

Determinación de Evapotranspiración de Referencia a partir de Modelos de Inteligencia Artificial

Jose Juarez¹, Gustavo Trentin², y Milton Heinen³

¹ Unipampa, Estudiante de Post-Grado en Computación Aplicada, Bagé - RS, Brasil
jose.juarez@science.com.pe

² Embrapa Pecuária Sul, Investigador A, Bagé - RS, Brasil
gustavo.trentin@embrapa.br

³ Unipampa, Profesor de Post-Grado en Computación Aplicada, Bagé - RS, Brasil
milton.heinen@unipampa.edu.br

Resumen. La evapotranspiración de referencia (ET_o) es un indicador importante para la gestión y planificación de los recursos hídricos. El objetivo principal de este estudio fue evaluar 4 modelos de inteligencia artificial (IA) para estimar la ET_o en función al número de variables de entrada. Para tal fin, se aplicó K-Nearest Neighbors (K-NN), Naïve Bayes (BAYES), Support Vector Machines (SVM) y Artificial Neural Network (ANN), sobre una base de datos de 3 años. Los resultados presentados en este estudio, indican que los modelos de IA podrían utilizarse para la aplicación práctica de la planificación y gestión de los recursos hídricos.

Palabras Claves: Penman-Monteith, meteorología, support vector machines, artificial neural network.

1 Introducción

La evapotranspiración de referencia (ET_o) es un indicador importante para la gestión y planificación de los recursos hídricos, su estimación debe ser realizada por el método que mejor caracterice las condiciones climáticas locales. Así el método de estimación de ET_o de Penman-Monteith fue estandarizado y recomendado por la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) [1].

Sobreestimar la ET_o podría resultar en pérdidas de agua, y subestimar la ET_o podría generar estrés hídrico en los cultivos, repercutiendo en reducción de la producción. De aquí la importancia de estimar la ET_o para planificar y gestionar los recursos hídricos.

El método Penman-Monteith, considerado como patrón, integra distintas variables meteorológicas para su estimación, las cuales no siempre se encuentran disponibles. De este modo, se precisa de modelos que utilicen el menor número de variables, pero sobre todo, con las que se consiga generalizar la estimación de la ET_o bajo distintas condiciones climáticas locales.

En ese sentido los modelos de inteligencia artificial (IA) pueden ser utilizadas para la estimación de la ET_o . Esto debido a su capacidad de aprendizaje mediante entrenamiento y corrección de errores, la que le permite especializarse en tareas [2], [3], [4], [5], [6]. En ese contexto, el objetivo principal de este estudio fue estimar la evapotranspiración mediante 4 modelos de IA, en función al número de variables de entrada y analizar la precisión de dicha estimación.

2 Materiales y métodos

La base de datos de colecta automática para este estudio corresponde a la Estación Meteorológica Principal de Embrapa Pecuária Sul, Bagé – RS, situada entre las coordenadas -31.347801° , -54.013292° (latitud, longitud) a 226 msnm. Esta base de datos meteorológicos incluye observaciones diarias de temperatura, humedad, viento, presión, precipitación y radiación solar que son presentados en la Tabla 1.

Los valores de referencia de la ET_o , se calcularon sobre la base diaria del método Penman-Monteith para el periodo de 30/03/2015 al 26/03/2018 (aproximadamente 3 años), siguiendo las recomendaciones en [1]:

$$ET_o = \frac{0,408 \Delta (R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0,34 u_2)} \quad (1)$$

Donde:

ET_o	evapotranspiración de referencia (mm día-1) calculada usando el método Penman-Monteith,
R_n	radiación neta en la superficie del cultivo (MJ m-2 día-1),
G	flujo del calor de suelo (MJ m-2 día-1),
T	temperatura media del aire a 2 m de altura (°C),
u_2	velocidad del viento a 2 m de altura (m s-1),
e_s	presión de vapor de saturación (kPa),
e_a	presión real de vapor (kPa),
$e_s - e_a$	déficit de presión de vapor (kPa),
Δ	pendiente de la curva de presión de vapor (kPa °C-1), y finalmente
γ	constante psicrométrica (kPa °C-1).

