی یکی از معیارهای اندازه‌گیری میزان عدم قطعیت یک متغیر به حساب می‌آید که به‌ازای متغیر $X=[x_1, x_2, \dots, x_n]$ به‌صورت زیر تعریف می‌شود:

$$H(X) = -\sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

به‌طوری که $p(x_i)$ احتمال رخداد مقدار x_i از مجموعه مقادیر ممکن برای متغیر X را نشان می‌دهد. حال آنتروپی توأم دو متغیر تصادفی $X=[x_1, x_2, \dots, x_m]$ و $Y=[y_1, y_2, \dots, y_m]$ که لزوماً از هم مستقل نیستند، عبارت است از

$$H(X, Y) = -\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log p(x_i, y_j) \quad (2)$$

$p(x, y)$ به احتمال توأم دو مقدار x و y اشاره می‌کند.

در نظریه اطلاعات، اطلاعات متقابل بین دو متغیر به مقداری از اطلاعات که بین دو متغیر به اشتراک گذاشته شده است، اشاره می‌کند که آن را می‌توان با استفاده از رابطه زیر به‌دست آورد:

یا گروهی به مجموعه داده‌ها اضافه می‌شوند می‌توان این الگوریتم‌ها را در دو دسته فردی و گروهی برخط دسته‌بندی کرد. در الگوریتم‌های انتخاب ویژگی فردی برخط فرض بر این است که ویژگی‌ها به‌صورت یکی‌یکی به مجموعه داده‌ها اضافه می‌شوند. در ادامه چند الگوریتم انتخاب ویژگی برخط فردی را معرفی می‌کنیم. برای اولین بار در [۲۰] نویسندگان یک روش جدید به منظور حل مسأله انتخاب ویژگی برخط فردی را بر اساس رویکرد کاهش گرایان خطا به نام grafting پیشنهاد کردند. این روش با هر دو مدل پیش‌بینی‌کننده خطی و غیرخطی به‌خوبی عمل می‌کند و هم برای طبقه‌بندی و هم برای رگرسیون می‌تواند استفاده شود. در [30] یک الگوریتم انتخاب ویژگی برخط فردی دیگر به نام Alpha-investing و نسخه تکامل‌یافته آن براساس رگرسیون جریانی برای مسأله انتخاب ویژگی برخط ارائه شد. این الگوریتم‌ها، می‌توانند تعداد نامشخص و حتی نامتناهی از ویژگی‌ها را سازماندهی کند. اما از آنجاکه ویژگی‌ها را دوباره ارزیابی نمی‌کند افزونگی میان ویژگی‌ها را در نظر نمی‌گیرد. در [۱۵] یک چهارچوب انتخاب ویژگی با دو الگوریتم به نام OSFS و fast-OSFS ارائه شد که با استفاده از نظریه احتمالات شرطی به فرایند انتخاب ویژگی‌های مهم پرداخته است. این الگوریتم‌ها ویژگی‌ها را به سه دسته مرتبط قوی، مرتبط ضعیف و نامربوط با استفاده از نظریه احتمالات شرطی دسته‌بندی می‌کند. در [۳۱] دو الگوریتم انتخاب ویژگی برخط فردی به نام OFS و OFSp به ترتیب ورودی کامل و ورودی جزئی ارائه شده است. در [۳۲] یک الگوریتم انتخاب ویژگی به نام SAOLA ارائه شده است. این الگوریتم دو چالش اساسی ابعاد بالا و مقیاس‌پذیری را در الگوریتم‌های انتخاب ویژگی برخط بررسی کرده است که با استفاده از بررسی دو به دوی ویژگی‌ها بر اساس اطلاعات متقابل، به‌صورت برخط به فرایند انتخاب ویژگی می‌پردازد. در [۳۳] یک روش انتخاب ویژگی برخط دیگر نیز به نام OSFSMI برای فرایند انتخاب ویژگی‌های مهم و حذف ویژگی‌های افزونه در جریان داده‌ها با استفاده از نظریه اطلاعات متقابل شرطی ارائه شد.

