

# Forecasting On Power Consumption Of Tetouan City

Daniel Molina, Dorian Jaramillo<sup>1</sup>

<sup>1</sup> Universidad de Antioquia CO, Departamento Ingeniería de Sistemas

Correos de los autores: Daniel Molina (e-mail: [daniel.molinay@udea.edu.co](mailto:daniel.molinay@udea.edu.co)), Dorian Jaramillo (e-mail: [dorian.jaramillo@udea.edu.co](mailto:dorian.jaramillo@udea.edu.co))

**ABSTRACT** La energía determina en gran medida el desarrollo y la sostenibilidad de un país, la forma en la que una sociedad administra la producción de energía, el consumo y su exportación son factores determinantes para su economía; por tanto la predicción del consumo de energía se convierte en un problema de vital importancia, es importante analizar el desbalance que existe entre producción y consumo para la toma de decisiones con el fin de cubrir la demanda requerida para un tiempo o temporada específica. En este trabajo se analiza la base de datos “Power consumption of Tetouan city Data Set”[1], tratándose de un problema de regresión se utilizarán los siguientes modelos: Redes neuronales, Regresión múltiple, Random Forest, Ventana de parzen y Máquinas de soporte vectorial. Los modelos serán aplicados variando algunos parámetros y seleccionando aquel para el cual se obtengan los mejores resultados.

**INDEX TERMS** Power consumption analysis, Machine learning for forecasting, Power consumption of Tetouan city

## I. INTRODUCCIÓN

La demanda de energía crece de forma vertiginosa debido a factores como: crecimiento poblacional, vehículos eléctricos, desarrollo económico, industrialización, centros de datos y minado de criptomonedas.

Aunque el avance tecnológico ha permitido la creación de ciudades inteligentes y dispositivos que autorregulan el consumo de energía, a su vez los nuevos avances vienen con nuevas formas de consumo masivo de energía como:

- Centros de cómputo para el minado de criptomonedas:  
dichos centros de minado en conjunto consumen más energía que países enteros.
- Vehículos eléctricos:  
Los vehículos eléctricos necesitan de estaciones de energía para su carga.

Debido a esto y otros factores se tiene como resultado un incremento anual del consumo de energía en la mayoría de los países desarrollados o en vía de desarrollo. se vuelve trascendente entonces contar con herramientas que permitan predecir el consumo de energía a futuro.

El problema abordado corresponde a un problema de regresión, donde se tratará de predecir el consumo de energía teniendo en cuenta los atributos fecha y hora, temperatura, humedad, velocidad del viento, flujo difuso general, flujo difuso y tres atributos de consumo de las zonas 1, 2 y 3.

Aunque se consideran distintos modelos, se hace énfasis en redes neuronales LSTM que funcionan bastante bien con series temporales, se tomó dicha decisión debido a que el consumo de energía depende mucho de la estación climática o temporadas especiales como navidad.

## II. ESTADO DEL ARTE

Se han encontrado autores que abordan el conjunto de datos del consumo de energía de la ciudad marroquí Tetouan. Uno de ellos es Waleed Abdu Zogaan[2] que propone el uso de métodos de Machine Learning como Random Forest, Máquinas de soporte vectorial y redes neuronales artificiales.

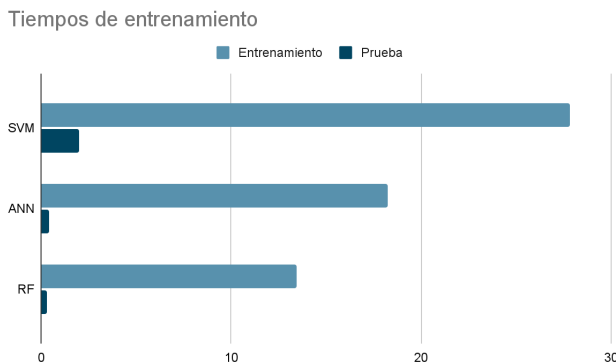
en la siguiente tabla se muestran los resultados obtenidos por el autor para cada uno de los modelos utilizados

