

Comparación de modelos de regresión para la eficiencia en el consumo de electricidad en las escuelas ecuatorianas: un caso de estudio

Alejandro Toapanta-Lema^{1,2}, Walberto Gallegos-Eras¹, Jefferson Rubio-Aguilar¹, Edilberto Llanes-Cedeño¹, Jorge Carrascal-García², Leticia García-Gallegos² and Paul D. Rosero-Montalvo³

¹ Universidad Internacional SEK, Quito-Ecuador,

² Instituto Superior Tecnológico 17 de Julio, Urcuquí-Ecuador

³ Universidad Técnica del Norte, Ibarra-Ecuador

Resumen Los modelos de pronóstico de consumo con su debida facturación permiten establecer estrategias para evitar sobre cargas en sistemas y penalizaciones por consumos elevados. Este trabajo presenta una comparación de modelos de predicción multivariante que permita detectar el coste final mensual del consumo eléctrico en relación con los diferentes parámetros de facturación. Como resultados relevantes, se obtuvo que los modelos a base de máquinas de soporte de decisión presentan una mejor sensibilidad al ser comparados con diferentes métricas que evalúan al error de predicción.

Palabras Claves: Modelos de regresión, predicción de consumo eléctrico.

Abstract. Consumption forecast models with their proper billing establish strategies to avoid overloads in systems and penalties for fuel consumption. This paper presents a comparison of multivariate prediction models that allows detecting the final monthly cost of electricity consumption in relation to the different billing parameters. The relevant results, it is obtained that the models based on decision support machines have a better sensitivity when compared with different metrics that evaluate the prediction error.

Keywords: regression models, electric consume prediction

1 Introducción

Los recursos naturales del planeta se deben conservar al máximo para su adecuado en los años venideros. Por tal motivo, la generación de electricidad está enfocada en energías renovables que satisfagan la creciente necesidad de las ciudades emergentes [1]. En consecuencia, el pronóstico de consumo es una tarea de planificación que permite a todos los integrantes del mercado energético cooperar eficazmente. También permite a los sectores de mayor consumo evitar sanciones [2]. Además, las empresas generadoras de electricidad puedan suministrar la cantidad adecuada, evitando sobrecargas en la generación del equipo. Sin embargo, el uso de electricidad en hogares e industrias muestran una alta variación que depende del estilo de vida de los usuarios y el tipo de bien o servicio que se genera. A todo esto hay que considerar el cambio climático [3]. Por ello, existe una gran variedad de estudios de consumo de energía. Entre los más importantes es la implementación de sensores que adquieran datos en tiempo real durante todo momento [4]. Con esta información, se busca la implementación de algoritmos de aprendizaje de máquina que permita el reconocimiento de patrones y la tendencia de crecimiento de la demanda energética del país. Un ejemplo son los trabajos [5][6][7] que plantean la metodología de pronóstico de las curvas de carga eléctrica

para las tecnologías de planificación a corto plazo. Para ello, los modelos matemáticos multivariantes y algoritmos como máquinas de soporte decisión (svm), regresiones matemáticas, descomposición empírica (EMD), redes neuronales (ANN), entre otros, han demostrado una gran exactitud en demostrar el fenómeno indicado.

El Ecuador, ha invertido fuertemente en proyectos hidroeléctricos con el objetivo de cambiar su matriz productiva y evitar la dependencia de la venta de petróleo crudo. Además, puede reducir toneladas de CO₂ al año. Las energías renovables en desarrollo son: (i) energía hidráulica, (ii) energía eólica, (iii) energía de biomasa, (iv) energía de biocombustibles, (v) Energía geotérmica, (vi) Energía mareomotriz, (vii) Energía fotovoltaica y (viii) Energía solar concentrada. Según la potencia efectiva, las centrales eléctricas representativas en la generación de electricidad son Hidráulica 58.53 %, Centrales térmicas 39.16 %, Biomasa 1.66 %, Fotovoltaica 0.32 \%, Viento 0.24 \%. No obstante, no existen trabajos que se relacione el consumo energético con modelos de datos con el objetivo de contar pronósticos de la creciente demanda, especialmente en el sector público. Actualmente, trabajos como [8][9] han presentado propuestas sobre el comportamiento de consumo y la demanda eléctrica en diferentes ciudades. Sin embargo, no se presenta un análisis robusto de información.