در مقابل روش‌های انتخاب ویژگی گروهی برخط، به فرایند انتخاب ویژگی در حالتی که ویژگی‌ها به‌صورت گروهی به مجموعه داده‌ها اضافه می‌شود پرداخته می‌شود. در [۳۴] یک روش انتخاب ویژگی برخط گروهی به نام GFSSF ارائه شده است که با استفاده از نظریه اطلاعات متقابل و آنتروپی قادر به انتخاب ویژگی‌های مهم و کارآمد هم در سطح فردی و هم در سطح گروهی است. این الگوریتم دارای دو فاز درون گروهی و برون گروهی است و بر اساس ارتباط میان ویژگی‌ها، آن‌ها را انتخاب می‌کند. در [۲۳] یک روش انتخاب ویژگی برخط گروهی دیگر به نام OGFS ارائه شده است. این روش نیز شامل دو مرحله انتخاب درون گروهی برخط و انتخاب میان گروهی برخط است. در انتخاب برخط درون گروهی، ویژگی‌ها با استفاده از معیار تجزیه و تحلیل طیفی و در انتخاب میان گروهی برخط، از مدل رگرسیون خطی به‌منظور انتخاب یک زیرمجموعه بهینه از ویژگی‌ها استفاده شده است. نویسندگان در [۲۴] یک روش دیگر به نام Group-SAOLA ارائه داده‌اند که نسخه

و آن ویژگی‌هایی که مقدار ارتباط آن‌ها با برجسب کلاس کمتر از یک مقدار آستانه است، از مجموعه حذف می‌شود و سپس سایر ویژگی‌ها به صورت نزولی و براساس مقدار ارتباطشان در لیست S به صورت نزولی مرتب می‌شوند (خطوط ۱ تا ۶ الگوریتم پیشنهادی در شکل ۱). حال بعد از این مرحله اگر در گروه G ویژگی باقی‌ماند الگوریتم وارد فاز دوم می‌شود. در غیراین صورت الگوریتم بدون اضافه کردن هیچ ویژگی جدیدی به مجموعه ویژگی‌های نهایی خاتمه می‌یابد.

فاز دوم (برون گروهی): این فاز هنگامی اجرایی می‌شود که حداقل یک ویژگی در گروه تازه از راه رسیده G باقی‌مانده باشد. در این فاز در ابتدا ویژگی‌های انتخابی از فاز اول به صورت مرتب و نزولی بر اساس میزان ارتباطشان با برجسب کلاس به مجموعه ویژگی‌های انتخابی از مراحل قبلی الگوریتم (S') اضافه می‌شوند. سپس براساس تعریف ۲، میزان افزونگی ویژگی‌ها نسبت به یکدیگر محاسبه می‌شود (خطوط ۱۱ تا ۱۹ الگوریتم پیشنهادی در شکل ۱). به این صورت که همانند خط ۹ از ویژگی‌ای که سر لیست S' قرار دارد (مثلاً Fi) شروع می‌کنیم و تمام ویژگی‌هایی که بعد از آن در لیست قرار دارند را از لحاظ میزان افزونگی طبق تعریف ۲ ارزیابی می‌کنیم. اگر Fj بعد از آن در لیست S' قرار داشته باشد و اطلاعات متقابل آن‌ها بیشتر از مقدار اطلاعات متقابل ویژگی Fi با برجسب کلاس باشد پس ویژگی Fj از مجموعه S' حذف می‌شود. این کار تا انتهای لیست S' انجام می‌شود. بعد از این که یک بار لیست S' به ازای ویژگی Fi پیمایش شد و تعدادی از ویژگی‌ها حذف شدند، مجدد دومین ویژگی که حال بعد از Fi در سر لیست S' قرار دارد انتخاب شده و ویژگی‌هایی که طبق تعریف ۲ با آن دارای افزونگی هستند شناسایی و از لیست S' حذف می‌شود. این عملیات تا جایی ادامه می‌یابد که برای تمام ویژگی‌ها یک بار لیست S' پیمایش شده باشد. در نهایت ویژگی‌های باقی‌مانده در S' به عنوان مجموعه ویژگی‌های نهایی انتخاب می‌شوند و الگوریتم منتظر ورود ویژگی‌های دیگر می‌ماند.

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (3)$$

مقدار رابطه (۳) معمولاً مثبت است. اگر اطلاعات متقابل میان دو ویژگی زیاد باشد به این معنی است که وابستگی دو متغیر نیز زیاد است.

