

## استفاده از خوارزمیک انتخاب ویژگی به منظور پیش‌بینی مقدار تبخیر و تعرق با کمترین داده

مرضیه مکرم<sup>1</sup>، عبدالرسول زارعی<sup>2\*</sup>

1. عضو هیأت علمی، دانشکده کشاورزی و منابع طبیعی داراب، دانشگاه شیراز، داراب، ایران.
  2. عضو هیأت علمی، گروه مهندسی منابع طبیعی (مرتع و آبخیزداری)، دانشکده کشاورزی، دانشگاه فسا، فسا، ایران.
- \* نویسنده مسئول: [Ar\\_Zareiee@Fasau.ac.ir](mailto:Ar_Zareiee@Fasau.ac.ir)

تاریخ دریافت: 1395/08/14 تاریخ پذیرش: 1396/06/29

### چکیده

در پژوهش حاضر نسبت به انتخاب مهم‌ترین متغیرهای مؤثر در پیش‌بینی تبخیر و تعرق با استفاده از خوارزمیک (الگوریتم) انتخاب ویژگی در شهرستان داراب واقع در جنوب غربی استان فارس اقدام شد. در این راستا ابتدا با استفاده از روش پنمن مونتیت فائو مقدار تبخیر و تعرق محاسبه، سپس به کمک روش انتخاب ویژگی از بین متغیرهای مؤثر در روش پنمن مونتیت فائو تأثیرگذارترین متغیرها در پیش‌بینی تبخیر و تعرق تعیین شدند. برای پیش‌بینی، از روش‌های Greedy-Stepwise, Best-First و Ranker به منظور انتخاب مؤثرترین متغیرها از بین 120 داده استفاده شد. در راستای بررسی خطای هر یک از این روش‌ها و انتخاب بهترین روش، از روش‌های طبقه بندی Naive Bayes, J48 و LMT استفاده شد. داده‌های مورد بررسی در این بررسی، برای محاسبه و پیش‌بینی تبخیر و تعرق عبارت از کمینه، میانگین و بیشینه دما، سرعت باد، کمینه ساعات آفتابی، بیشینه ساعات آفتابی، ارتفاع، تابش خورشیدی و رطوبت نسبی بودند. نتایج پژوهش حاضر نشان داد که روش Ranker با مدل Relief-F Attribute-Eval در هر سه روش طبقه‌بندی، دارای کمینه خطاست. این موضوع مبین این نکته است که بیشینه مقدار ساعات آفتابی، میانگین و بیشینه دما نسبت به سایر متغیرها در پیش‌بینی مقدار تبخیر و تعرق مؤثرترند. بنابراین با استفاده از خوارزمیک Feature selection می‌توان در مناطق دارای کمبود داده با استفاده از متغیرهای کمتر اقدام به پیش‌بینی تبخیر و تعرق نمود و در زمان و هزینه صرفه‌جویی کرد.

**واژگان کلیدی:** متغیرهای اقلیمی؛ تابع ارزیابی؛ تابع تولیدکننده؛ پنمن مونتیت فائو؛ تابع تعیین اعتبار

### n مقدمه

در تبخیر و تعرق دخالت دارند برآورد دقیق آن، کاری بسیار دشوار است. تبخیر و تعرق در چرخه هیدرولوژی نقش بسیار مهمی دارد. مقدار تبخیر و تعرق پتانسیل تا حد زیادی مبین مقدار رطوبت از دست رفته یک حوضه آبخیز است. در واقع پس از بارندگی و جذب رطوبت توسط گیاه از خاک، گیاه به‌منظور حفظ تعادل دما در ساختار

در بحث محاسبه نیاز آبی گیاهان، برآورد دقیق تبخیر و تعرق از پایه‌ای‌ترین رکن‌های طراحی و محاسبه مهندسی به‌شمار می‌رود (21). از این رو روش‌های محاسباتی تبخیر و تعرق گیاهان، جزو اولویت‌های تحقیقاتی کشورهای جهان است. چون عوامل بسیار زیادی

در مطالعه‌های دیگر، مقدار تأثیرگذاری متغیرهای اقلیمی مختلف در تغییر مقدار تبخیر و تعرق پتانسیل با بهره‌گیری از رگرسیون لجستیک و با استفاده از داده‌های چهار ایستگاه رشت، شهرکرد، شیراز و زابل در اقلیم‌های مختلف طی دوره آماری (1967-2014) ارزیابی شد. نتایج این بررسی نشان داد که در مناطق فراخشک و خشک (به ترتیب ایستگاه‌های زابل و شیراز) متوسط سرعت باد تأثیرگذارترین متغیر و متوسط ساعات آفتابی کم‌اهمیت‌ترین شاخص تأثیرگذار بر مقدار تبخیر و تعرق پتانسیل‌اند. در مناطق نیمه‌خشک و مرطوب (به ترتیب شهرکرد و رشت) متوسط بیشینه دمای ماهانه کم‌اهمیت‌ترین شاخص است. در این مناطق متوسط سرعت باد (شهرکرد) و متوسط ساعات آفتابی (رشت) مهم‌ترین متغیرهای تأثیرگذار بر مقدار تبخیر و تعرق بودند (23).

استفاده از روش پنمن به منظور تعیین مقدار تبخیر و تعرق گیاه زعفران در استان اصفهان، نشان داد که در ایستگاه‌های غرب و جنوب غرب استان به دلیل تأثیر دما تبخیر و تعرق در مقایسه با مناطق شرقی کمتر است (14). نتایج پژوهش در بررسی مقدار تبخیر و تعرق جنوب استان فارس در شهرستان داراب با استفاده از شبکه‌های عصبی خودسازمان‌ده یا Self Organizing Map نیز، نشان داد که با توجه به مقدار تبخیر و تعرق، منطقه مورد مطالعه در سه کلاس تبخیر و تعرق قرار می‌گیرد (12). همچنین پژوهش‌های متعددی به مطالعه روش انتخاب ویژگی در بررسی میزان تبخیر و تعرق پرداخته‌اند (9، 11، 18 و 20).