Durante este periodo de inversión en hidroeléctricas del país, el ministerio de educación se suscribió a la declaración del Milenio, en donde se establecen el conjunto de Metas de Desarrollo del Milenio (MDG). Entre las más importantes se destacan en este campo de educación que se basa en que todos los niños y niñas del mundo completen la educación primaria de forma igualitaria. Por tal motivo. El Ecuador, estableció criterios de ubicación en relación con los sectores relegados en conjunto con la demanda estudiantil y los bajos resultados académicos. Todas estas infraestructuras son llamadas Escuelas de Educación del Milenio (MES), las mismas que tienen una capacidad de 1140 o 570 estudiantes por jornada (matutina y vespertina) con bloques de aulas, educación inicial, laboratorios, uso múltiple, bar y administración, entre otros. En todo el país, se cuentan con 97 MES activas [10][11].

Cada MES cuenta con laboratorios, aulas, campos deportivos, que tienen una carga instalada significativa de 207.07KW. No obstante, en la actualidad no se tiene un estudio de consumo energético en este sector educativo y sus pronósticos de consumo en el futuro para una planificación de suministro. Por esta razón, se presenta un análisis de pronóstico de consumo y su comparación con diferentes modelos matemáticos con el fin de determinar el adecuado que permita regular el consumo eléctrico y establecer políticas de consumo. Como resultados relevantes, se determina que el modelo de regresión multivariante de máquinas de soporte de decisión cuenta con un menor error de predicción (alrededor del 3%) con una matriz de entrenamiento optimizada bajo el criterio de backward elimination.

El resto del documento se encuentra estructurado de la siguiente forma: la sección 2, muestra la metodología empleada y los modelos matemáticos utilizado. La sección

3, presentan los resultados de los modelos y la selección de este. Finalmente, la sección 4 muestra las conclusiones y trabajos futuros.

2 Metodología y modelos matemáticos

En esta sección se presentan la descripción de la base de datos obtenida para el análisis y el posterior uso de los algoritmos de predicción.

2.1 Descripción del conjunto de datos

El conjunto de datos es adquirido de la recopilación de consumo eléctrico durante los últimos 3 años de funcionamiento de la MES SUMAK YACHANA WASI, ubicada en el catón Cotacachi-Imbabura. La misma que cuenta con 997 estudiantes, 48 docentes en los horarios Matutino y vespertino. Hay que tomar en cuenta que el consumo está relacionado con la potencia activa y reactiva en los horarios de uso. Como resultado se obtuvo una matriz $Y \in R^{m \times n}$, donde m es el número de muestras recolectadas y n son los atributos de cada m . En este caso $m=36$ y $n=8$. Además, hay que tener en consideración tomar datos para validar los modelos. Por esta razón, se planeta el 80% de la base de datos para entrenamiento y el 20% para pruebas. Sin embargo, con el objetivo de encontrar los atributos que se encuentren estrechamente relacionados a nuestra variable objetivo, en este caso el rubro total de pago, se utiliza el criterio de **backward elimination**. El mismo que permite eliminar variables que no brinden demasiada cantidad de información al modelo matemático. En este sentido mediante el uso del *p-value* y *R-squared* brinda información necesaria para la selección de atributos [12]. En la Fig. 1 se muestran los valores significancia para el modelo.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  34.938803  13.354928  2.616 0.017494 *
P..ACTIVE.8_18.Kwh  0.054831  0.007963  6.886 1.93e-06 ***
P.ACTIVE.18_22.Kwh  0.087445  0.023630  3.701 0.001637 **
P.ACTIVE.22_8.Kwh  0.053124  0.010780  4.928 0.000109 ***
P.REACTIVE.00_24.KVARh -0.031909  0.017829 -1.790 0.090324 .
DEMAND.8_18.KW  1.196235  0.906645  1.319 0.203570
DEMAND.18_22.KW  2.556066  0.856984  2.983 0.007981 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.8 on 18 degrees of freedom
Multiple R-squared:  0.9581,    Adjusted R-squared:  0.9441
F-statistic: 68.55 on 6 and 18 DF,  p-value: 2.04e-11

```

Fig. 1. Significance importance for model prediction

Como se puede apreciar, las variables de potencia reactiva y la demanda en horario de 8 a 18 horas no aporta valiosa información al modelo. Con ello se valora nuevamente el valor significancia de cada variable como se muestra en la Fig. 2

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.098136  14.160240  2.691 0.014067 *
P..ACTIVE.8_18.Kwh  0.057864  0.005946  9.731 4.99e-09 ***
P.ACTIVE.18_22.Kwh  0.104660  0.023898  4.379 0.000290 ***
P.ACTIVE.22_8.Kwh   0.042576  0.010362  4.109 0.000545 ***
DEMAND.18_22.KW    3.198316  0.794444  4.026 0.000662 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.57 on 20 degrees of freedom
Multiple R-squared:  0.9465,    Adjusted R-squared:  0.9358
F-statistic: 88.52 on 4 and 20 DF,  p-value: 1.997e-12