۳. راهکار پیشنهادی

در این بخش چهار چوب اصلی روش پیشنهادی با استفاده از معیار اطلاعات متقابل که در فصل قبلی معرفی شده ارائه می‌شود. بطور کلی می‌توان گفت الگوریتم پیشنهادی دارای دو فاز اصلی است. فاز اول به ازای ورود هر گروه G از ویژگی‌های تازه وارد اجرایی می‌شود. در این فاز ابتدا ویژگی‌های مهم و مرتبط با برجسب کلاس شناسایی و سپس ویژگی‌هایی که با برجسب کلاس مرتبط نیستند از گروه حذف می‌شوند. در صورتی که بعد از فاز اول، همچنان ویژگی در گروه G باقی‌مانده باشد، الگوریتم وارد فاز دوم می‌شود. در فاز دوم ویژگی‌های انتخابی از گروه G به مجموعه ویژگی‌های انتخابی از مراحل قبلی الگوریتم (مجموعه S') اضافه و سپس میزان افزونگی آن‌ها نسبت به یکدیگر و برجسب کلاس محاسبه و ویژگی‌های اضافی از مجموعه S' حذف می‌شوند. بعد از این فاز الگوریتم همچنان منتظر ورود گروه‌های جدید از ویژگی‌ها می‌ماند. ویژگی‌هایی که از فاز اول و فاز دوم حذف می‌شوند، در مراحل بعدی اجرای الگوریتم حساب نمی‌شوند. در ادامه دو معیار برای محاسبه میزان ارتباط و افزونگی ویژگی‌ها معرفی شده و سپس با استفاده از این معیارها مراحل فازهای الگوریتم پیشنهادی را با جزئیات بیشتر معرفی می‌کنیم. علاوه بر این شبه کد الگوریتم پیشنهادی در شکل ۱ آمده است.

تعریف ۱ (ارتباط ویژگی): میزان ارتباط و تأثیر یک ویژگی $F = [f_1, f_2, \dots, f_n]$ با برجسب کلاس C با استفاده از رابطه زیر به دست می‌آید:

$$I(F; C) = \sum_{f_i \in F} \sum_{c \in C} p(f_i, c) \log \frac{p(f_i, c)}{p(f_i)p(c)} \quad (4)$$

به طوری که $I(X, C)$ به اطلاعات متقابل میان ویژگی X و برجسب C که از رابطه ۲ می‌توان آن را محاسبه کرد، اشاره می‌کند.

تعریف ۲ (افزونگی ویژگی): ویژگی F2 نسبت به ویژگی F1 اضافی در نظر گرفته می‌شود اگر و تنها اگر

$$I(F_1; F_2) > I(F_1; C), \quad I(F_2; C) < I(F_1; C) \quad (5)$$

فاز اول (درون گروهی): این فاز به ازای ورود یک گروه تازه وارد از ویژگی‌ها به نام G اجرایی می‌شود. در قدم اول در این فاز، مقدار ارتباط هر کدام از ویژگی‌ها با برجسب کلاس با استفاده از تعریف ۱ محاسبه شده

۱.۴. ارزیابی روش پیشنهادی از لحاظ دقت دسته‌بندی

در این بخش به ارزیابی روش پیشنهادی در مقایسه با سایر روش‌های انتخاب ویژگی [۳۵] FNE_OGSFS، Group-SAOLA [۲۴] و OGSFS [۲۲] از لحاظ دقت دسته‌بندی می‌پردازیم. در ادامه نتایج به‌دست‌آمده برای هر کدام از این دسته‌بندی‌ها و داده‌های آموزشی به همراه الگوریتم‌های انتخاب ویژگی برخط گروهی در جدول‌های ۲ تا ۴ نشان داده شده‌است. در هر جدول سطرها نشان‌دهنده یک داده آموزشی و ستون‌ها مربوط به یک الگوریتم انتخاب ویژگی برخط گروهی است. همچنین بیشترین دقت به‌دست‌آمده به‌ازای هر داده آموزشی و هر الگوریتم انتخاب ویژگی به‌صورت ضخیم و زیر خط‌دار نشان داده شده‌است. نتایج به‌دست‌آمده به‌ازای دسته‌بند SVM در جدول شماره ۲ نشان داده شده‌است. همان‌طور که نتایج نشان می‌دهد ویژگی‌های انتخاب شده به‌ازای الگوریتم پیشنهادی و همه داده‌های آموزشی همواره به‌جز یک مورد بیشترین دقت را به‌دست آورده‌است. به‌عنوان مثال، به‌ازای داده آموزشی ALLAML، دقت به‌دست‌آمده به‌ازای الگوریتم پیشنهادی، FNE_OGSFS، Group-SAOLA و OGSFS به‌ترتیب ۹۸،۹۸، ۲۱،۷۱، ۶۲،۵۸ و ۷۰،۷۳ است. که نشان می‌دهد الگوریتم پیشنهادی تقریباً ۷۸٪، ۳۶٪ و ۲۸٪ نسبت به سایر الگوریتم‌ها بهتر عمل کرده‌است. در سطر آخر جدول ۲، مقادیر میانگین به‌ازای همه داده‌های آموزشی آمده‌است. طبق نتایج الگوریتم پیشنهادی، FNE_OGSFS، Group-SAOLA و OGSFS به‌ترتیب به‌طور متوسط روی همه داده‌های آموزشی دارای دقت ۸۰،۱۶٪، ۴۷،۶٪، ۶۲،۶٪ و ۷۱،۳٪ می‌باشد. این نتایج نشان‌دهنده برتری روش پیشنهادی نسبت به سایر روش‌ها است.