Tabla 1 - Resultados para diferentes métricas de error

Model	MAE	MSE	R <sup>2</sup>
SVM	0.7073	0.1952	0.9765
ANN	0.8441	13.5639	0.7335
RF	0.2166	0.0987	0.9013

- Teniendo en cuenta dichos resultados se logra evidenciar que el modelo RF obtuvo el error más bajo tanto en MAE como en MSE pero las máquinas de soporte vectorial SVM obtuvieron mejor R<sup>2</sup>.

Figura 1 - Tiempos de entrenamiento



En la figura anterior se puede evidenciar que el modelo con menor costo durante el entrenamiento es RF.

Los otros autores en cuestión son Abdulwahed Salam y Abdelaziz El Hibaoui [3], dichos autores proponen un modelo de deep learning haciendo variaciones en las redes neuronales utilizadas, obteniendo como resultado los siguientes datos:

Tabla 2 - Medidas obtenidas por algoritmo

Model	Median RMSE	Ensemble RMSE
DFNN	7208	6611.7
DFNN-ResNet	7397	7071.3
CNN	7191.9	7079.9
CNN-ResNet	6874	6365.6

CNN LSTM	6744.1	6645.2
CNN-ResNet LSTM	6429.1	6424.6
DFNN LSTM	6547.5	6944.5
DFNN ResNet LSTM	6941.4	6538
DENSENET	10220	9978
DENSENET LSTM	7443.9	7238.3
EECP-CBL	8146.2	6604.97

### III. EXPERIMENTOS

La base de datos que se está analizando representa el consumo de energía en tres zonas de la ciudad de Tetouan [1] ubicada al norte de Marruecos, dicha base de datos cuenta con un total de 52,417 instancias.

El problema a resolver es multivariable, cuenta con un total de 9 variables continuas de las cuales 6 de ellas son independientes y corresponden a los datos de entrada y tres de ellas son dependientes y corresponden a los datos de salida.

Tabla 3 - Variables de entrada del modelo

ATRIBUTOS	DESCRIPCIÓN
<b>DateTime</b>	Representa la fecha y la hora en la que fueron tomadas las muestras. (La recolección fue realizada cada 10 minutos)
<b>Temperature</b>	Temperatura de la ciudad de Tetouan.
<b>Humidity</b>	Humedad de la ciudad Tetouan.
<b>Wind Speed</b>	Velocidad del viento de la ciudad de Tetouan.
<b>general diffuse flows</b>	Flujo difuso general
<b>diffuse flows</b>	Flujo difuso

Tabla 4 - Variables de salida del modelo

VARIABLE	DESCRIPCIÓN
<b>power consumption of zone 1 of Tetouan city</b>	Consumo de energía para la zona 1
<b>power consumption of zone 2 of Tetouan city</b>	Consumo de energía para la zona 2
<b>power consumption of zone 3 of Tetouan city</b>	Consumo de energía para la zona 3

La metodología de validación elegida es la “validación cruzada” (K-Fold), se descarta la validación simple porque los datos destinados para pruebas no son tenidos en cuenta para el entrenamiento, prescindiendo de datos importantes para mejorar el modelo, de igual forma se descarta K-Fold estratificado debido a que nuestro problema no corresponde a un problema de clases desbalanceado, y por último se descarta “K-Fold leave one out” debido a su excesivo coste durante el entrenamiento.

El numero de particiones usadas para la validación cruzada fue de tres, se optó por este valor con el fin de tener un equilibrio entre el tiempo que toma el entrenamiento y la eficiencia del modelo resultante; dicha configuración fue aplicada para todos los algoritmos usados durante el entrenamiento.

El problema en cuestión es un problema de regresión por tanto no se presenta el inconveniente de las clases desbalanceadas.

A continuación se describen algunas medidas descriptivas de las características que se usaran para el entrenamiento de los modelos.