```

Fig. 2. Significance importance for model prediction

Con esta información se plantean dos bases de datos de entrenamiento de los modelos. Por un lado, la matriz $Y \in R^{m \times n}$ antes comentada con todas las variables del consumo eléctrico. Por otro lado, la matriz $X \in R^{m \times p}$, donde p solo cuenta con 6 valores.

2.2 Modelos de regresión.

Con el fin de generar un pronóstico de consumo, existen diferentes modelos para realizar un análisis multivariante de regresión lineal. En este sentido se plantan los más importantes en relación con la literatura encontrada como: (i) covariance matrix, (ii) similarity Functions, (iii) Frequency table and others. El primer criterio (covariance matrix) es un modelo lineal simple en base a la ecuación:

$$y = b_0 + \sum_{i=0}^m x_i b_i \quad (1)$$

Donde los valores de x_i son los atributos de las matrices Y y X . El resultado gráfico de este modelo se encuentra en la Fig. 3

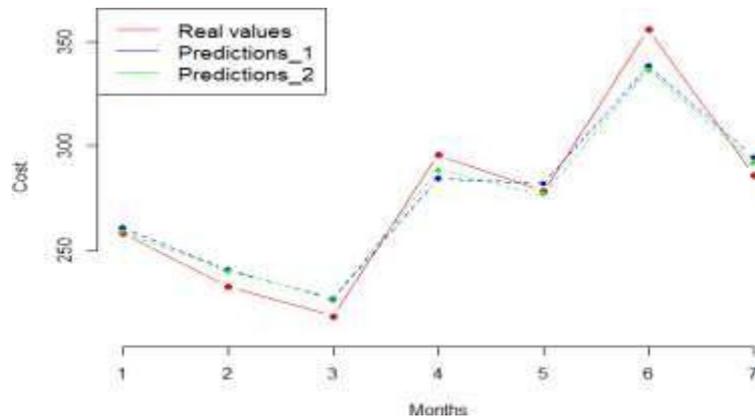


Fig. 3. Simple multivariate regression model. Predictions 1: Matrix Y , Predictions 2: Matrix X

Para mejorar el ajuste de un modelo lineal, se puede elevar al cuadrado sus valores de los atributos de cada matriz y encontrar una regresión multivariante polinómica de ecuación:

$$y = b_0 + \sum_{j=0}^r \sum_{i=0}^m x^j b_i \quad (2)$$

En la Fig. 4 se puede apreciar el ajuste del modelo con respecto al conjunto de entrenamiento.

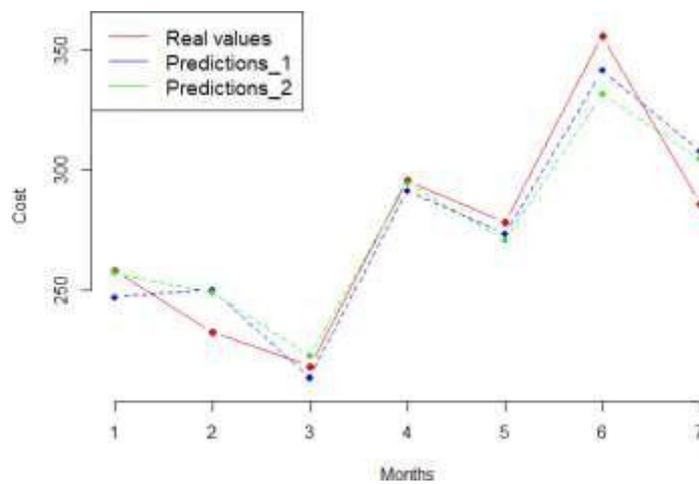


Fig. 4. Polynomial multivariant regression model. Predictions 1: Matrix Y , Predictions 2: Matrix X