جدول ۲: دقت دسته‌بند SVM به‌ازای الگوریتم‌های انتخاب ویژگی برخط گروهی مختلف

Proposed_method	FNE_OGSFS	Group-SAOLA	OGSFS
ALLAML	98.98	21.72	62.58
Colon	87.98	63.86	62.51
SMK-CAN-1	79.02	34.31	62.51
credit-g	72.25	46.51	69.85
Sonar	75.25	52.93	52.39
breast.canc	70.11	70.5	69.72
Avg	80.16	47.6	62.6

دقت ویژگی‌های انتخاب شده به‌ازای دسته‌بند NB در جدول ۳ نشان آمده‌است. طبق نتایج، الگوریتم پیشنهادی در همه موارد دقت بالاتری نسبت به سایر الگوریتم‌های انتخاب ویژگی برخط گروهی به‌دست آورده‌است. به‌عنوان مثال، دقت الگوریتم پیشنهادی به‌ازای داده آموزشی sonar عبارت است از ۷۰،۷۳٪ که نشان می‌دهد، الگوریتم پیشنهادی تقریباً ۲۴٪، ۲۵٪ و ۸٪ از سایر الگوریتم‌های انتخاب ویژگی گروهی بهتر عمل کرده‌است.

Input: new arrival group of features G
Output: the final selected feature set S

```

# Phase 1:
1. For each feature F in new arrival group of features G
2.   Calculate MI=I(F, C)
3.   If ( MI > threshold)
4.     Insert F into selected feature set S in descending order
5.   End if
6. End For

#Phase 2:
7. If (S is not empty)
8.   insert the features in S into S' in descending order
9.   Fj= First _ Element (S) // select the first feature in sorted list S
10.  Do
11.    Fi= Next _ Element (S, Fj) //find feature fi in S which is next to Fj
12.    Do
13.      If I( Fi, Fj)>I( Fi, C)
14.        Remove fi from S
15.        Fi= Next _ Element ( S, Fj)
16.      End If
17.    End Until (Fi==Null);
18.    Fj= Next _ Element (S, Fj)
19.  End Until (Fj==Null)
20. End
21. End
22. Return S
    
```

شکل ۱: شبه کد الگوریتم پیشنهادی

۴. نتایج آزمایشگاهی

در این بخش به بررسی کارایی روش پیشنهادی در مقایسه با سه الگوریتم انتخاب ویژگی برخط FNE_OGSFS [۳۵]، Group-SAOLA [۲۴] و OGSFS [۲۲] که در حالت گروهی به فرایند انتخاب ویژگی‌ها می‌پردازند و ۶ داده آموزشی معروف و پرکاربرد که مشخصات آن‌ها در جدول ۱ آمده، می‌پردازیم. برای این کار از سه معیار ارزیابی تعداد ویژگی‌های انتخابی هر الگوریتم، دقت سه دسته‌بندها [36] KNN، SVM [۳۷] و NB [۳۸] و زمان اجرا برای ارزیابی الگوریتم‌ها استفاده شده‌است. هرچقدر تعداد ویژگی‌های انتخاب شده و زمان اجرا کمتر باشد الگوریتم مطلوب‌تر است. از طرف دیگر هرچقدر دقت به‌دست‌آمده به‌ازای انواع دسته‌بندی‌ها برای یک الگوریتم بیشتر باشد پس مطلوبیت الگوریتم بیشتر است.

جدول ۱: داده‌های آموزشی

کلیود	تعداد کلاس‌ها	تعداد نمونه‌های آموزشی	تعداد ویژگی‌ها
Bio	2	72	7129
Bio	2	62	2000
Microarray, Bio	2	187	19993
credit risk	2	20	1000
signal	2	60	208
Bio	2	9	286

برای راحتی کار، اندازه گروه‌ها در همه الگوریتم‌ها ۵ در نظر گرفته شده‌است. به این معنی که هر بار یک گروه ۵ تایی از ویژگی‌ها به مجموعه اضافه می‌شود. همچنین از آنجا که ویژگی‌ها با ترتیب‌های مختلفی به مجموعه اضافه می‌شوند پس، نتایج حاصل روی ۵ بار اجرای هر کدام از الگوریتم‌ها و به‌ازای ترتیب‌های مختلف اجرا شده‌است. علاوه بر این تمام آزمایش‌ها روی یک سیستم با سیستم عامل ویندوز ۱۱، 8 RAM و گرافیک RTX 3050 انجام شده‌است.