Tanto para el tratamiento de datos, como para la implementación de los modelos de IA se hizo el uso de la herramienta Python 3.6.5 [7] y bibliotecas de código abierto pandas [8], numpy [9], scipy [10], sklearn [11], tensorflow [12] y keras [13], destinadas a apoyar el objetivo de este trabajo.

Posteriormente, los casos de la base de datos fueron divididos en 80% para el conjunto de entrenamiento y 20% para el conjunto de prueba. El conjunto de entrenamiento utiliza la técnica de validación cruzada, la cual fue subdividida en 3 subconjuntos, la que permite evaluar los resultados y garantizar la independencia de la partición entre datos de entrenamiento y prueba [14].

La distribución de frecuencia de la ET_o del conjunto de entrenamiento no difiere al conjunto de prueba. Por lo tanto las series son parte de la misma población. Esto es relevante porque se utiliza la técnica de validación cruzada en el proceso de entrenamiento, y este método es sensible a la forma de cómo es dividido los datos [14].

La variable ET_o fue redondeada a una cifra significativa para que pueda ser considerada como clase, obteniéndose de esta manera 9 clases: 0, 1, 2, 3, 4, 5, 6, 7 y 8. Los 4 modelos de IA fueron implementados y utilizados para estimar la ET_o en función al número de variables de entrada.

Tabla 1. Variables: mes, meteorológicas y ET_o

Nº	Variable	Descripción	Variable en el sistema
1	Mes	Expresada en número entero entre 1 y 12	Entrada
2	Temperatura	Temperatura media del aire (°C)	Entrada
3	Humedad	Humedad relativa media del aire (%)	Entrada
4	Viento	Viento (km.día ⁻¹)	Entrada
5	Presión	Presión (hPa)	Entrada
6	Precipitación	Precipitación pluviométrica (mm)	Entrada
7	Radiación	Radiación solar (W.m ⁻²)	Entrada
8	Evapotranspiración	Evapotranspiración de referencia (ET_o) (mm), método Penman-Monteith	Salida

Al observar que se tiene distintas variables de entrada, se decidió realizar la correlación por el método de Pearson (Figura 1) con el objetivo de determinar el grado de correlación o interdependencia de las variables de entrada con la ET_o [15].