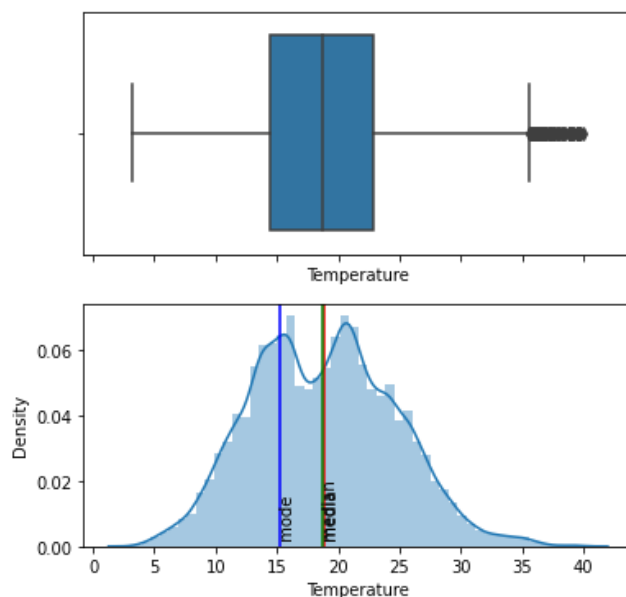
Tabla 5 - Análisis de los características

VARIABLE	RANGO	MEDIA	MODA	MEDIANA	INTERCUARTIL
Temperatura	83.46	68.25	85.9	69.86	8.48
Humedad	83.46	68.25	85.9	69.86	23.09
Velocidad del viento	64.433	1.69	0.08	0.09	4.837
Flujo difuso general	1162.99	6	0.05	05.03	319.53
Flujo difuso	935.98	75.02	0.11	4.45	100.87

En las siguientes figuras se muestran los diagramas de caja de bigotes y la FDP de las variables Temperatura, Humedad y velocidad del viento.

Figura 2: Caja de bigotes y FDP de la variable temperatura

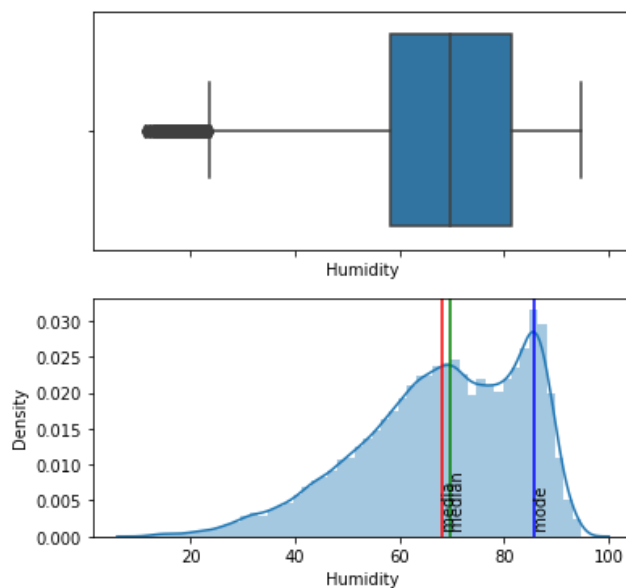
Descripción estadística de la variable Temperature



Para la variable temperatura la distribución es bimodal, en este caso no se presentan outliers representativos.