جدول ۳: دقت دسته‌بند NB به‌ازای الگوریتم‌های انتخاب ویژگی برخط گروهی مختلف

	Proposed_method	FNE_OGSFS	Group-SAOLA	OGSFS
ALLAML	96.96	21.71	53.03	63.65
Colon	87.94	56.65	54.6	76.6
SMK-CAN-1	77.43	38.37	59.95	59.31
credit-g	73.55	46.33	59.18	66.96
Sonar	70.73	53.54	52.46	64.59
breast.canc	71.02	70.89	70.5	70.63
Avg	79.7	47.33	67.8	66.3

در ادامه در جدول ۴ نتایج به‌دست‌آمده به‌ازای دسته‌بند KNN آورده شده‌است. نتایج به‌دست‌آمده به‌ازای این دسته‌بند نیز همانند نتایج به‌دست‌آمده به‌ازای دو دسته‌بند قبلی است. همان‌طور که در جدول ۴ آمده‌است الگوریتم پیشنهادی به‌ازای داده‌های آموزشی ALLAML، colon، SMK-CAN-187، credit-g، sonar و breast.cancer بیشترین دقت را به‌دست آورده‌است.

جدول ۴: دقت دسته‌بند knn به‌ازای الگوریتم‌های انتخاب ویژگی برخط گروهی مختلف

	Proposed_method	FNE_OGSFS	Group-SAOLA	OGSFS
ALLAML	98.45	20.51	64.15	69.68
Colon	88.01	57.46	63.47	78.46
SMK-CAN-1	71.43	50.56	60.96	66.13
credit-g	70.03	20.66	63.14	68.4
Sonar	76.75	58.95	51.48	74.66
breast.canc	66.44	69.55	67.35	70.59
Avg	78.16	45.66	61.3	70.8

میانگین نتایج حاصل به‌ازای همه داده‌های آموزشی و هر الگوریتم انتخاب ویژگی برخط گروهی در سطر آخر آمده‌است. طبق نتایج، به‌ازای هر سه دسته‌بند دقت به‌دست‌آمده توسط الگوریتم پیشنهادی از سایر الگوریتم‌ها بهتر بوده‌است. همچنین می‌توان گفت، OGSFS در رتبه دوم، الگوریتم Group-SAOLA در رتبه سوم و در نهایت الگوریتم FNE-OGSFS در رتبه سوم قرار دارد.

۲.۴. ارزیابی روش پیشنهادی از لحاظ تعداد ویژگی‌های انتخابی
در این بخش به مقایسه الگوریتم پیشنهادی و سایر الگوریتم‌های انتخاب ویژگی گروهی برخط بر اساس تعداد ویژگی‌های انتخاب شده، می‌پردازیم. نتایج حاصل بعد از ۵ بار اجرای الگوریتم‌های انتخاب ویژگی مختلف در جدول ۵ آمده‌است. همان‌طور که نمودارها نشان می‌دهد به‌ازای داده آموزشی ALLAML، الگوریتم پیشنهادی، FNE-OGSFS، Group-SAOLA و OGSFS به ترتیب ۲۲، ۲۵، ۶۶، ۱۱ و ۳۳، ۲۴ ویژگی را انتخاب کرده‌اند. پس الگوریتم FNE-OGSFS نسبت به سایر الگوریتم‌ها تعداد کمتر و OGSFS بیشترین ویژگی را انتخاب کرده‌است. نتایج به‌ازای سایر داده‌های آموزشی نیز تقریباً به همین صورت است. به‌طور کلی می‌توان گفت الگوریتم FNE-OGSFS معمولاً تعداد کمتر، الگوریتم group-SAOLA دومین کمترین و الگوریتم پیشنهادی سومین کمترین

تعداد ویژگی‌ها را انتخاب کرده‌است. از این میان، الگوریتم OGSFS معمولاً تعداد بیشتری ویژگی انتخاب کرده‌است.