با توجه به اینکه محاسبه و پیش‌بینی تبخیر و تعرق نیاز به متغیرهای متعددی دارد و از طرفی اندازه‌گیری همه این متغیرها نیاز به صرف وقت و هزینه بالا دارد، استفاده از روشی ساده‌تر که از دقت کافی نیز برخوردار باشد، به منظور انتخاب مؤثرترین متغیرها برای پیش‌بینی تبخیر و تعرق مهم و کاربردی به نظر می‌رسد. هدف از پژوهش حاضر استفاده از خوارزمیک Feature Selection و انتخاب بهترین مدل به منظور انتخاب مؤثرترین و مهم‌ترین متغیرها برای پیش‌بینی تبخیر و تعرق است.

خودمقداری از رطوبت را از طریق برگ‌ها طی عمل تبخیر و تعرق به اتمسفر برمی‌گرداند (16 و 19). این چرخه و بازگشت رطوبت به اتمسفر در سالم‌ماندن گیاه به‌ویژه در مناطق خشک و نیمه‌خشک و همچنین مقدار عملکرد و تولید محصولات کشاورزی مهم و مؤثر است.

با توجه به اهمیت این متغیر و لزوم اندازه‌گیری مستقیم و سریع آن روش‌های مختلفی جهت اندازه‌گیری و طبقه‌بندی آن ارائه شده است (5، 14 و 15). نتایج بهره‌گیری از داده‌های جوی همچون ساعات آفتابی بر پایه مدل ساده‌ای برای تخمین و طبقه‌بندی مقدار تبخیر و تعرق پتانسیل در سطح افقی توسط پژوهشی پیشنهاد شده است (4). برخی محققان، مؤلفه‌های نور خورشید را به‌طور مجزا برای افزایش دقت مدل خود در تخمین و طبقه‌بندی تبخیر و تعرق پتانسیل مد نظر قرار دادند (6 و 12) و در مدل ساده شده دیگری، مقادیر تابش مستقیم و مؤثر روزانه را با استفاده از نسبت ساعات آفتابی به ساعات روشنایی و زاویه سمت‌الراس خورشید محاسبه نمودند (13). همچنین پژوهشی نشان داده‌است که به منظور تخمین  $ET_0$  در اقلیم‌های خشک و نیمه‌خشک می‌توان از برخی مدل‌های تابش استفاده کرد (8). شرقی و همکاران (2010) از روش پنمن مونتیت فائو به منظور بررسی تبخیر و تعرق گیاه مرجع از 29 ایستگاه کلیماتولوژی و سینوپتیک در استان یزد استفاده کردند. سپس با برقراری ارتباط بین متغیر تبخیر و تعرق با ارتفاع و همچنین روش عکس مجذور فاصله نقشه‌های مقادیر تبخیر و تعرق گیاه مرجع را در سطح استان تهیه نمودند. نتایج نشان داد که مناطق مرکزی، جنوبی و برخی نواحی غربی استان از شدت تبخیر و تعرق بیشتری نسبت به مناطق شمالی و شرقی برخوردارند (15).

در پژوهشی اقدام به ارزیابی و پیش‌بینی تغییرات مقدار تبخیر و تعرق گیاه مرجع با استفاده از زنجیره مارکوف در ایستگاه شیراز شد. نتایج این مطالعه پیش‌بینی نمود که طی 16 سال (2015 الی 2030)، 25/52 درصد از ماه‌ها دارای کلاس تبخیر و تعرق کم، 52/08 درصد از ماه‌ها دارای کلاس تبخیر و تعرق متوسط و 22/39 درصد از ماه‌ها دارای کلاس تبخیر و تعرق زیاد هستند (22).

**n مواد و روش ها**

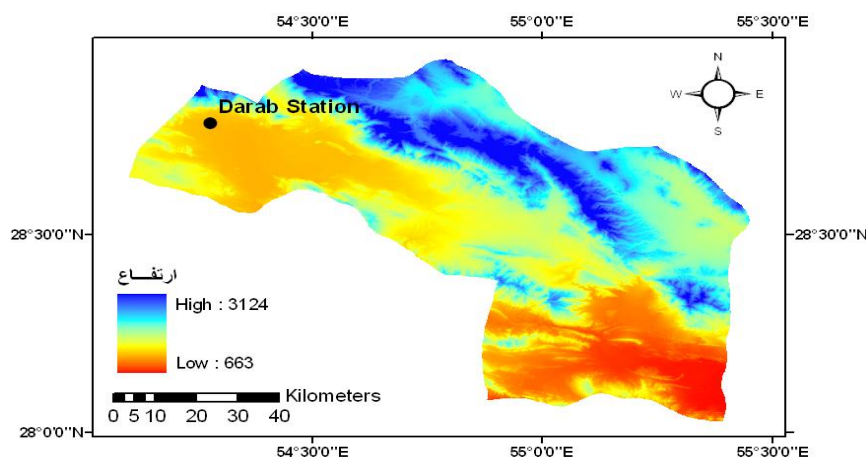
**منطقه مورد مطالعه**

منطقه مورد مطالعه در پژوهش حاضر شهرستان داراب، در جنوب غربی استان فارس، با طول جغرافیایی 28° 01' تا 54° 06' شرقی و عرض جغرافیایی 28° 56' تا 54° 06' شمالی با مساحتی برابر با 6560/32 کیلومتر مربع است (شکل 1). در پژوهش حاضر از

متغیرهای اقلیمی مؤثر در محاسبه تبخیر و تعرق بر مبنای روش پنمن مونتیتش فائو در این ایستگاه طی سال های 1984 تا 2014 استفاده شد. برخی از خصوصیات اقلیمی منطقه مورد مطالعه از جمله: کمینه، میانگین و بیشینه دما، ساعات آفتابی، سرعت باد، بیشینه ساعات آفتابی، رطوبت نسبی و مقدار تشعشع که در جدول 1 ارائه شده است.