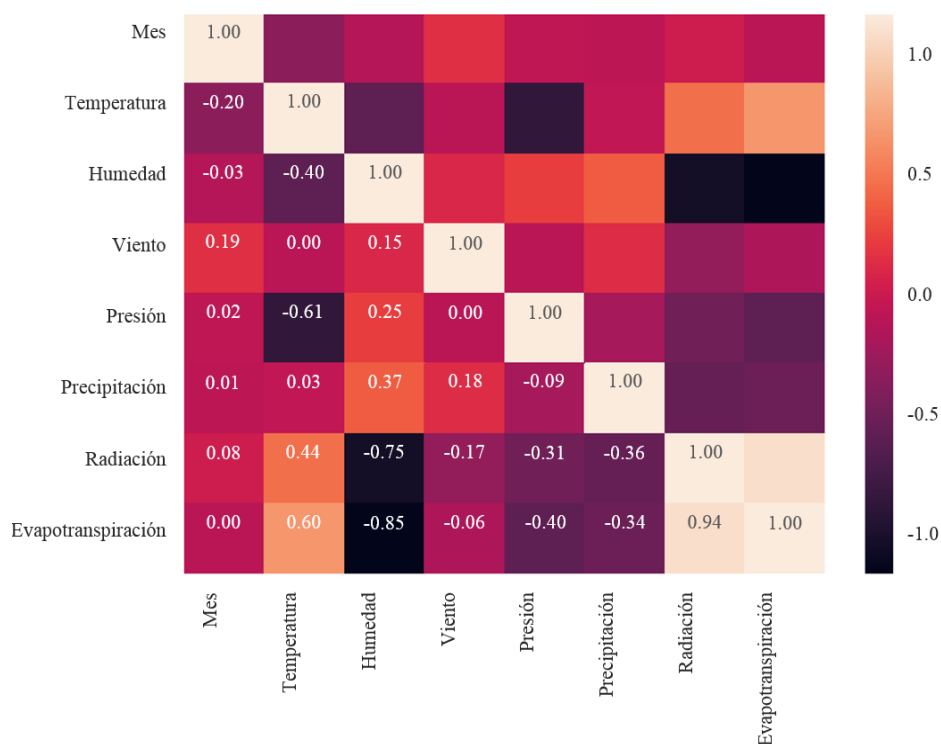


Fig. 1. Correlación de variables (mes y meteorológicas) en relación a la variable ET_o . Presentada según el mayor grado de correlación: Radiación(0.94), Humedad(-0.85), Temperatura(0.60), Presión(-0.40), Precipitación(-0.34), Viento(-0.06), Mes(0.00), respectivamente para los datos diarios pertenecientes a 3 años, de la Estación Meteorológica Principal de Embrapa Pecuária Sul, Bagé – RS.

Esto permitió secuenciar las variables con mayor correlación, y observar el comportamiento de un menor número de variables de entradas [16], [17] con la finalidad de estimar la ET_o , utilizando los 4 modelos de IA, tal como se describe a continuación.

El algoritmo K-Nearest Neighbors (K-NN), k-vecinos más cercanos, es un modelo de predicción no paramétrico. Busca los vectores de características más similares dentro de la base de datos histórica para predecir valores futuros. El modelo tiene una estructura simple y una alta eficiencia de cálculo [18], criterios relevantes para su elección en este trabajo. Este modelo usa la librería `sklearn.neighbors` y la función `KNeighborsClassifier` [11].

El clasificador Naïve Bayes (BAYES), clasificador bayesiano ingenuo, es una extensión del teorema de Bayes, que supone independencia condicional entre todas las características. La ventaja única que ofrece en comparación con otros algoritmos de clasificación no probabilísticos, es el uso de probabilidades en los algoritmos a eventos repetitivos como es característico en nuestra de base de datos [19]. Este modelo usa la librería `sklearn.naive_bayes` y la función `GaussianNB` [11].

El modelo Support Vector Machines (SVM), máquinas de soporte vectorial, permite obtener resultados a problemas de clasificación y regresión aplicados a grandes conjuntos de datos complejos con ruido, es decir con variables inherentes al modelo que para otros modelos aumentan la posibilidad de error en los resultados [20]. De hecho, la ET_o es un proceso complejo y no lineal debido a que depende de un gran número de variables meteorológicas. En la última década, se ha prestado mucha atención a la estimación de la ET_o , utilizando enfoques de IA, como se demuestra en [5], [21]. Este modelo usa la librería `sklearn.svm` y la función `SVC` [11].

El uso de la Artificial Neural Network (ANN), red neural artificial, es un método que emula el sistema nervioso biológico. Las estructuras ANN incluyen una capa de entrada, una capa oculta y una capa de salida, cada una con sus propias neuronas conectadas entre sí, con conexiones que tienen pesos numéricos que pueden ajustarse durante el entrenamiento, estos pesos son conocidos como pesos sinápticos, se utilizan para almacenar el conocimiento adquirido, esta capacidad por tanto le permite aprender y generalizar resultados razonables. Este enfoque es capaz de identificar algunos procesos previamente desconocidos que dominan la variación de ET_o a nivel de ecosistema y por tanto pueden estimar la ET_o con bastante precisión, demostrado en [3], [5], [4], [6]. Este modelo usa la librería `keras.models`, `keras.layers` y `keras.optimizers` y las funciones `sequential`, `dense` y `adam`. El modelo hace uso de 2 capas ocultas, cada una con 10 neuronas conectadas entre sí, activadas con la función `rectified linear unit (ReLU)`, la capa de salida es activada con la función `softmax` [12], [13].

Se seleccionó la precisión como medida de rendimiento de los modelos generados para compararlos. Esta medida se utilizó en [22], [23] y está dada por la proporción de los resultados verdaderos sobre el total de casos y convertida a porcentual.