Las máquinas de soporte de decisión (SVM) puede ser utilizadas como métodos de regresión. SVM establece márgenes de tolerancia para minimizar el error, individualizando el hiperplano que maximiza el margen de predicción. Este criterio viene dado por la ecuación 3 al usar un método kernel lineal [13].

$$y = b_0 + \sum_{i=0}^m (\alpha_i - \alpha^*)_i (x_i, x) + b \quad (3)$$

Como resultado de la implementación de este método kernel su representación gráfica se muestran en la Fig. 5

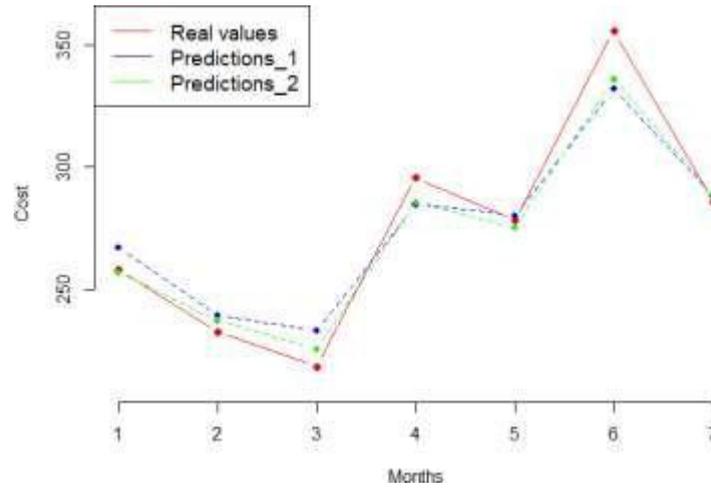


Fig. 5. SVM multivariant regression model. Predictions 1: Matrix Y , Predictions 2: Matrix X

Los algoritmos basados en tablas de frecuencia como árboles de decisión permiten tener un modelo de regresión lineal complejo y altamente no lineal. El mismo que se basa en la ecuación 4. Como resultado, aprende regresiones lineales locales que se aproximan a la curva sinusoidal [14].

$$y = \frac{1}{B} + \sum_{i=0}^m f_i * x, B = \text{numbers of trees} \quad (4)$$

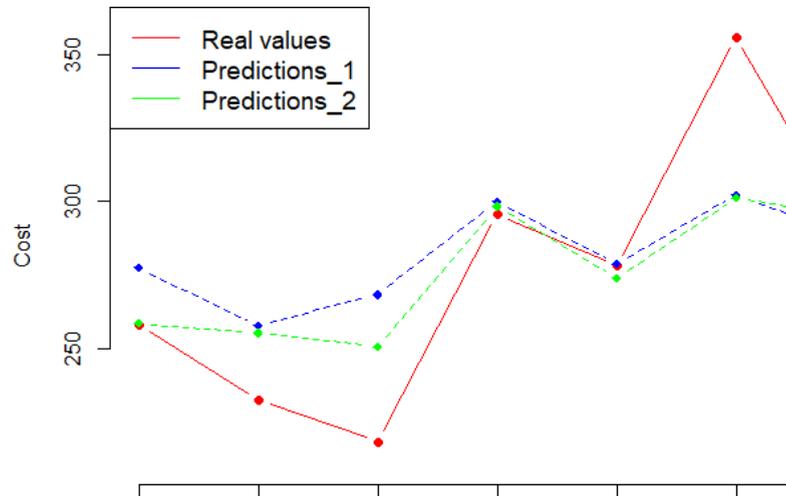


Fig. 5. Random Forest multivariant regression model. Predictions 1: Matrix Y , Predictions 2: Matrix X

Una implementación simple de la regresión KNN es calcular el promedio del objetivo numérico de los K vecinos más cercanos. Otro enfoque utiliza un promedio ponderado de distancia inversa de los K vecinos más cercanos. La regresión KNN utiliza las mismas funciones de distancia que la clasificación KNN. Las tres medidas de distancia anteriores solo son válidas para variables continuas [15]. Su modelo se basa en la distancia entre dos puntos como la ecuación 5.