جدول 1. ویژگی های آماری هریک از متغیرهای مؤثر در تعیین تبخیر و تعرق پتانسیل منطقه مورد مطالعه

متغیر	انحراف معیار	میانگین	بیشینه	کمینه
کمینه دما (درجه سانتی گراد)	7/83	14/35	27/3	2/3
بیشینه دما (درجه سانتی گراد)	9/09	29/97	43/1	14
دمای میانگین (درجه سانتی گراد)	8/39	22/18	34/3	8/6
رطوبت نسبی (درصد)	14/03	39/62	71	19
ساعات آفتابی	1/38	11/3	13/2	8/1
بیشینه ساعات آفتابی	1/34	12/15	14	10/2
سرعت باد	5/74	11/29	40	3
مقدار تشعشع	140/12	601/58	793	371
ارتفاع (متر)	1/1	1870	2100	1098



شکل 1. منطقه مورد مطالعه

**روش پژوهش**

**محاسبه تبخیر و تعرق**

در این مطالعه به منظور برآورد مقدار تبخیر و تعرق پتانسیل از روش پنمن مونتیتش فائو استفاده شده است. این روش به عنوان روش استاندارد برآورد تبخیر و تعرق

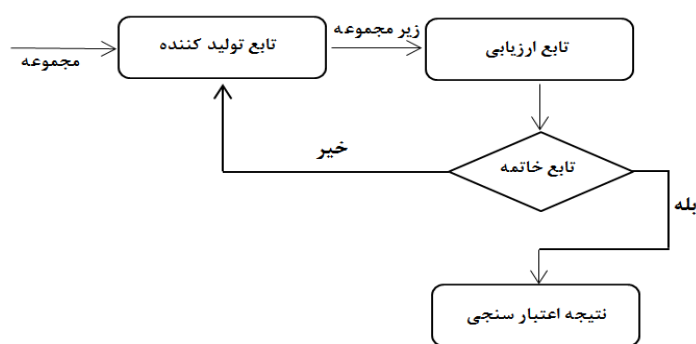
گیاه مرجع محسوب می شود. در رابطه 1 روش محاسبه آن ارائه شده است (2, 3, 24, 25 و 26).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

مجموعه‌های جواب به‌صورت توانی با  $T$  افزایش پیدا می‌کنند، پیدا کردن جواب بهینه مشکل و در  $T$  های متوسط و بزرگ بسیار پرهزینه است. برای اینکه بتوانیم تقسیم‌بندی درستی از روش‌های مختلف انتخاب ویژگی داشته باشیم، به این صورت عمل می‌شود که فرآیند انتخاب ویژگی در تمامی روش‌ها را به این بخش‌ها تقسیم می‌کنیم (7، 17 و 18):

1. تابع تولیدکننده یا Generation procedure: این تابع زیرمجموعه‌های کاندید را برای روش موردنظر پیدا می‌کند.
2. تابع ارزیابی یا Evaluation function: زیرمجموعه موردنظر را بر اساس روش داده‌شده، ارزیابی و یک عدد به‌عنوان مقدار خوبی روش باز می‌گرداند. روش‌های مختلف سعی در یافتن زیرمجموعه‌ای دارند که این مقدار را بهینه کند.
3. شرط خاتمه: برای تصمیم‌گیری در مورد زمان توقف خوارزمیک.

به کمک تابع تعیین اعتبار یا Validation procedure مشخص می‌شود که آیا زیرمجموعه انتخاب شده معتبر است یا خیر. روند تعیین اعتبار به کمک این تابع در شکل 2 ارائه شده است:



شکل 2. مراحل خوارزمیک Feature selection و انتخاب ویژگی (5)

مختلف ارزیابی روش انتخاب ویژگی در جدول 2 ارائه شده است.

در رابطه 1 عوامل عبارتند از:  $ET_0$ ؛ تبخیر و تعرق گیاه مرجع برحسب میلی‌متر بر روز،  $\Delta$ ؛ شیب منحنی فشار بخار در دمای  $T$  برحسب کیلو پاسکال بر درجه سانتی‌گراد،  $R_n$ ؛ تابش خالص در سطح گیاه برحسب مگاژول بر مترمربع در روز،  $G$ ؛ شار گرمای خاک برحسب مگاژول بر مترمربع در روز،  $T$ ؛ دمای هوا در ارتفاع 2 متری برحسب سلسیوس،  $u_2$ ؛ سرعت باد در ارتفاع 2 متری برحسب متر بر ثانیه،  $e_0(T) = e_s$ ؛ فشار بخار اشباع برحسب کیلو پاسکال،  $e_a = e_0(T_{dew})$ ؛ فشار بخار واقعی برحسب کیلو پاسکال،  $(e_s - e_a)$ ؛ کمبود فشار بخار اشباع برحسب کیلو پاسکال،  $\gamma$ ؛ ضریب ثابت سایکرومتری برحسب کیلو پاسکال بر درجه سانتی‌گراد است.