جدول ۵: تعداد ویژگی‌های انتخابی توسط الگوریتم‌های انتخاب ویژگی مختلف

	Proposed_method	FNE_OGSFS	Group-SAOLA	OGSFS
ALLAML	22	2.5	11.66	24.33
Colon	5	1	4.6	36
SMK-CAN-187	12.66	1.33	47.33	59
credit-g	4	2.33	3.33	5
Sonar	3.66	2	2.66	18.66
breast.cancer	2.33	1	2	3.33
Avg	8.27	1.69	11.95	24.38

۳.۴. ارزیابی روش پیشنهادی از لحاظ زمان اجرا

متوسط زمان اجرای الگوریتم‌های انتخاب ویژگی برخط گروهی به‌ازای ۵ بار اجرای مستقل در جدول ۶ آمده‌است. طبق نتایج حاصل، متوسط زمان اجرای الگوریتم پیشنهادی معمولاً از سایر الگوریتم‌ها کمتر است. بعد از الگوریتم پیشنهادی، الگوریتم Group-SAOLA زمان اجرای کمتری داشته‌است. در نهایت الگوریتم‌های OGSFS و FNE-OGSFS در رتبه سوم و چهارم قرار دارند.

جدول ۶: زمان اجرای الگوریتم‌های انتخاب ویژگی برخط گروهی مختلف

	Proposed_method	FNE_OGSFS	Group-SAOLA	OGSFS
ALLAML	7.0392	12.8143	0.2743	1.136
Colon	0.1866	2.3559	0.2329	0.4299
SMK-CAN-1	13.9119	242.0667	40.3233	124.1812
credit-g	0.0598	10.157	0.0668	0.0827
Sonar	0.0625	2.1401	0.1231	0.1274
breast.cancer	0.0406	0.5778	0.0376	0.089
Avg	3.3	44.8	6.76	20.5

۴.۴. پیچیدگی محاسباتی

همان‌طور که گفته شد، الگوریتم پیشنهادی با ورود هر گروه از ویژگی‌ها اجرایی می‌شود. در ابتدا در فاز اول، میزان ارتباط هر کدام از ویژگی‌های موجود در گروه طبق معادله ۴ محاسبه می‌شود. از آنجاکه این مقدار یک مقدار ثابت است پس می‌توان با m به آن اشاره کرد. حال اگر تعداد ویژگی‌های موجود در گروه تازه‌وارد $|G|$ باشد پس می‌توان گفت این فاز از مرتبه زمانی m/G^2 است. حال در فاز دوم، میزان افزونگی هر ویژگی با سایر ویژگی‌های دیگر در مجموعه S بر اساس رابطه (۵) محاسبه می‌شود. اگر پیچیدگی محاسباتی رابطه ۲ را نیز n در نظر بگیریم می‌توان گفت این فاز نیز از مرتبه زمان $n|S|^2$ است. پس به‌طور کلی مرتبه اجرایی الگوریتم پیشنهادی عبارت است از

$$O(m|G|^2 + n|S|^2) \square O(\max(|G|^2, |S|^2))$$

- [۳] M. Hariri and H. Najafy, "Improve the Quality of Mammogram Images by Image Processing Techniques," *Intelligent Multimedia Processing and Communication Systems (IMPCS)*, vol. 3, no. 1, pp. 57-69, 2022. [Online]. Available: https://impcs.zanjan.iau.ir/article_696998_4bc75e4c5155a48f9fb7b94ac977a892.pdf.
- [۴] M. Köppen, "The curse of dimensionality," in *5th online world conference on soft computing in industrial applications (WSC5)*, 2000, vol. 1, pp. 4-8.
- [۵] M. Najafi, M. Afzali, and M. Moradi, "Use data mining to identify factors affecting students' academic failure," *Intelligent Multimedia Processing and Communication Systems (IMPCS)*, vol. 2, no. 1, pp. 23-33, 2021. [Online]. Available: https://impcs.zanjan.iau.ir/article_682093_7de40314874a6374a5cdc7b20b69a455.pdf.
- [۶] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, 2015: Ieee, pp. 1200-1205.
- [۷] L. Sun, S. Fu, and F. Wang, "Decision tree SVM model with Fisher feature selection for speech emotion recognition," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 1, pp. 1-14, 2019.
- [۸] Y. Liu, F. Tang, and Z. Zeng, "Feature selection based on dependency margin," *IEEE Transactions on Cybernetics*, vol. 45, no. 6, pp. ۱۲۲۱-۱۲۰۹, ۲۰۱۴.
- [۹] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Covert cyber assault detection in smart grid networks utilizing feature selection and euclidean distance-based machine learning," *Applied Sciences*, vol. 8, no. 5, p. 772, 2018.
- [۱۰] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural computing and applications*, vol. 24, pp. 175-186, 2014.
- [۱۱] S. Maldonado and R. Weber, "A wrapper method for feature selection using support vector machines," *Information Sciences*, vol. 179, no. 13, pp. 2208-2217, 2009.
- [۱۲] H. Liu, M. Zhou, and Q. Liu, "An embedded feature selection method for imbalanced data classification," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 703-715, 2019.
- [۱۳] T. N. Lal, O. Chapelle, J. Weston, and A. Elisseeff, "Embedded methods," *Feature Extraction: Foundations and Applications*, pp. 137-165, 2006.
- [۱۴] W. Ding et al., "Subkilometer crater discovery with boosting and transfer learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 4, pp. 1-22, 2011.
- [۱۵] X. Wu, K. Yu, W. Ding, H. Wang, and X. Zhu, "Online feature selection with streaming features," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 5, pp. ۱۱۶۲-۱۱۷۸, ۲۰۱۳.
- [۱۶] D. You et al., "Online feature selection for multi-source streaming features," *Information Sciences*, vol. 590, pp. 267-295, 2022/04/01/ 2022, doi: <https://doi.org/10.1016/j.ins.2022.01.008>.
- [۱۷] D. Paul, A. Jain, S. Saha, and J. Mathew, "Multi-objective PSO based online feature selection for multi-label classification," *Knowledge-Based Systems*, vol. 222, p. 106966, 2021.
- [۱۸] S. Eskandari and M. M. Javidi, "Online streaming feature selection using rough sets," *International*