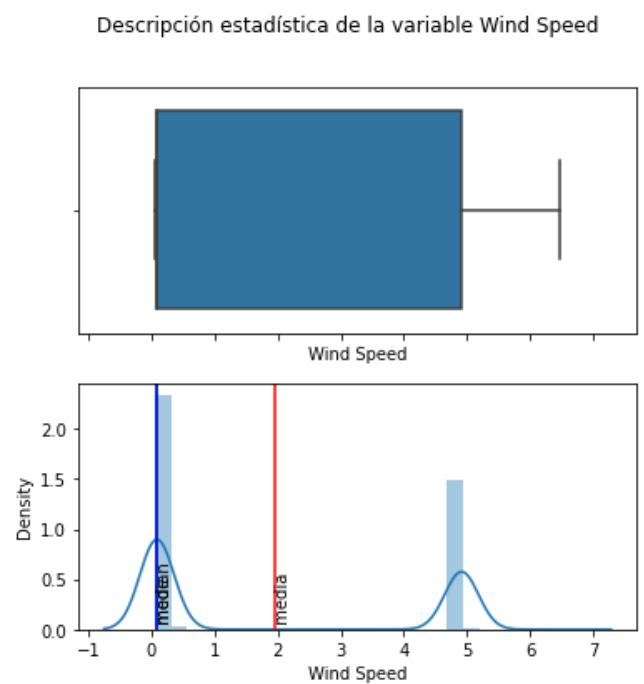
Figura 3: Caja de bigotes y FDP de la variable humedad

Descripción estadística de la variable Humidity



Para la variable humedad la distribución es asimétrica negativa y hay una gran cantidad de outliers a la izquierda, representando humedades extremadamente bajas, posible consecuencia de un aumento en la temperatura.

Figura 4: Caja de bigotes y FDP de la variable velocidad del viento



Para la variable velocidad del viento el tipo de distribución es bimodal, con una moda muy cerca de "0" y otra moda muy cerca de "5". En este caso no cuenta con outliers.

Medidas de desempeño

Entre las medidas de desempeño usadas se encuentran:

**R<sup>2</sup>:** usada con el fin de determinar el porcentaje de varianza que puede explicar el modelo (medida principal de desempeño).

**MAPE:** se usa MAPE en lugar de MAE con el fin de obtener una medida en porcentaje del error dado por el modelo.

**Intervalo de confianza:** dado por la desviación estándar del MAPE, nos ayuda a identificar qué tanta variación existe entre las medidas de error del modelo.

**Duración del entrenamiento:** Indica la cantidad de minutos que tomó el modelo en entrenar completamente dado un conjunto específico de parámetros.

IV. MODELOS DE APRENDIZAJE

Por tratarse de un problema de regresión, en el cual se trata de predecir el consumo de energía en tres zonas de la ciudad de Tetouan, la base de datos será abordada por los

modelos de aprendizaje regresión múltiple, ventana de parzen, random forest, redes neuronales artificiales y regresión por vectores de soporte con kernel RBF

Regresión múltiple

Tabla 6 - Resultados entrenamiento para regresión múltiple

ZONA	R <sup>2</sup>	MAPE	INTERVALO DE CONFIANZA
1	0.523610	0.125662	0.012095
2	0.327768	0.145760	0.013714
3	-1.964040	0.391489	0.238354

El modelo de regresión múltiple en específico no contaba con hiperparámetros, por tanto al no tener variaciones en los resultados de entrenamiento se tiene que los mejores resultados de R<sup>2</sup> para cada una de las zonas fueron:

Tabla 7 - Mejor resultado de entrenamiento para regresión

	Zona 1	Zona 2	Zona 3
R <sup>2</sup>	0.52	0.32	-1.96

Se infiere de dichos resultados que el modelo con mayor porcentaje de varianza explicada es el modelo de la zona 1.

Ventana de parzen

Tabla 8 - Resultados entrenamiento para ventana de parzen

ZON A	H	R <sup>2</sup>	MAPE	I.C	DURACIÓ N
1	1	0.507809	0.129746	0.015187	20.3779796
	2	0.214197	0.167071	0.010042	214.460829
	3	0.038492	0.184848	0.011217	206.747262
	5	0.195756	0.195756	0.011346	202.737921
2	1	0.459666	0.140867	0.000608	20158059
	2	0.097628	0.189485	0.031842	213.262206
	3	-0.115388	0.212893	0.043641	208.042370
	5	-0.255256	0.226903	0.048833	209.088500
3	1	-2.48606	0.414835	0.28658	204.211580