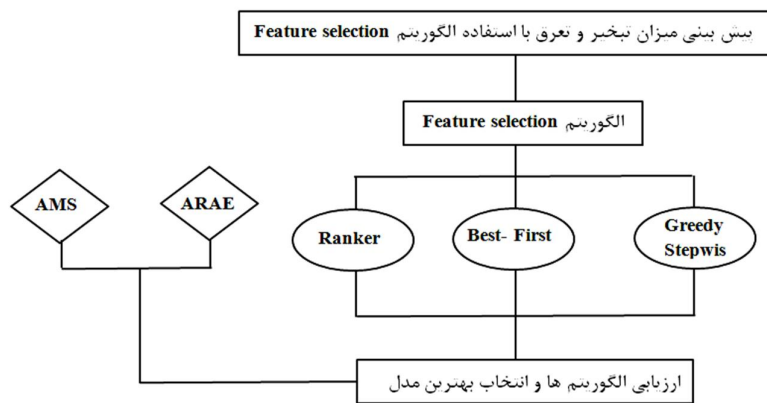
### روش Feature selection یا انتخاب ویژگی

روش‌های مختلف انتخاب ویژگی، تلاش می‌کنند تا از میان  $2^T$  زیر مجموعه موردنظر، بهترین زیرمجموعه را پیدا کنند. در تمام این روش‌ها بر اساس کاربرد و نوع تعریف، زیرمجموعه‌ای به‌عنوان جواب انتخاب می‌شود که به‌تواند مقدار تابع ارزیابی را بهینه کند. باوجوداینکه هر روشی سعی می‌کند بهترین ویژگی‌ها را انتخاب کند، اما با توجه به وسعت جواب‌های ممکن و اینکه این

مراحل مختلف به‌منظور انتخاب مهم‌ترین فاکتورهای مؤثر در پیش‌بینی تبخیر و تعرق در شکل 3 نشان داده شده است (11 و 19). لازم به ذکر است که مدل‌های

جدول 2. مهم‌ترین ویژگی‌های انتخاب‌شده توسط روش Feature selection به منظور پیش‌بینی مقدار تبخیر و تعرق

نام روش	نام مدل ارزیابی ویژگی
Best-First	CFSSubsetEva
Greedy-Stepwise	CFS-Subset-Eval
	Info-Gain-Attribute-Eval
	Gain-Ratio-Attribute-Eval
Ranker	Symmetrizer-Attribute-Eval
	Relief-FAttribute-Eval
	Principal-Components



شکل 3. روند خوارزمیک Feature selection برای انتخاب بهترین متغیرهای مؤثر در پیش‌بینی تبخیر و تعرق

$$ARAE = \frac{\sum_{i=1}^n RAE_i}{N} \quad (3)$$

که در رابطه 3: ARAE: میانگین خطای مطلق نسبی،  $RAE_i$  تعداد کلاس‌های نادرست پیش‌بینی‌شده توسط مدل و  $N$  تعداد کل داده‌هاست.

بررسی دقت هریک از روش‌ها

به منظور مشخص کردن دقت هریک از روش‌ها از خوارزمیک‌های زیر استفاده شد:

خوارزمیک Naive Bayes

در این خوارزمیک از حالت 10 Fold برای طبقه‌بندی کلیه روش‌ها استفاده شد. همچنین مدل طبقه‌بندی مجموعه، آموزش کل (Full training set) است. در این روش برای انتخاب بهترین روش اطلاعات مربوط به ضریب

ارزیابی عملکرد

در روش AMS برای بررسی دقت مدل از رابطه 2 استفاده شد. در این روش نسبت بین میانگین تعداد طبقه‌های نادرست پیش‌بینی‌شده توسط مدل به تعداد کل داده‌ها استفاده می‌شود (5).

$$AMS = \frac{\sum_{i=1}^n MS_i}{N} \quad (2)$$

که در رابطه 2: AMS میانگین کلاس‌های نادرست پیش‌بینی‌شده،  $MS_i$  تعداد کلاس‌های نادرست پیش‌بینی‌شده توسط مدل و  $N$  تعداد کل داده‌هاست. در روش ARAE از نسبت میانگین خطای مطلق نسبی به کل داده‌ها بر اساس رابطه 3 استفاده شد (5).

### خوارزمیک<sup>1</sup> LMT

با استفاده از این خوارزمیک از بین روش‌های Best-First، Greedy-Stepwise و Ranker بهترین روش‌ها برای پیش‌بینی مقدار تبخیر و تعرق تعیین شد. برای ارزیابی دقت روش‌های مورد بررسی در تعیین مؤثرترین داده‌ها بر مقدار تبخیر و تعرق از دو روش AMS و ARAE استفاده شد (5).

### n نتایج

نتایج حاصل از انتخاب کمینه داده‌ها برای محاسبه تبخیر و تعرق بر مبنای روش‌های مختلف انتخاب ویژگی از جمله: روش Best-First با استفاده از مدل CFS-Subset-Eva، روش Greedy-Stepwise با استفاده از مدل CFS-Subset-Eva و همچنین مدل‌های مختلف روش Ranker در جدول 3 ارائه شده است.

جدول 3. مهم‌ترین ویژگی‌های انتخاب‌شده توسط روش Feature selection برای پیش‌بینی مقدار تبخیر و تعرق

نام روش	نام مدل ارزیابی ویژگی	متغیر انتخاب شده
Best-First	CFS-Subset-Eva	کمینه دما، بیشینه دما، رطوبت نسبی
Greedy-Stepwise	CFS-Subset-Eval	کمینه دما، بیشینه دما، رطوبت نسبی
	Info-Gain-Attribute-Eval	بیشینه ساعات آفتابی، مقدار تشعشعات، بیشینه دما
	Gain-Ratio-Attribute-Eval	دمای میانگین، ساعات آفتابی، رطوبت نسبی
Ranker	Symmetrizer-Attribute-Eval	بیشینه ساعات آفتابی، میانگین دما، کمینه دما
	Relief-F Attribute-Eval	بیشینه ساعات آفتابی، میانگین دما، بیشینه دما
	Principal-Components	بیشینه دما، دمای میانگین، کمینه دما

ضریب همبستگی، میانگین خطای مطلق، ریشه میانگین مربع خطا، خطای مطلق نسبی، ریشه مربع خطای نسبی، تعداد طبقات پیش‌بینی‌شده درست و غلط مربوط به کلیه روش‌های مورد بررسی با استفاده از خوارزمیک Naive Bayes در جدول 4 ارائه گردید.