$$\sqrt{\sum_{i=0}^m (x_i - y_i)^2}$$

En la Fig. 6 se muestran las predicciones del modelo a base de similitudes k-NN

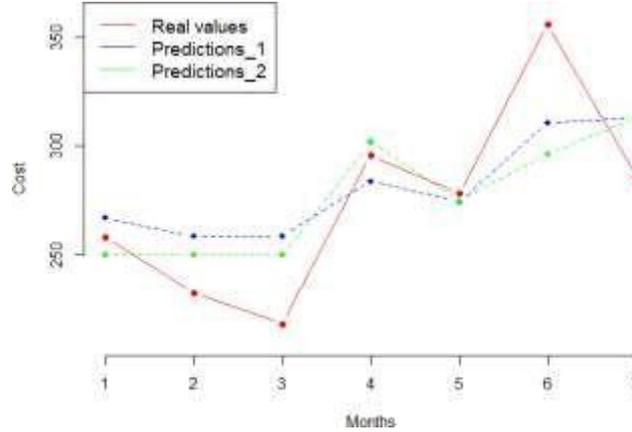


Fig. 6. k-NN multivariate regression model. Predictions 1: Matrix Y , Predictions 2: Matrix X

3 Resultados

Con los modelos implementados, los diferentes tipos de error permiten determinar el adecuado que represente la problemática planteada. Por esta razón, se analizó primero la suma acumulativa de errores de pronóstico (CFE) que proporciona información adicional como el pronóstico de desviación. En segundo lugar, la desviación del error medio (MAD) que da el valor absoluto de la diferencia entre real y valor estimado. Tercero, error cuadrático medio (RMSE) que da la dispersión del error predicho. Finalmente, porcentaje absoluto medio error (MAPE) que proporciona el error en porcentaje dado por el modelo y los datos reales.

Table 1. Error Analysis

Error	Linear		Polynomial		SVM		Random Forest		k-NN	
	Y	X	Y	X	Y	X	Y	X	Y	X
CFE	0.36	-0.67	0.12	0.87	0.08	-2.71	6.84	1.13	6	1.75
MAD	1.23	1.05	1.59	1.49	1.43	0.99	3.16	2.55	2.55	3.32
RMSE	0.01	0.06	0.001	0.11	0.001	1.05	6.69	0.18	5.14	0.43
MAPE	4.3	3.7	4.8	5.2	5.1	3.4	11.9	9.2	12.3	11.1

Según los análisis de errores se puede apreciar el menor error de pronóstico es el modelo lineal. Sin embargo, la dispersión del error (menor variabilidad) presenta el modelo SVM. Además, el error RMSE permite elevar el error al cuadrado y genera un mayor análisis del error. Finalmente, el error MAPE presenta en porcentaje un menor error de predicción. Como resultado, SVM con la matriz de datos X ha demostrado tener un

mejor rendimiento de predicción. En consecuencia, el modelo es validado en los próximos 6 meses, demostrando un ajuste muy significativo con un error de 3.2%. Esto en cuestiones monetarias es un error aproximado de 3.75 dólares. En a Fig.

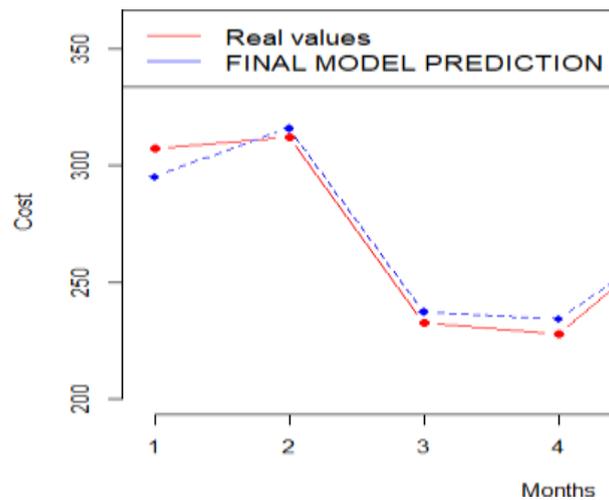


Fig. 7. Final regression model

4 Conclusiones y trabajos futuros.

El presente trabajo establece una base sólida para la planificación energética en centros de educación a lo largo del país. Por un lado, se pueden realizar planes de ahorro de consumo y validarlo en entornos reales. Por otro lado, en algunos sectores la lectura de los medidores es de forma manual, en donde existen muchos errores de lectura y cobros indebidos a los usuarios. Con la implementación de estos modelos de predicción se puede realizar seguimientos de consumo y detectar inconvenientes o fugas de corriente.

