

---

## Pronóstico de variables monitoreadas con enfoque *Data Mining* y *Big Data*

---

Vanderley Espínola<sup>1</sup>, Enrique Chaparro<sup>1</sup>, Eustaquio Martínez<sup>2</sup>

<sup>1</sup>Itaipu Binacional, <sup>2</sup>Facultad Politécnica de la UNE (FPUNE)

Paraguay

### RESUMEN

La aparición de nuevas tecnologías en el campo de las telecomunicaciones y la computación ha dado origen al concepto de *Big Data* que por definición posee las siguientes características: gran volumen de datos estructurados (hojas de Excel, formularios web, etc.) y no estructurados (audio, imágenes, video, pdf, datos de redes sociales, etc.), variedad de formatos a procesar y combinar, y la necesidad de gran velocidad de procesamiento de datos.

En los sistemas eléctricos de potencia (SEP) las variables generan datos que llegan a los centros de operación y control también con las características del *Big Data*. Una de las variables más significativas en los SEP es el comportamiento de la carga o de la *Demanda de Energía Eléctrica* (DEE). El comportamiento de la DEE depende de factores internos o inherentes y externos o independientes al SEP [1]. Estos factores contribuyen para determinar la energía eléctrica demandada a cada instante por los clientes.

La contribución de los factores internos y externos sobre la carga no sigue el principio de superposición, es decir, no obedece a una combinación lineal de variables independientes. Sino que el problema de pronóstico de DEE es un problema no lineal, con varias correlaciones, interacciones e interdependencias entre sus factores. En ese sentido, el pronóstico con técnicas de minería de datos (*Data Mining*, DM) es un campo de gran interés para ayudar resolver ese problema no lineal [1]. La DM es un campo de estudio que permite analizar, explorar y descubrir patrones en grandes volúmenes de datos, con el fin de generar información para la toma de decisiones [2]. En otras palabras, DM permite tornar datos disponibles para los operadores de modo a convertir la avalancha actual de datos (*Big Data*), en informaciones útiles para la toma de decisiones.

De esa manera, el presente trabajo explora por una parte el potencial de las técnicas del *Big Data* para resolver la necesidad de procesar gran variedad y volumen de datos de manera confiable y a gran velocidad; y, por otra parte, explora el potencial de las técnicas de la minería de datos para realizar el pronóstico de variables monitoreadas ya sea que los datos sean estructurados o no estructurados.

Para la prueba de concepto y validación de la propuesta se utilizan fuentes de datos abiertos del área de finanzas. A partir de los resultados obtenidos se realiza un paralelismo de su potencial aplicación para pronosticar la DEE utilizando técnicas de minería de datos en el contexto del *Big Data*. Así, el trabajo surge como una nueva alternativa para transformar los datos en informaciones útiles para la toma eficiente de decisiones en el sector eléctrico. El siguiente aporte de las técnicas de *Big Data* y *Data Mining*, que ya está siendo elaborado y es propuesto en este artículo, será la aplicación directa del procedimiento propuesto para la predicción de la DEE en un sistema académico y en contexto real.

### PALABRAS CLAVES

*Pronóstico, Natural Language Processing, Data Mining, Text Mining, Big Data, Inteligencia Artificial.*

XIV SEMINARIO DEL SECTOR ELECTRICO PARAGUAYO - CIGRE  
23 y 24 de Junio 2022

## 1. INTRODUCCIÓN Y CONTEXTUALIZACIÓN

### 1.1 Pronóstico de Demanda de Energía Eléctrica (DEE)

El pronóstico de la DEE es un procedimiento sistemático que permite definir cuantitativamente la demanda futura que puede ser representada en términos de energía (en unidades de Wh – *Watt hora*) y en términos de potencia (en unidades de W). Dependiendo del periodo de tiempo adoptado, el pronóstico puede ser de muy corto plazo (minutos a 1 hora), corto plazo (horas, días, semanas), mediano plazo (2 semanas a 3 años) y largo plazo (3 a 50 años) [3]. La importancia del pronóstico de DEE radica en que los *costos de errores de previsión* pueden llegar a ser significativos [4]. En ese sentido, de acuerdo con [5], la economía en los costos operacionales es altamente sensible a los errores de pronóstico. Algunas consecuencias de los errores de pronóstico se muestran a continuación [6], [7].