همبستگی، میانگین خطای مطلق، ریشه میانگین مربع خطا، خطای مطلق نسبی، ریشه مربع خطای نسبی، تعداد طبقات پیش‌بینی‌شده درست و غلط برای هر یک از روش‌ها تعیین می‌شود (10).

### خوارزمیک J48

با استفاده از خوارزمیک J48 اقدام به تولید درخت تصمیم C4.5 شد که با کم کردن شاخه‌ها و یا بدون کم کردن آنها اجرا گردید. در این خوارزمیک داده‌ها با استفاده از اطلاعات نرمال به بخش‌هایی به‌عنوان زیرمجموعه تقسیم گردید. اصطلاحاً در این روش از ریشه به سمت برگ محاسبه شد و کلاس‌های مورد نظر تولید گردید. در طبقه‌بندی از حالت 10 Fold و مجموعه آموزشی کل استفاده شد (17).

### بررسی دقت هر یک از روش‌ها

#### خوارزمیک Naive Bayes

بر مبنای خوارزمیک Naive Bayes روش Ranker با مدل Relief-F Attribute-Eval به‌عنوان بهترین روش پیش‌بینی مقدار تبخیر و تعرق تعیین شد. نتایج مربوط به

<sup>1</sup> logistic model tree

جدول 4. ارزیابی طبقات با استفاده از Naive Bayes

نام روش	نام مدل ارزیابی ویژگی	ضریب همبستگی	میانگین خطای مطلق	ریشه میانگین مربع خطا	خطای مطلق نسبی	ریشه مربع خطای نسبی	تعداد طبقات پیش‌بینی شده	
							درست	غلط
Best-First	CFSSubsetEva	76/66	0/105	0/26	30/05	63/03	92	28
Greedy-Stepwise	CFS-Subset-Eval	76/66	0/105	0/26	30/05	63/03	92	28
	Info-Gain-Attribute-Eval	93/33	0/05	0/15	14/02	36/78	112	8
	Gain-Ratio-Attribute-Eval	87/5	0/08	0/21	21/91	50/07	105	15
Ranker	Symmetrizer-Attribute-Eval	95	0/04	0/15	11/38	35/97	114	6
	Relief-F Attribute-Eval	98	0/04	0/15	11/38	35/97	114	6
	Principal-Components	80	0/11	0/25	29/97	59/91	62	24

آمارهای روش‌های مورد بررسی در این خوارزمیک شامل ضریب همبستگی، میانگین خطای مطلق، ریشه میانگین مربع خطا، خطای مطلق نسبی، ریشه مربع خطای نسبی، تعداد طبقات پیش‌بینی شده درست و غلط در جدول 5 نشان داده شده است.

#### خوارزمیک J48

بر مبنای نتایج حاصل از خوارزمیک J48 همچون روش Naive Bayes روش Ranker با مدل Relief-F Attribute-Eval به‌عنوان بهترین روش پیش‌بینی مقدار تبخیر و تعرق تعیین شد. نتایج مربوط به آنالیز

جدول 5. ارزیابی طبقات با استفاده از J48

نام روش	نام مدل ارزیابی ویژگی	ضریب همبستگی	میانگین خطای مطلق	ریشه میانگین مربع خطا	خطای مطلق نسبی	ریشه مربع خطای نسبی	تعداد طبقات پیش‌بینی شده	
							درست	غلط
Best-First	Eval-Subset-CFS	95	0/045	0/15	12/9	35/96	114	6
Greedy-Stepwise	CFS-Subset-Eval	95	0/045	0/15	12/9	35/96	114	6
	Info-Gain-Attribute-Eval	98	0/016	0/09	4/63	21/54	118	2
	Gain-Ratio-Attribute-Eval	96/66	0/029	0/21	8/44	29/09	116	4
Ranker	Symmetrizer-Attribute-Eval	98/33	0/016	0/09	9/49	21/22	118	2
	Relief-F Attribute-Eval	98/33	0/016	0/09	9/49	21/22	118	2
	Principal-Components	89/16	0/082	0/20	23/63	48/67	107	13

خطای مطلق، ریشه میانگین مربع خطا، خطای مطلق نسبی، ریشه مربع خطای نسبی، تعداد طبقات پیش‌بینی شده درست و غلط مربوط به کلیه روش‌های مورد بررسی با استفاده از خوارزمیک Naive Bayes در جدول 6 ارائه شده است.

#### خوارزمیک LMT

بر مبنای نتایج حاصل خوارزمیک LMT در بحث انتخاب بهترین روش پیش‌بینی مقدار تبخیر و تعرق، روش Ranker با مدل Relief-F Attribute-Eval به‌عنوان بهترین روش پیش‌بینی مقدار تبخیر و تعرق انتخاب شد. نتایج مربوط به ضریب همبستگی، میانگین

جدول 6. ارزیابی طبقات با استفاده از LMT

نام روش	نام مدل ارزیابی ویژگی	ضریب همبستگی	میانگین خطای مطلق	ریشه میانگین مربع خطا	خطای مطلق نسبی	ریشه مربع خطای نسبی	تعداد طبقات پیش‌بینی شده	
							درست	غلط
Best-First	CFSSubsetEva	88/32	0/091	0/21	25/90	49/57	106	14
Greedy-Stepwise	CFS-Subset-Eval	88/32	0/091	0/21	25/90	49/57	106	14
	Info-Gain-Attribute-Eval	96/66	0/080	0/15	22/83	36/86	116	4
	Gain-Ratio-Attribute-Eval	92/50	0/083	0/18	23/94	43/68	111	9
Ranker	Symmetricer-Attribute-Eval	97/50	0/103	0/18	29/61	43/43	117	3
	Relief-F Attribute-Eval	98/50	0/103	0/18	29/61	43/43	117	3
	Principal-Components	84/16	0/126	0/24	36/14	57/22	101	19

پیش‌بینی مقدار تبخیر و تعرق در جدول 7 ارائه شده است.