		4			
	2	-1.94215 2	0.409319	0.230391	207.747027
	3	-1.82599 9	0.403060	0.213083	201.266425
	5	-1.77150 6	0.399480	0.202722	200.833213

Analizando los hiperparametros se observa que el mejor resultado de R2 se obtiene cuando H es igual a 1:

Tabla 9 - Mejores resultados para ventana de parzen

ZONA	H	R2
1	1	0.50
2	1	0.45
3	1	-2.48

## Random Forest

Tabla 10 - Resultados usando random forest

Z O N A	M A X- C A R	N- E S T I M A D O R	R <sup>2</sup>	MAPE	I-C	DUR A C I O N
1	2.0	5	0.950375	0.034587	0.000522	0.014
	3.0	5	0.963139	0.029403	0.000186	0.018
	4.0	5	0.965296	0.028320	0.000139	0.022
	5.0	5	0.967137	0.027528	0.000240	0.024
	2.0	10	0.961768	0.030733	0.000140	0.029
	3.0	10	0.968551	0.027576	0.000215	0.036
	4.0	10	0.970446	0.026341	0.000355	0.041
	5.0	10	0.971673	0.025699	0.000322	0.048
	2.0	15	0.965332	0.029459	0.000140	0.041 372
	3.0	15	0.971065	0.026544	0.000227	0.052

						927
	4.0	15	0.972195	0.025527	0.000338	0.063 782
	5.0	15	0.973404	0.024978	0.000073	0.072 232
	2.0	20	0.967088	0.028638	0.000381	0.056 357
	3.0	20	0.971107	0.026456	0.000139	0.071 063
	4.0	20	0.973366	0.025058	0.000184	0.083 253
	5.0	20	0.973956	0.024669	0.000130	0.096 852
	2.0	5	0.953145	0.038705	0.000515	0.014 051
	3.0	5	0.962816	0.033793	0.000072	0.017 682
	4.0	5	0.967443	0.031125	0.000418	0.021 346
2	5.0	5	0.969107	0.030184	0.000176	0.025 059
	2.0	10	0.962952	0.034602	0.000692	0.028 541
	3.0	10	0.970313	0.030375	0.000152	0.034 979
	4.0	10	0.972724	0.028672	0.000089	0.041 454
	5.0	10	0.973706	0.027990	0.000200	0.048 595
	2.0	15	0.967351	0.032872	0.000259	0.043 113
	3.0	15	0.972802	0.029362	0.000194	0.051 396
	4.0	15	0.974256	0.027871	0.000244	0.061 788
	5.0	15	0.974991	0.027386	0.000420	0.072 163
	2.0	20	0.967677	0.032652	0.000075	0.054 899
	3.0	20	0.973644	0.028810	0.000179	0.068 217
	4.0	20	0.975382	0.027442	0.000026	0.081 570
	5.0	20	0.976194	0.026663	0.000063	0.093 834
	2.0	5	0.967523	0.044938	0.002150	0.014
	3.0	5	0.976558	0.037859	0.000219	0.017

	4.0	5	0.978745	0.035328	0.000420	0.021
	5.0	5	0.980440	0.033861	0.000417	0.024
	2.0	10	0.975204	0.040381	0.001051	0.028
	3.0	10	0.980331	0.034599	0.000415	0.034
	4.0	10	0.982418	0.032714	0.000501	0.041
	5.0	10	0.983030	0.031957	0.000264	0.047
	2.0	15	0.978017	0.037860	0.000747	0.040
	3.0	15	0.981776	0.033687	0.000542	0.051
	4.0	15	0.983006	0.031930	0.000079	0.060
	5.0	15	0.983805	0.031070	0.000089	0.070
	2.0	20	0.978611	0.037317	0.000203	0.057
	3.0	20	0.982646	0.033044	0.000282	0.073
	4.0	20	0.983944	0.031326	0.000235	0.084
	5.0	20	0.984315	0.030803	0.000371	0.100