Si el pronóstico es mayor que la demanda real (*overforecasts*):

- Las unidades generadoras (UG) pudieron haber sido iniciadas sin necesidad, incrementando los costos de producción;
- La energía pudo haber sido contratada sin necesidad, a precios mayores;
- El agua pudo haber sido desperdiciada en la producción de energía hidroeléctrica no utilizada.

Si el pronóstico es menor que la demanda real (*underforecasts*):

- Pueden faltar recursos para atender las necesidades energéticas;
- Puede ser necesario comprar energía en el mercado, o apelar para unidades generadores más costosas;
- La energía producida puede tener una menor fijación de precio.

Según [8], [9], [10], los factores que influyen en el pronóstico de demanda de energía eléctrica son:

- Datos históricos de la demanda;
- Densidad de carga;
- Planes de desarrollo urbano;
- Producto Interno Bruto (PIB);
- El clima;
- Factores geográficos;
- Tarifas de energía eléctrica;
- Hábitos de consumo;
- Planes industriales;
- Factores demográficos;
- Planes de expansión de energías renovables.

Cada uno de estos factores a su vez pueden estar asociados a variables internas o externas. Las variables internas o variables endógenas son aquellas inherentes al propio sistema eléctrico y son dependientes de otras variables; por otra parte, las variables externas o exógenas son aquellas ajenas al sistema eléctrico y en general no dependen de otras variables [11]. La variable demanda de energía por ejemplo es endógena con respecto al Producto Interno Bruto, ya que depende de la variable PIB. Si se tiene un buen año con un buen crecimiento del PIB es de esperar que la DEE futura se incremente.

Entre las variables exógenas se pueden citar las variables climáticas (como la temperatura, radiación solar, humedad, velocidad del viento, nubosidad, duración de la luz solar, precipitaciones, etc.). En el caso del incremento de la variable temperatura, es de esperar un aumento de la DEE. Luego, la variable temperatura es exógena porque es externa al SEP y no depende de la demanda de energía eléctrica.

**XIV SEMINARIO DEL SECTOR ELECTRICO PARAGUAYO - CIGRE**  
**23 y 24 de Junio 2022**

La cantidad de variables internas y externas a utilizar depende de diferentes factores como la correlación con la DEE, el horizonte de pronóstico o el plazo seleccionado, el nivel de detalle y precisión que se desea lograr, la disponibilidad y fiabilidad de los datos, y las herramientas computacionales con las que se cuenta para el procesamiento eficiente de los datos. Cada factor se puede subdividir prácticamente de forma indefinida, ya que cada variable también es afectada por otras variables internas y externas. Cabe entonces al diseñador del sistema de pronóstico seleccionar los factores y variables más adecuadas a utilizar.

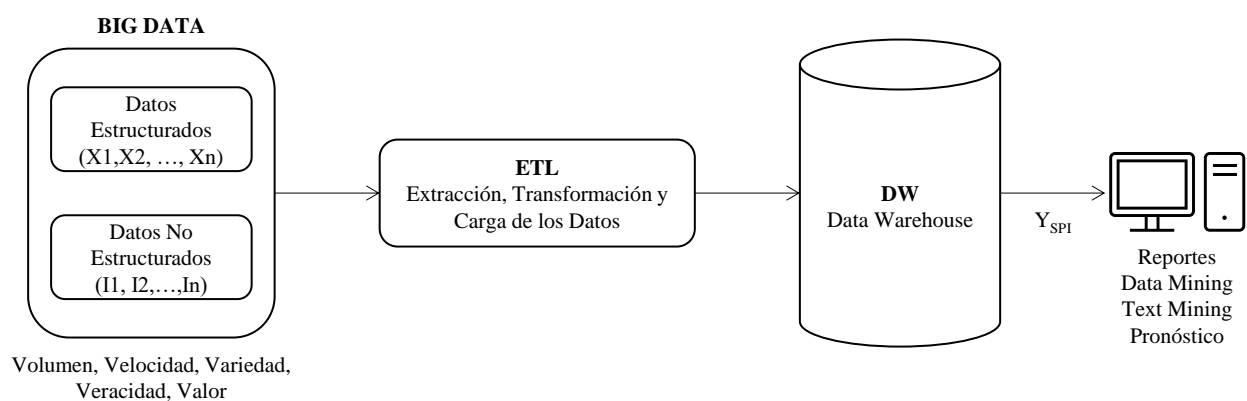
Los datos sobre la variable DEE se presentan en forma de series temporales. El pronóstico de los valores futuros de esas series puede ser realizado a través de diferentes métodos, técnicas y abordajes. Están los métodos basados en modelos univariantes, en que la DEE de un instante futuro es función del histórico de la DEE, y donde el efecto de las variables exógenas se considera implícito en los datos; o multivariantes, en que la DEE futura es función del histórico de la DEE y de otras variables exógenas, principalmente las relacionadas al clima y a factores socio-económicos [12].