نتایج حاصل از روش‌های AMS و ARAE برای تعیین مقدار خطای هر یک از روش‌ها به منظور اطمینان از انتخاب بهترین روش در انتخاب مؤثرترین داده‌ها برای

جدول 7. دقت کلاس‌های مختلف Feature selection

روش ارزیابی						نام مدل ارزیابی ویژگی	نام روش
Trees				Bayes			
Naive Bayes		J48		LMT			
AMS	ARAE	AMS	ARAE	AMS	ARAE		
23/34	25/04	5	10/75	11/68	21/58	CFS-Subset-Eva	Best-First
23/34	25/04	5	10/75	11/68	21/58	CFS-Subset-Eval	Greedy-Stepwise
6/67	11/68	2	3/86	3/34	19/03	Info-Gain-Attribute-Eval	
12/5	18/26	3/34	7/03	7/5	19/95	Gain-Ratio-Attribute-Eval	
5	9/48	1/67	3/74	2/5	24/68	Symmetricer-Attribute-Eval	Ranker
4	8/48	1/67	3/74	1/5	24/68	Relief-F Attribute-Eval	
20	24/98	10/84	19/69	15/84	30/12	Principal-Components	

تخمین مقدار تبخیر و تعرق، از 20 داده به‌عنوان نمونه آزمایشی استفاده شد که نتایج مربوط به آن در جدول 8 ارائه شده است.

در نهایت به‌منظور ارزیابی داده‌های انتخاب‌شده با استفاده از روش خوارزمیک ویژگی بر اساس روش Ranker با مدل Relief-F Attribute-Eval برای



جدول 8. مقادیر تبخیر و تعرق پیش‌بینی شده با استفاده از داده‌های انتخاب شده توسط روش انتخاب ویژگی و کل داده‌ها

داده	مقدار تبخیر و تعرق		داده	مقدار تبخیر و تعرق	
	داده‌های انتخاب شده توسط خوارزمیک انتخاب ویژگی	کل داده‌ها		داده‌های انتخاب شده توسط خوارزمیک انتخاب ویژگی	کل داده‌ها
1	0/881	0/892	11	1/45	1/400
2	0/921	0/946	12	1/51	1/541
3	0/968	0/974	13	1/5	1/556
4	1/01	1/035	14	1/56	1/575
5	1	1/077	15	1/6	1/635
6	1	1/081	16	1/55	1/655
7	1/23	1/112	17	1/6	1/675
8	1/11	1/214	18	1/5	1/788
9	1/20	1/244	19	1/7	1/821
10	1/32	1/372	20	1/56	1/860

## n بحث و نتیجه‌گیری

میانگین دما، بیشینه دما مؤثرتر بودند. خروجی‌ها نشان داد که بر مبنای روش Ranker با استفاده از مدل Principal-Components بیشینه دما، دمای میانگین، کمینه دما مؤثرترین متغیرها محسوب می‌شوند.

بر مبنای نتایج حاصل از ارزیابی مقدار دقت روش‌های پیش‌بینی تبخیر و تعرق بر اساس خوارزمیک Naive Bayes روش Ranker با مدل Relief-FAttribute-Eval با ضریب همبستگی معادل 98% و همچنین کمینه مقدار میانگین خطای مطلق، کمینه مقدار ریشه میانگین مربع خطا، کمینه مقدار خطای مطلق نسبی و کمینه مقدار ریشه مربع خطای نسبی در بین روش‌های مورد بررسی به‌عنوان بهترین روش پیش‌بینی تبخیر و تعرق تعیین شد. بر مبنای نتایج دقت‌سنجی روش‌های پیش‌بینی تبخیر و تعرق با استفاده از خوارزمیک J48 نیز روش Ranker با مدل Relief-FAttribute-Eval با ضریب همبستگی معادل 98/33% و همچنین کمینه مقدار میانگین خطای مطلق، کمینه مقدار ریشه میانگین مربع خطا، کمینه مقدار خطای مطلق نسبی و کمینه مقدار ریشه مربع خطای نسبی در بین روش‌های مورد بررسی به‌عنوان بهترین روش پیش‌بینی تبخیر و تعرق تعیین گردید. بر مبنای خوارزمیک LMT روش Ranker با مدل Relief-FAttribute-Eval با ضریب همبستگی معادل 98/5% به‌عنوان بهترین روش پیش‌بینی تبخیر و تعرق تعیین شد. نتایج حاصل از روش‌های AMS و