Con base en los resultados se observa que la mejor configuración de hiperparametros por zona utilizando random forest corresponde a:

Tabla 11 - Mejores resultados con Random Forest

Zona	MAX-CAR	N-ESTIMA DOR	R <sup>2</sup>
1	5	20	0.973956
2	5	20	0.976194
3	5	20	0.984315

## Máquina de soporte vectorial

Tabla 12 - Resultados usando SVM

Zona	KERNEL	GAMMA	R <sup>2</sup>	MAP E	I-C	DURACIÓN
1	RBF	Auto	0.287317	0.156596	0.001114	5.737755
	RBF	Scalar	0.287669	0.156565	0.001054	5.916860
	Lineal	Auto	0.627	0.107	0.000	2.130

2			081	648	731	983
	Lineal	Escalar	0.626872	0.107682	0.000142	2.131243
	Poly	Auto	0.140145	0.159721	0.001615	3.050227
	Poly	Escalar	0.240263	0.159724	0.000641	3.027026
	RBF	Auto	0.332531	0.167998	0.001421	5.886628
	RBF	Escalar	0.332424	0.168060	0.000568	5.932536
	Lineal	Auto	0.586628	0.128606	0.001422	2.12.8587
	Lineal	Escalar	0.586539	0.128620	0.001110	2.127127
	Poly	Auto	0.269322	0.176996	0.001037	2.296549
	Poly	Escalar	0.269201	0.176997	0.001098	2.592007
	RBF	Auto	0.158605	0.246720	0.000517	5.951777
	RBF	Escalar	0.158588	0.246695	0.000844	5.944983
	Lineal	Auto	0.538520	0.187995	0.007033	2.099771
	Lineal	Escalar	0.538536	0.187960	0.001000	2.239953
	Poly	Auto	0.186008	0.248367	0.001946	2.391817
	Poly	Escalar	0.186110	0.248369	0.000574	2.298353

Tabla 13 - Mejores resultados usando SVM

ZONA	KERNEL	GAMMA	R <sup>2</sup>
1	Linear	Auto	0.627081
2	Linear	Auto	0.586628
3	Linear	Escalar	0.868536

Se observa que en general al aplicar “Linear” como kernel y “Auto” como Gamma se obtienen los mejores resultados..

## Redes neuronales

Tabla 14 - Resultados con redes neuronales

Zo na	F. Acti vac ión	Neu rona s	Ca pas ocu ltas	R <sup>2</sup>	MAP E	I.C	Dura ción
1	Logi stica	8.0	1.0	-21.34 3490	0.984 538	0.000 838	0.555 211
	Logi stica	8.0	3.0	-21.45 940	0.984 596	0.000 852	1.113 646
	Logi stica	16.0	1.0	-20.78 0479	0.970 786	0.001 576	0.716 479
	Tanh	8.0	1.0	-21.33 8900	0.984 428	0.000 833	-0.42 1413
	Tanh	8.0	3.0	-21.34 2160	0.984 506	0.000 850	0.776 151
	Tanh	16.0	1.0	-20.77 1740	0.970 568	0.001 629	0.492 319
	Tanh	16.0	3.0	-20.77 9886	0.970 770	0.001 613	0.992 903
	Relu	8.0	1.0	-1.248 885	0.265 802	0.048 105	0.423 094
	Relu	8.0	3.0	0.167 610	0.165 239	0.065 884	0.813 376
	Relu	16.0	1.0	-0.123 275	0.186 627	0.047 048	0.479 063
	Relu	16.0	3.0	-0.073 638	0.182 790	0.076 057	1.000 308
	Logi stica	8.0	1.0	-17.80 4015	0.975 909	0.002 139	0.533 154
2	Logi stica	8.0	3.0	-17.80 6879	0.975 990	0.002 112	1.051 07
	Logi stica	16.0	1.0	-17.07 3256	0.954 478	0.004 017	0.712 295
	Logi stica	16.0	3.0	-17.08 1727	0.954 730	0.004 054	1.659 629
	Tanh	8.0	1.0	-17.79 7045	0.975 707	0.002 130	0.421 433
	Tanh	8.0	3.0	-17.80 2543	0.975 865	0.002 142	0.773 387
	Tanh	16.0	1.0	-17.06	0.954	0.004	0.004