Los modelos multivariantes más comunes son los modelos de regresión lineal, y aquellos en que se descomponen la carga en un componente básico lineal y en un componente no lineal dependiente de las variables exógenas. Para la previsión del componente no lineal, gran parte de la literatura utiliza técnicas de Inteligencia Artificial [13], [7].

Algunos modelos utilizados en la literatura para el pronóstico de la DEE son: ARIMAX [14], los basados en técnicas de DM como Regresión Lineal [15], *Support Vector Machine* [16], Redes Neuronales Artificiales (RNA) [13], [17], los basados en Lógica Difusa [18], y en modelos híbridos que utilizan algoritmos de agrupamiento para combinar datos con características similares [19], [20]. En este trabajo se propone realizar el pronóstico combinando técnicas (modelo híbrido) de *Data Mining* en el contexto del *Big Data*.

## 1.2 Arquitectura Propuesta

A continuación, se muestra la arquitectura propuesta.



**Figura 1:** Arquitectura propuesta.

En la **Figura 1**, en el contexto del *Big Data*, los datos son extraídos, transformados y cargados en el *Data Warehouse* (Almacén de Datos). Posteriormente se aplican las técnicas de DM y de *Minería de Textos* para realizar el pronóstico. El sistema también permite generar reportes. El análisis de datos en el contexto del *Big Data* exige la necesidad de herramientas computacionales apropiadas. Para atender esa necesidad, surgió el área del conocimiento denominado KDD (*Knowledge Discovery in Databases*). La DM es la principal etapa del proceso KDD y envuelve la aplicación de algoritmos sobre los datos en busca de conocimientos implícitos y útiles [21].

## XIV SEMINARIO DEL SECTOR ELECTRICO PARAGUAYO - CIGRE 23 y 24 de Junio 2022

La *Minería de Textos (Text Mining, TM)*, también forma parte del proceso KDD y de la DM, y describe una gama de tecnologías para analizar el procesamiento de datos de textos semiestructurados y no estructurados [22]. El tema unificador detrás de cada una de estas tecnologías es la necesidad de "convertir textos en números" para poder aplicar algoritmos a grandes bases de datos de documentos y así identificar patrones o analizar sentimientos asociados al texto (*Data Analysis*).

El TM es una nueva forma de aprendizaje automático para extraer la polaridad de la opinión (positiva, negativa, neutral) de un segmento de texto escrito para cualquier producto, organización, persona o entidad. El TM se puede utilizar para predecir el estado de ánimo de las personas, por lo que puede ayudar a predecir el comportamiento de variables en el tiempo [22].

El Procesamiento en Lenguaje Natural (*Natural Language Processing, NLP*) consiste en transformar el lenguaje natural en un lenguaje formal, por ejemplo, el de la programación, para que los procesadores puedan procesar esa información. Esta propiedad convierte el NLP en una herramienta para el análisis de textos. De manera que, el NLP es una metodología para hacer TM, al permitir estructurar la información en texto de la que disponemos y facilitar su análisis y la identificación de patrones que son recursos propios del DM [23]. En este trabajo se utilizan técnicas de DM y TM para el análisis y procesamiento de los datos.

## 2. OBJETIVOS

### 2.1 Objetivo General

Evaluar el potencial de la minería de datos para el pronóstico combinado de variables monitoreadas en el contexto del *Big Data*.

### 2.2 Objetivos Específicos

- Seleccionar las variables a ser analizadas.
- Aplicar el proceso ETL para obtener información de las variables monitoreadas.
- Utilizar técnicas de DM para el pronóstico de las variables seleccionadas en el contexto del *Big Data*.
- Evaluar los resultados de los pronósticos obtenidos.
- Proponer la metodología propuesta al pronóstico de la DEE en el contexto del *Big Data*.

## 3. METODOLOGÍA

Para el desarrollo de la *prueba de concepto* descrito en la **Figura 1**, se ha seleccionado el mercado bursátil del área de finanzas por contar con datos abiertos que son de uso libre y público. El trabajo analiza el comportamiento del precio de acciones en el mercado bursátil a partir de tres variables de entrada (X1, I1, I2) obtenidas empleando técnicas de *Web Scraping* y APIs (*Application Programming Interface*) de datos abiertos; y una variable de salida ( $Y_{SP1}$ ).