در این بررسی مهم‌ترین متغیرهای اقلیمی مؤثر در پیش‌بینی تبخیر و تعرق پتانسیل با استفاده از روش Feature selection در ایستگاه داراب، واقع در جنوب استان فارس، مورد بررسی و انتخاب قرار گرفتند. متغیرهای اقلیمی مورد بررسی در این مطالعه شامل: کمینه، میانگین و بیشینه دما، ساعات آفتابی، سرعت باد، بیشینه ساعات آفتابی، ارتفاع، مقدار تشعشعات و رطوبت نسبی است. نتایج حاصل از روش Best-First با استفاده از مدل CFS-Subset-Eva نشان داد که مؤثرترین متغیرها جهت پیش‌بینی تبخیر و تعرق کمینه دما، بیشینه دما و رطوبت نسبی است. نتایج به‌دست‌آمده از روش Greedy-Stepwise مشابه با روش Best-First است. براساس نتایج حاصل از روش Ranker بر مبنای مدل Info-Gain-Attribute-Eval مؤثرترین متغیرها بیشینه ساعات آفتابی، مقدار تشعشعات، بیشینه دما بودند درحالی‌که در همین روش با استفاده از مدل Gain-Attribute-Eval دمای میانگین، ساعات آفتابی، رطوبت نسبی مؤثرترین متغیرها بودند. نتایج روش Ranker بر مبنای مدل Symmetricer-Attribute-Eval نشان‌دهنده تأثیرگذاری بیشتر متغیرهای بیشینه ساعات آفتابی، میانگین دما، کمینه دما مؤثرترین پارمترها بود درحالی‌که در همین روش بر مبنای مدل Relief-FAttribute-Eval متغیرهای بیشینه ساعات آفتابی،

سرعت باد و در ایستگاه رشت متوسط ساعات آفتابی مهم‌ترین متغیرهای تأثیرگذار بر مقدار تبخیر و تعرق هستند (22). نتایج حاصل از پژوهش حاضر با نتایج پژوهش زارعی (2017) در مناطق نیمه‌خشک هم‌خوانی دارد. پیش‌بینی تبخیر و تعرق با کاربرد شبکه عصبی مصنوعی و کمینه داده‌های هواشناسی نشان می‌دهد که سه متغیر دمای متوسط، ساعت آفتابی و سرعت باد، مؤثرترین متغیرها در پیش‌بینی خشکسالی است (1)؛ که با نتایج پژوهش حاضر در بحث متوسط دما و ساعات آفتابی هم‌خوانی دارد و به‌نوعی مبین مناسب بودن روش Feature selection در انتخاب مهم‌ترین متغیرهای اقلیمی مؤثر در پیش‌بینی تبخیر و تعرق است. بنابراین با بهره‌گیری از خوارزمیک Feature selection با انتخاب متغیرهای کمتر می‌توان از اندازه‌گیری همه داده‌ها صرف‌نظر و در زمان و هزینه صرفه‌جویی کرد. طبیعی است استفاده از خوارزمیک Feature selection در انتخاب متغیرهای کمتر در رابطه با دیگر زمینه‌های پژوهشی از قبیل ارزیابی تناسب اراضی، حاصلخیزی خاک، رطوبت نسبی و ... نیز می‌تواند کارایی داشته باشد.

ARAE نشان داد که بهترین روش برای انتخاب مؤثرترین داده‌ها برای پیش‌بینی مقدار تبخیر و تعرق روش J48 با کمترین مقدار AMS و ARAE است. نتایج پژوهش حاضر نشان داد که می‌توان با کاهش داده‌های مؤثر در برآورد تبخیر و تعرق نتایج مشابه‌ای به‌دست آورد؛ بنابراین با توجه به اینکه کاهش تعداد داده‌ها منجر به کاهش حجم اطلاعات و تسریع در تجزیه و تحلیل داده‌ها می‌شود، لذا استفاده از خوارزمیک انتخاب ویژگی به‌منظور کاهش داده‌ها مهم، کاربردی و تأثیرگذار است.

ارزیابی مقدار تأثیرگذاری متغیرهای اقلیمی مختلف در تغییر مقدار تبخیر و تعرق پتانسیل با استفاده از رگرسیون لجستیک با استفاده از داده‌های چهار ایستگاه رشت، شهرکرد، شیراز و زابل نیز نشان داد که در مناطق فراخشک مانند ایستگاه زابل و مناطق خشک مانند ایستگاه شیراز متوسط سرعت باد تأثیرگذارترین متغیر و متوسط ساعات آفتابی کم‌اهمیت‌ترین شاخص تأثیرگذار بر مقدار تبخیر و تعرق پتانسیل است. در مناطق نیمه‌خشک و مرطوب، به‌ترتیب شامل ایستگاه‌های شهرکرد و رشت متوسط بیشینه دمای ماهانه، کم‌اهمیت‌ترین شاخص است. در ایستگاه شهرکرد متوسط

## References

1. Aligholnia, T., Rasouli Majd, N., Rezaie, H., & Jabbari A. (2016) Application of artificial neural network to predict the evapotranspiration with the minimum meteorological data, *Journal of Natural Environment and Water Engineering*, 2(2): 122-135.
2. Allen, R., Raes, G., & Smith, M. (1998). Crop evapotranspiration guidelines for computing crop water requirements, *FAO Irrigation and Drainage*, 56, FAO, Rome, Italy. 301P.
3. Allen, R., & Pruitt, W.D. (1986). Rational use of the FAO Blany-Criddle Formula, *Journal of Irrigation and Drainage Engineering*, ASCE, 2, 139-155.
4. Angstrom, A. (1924). Solar and terrestrial radiation, *Journal Meteorological Society*, 50, 121-136.
5. Colkesen, I., Kavzoglu T (2017) The use of logistic model tree (LMT) for pixel- and object-based classifications using high-resolution WorldView-2 imagery, *International Journal of Geocarto*, 32(1): 71-86.
6. Daneshyar, M. (1978). Solar radiation statistics for Iran, *Sol Energy*, 21, 345-9.
7. Dash, M., & Liu, H. (2003). Consistency-based Search in Feature Selection, *Artificial Intelligence*, 151, 155-176.
8. Irmak, A., Allen, R., & Jones, JW. (2003). Solar and net radiation based equations to estimate reference evapotranspiration in humid climates, *Journal of Irrigation Draining Engineering*, 5, 336-347.
9. Jensen, R. (2008). Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches,