می توان گفت، الگوریتم پیشنهادی در یک زمان خطی که وابسته به تعداد ویژگی های انتخاب شده تا آن لحظه یا اندازه گروه است به فرایند انتخاب ویژگی می پردازد. بنابراین از آنجا که مرتبه زمانی انتخاب ویژگی ها در الگوریتم پیشنهادی به صورت خطی است می توان گفت الگوریتم پیشنهادی مقیاس پذیر است.

۵. نتیجه گیری و کارهای آینده

یکی از چالش های اساسی که این روزها با آن سروکار داریم جریان بودن داده های آموزشی است که باعث بروز مشکلات زیادی شده است. به این معنی که داده های آموزشی به صورت تدریجی و با مرور زمان به مجموعه اضافه شده و همین امر باعث افزایش بی رویه داده های آموزشی شده است. از طرف دیگر در این داده ها معمولاً مسأله ازدحام ابعاد وجود دارد به این معنی که که داده های آموزشی تحت تأثیر تعداد بالای ویژگی ها قرار گرفته و باعث می شوند الگوریتم های یادگیری ماشین با مشکل روبرو شوند. این امر به این دلیل است که تعداد زیادی از این ویژگی ها نامربوط و یا اضافی هستند. تاکنون الگوریتم های انتخاب ویژگی زیادی در جریان داده های گروهی ویژگی ها ارائه شده است. اما هیچ کدام از آن ها نتوانسته است همه چالش های مربوط به این دسته از داده ها را برطرف کند. به همین منظور، در این مقاله یک راهکار انتخاب ویژگی در جریان داده های گروهی ویژگی ارائه شد. راهکار پیشنهادی با استفاده از اطلاعات متقابل طی دو فاز، ویژگی های غالب را شناسایی و ویژگی های اضافی را حذف می کند. در فاز اول ویژگی های مرتبط بر اساس اطلاعات متقابل انتخاب و در فاز دوم بر اساس اطلاعات متقابل ویژگی های اضافی حذف می شوند. سپس روش پیشنهادی در مقایسه با سه الگوریتم انتخاب ویژگی برخط گروهی و ۶ داده آموزشی معروف و پر کاربرد با استفاده از چندین معیار مختلف مانند تعداد ویژگی های انتخاب شده، زمان اجرا و دقت انواع دسته بندیها مورد ارزیابی قرار گرفت. نتایج نشان داد الگوریتم پیشنهادی معمولاً تعداد کمتری ویژگی و در زمان کمتری با دقت دسته بندی بالا انتخاب می کند. همچنین الگوریتم پیشنهادی نسبت به سایر الگوریتم ها مقیاس پذیرتر است. در آینده قصد داریم، الگوریتم پیشنهادی را به گونه ای تغییر دهیم که در داده های آموزشی چندبرجسی نیز به خوبی بتواند کارایی خود را حفظ کند.

References

- [۱] Z. Fang, J.-N. Hwang, X. Huo, H.-J. Lee, and J. Denzler, "Emergent Techniques and Applications for Big Visual Data," *International Journal of Digital Multimedia Broadcasting*, vol. 2017, p. 6468502, 2017/10/31 2017, doi: 10.1155/2017/6468502.
- [۲] T. Zhang and B. Yang, "Big data dimension reduction using PCA," in *2016 IEEE international conference on smart cloud (SmartCloud)*, 2016: IEEE, pp. 152-157.