La variable X1 (datos estructurados) representa datos históricos, donde los indicadores son: El valor en dólares de la *acción* en función del tiempo y el volumen de transacciones realizadas. Una vez extraídos, transformados y cargados los datos de entrada (proceso ETL), se aplican técnicas de DM y TM para realizar el pronóstico de los precios de las acciones. La variable I1 (datos no estructurados) representa las opiniones financieras extraídas de la API de *Twitter* acerca de hilos de comentarios sobre empresas asociadas a los portafolios analizados. La polaridad de opiniones sobre las empresas puede ser neutra, positiva o negativa. La variable I2 (datos no estructurados) representa datos extraídos de informes financieros de las empresas analizadas, en la WEB; la expectativa asociada a variables financieras de la empresa puede ser optimista, pesimista o neutra.

**XIV SEMINARIO DEL SECTOR ELECTRICO PARAGUAYO - CIGRE**  
23 y 24 de Junio 2022

La variable de salida  $Y_{SPI}$  (*Sistema de Pronóstico Integrado, SPI*), representa el resultado de combinar los tres pronósticos que tienen como base de datos en cada caso a las variables  $X1$ ,  $I1$  e  $I2$ , respectivamente. Para el proceso de minería de datos, minería de textos, ETL y generación de reportes se utiliza el *software* Pentaho (plataforma de inteligencia empresarial orientada al análisis y procesamiento en el contexto del *Big Data*).

Existen diversos algoritmos que pueden ser utilizados para pronóstico de series temporales de datos financieros, en este trabajo se han evaluado ARIMA (*Autoregressive Integrated Moving Average*) [24], ANN (*Artificial Neural Network*) [25], y el algoritmo *Prophet*. Se ha optado por utilizar los resultados del algoritmo *Prophet* por mostrar mejor desempeño para el pronóstico de los datos financieros [26]. El algoritmo de *Prophet* se basa en un modelo aditivo que puede tener tendencias tanto lineales como no lineales y tiene ventajas con respecto a otros métodos de pronóstico por: ser de código abierto, poder manejar valores faltantes, ser rápido y preciso y poder manejar oscilaciones repentinas en series de tiempo [26]. La expresión general para pronóstico con el algoritmo de *Prophet* se muestra a continuación:

$$Y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

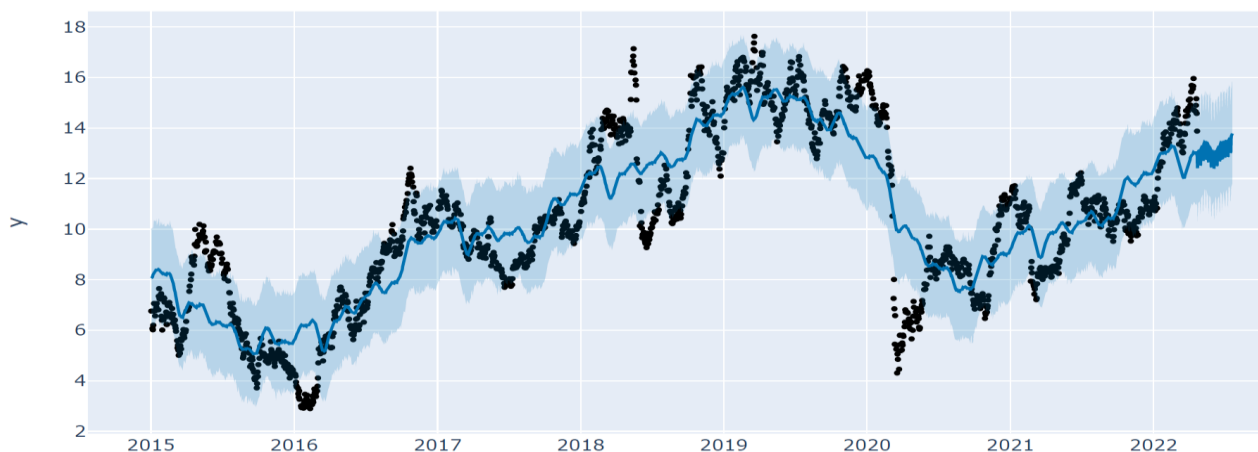
Donde:

- $y(t)$ : pronóstico al que llamaremos *Prophet*
- $g(t)$ : tendencia lineal o logística (cambios no periódicos)
- $s(t)$ : estacionalidad (cambios periódicos)
- $h(t)$ : efecto de oscilaciones o eventos repentinos en series de tiempo
- $\epsilon_t$ : error residual

Para validación de la *prueba de concepto*, se evaluaron precios de las acciones de las siguientes empresas que cotizan en la bolsa de valores de New York (NYSE): Embraer S.A. (ERJ), Itaú Unibanco (ITUB), Petrobrás (PBR), Ambev S.A. (ABEV), Compañía Paranense de Energía (ELP) y la Electrobrás (EBR). El periodo de análisis elegido corresponde a los últimos 7 años, considerado tiempo suficiente para el análisis del comportamiento del precio de las acciones, pues considera la situación antes y durante la pandemia del COVID-19.

#### 4. RESULTADOS

Los resultados obtenidos se muestran a continuación.



**Figura 2:** Pronóstico del precio de acciones de Petrobrás.

**XIV SEMINARIO DEL SECTOR ELECTRICO PARAGUAYO - CIGRE**  
23 y 24 de Junio 2022

En la **Figura 2** anterior, los valores del eje de ordenadas “Y” están en dólares americanos y representan los valores del precio de acciones de la Petrobrás, mientras que en el eje de abscisas se representa el periodo de análisis en años (2015 a 2022). Los valores pronosticados se muestran con la línea de color azul; mientras que los valores reales del precio de acciones se muestran con puntos de color negro.

Los resultados obtenidos con el *Prophet* se muestran a continuación (*Error Absoluto Medio*, EAM).

**Tabla I** – Pronóstico con *Prophet* a partir de datos extraídos de <https://finance.yahoo.com/>

Fecha	ERJ			ITUB			PBR			ABEV			ELP			EBR		
	Real	Prophet	EAM	Real	Prophet	EAM	Real	Prophet	EAM	Real	Prophet	EAM	Real	Prophet	EAM	Real	Prophet	EAM
4-mar	13,01	17,34	1,92	4,84	4,84	0,47	14,51	12,74	1,27	2,73	3,01	0,26	6,79	6,36	0,55	6,82	7,37	0,83

Para el análisis de sentimiento en *Twitter* se utilizaron bibliotecas de Python como *tweepy* (para rastrear *tweets*), *textblob* (realiza tareas de NLP como análisis de sentimiento), entre otros. Los resultados obtenidos se muestran en la **Tabla II**.

**Tabla II** – Datos extraídos de *Twitter* entre el 28-feb a 04 de mar

ERJ				ITUB				PBR				ABEV				ELP				EBR			
Neutro	-	+	Suma	Neutro	-	+	Suma	Neutro	-	+	Suma	Neutro	-	+	Suma	Neutro	-	+	Suma	Neutro	-	+	Suma
65	12	23	11	100	0	0	0	14	59	27	-32	27	30	43	13	35	7	58	51	39	27	34	7

Para el análisis de sentimiento de noticias financieras fueron utilizadas bibliotecas del Python como *BeautifulSoup* (para extraer información de contenido en formato HTML o XML), *NLTK* (*Natural Language Toolkit*, para análisis de texto) entre otros [27]. Los resultados se muestran en la **Tabla III**.

**Tabla III** – Datos extraídos de <https://finviz.com/> entre 28-feb a 04-mar

ERJ	ITUB	PBR	ABEV	ELP	EBR
0,3306	0,1763	-0,0518	0,4727	0,0000	0,0023

Los valores pronosticados del SPI se obtienen según la siguiente expresión matemática:

$$SPI = Y_{SPI} = Prophet - COM * EAM \quad (2)$$

Para el valor combinado (COM) de los datos no estructurados se aplica *lógica financiera del mercado de valores*. Los resultados pronosticados por el SPI para el día 04 de marzo de 2022 se muestran en la **Tabla IV**.