Con respecto a la implementación de algoritmos de predicción, se pudo demostrar que los modelos lineales se ajustan de mejor forma a otros criterios establecidos con sus algoritmos de Random Forest o k-NN. Además, el criterio de Backward Elimination permitió reducir la base de datos de entrenamiento y enfocarse en los atributos que brindan más información a la generación del modelo. Con ello, criterios estadísticos como el p-valor o el r-cuadrado ayuda a un análisis adecuado de variables.

Como trabajos futuros, se plantean instalar sensores de medición de corriente para generar una propia medida de consumo y establecer recomendaciones y políticas de corrección de factor de potencia para contar con rubros adecuados y sujetos a mejoras.

5 Referencias

- [1] T. Ishtiak, R. M. Orpon, N. Mashnoor, M. Ahmed, and M. A. Nazim, “An advanced application to decrease household power consumption and save energy detecting the weather condition,” in *2017 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2017*, 2017, pp. 622–627.
- [2] K. Dmitri, A. Maria, and A. Anna, “Comparison of regression and neural network approaches to forecast daily power consumption,” in *Proceedings - 2016 11th International Forum on Strategic Technology, IFOST 2016*, 2017, pp. 247–250.
- [3] X. M. Zhang, K. Grolinger, M. A. M. Capretz, and L. Seewald, “Forecasting Residential Energy Consumption: Single Household Perspective,” in *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 2019, pp. 110–117.
- [4] Y. Sun, W. Gu, J. Lu, and Z. Yang, “Fuzzy clustering algorithm-based classification of daily electrical load patterns,” in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015*, 2016, pp. 50–54.
- [5] B. Yildiz, J. I. Bilbao, J. Dore, and A. B. Sproul, “Recent advances in the analysis of residential electricity consumption and applications of smart meter data,” *Applied Energy*, vol. 208. Elsevier Ltd, pp. 402–427, 15-Dec-2017.
- [6] Y. Liu, W. Wang, and N. Ghadimi, “Electricity load forecasting by an improved forecast engine for building level consumers,” *Energy*, vol. 139, pp. 18–30, 2017.
- [7] H. X. Zhao and F. Magoulès, “A review on the prediction of building energy consumption,” *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6. pp. 3586–3592, Aug-2012.
- [8] M. Baquero and F. Quesada, “Eficiencia energética en el sector residencial de la Ciudad de Cuenca, Ecuador,” *MASKANA*, vol. 7, no. 2, pp. 147–165, Dec. 2016.
- [9] D. Gabriel and T. Urgilés, “UNIVERSIDAD POLITECNICA SALESIANA

SEDE CUENCA FACULTAD DE INGENIERIAS CARRERA DE
INGENIERÍA ELÉCTRICA.”

- [10] “UEM en Funcionamiento – Ministerio de Educación.” [Online]. Available: <https://educacion.gob.ec/uem-en-funcionamiento/>. [Accessed: 24-Sep-2019].
- [11] “Ministerio de Electricidad y Energía Renovable – Ente rector del Sector Eléctrico Ecuatoriano.” [Online]. Available: <http://historico.energia.gob.ec/>. [Accessed: 24-Sep-2019].
- [12] P. D. Rosero-Montalvo *et al.*, “Wireless Sensor Networks for Irrigation in Crops Using Multivariate Regression Models,” in *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, 2018, pp. 1–6.
- [13] X. Huang and S. Wang, “Prediction of bottom-hole flow pressure in coalbed gas wells based on GA optimization SVM,” in *Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018*, 2018, pp. 138–141.
- [14] X. Ye, X. Wu, and Y. Guo, “Real-time Quality Prediction of Casting Billet Based on Random Forest Algorithm,” in *Proceedings of the 2018 IEEE International Conference on Progress in Informatics and Computing, PIC 2018*, 2019, pp. 140–143.
- [15] L. He, Q. Song, and J. Shen, “K-NN numeric prediction using bagging and instance-relevant combination,” in *Proceedings - 2nd International Symposium on Data, Privacy, and E-Commerce, ISDPE 2010*, 2010, pp. 3–8.