A partir del análisis de los resultados de la ejecución con todas las variables de cada modelo, se procedió reducir el número de variables de entrada para estimar la ET_o , escogiendo secuencialmente las variables que poseen mayor correlación con la ET_o .

3 Resultados y discusión

Los resultados observados al utilizar la totalidad de variables (Tabla 1), muestran que de los 4 modelos utilizados: SVM y ANN tienen una mejor aproximación al estimar la ET_o , destacándose ANN por sobre los otros 3 modelos (Tabla 2).

Tabla 2. Comparación de precisión al estimar la ET_o con todas las variables disponibles.

Modelos	Penman-Monteith	K-NN	BAYES	SVM	ANN
Precisión	-	68.98%	69.59%	79.17%	85.37%

Los resultados obtenidos tras aplicar los 4 modelos de IA, sobre un número de variables limitadas e incrementadas según su correlación con la ET_o , indican que, las variables radiación, humedad, temperatura y viento son las más relevantes en la estimación de la ET_o , estas variables de entrada coinciden con el análisis de [4] y [6], puesto que con ellas se consigue un incremento considerable en la estimación de la ET_o , destacándose los modelos SVM y ANN independientemente del número de variables de entrada (Tabla 3).

Nótese también que el modelo BAYES característico por el uso de probabilidades en eventos repetitivos, al hacer uso de las 4 primeras variables solo consigue una aproximación a los demás modelos, mostrando ineficiencia al incrementar las variables de entrada en la estimación de la ET_o . Efecto similar con el modelo K-NN que al utilizar las 3 primeras variables se destaca de los demás modelos, sin embargo al incrementar las variables de entrada este muestra ineficiencia al estimar la ET_o (Tabla 3).

Tabla 3. Resultados y comparativa entre modelos de IA

Modelos	Precisión en relación a cantidad de variable usada por los modelos						
	1*	2*	3*	4*	5*	6*	7*
Penman-Monteith	-	-	-	-	-	-	-
K-NN	53.70%	59.26%	75.46%	71.30%	69.91%	68.98%	68.98%
BAYES	56.01%	64.35%	66.67%	70.37%	67.59%	68.06%	69.59%
SVM	56.02%	64.81%	71.30%	73.15%	75.93%	81.02%	79.17%
ANN	60.65%	66.67%	71.99%	72.92%	75.15%	80.32%	85.37%

1*=(Radiación); 2*=(Radiación y Humedad); 3*=(Radiación, Humedad y Temperatura); 4*=(Radiación, Humedad, Temperatura y Presión); 5*=(Radiación, Humedad, Temperatura, Presión y Precipitación); 6*=(Radiación, Humedad, Temperatura, Presión, Precipitación y Viento); 7*=(Radiación, Humedad, Temperatura, Presión, Precipitación, Viento y Mes).

De la Tabla 3 y la Figura 2, se puede observar que el conjunto de datos de entrada afecta significativamente la predicción de ET_o . También se examinó una mayor capacidad con las ANN al predecir valores diarios de ET_o , independientemente al número de variables de entrada.

Varios investigadores trabajaron con métodos de estimación de la ET_o basados en radiación como Priestly-Taylor [17], o basados en temperatura como Hargreaves [16]. El modelo de Hargreaves encontró resultados más precisos en todos los escenarios, seguidos por el modelo calibrado de Priestley-Taylor. En 2010, Traore et al. [24] demostraron que el modelo basado en la temperatura de ANN tiene un mejor rendimiento en comparación con el método empírico de Hargreaves.

Como sostienen Antonopoulos et al. [4] la capacidad de las ANN superan a los modelos empíricos en la predicción de valores diarios de la ET_o . Trabajos más recientes consideran eficientes a los modelos de SVM por la consistencia en sus fundamentos matemáticos y estadísticos [5], [21] y ANN [4], [5], [6]. Dentro de ellas se destaca el modelo de ANN, debido a su capacidad de aprendizaje mediante entrenamiento y

corrección de errores, que logra ajustarse a las variables meteorológicas e incluso ajustarse a la variable mes.