- 2014 *IEEE International Conference on Data Mining*, 2014: IEEE, pp. 660-669.
- [۳۳] M. Rahmaninia and P. Moradi, "OSFSMI: Online stream feature selection method based on mutual information," *Applied Soft Computing*, vol. 68, pp. 733-746, 2018/07/01/ 2018, doi: <https://doi.org/10.1016/j.asoc.2018.07.034>.
- [۳۴] H. Li, X. Wu, Z. Li, and W. Ding, "Group feature selection with streaming features," in *2013 IEEE 13th International Conference on Data Mining*, 2013: IEEE, pp. 1109-1114.
- [۳۵] J. Xu, Y. Sun, K. Qu, X. Meng, and Q. Hou, "Online group streaming feature selection using entropy-based uncertainty measures for fuzzy neighborhood rough sets," *Complex & Intelligent Systems*, vol. 8, no. 6, pp. 5309-5328, 2022.
- [۳۶] E. Fix and J. L. Hodges, "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties," *International Statistical Review / Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238-247, 1989, doi: 10.2307/1403797.
- [۳۷] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998/06/01 1998, doi: 10.1023/A:1009715923555.
- [۳۸] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2۰۰۹
- Journal of Approximate Reasoning*, vol. 69, pp. 35-57, 2016.
- [۱۹] J. Zhou, D. Foster, R. Stine, and L. Ungar, "Streaming feature selection using alpha-investing," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 2005, pp. 384-393.
- [۲۰] S. Perkins, K. Lacker, and J. Theiler, "Grafting: Fast, incremental feature selection by gradient descent in function space," *The Journal of Machine Learning Research*, vol. 3, pp. 1333-1356, 2003.
- [۲۱] M. Wang, H. Li, D. Tao, K. Lu, and X. Wu, "Multimodal Graph-Based Reranking for Web Image Search," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4649-4661, 2012, doi: 10.1109/TIP.2012.2207397.
- [۲۲] P. Zhou, N. Wang, and S. Zhao, "Online group streaming feature selection considering feature interaction," *Knowledge-Based Systems*, vol. 226, p. 107157, 2021/08/17/ 2021, doi: <https://doi.org/10.1016/j.knsys.2021.107157>.
- [۲۳] J. Wang *et al.*, "Online feature selection with group structure analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3029-3041, 2015.
- [۲۴] K. Yu, X. Wu, W. Ding, and J. Pei, "Scalable and accurate online feature selection for big data," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 2, pp. 1-3.۲۰۱۶, ۹
- [۲۵] T. Dokeroglu, A. Deniz, and H. E. Kiziloz, "A comprehensive survey on recent metaheuristics for feature selection," *Neurocomputing*, vol. 494, pp. 269-296, 2022/07/14/ 2022, doi: <https://doi.org/10.1016/j.neucom.2022.04.083>.
- [۲۶] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1060-1073, 2022/04/01/ 2022, doi: <https://doi.org/10.1016/j.jksuci.2019.06.012>.
- [۲۷] M. Vahmiyan, M. Kheirabadi, and E. Akbari, "Feature selection methods in microarray gene expression data: a systematic mapping study," *Neural Comput. Appl.*, vol. 34, no. 22, pp. 19675-19702, 2022, doi: 10.1007/s00521-022-07661-z.
- [۲۸] M. H. Nadimi-Shahraki, H. Zamani, and S. Mirjalili, "Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study," *Computers in Biology and Medicine*, vol. 148, p. 105858, 2022/09/01/ 2022, doi: <https://doi.org/10.1016/j.compbiomed.2022.105858>.
- [۲۹] M. Zivkovic, C. Stoean, A. Chhabra, N. Budimirovic, A. Petrovic, and N. Bacanin, "Novel Improved Salp Swarm Algorithm: An Application for Feature Selection," *Sensors*, vol. 22, no. 5, p. 1711, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/5/1711>
- [۳۰] M. M. Javidi and S. Eskandari, "Streamwise feature selection: a rough set method," *International Journal of Machine Learning and Cybernetics*, vol. 9, pp. 667-676, 2018.
- [۳۱] J. Wang, P. Zhao, S. C. Hoi, and R. Jin, "Online feature selection and its applications," *IEEE Transactions on knowledge and data engineering*, vol. 26, no. 3, pp. 698-710, 2013.
- [۳۲] K. Yu, X. Wu, W. Ding, and J. Pei, "Towards scalable and accurate online feature selection for big data," in