3				2824	169	054	054
	Tanh	16.0	3.0	-17.07 4277	0.954 508	0.003 990	0.994 695
	Relu	8.0	1.0	-0.161 953	0.200 959	0.019 011	0.425 889
	Relu	8.0	3.0	0.063 973	0.188 824	0.021 919	0.741 422
	Relu	16.0	1.0	0.123 599	0.182 847	0.017 538	0.474 941
	Relu	16.0	3.0	0.107 827	0.182 847	0.017 538	0.474 941
	Logi stica	8.0	1.0	-12.44 6213	0.969 766	0.006 634	0.535 273
	Logi stica	8.0	3.0	-12.44 7620	0.969 823	0.006 609	1.112 100
	Logi stica	16.0	1.0	-11.80 1881	0.942 820	0.012 597	0.714 911
	Logi stica	16.0	3.0	-11.81 1329	0.943 209	0.012 597	1.661 174
	Tanh	8.0	1.0	-12.43 9291	0.969 484	0.006 689	0.421 408
	Tanh	8.0	3.0	-12.44 4032	0.969 670	0.006 635	0.779 339
	Tanh	16.0	1.0	-11.79 1182	0.942 378	0.012 649	0.489 715
	Tanh	16.0	3.0	-11.80 2997	0.942 870	0.012 564	0.990 024
	Relu	8.0	1.0	-4.011 914	0.462 752	0.324 172	0.423 679
	Relu	8.0	3.0	-2.272 303	0.392 556	0.256 147	0.797 451
	Relu	16.0	1.0	-2.862 484	0.435 840	0.300 254	0.482 604
	Relu	16.0	3.0	-3.719 198	0.451 383	0.340 725	0.919 624

Tabla 15 - Mejores resultados usando Redes neuronales

ZONA	F. ACTIVA CIÓN	NEURO NAS	CAPAS OCULTA S	R <sup>2</sup>
1	Relu	8.0	3.0	0.167610
2	Relu	8.0	3.0	0.063973
3	Relu	8.0	3.0	-2.272303

En general se observa un resultado poco óptimo para R<sup>2</sup> con el uso de redes neuronales, sin embargo se encuentra que usando “Relu” como función de activación, 8 neuronas y 3 capas ocultas se obtiene el mejor resultado posible.

## V. CONCLUSIÓN

El problema en cuestión es bastante interesante ya que es de gran envergadura por ser la energía y su administración algo inherente a una sociedad entera. En base al estado del arte se tiene cierta predisposición a redes neuronales, sin embargo se encuentra que Random Forest nos dio los mejores resultados con un porcentaje de variabilidad explicada R<sup>2</sup> de 0.97, un porcentaje de error MAPE de 0.03 y un I.C de 0.0001

## REFERENCES

- [1] Machine Learning Repository “Power consumption of Tetouan city Data Set” Dic 2018  
<https://archive.ics.uci.edu/ml/datasets/Power+consumption+of+Tetouan+city>
- [2] W. A. Zogaan, “Power Consumption prediction using Random Forest model”, Vol 7, pp 329 - 341,  
[https://kalaharijournals.com/resources/Special\\_Issue\\_April\\_May\\_43.pdf](https://kalaharijournals.com/resources/Special_Issue_April_May_43.pdf)
- [3] A. Salam, A. El Hibaoui, "Energy consumption prediction model with deep inception residual network inspiration and LSTM", May. 2021  
<https://www.sciencedirect.com/science/article/pii/S0378475421001774>