**Tabla IV** – Resultados del SPI

	Real	Prophet	EAM	SPI	Error = Real - Prophet		Error* = Real - SPI	
					Error	Error (%)	Error*	Error* (%)
ERJ	13,01	17,34	1,92	15,42	-4,33	-33%	-2,41	-19%
ITUB	4,84	4,84	0,47	4,84	0,00	0%	0,00	0%
PBR	14,51	12,74	1,27	14,01	1,78	12%	0,50	3%
ABEV	2,73	3,01	0,26	2,75	-0,28	-10%	-0,02	-1%
ELP	6,79	6,36	0,55	6,36	0,43	6%	0,43	6%
EBR	6,82	7,37	0,83	6,54	-0,55	-8%	0,28	4%

Es posible observar que la metodología propuesta, en el intervalo analizado, empleando técnicas de DM para realizar el pronóstico combinado (SPI), permite reducir el error porcentual, cuando se compara con el pronóstico realizado únicamente con el algoritmo de pronóstico de series temporales (*Prophet*). Además, la propuesta proporciona la estructura necesaria para trabajar de manera eficiente en el contexto del *Big Data*.

XIV SEMINARIO DEL SECTOR ELÉCTRICO PARAGUAYO - CIGRE  
23 y 24 de Junio 2022

## 5. APLICACIÓN EN EL SECTOR ELÉCTRICO

Para proyección adecuada del SEP es necesario el planeamiento y control de la producción en términos de plazo. Decisiones de largo plazo tienen que ver con la ampliación con la capacidad de producción, transmisión y distribución de energía; a mediano plazo, pueden ser resueltas cuestiones sobre utilización de la capacidad productiva para satisfacer una determinada demanda de energía eléctrica; y a corto o muy corto plazo, son tomadas decisiones respecto a la optimización del uso de unidades generadoras e instalaciones [7].

Debido a la dificultad de almacenar la energía eléctrica generada, la producción de energía eléctrica debe atender a la demanda sin que haya desperdicios, y sin provocar falta de abastecimiento. En ese contexto, el pronóstico de la demanda de energía eléctrica es fundamental para el planeamiento de las empresas del sector eléctrico [28], pues permite optimizar la utilización de recursos económicos [7].

Dependiendo de los plazos, algunos de los trabajos revisados como en [20], han utilizado hasta 32 variables de base de datos para realizar el pronóstico de la DEE, como; balance energético de energía, capacidad de generación, velocidad del viento, índice de lluvias, promedio de presión y temperatura anual, humedad relativa anual promedio, tipo de cambio con relación al dólar estadounidense, producto interno bruto, crecimiento poblacional, tasa de crecimiento del nivel de precios, inflación porcentual, variación del salario mínimo, variación de la tasa de interés, tarifas de energía eléctrica aplicadas entre otros.

La metodología presentada propone el pronóstico de la DEE utilizando indicadores socioeconómicos, indicadores climáticos e históricos de demanda de energía eléctrica. En ese sentido es un modelo multivariante. Por otra parte, es un modelo híbrido, ya que combina diferentes técnicas de DM y TM para realizar el pronóstico. En cuanto a los tipos de datos, como la estructura está planteada para el contexto del *Big Data* puede obtener y procesar datos estructurados y no estructurados.

Considerando la arquitectura propuesta en la **Figura 1**, seguidamente se presentan los parámetros y variables que serán utilizados como base de datos para realizar el pronóstico de la DEE:

- Definición del plazo: el sistema puede ser utilizado para pronósticos de muy corto, corto, mediano y largo plazo. Generalmente se utilizan las variables más influyentes para cada plazo de pronóstico seleccionado. La definición del plazo o intervalo de análisis además está condicionada por la disponibilidad de los datos;
- Periodo de los datos históricos de la carga: se requieren valores horarios para abarcar pronósticos a muy corto y corto plazo. El intervalo ideal de datos es del 2000 a 2022. Aunque con una base de datos de 10 años ya se podría pronosticar con buena precisión;
- El proceso ETL de las variables se realiza con la plataforma Pentaho;
- Las variables con fuente de datos estructurados seleccionadas son: histórico de carga eléctrica (X1), temperatura horaria promedio (X2), humedad relativa (X3), velocidad horaria promedio del viento (X4), precipitación (X5), cotización del dólar (X6), PIB (X7), crecimiento poblacional (X8), y tarifas de energía eléctrica (X9);
- Las variables con fuente de datos no estructurados seleccionadas son: inflación (I1), y tasas de interés (I2). Los valores se obtienen de la encuesta sobre expectativas de variables económicas realizada mensualmente por el Banco Central del Paraguay;
- Valores pronosticados  $Y_{SPI}$ : con técnicas de data mining de forma independiente y combinada, posteriormente se evaluará el error absoluto medio (EAM) de los pronósticos realizados. El EAM permitirá validar el modelo a ser utilizado.