- Wiley-IEEE Press.
10. Khairina, DM., Maharani, S., Ramadiani, H., & Heliza R. (2017) Decision Support System for Admission Selection and Positioning Human Resources by Using Naive Bayes Method, *Advanced Science Letters*, 23(3): 2495-2497.
  11. Lin, S.W., Ying, K.C., Chen, S.C., & Lee, Z.J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines, *Expert Systems with Applications*, 35, 1817-1824.
  12. Mokarram, M & Zarei, A.R. (2015). Classification of potential evapotranspiration using Self Organizing Map (SOM), *Journal of Water and Irrigation Engineering*, 6 (22), 128-137. (In Farsi)
  13. Paltridge, G.W., & Proctor, D. (1976). Monthly mean solar radiation statistics for Australia. *Sol Energy*, 3, 235-43.
  14. Rajabi, Z., Behyar, M., Ghauor, B., Ezatiyan, H., & Gandomkar, A. (2015). Determination of evapotranspiration in Zaferan using FAO pen in Esfahan province, *Geographical Research Journal*, 30 (1), 23-34. (In Farsi)
  15. Sharghi, T., Barei Abarghoei, H., Asadi, M., & Kosari M. (2010). Estimation of reference evapotranspiration using FAO-Penman-Monteith method and its zonation in Yazd province, *Arid Biom Scientific and Research Journal*, 1 (1), 25-33. (In Farsi)
  16. Soltani koupaei, S., & Morid S. (2005). Solar radiation estimation using temperature-based and artificial neural networks approaches, *Journal of Agricultural Science (University of Tabriz)*, 15 (1), 69-77. (In Farsi)
  17. Veeraswamy, A., Sudhakara, AVS., & Krishna, TK. (2017) Feature Ranking Based Data Classification for Large Scale Data Analysis Using J48 Algorithm, *International Journal for Modern Trends in Science and Technology*, 3 (1): 104- 107.
  18. Visalakshi, S., & Radha, V. (2015). Wrapper based Feature Selection and Classification for Real Time Dataset. *International Journal of Emerging Technologies in Computational and Applied Sciences*, 3, 306-311.
  19. HWRP, WMO (2014). Statement of guidance for hydrology and water resources. *SOG For Hydrology and Water Resources*, 1, 1-13.
  20. Xue, B., Qin, A.K., & Zhang, M. (2014). An archive based particle swarm optimization for feature selection in classification. *In Proceedings of 2014 Congress on Evolutionary Computation*, Beijing, IEEE, 3119- 3126.
  21. Zarei, A.R., Jafari, M., Azarnivand, H., & Zehtabian, G.H. (2014). *Determination of crop coefficient (Kc) and evaluation of effect of aridity and salinity stresses in species of Atriplex lentiformis, Tamarix aphylla, Acacia victoria and Zizyphus spina- christi in order to select of suitable plant for reclamation of arid regions (Case study: Garebayegan, Fasa)*, PhD Thesis, Tehran University. (In Farsi)
  22. Zarei, A.R. (2017). Evaluation and prediction of reference evapotranspiration using Markov chain (Case study: Shiraz station), *2th national conference of hydrology*, 12 and 13 July, Shahr Kord University. (In Farsi)
  23. Zarei, A.R. (2017). Evaluation of effect of climate parameters in changes of potential evapotranspiration using logistic regression, *2th national conference of hydrology*, 12 and 13 July, Shahr Kord University. (In Farsi)
  24. Zarei, A. R., Mahmoudi, M. R. (2017). Evaluation of changes in RDIst index effected by different Potential Evapotranspiration calculation methods, *Water Resources Management*, 31 (15), 4981-4999.
  25. Zarei, A. R., Moghimi, M. M., Mahmoudi, M. R. (2016). Analysis of changes in spatial pattern of drought using RDI index in south of Iran, *Water Resources Management*, 30 (11), 3723-3743.
  26. Zarei, A. R., Moghimi, M. M., Mahmoudi, M. R. (2016). Parametric and Non-Parametric Trend of Drought in Arid and Semi-Arid Regions Using RDI Index, *Water Resources Management*, 30, 5479-5500.

## **Using feature selection algorithm for prediction of evapotranspiration with the lowest data**

M. Mokarram<sup>1</sup>, A.R. Zarei<sup>2\*</sup>

1. Faculty member, College of Agriculture and Natural Resources of Darab, Shiraz University, Darab, Iran.
2. Faculty member, College of Agriculture, University of Fasa, Fasa, Iran.

\* Corresponding Author: Ar\_Zareiee@Fasau.ac.ir

Received date: 04/12/2016

Accepted date: 20/09/2017

### **Abstract**

In the present study the most effective climatological parameters for prediction of evapotranspiration using feature selection algorithm in Darab city located in southwest of Fars province was selected. In the first stage, the values of evapotranspiration were calculated based on FAO Penman-Montith method, then using feature selection method, the most effective parameters were selected among all effective parameters to evapotranspiration prediction based on FAO Penman-Montith method. Using Best First, Greedy Stepwise and Ranker as the most famous methods of feature selection the most effective parameters from 120 data were selected. Also in order to investigate of the error of each method to choose the best method Naïve Bayes, J48 and LMT was used. Using minimum, maximum and average temperatures, relative humidity, sunshine and maximum sunshine hours, wind speed, clear sky solar radiation ( $R_{so}$ ), (75% of solar radiation in the upper atmosphere), the evapotranspiration was predicted. The results show that Ranker method with Relief- Attribute-Eval in Naïve Bayes, J48 and LMT method had the lowest error. So maximum sunshine hours, maximum and average temperatures were found to be the most effective parameters for prediction of evapotranspiration. Using feature selection algorithm can be useful to predict of evapotranspiration in regions with limited data and save time and money.

**Keywords:** Climatic parameters; Generation procedure; Evaluation function; FAO Penman-Montith; Validation procedure