En la Figura 2 se presenta el número de variables utilizadas por los modelos de IA y, su precisión en % al estimar la ET_o .

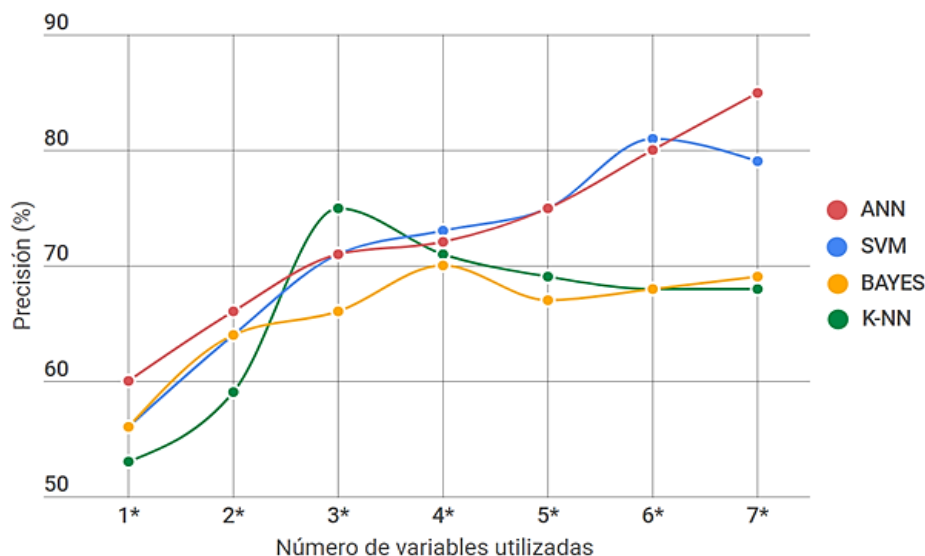


Fig. 2. Comparación entre la precisión de los modelos de IA y la cantidad de variables de entrada al estimar la ET_o , respectivamente para los datos diarios pertenecientes a 3 años, de la Estación Meteorológica Principal de Embrapa Pecuária Sul, Bagé – RS.

1* = (Radiación); 2* = (Radiación y Humedad); 3* = (Radiación, Humedad y Temperatura); 4* = (Radiación, Humedad, Temperatura y Presión); 5* = (Radiación, Humedad, Temperatura, Presión y Precipitación); 6* = (Radiación, Humedad, Temperatura, Presión, Precipitación y Viento); 7* = (Radiación, Humedad, Temperatura, Presión, Precipitación, Viento y Mes).

4 Conclusiones

En este trabajo se compararon 4 modelos de IA en relación al número de variables de entrada, a fin de evaluar su capacidad al estimar la ET_o . A partir de los resultados obtenidos, se determinó que las variables principales para la estimación de la ET_o son radiación, humedad, temperatura y viento.

El trabajo desarrollado demuestra, que para los datos utilizados y el área de influencia de la localización de la estación meteorológica, la estimación de la ET_o por el método K-NN con las tres principales variables ha sido aceptable. Sin embargo este modelo y el modelo BAYES, al incrementarles el número de variables de entrada, muestran ineficiencia al estimar la ET_o , probablemente debido a que no consiguen

interpretar los fenómenos de las variables presión, precipitación, viento y mes perdiendo precisión al estimar la ET_o .

Por otra parte, los modelos de aprendizaje SVM y ANN, logran adaptarse mejor a las variables, incrementando su precisión al estimar la ET_o , la cual es directamente proporcional al número de variables de entrada, demostrando robustez y flexibilidad, destacándose el modelo ANN.

Finalmente los modelos SVM y ANN podrían ser considerados en la aplicación práctica para estimar la ET_o , consecuentemente en la planificación y gestión de recursos hídricos, en caso de contar al menos con las tres variables de entrada principales.