La aplicación de la metodología propuesta al pronóstico de la demanda de energía eléctrica se encuentra actualmente en desarrollo.

XIV SEMINARIO DEL SECTOR ELÉCTRICO PARAGUAYO - CIGRE  
23 y 24 de Junio 2022

## 6. CONCLUSIONES

En la *prueba de concepto* y validación de la propuesta, una vez seleccionadas las variables de análisis, se realizó el proceso ETL empleando técnicas de *Web Scraping* y APIs de datos abiertos. Además, se utilizaron técnicas de DM para el pronóstico de las variables financieras monitoreadas, y finalmente se construyó el SPI combinando los resultados de los pronósticos obtenidos.

Fue posible observar que la metodología empleando técnicas de DM para realizar el pronóstico combinado (SPI) permitió reducir el error porcentual en el intervalo analizado, cuando comparado con el pronóstico realizado únicamente con el algoritmo de pronóstico de series temporales. Además, la propuesta proporciona la estructura necesaria para trabajar de manera eficiente en el contexto del *Big Data*. Como un trabajo futuro se propone evaluar la metodología de pronóstico con datos no estructurados de Twitter para intervalos de pronósticos mayores.

Además, se presentó el avance del desarrollo de la propuesta para su aplicación en el pronóstico de la DEE como una herramienta de apoyo a la toma de decisiones en el sector eléctrico, que permitirá definir con mejor precisión en el corto plazo la cantidad de energía a generar, programar de manera eficiente la operación de unidades generadoras, realizar el monitoreo del comportamiento de la red con el fin de identificar puntos críticos que en el futuro podrían ocasionar fallas; y en el mediano y largo plazo permitiría detectar los periodos para los cuales existen excesos y faltantes de capacidad de generación con el fin de planificar de manera más precisa las inversiones en la expansión del sistema eléctrico de potencia.

Finalmente, los principales aportes del trabajo son por una parte mostrar el potencial de las técnicas del *Big Data* para resolver la necesidad de procesar gran variedad y volumen de datos de manera confiable y a gran velocidad; y, por otra parte, mostrar el potencial de las técnicas de la minería de datos para realizar el pronóstico de variables monitoreadas ya sea que los datos sean estructurados o no estructurados.

## BIBLIOGRAFÍA

- [1] C. I. Costa, “Aplicação de técnicas de Big Data à Previsão da Carga Elétrica,” Universidade Federal de Itajubá, 2017.
- [2] R. Goldschmidt, *Data mining: um guia prático*, 4th ed. Rio de Janeiro: Elsevier Editora Ltda, 2005.
- [3] K. Zor, O. Timur, and A. Teke, “A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting,” *2017 6th Int. Youth Conf. Energy, IYCE 2017*, no. May 2018, 2017, doi: 10.1109/IYCE.2017.8003734.
- [4] D. W. Bunn, “Forecasting loads and prices in competitive power markets,” *Proc. IEEE*, vol. 88, no. 2, pp. 163–169, 2000, doi: 10.1109/5.823996.
- [5] D. Bunn and E. D. Farmer, *Comparative models for electrical load forecasting*. John Wiley & Sons Ltd, 1985.
- [6] B. F. Hobbs, “Analysis of the value for unit commitment of improved load forecasts,” *IEEE Trans. Power Syst.*, vol. 14, no. 4, pp. 1342–1348, 1999, doi: 10.1109/59.801894.
- [7] A. C. M. Carneiro, “Previsão do consumo de energia elétrica a curto prazo, usando combinações de métodos univariados,” Universidade Federal de Juiz de Fora, 2014.
- [8] T. Gonen, *Electric Power Distribution Engineering*, 3rd ed. CRC Press, 2014.
- [9] S. Khatoon, Ibraheem, A. K. Singh, and Priti, “Effects of various factors on electric load forecasting: An overview,” *Proc. 6th IEEE Power India Int. Conf. PIICON 2014*, 2014, doi: 10.1109/34084POWERI.2014.7117763.
- [10] M. U. Fahad and N. Arbab, “Factor Affecting Short Term Load Forecasting,” *J. Clean Energy Technol.*, vol. 2, no. 4, pp. 305–309, 2014, doi: 10.7763/jocet.2014.v2.145.
- [11] R. Christen, L. Mazzola, A. Denzler, and E. Portmann, “Exogenous data for load forecasting: A review,” *IJCCI 2020 - Proc. 12th Int. Jt. Conf. Comput. Intell.*, no. Ijcci, pp. 489–500, 2020, doi: 10.5220/0010213204890500.
- [12] S. Sargunraj, D. P. Sen Gupta, and S. Devi, “Short-term load forecasting for demand side management,” *IEE Proc.*