Referencias

- [1] R. G. Allen, Evapotranspiración del cultivo: guías para la determinación de los requerimientos de agua de los cultivos, Food & Agriculture Org, 2006.
- [2] J. H. Holland, Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence., MIT press, 1992.
- [3] S. S. Haykin, Neural networks and learning machines, Pearson Upper Saddle River, NJ, USA, 2009.
- [4] V. Z. Antonopoulos y A. V. Antonopoulos, «Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables,» *CEA*, vol. 132, pp. 86-96, 2017.
- [5] X. Dou y Y. Yang, «Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems,» *Computers and Electronics in Agriculture*, vol. 148, pp. 95-106, 2018.
- [6] A. Irigoyen, «Knowledge extraction from artificial neural networks: Case study on reference evapotranspiration in southeastern of rolling pampas of Argentina,» *VIII Congreso Argentino de AgroInformática (CAI-2016)-JAIIO 45*, 2016.
- [7] Python, «python.org,» Python Software Foundation, [En línea]. Available: <https://www.python.org>. [Último acceso: 7 Julio 2018].
- [8] pandas, «pandas.pydata.org,» The pandas project, [En línea]. Available: <https://pandas.pydata.org>. [Último acceso: 7 Julio 2018].
- [9] NumPy, «numpy.org,» [En línea]. Available: <http://www.numpy.org>. [Último acceso: 7 Julio 2018].
- [10] SciPy, «scipy.org,» [En línea]. Available: <https://www.scipy.org/>. [Último acceso: 7 Julio 2018].
- [11] Scikit-Learn, «Scikit-learn.org,» [En línea]. Available: <http://scikit-learn.org>. [Último acceso: 7 Julio 2018].
- [12] Google, «tensorflow.org,» [En línea]. Available: [tensorflow.org](https://www.tensorflow.org/). [Último acceso: 7 Julio 2018].
- [13] Keras, «keras.io,» [En línea]. Available: <https://keras.io/>. [Último acceso: 7 Julio 2018].
- [14] R. Kohavi, «A study of cross-validation and bootstrap for accuracy estimation and model selection,» *Appears in the International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 14, n° 2, pp. 1137-1145, 1995.

- [15] J. Hauke y T. Kossowski, «Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data.,» *Quaestiones geographicae*, v. 30, n. 2, 2011, pp. 89-93.
- [16] G. H. Hargreaves y Z. A. Samani, «Reference crop evapotranspiration from temperature,» *Applied engineering in agriculture*, vol. 1, n° 2, pp. 96-99, 1985.
- [17] C. Priestley y R. Taylor, «On the assessment of surface heat flux and evaporation using large-scale,» *Monthly weather review*, vol. 100, n° 2, pp. 81-92, 1972.
- [18] N. S. Altman, «An introduction to kernel and nearest-neighbor nonparametric regression.,» *The American Statistician* v.46, n. 3, pp. 175-185, 1992.
- [19] A. Yahyaoui y I. Yahyaoui, «Machine Learning Techniques for Data Classification,» de *Advances in Renewable Energies and Power Technologies*, Elsevier, 2018, pp. 441-450.
- [20] N. Cristianini y J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press, 2000.
- [21] D. Ravinesh, S. Salcedo, L. Carro y B. Saavedra, *Drought prediction with standardized precipitation and evapotranspiration index and support vector regression models*, Elsevier, 2018, pp. 151-174.
- [22] M. Sokolova, N. Japkowicz y S. Szpakowicz, «Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation,» de *Australasian joint conference on artificial intelligence*, Springer, 2006, pp. 1015-1021.
- [23] N. Japkowicz, «Why question machine learning evaluation methods,» de *AAAI Workshop on Evaluation Methods for Machine Learning*, 2006, pp. 6-11.
- [24] S. Traore, Y.-M. Wang y T. Kerh, «Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone,» *Agricultural water management*, vol. 97, n° 5, pp. 707-714, 2010.
- [25] K. Loague y R. E. Green, «Statistical and graphical methods for evaluating solute transport models: overview and application,» *Journal of contaminant hydrology*, vol. 7, n° 1-2, pp. 51-73, 1991.