**XIV SEMINARIO DEL SECTOR ELÉCTRICO PARAGUAYO - CIGRE**  
**23 y 24 de Junio 2022**

- 
- Gener. Transm. Distrib.*, vol. 144, no. 1, pp. 68–73, 1997, doi: 10.1049/ip-gtd:19970599.
- [13] H. S. Hippert, C. E. Pedreira, and R. C. Souza, “Neural networks for short-term load forecasting: A review and evaluation,” *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 44–55, 2001, doi: 10.1109/59.910780.
- [14] C. Bennett, R. A. Stewart, and J. Lu, “Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks,” *Energies*, vol. 7, no. 5, pp. 2938–2960, 2014, doi: 10.3390/en7052938.
- [15] S. A. Soliman, S. Persaud, K. El-Nagar, and M. E. El-Hawary, “Application of least absolute value parameter estimation based on linear programming to short-term load forecasting,” *Int. J. Electr. Power Energy Syst.*, vol. 19, no. 3, pp. 209–216, 1997, doi: 10.1016/s0142-0615(96)00048-8.
- [16] B. J. Chen, M. W. Chang, and C. J. Lin, “Load forecasting using support vector machines: A study on EUNITE Competition 2001,” *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, 2004, doi: 10.1109/TPWRS.2004.835679.
- [17] H. S. Hippert, D. W. Bunn, and R. C. Souza, “Large neural networks for electricity load forecasting: Are they overfitted?,” *Int. J. Forecast.*, vol. 21, no. 3, pp. 425–434, 2005, doi: 10.1016/j.ijforecast.2004.12.004.
- [18] V. H. Hinojosa and A. Hoese, “Short-term load forecasting using fuzzy inductive reasoning and evolutionary algorithms,” *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 565–574, 2010, doi: 10.1109/TPWRS.2009.2036821.
- [19] C. López, W. Zhong, and M. Zheng, “Short-term Electric Load Forecasting Based on Wavelet Neural Network, Particle Swarm Optimization and Ensemble Empirical Mode Decomposition,” *Energy Procedia*, vol. 105, pp. 3677–3682, 2017, doi: 10.1016/j.egypro.2017.03.847.
- [20] E. Lopes de Oliveira, “Previsão de demanda de energia elétrica com redes neurais artificiais e análise por série de Fourier,” 2016, [Online]. Available: [http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/BUBD-AN6PSF/dissertacao\\_de\\_mestrado\\_923.pdf?sequence=1](http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/BUBD-AN6PSF/dissertacao_de_mestrado_923.pdf?sequence=1).
- [21] F. Pfeffer, “El concepto de control societario, la administración de la sociedad anónima, los conflictos de interés y la potestad punitiva de la superintendencia de valores y seguros en el contexto del denominado ‘caso Chispas,’” *Rev. Chil. Derecho*, vol. 32, pp. 501–537, 2005.
- [22] G. Miner *et al.*, *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*. Waltham: Elsevier Inc., 2012.
- [23] M. Brojo Kshore and K. Raghvendra, *Natural Language Processing in Artificial Intelligence*, 1st ed. Palm Bay: Apple Academic Press, 2020.
- [24] A. A. Adebisi, A. O. Adewumi, and C. K. Ayo, “Stock price prediction using the ARIMA model,” *Proc. - UKSim-AMSS 16th Int. Conf. Comput. Model. Simulation, UKSim 2014*, vol. 2, no. 2, pp. 106–112, 2014, doi: 10.1109/UKSim.2014.67.
- [25] M. Gurjar, P. Naik, G. Mujumdar, and T. Vaidya, “Stock Market Prediction Using ANN,” *Int. Res. J. Eng. Technol.*, vol. 05, no. 03, pp. 2758–2761, 2018.
- [26] C. R. Madhuri, M. Chinta, and V. V. N. V. P. Kumar, “Stock market prediction for time-series forecasting using prophet upon ARIMA,” *2020 7th Int. Conf. Smart Struct. Syst. ICSSS 2020*, pp. 1–5, 2020, doi: 10.1109/ICSSS49621.2020.9202042.
- [27] R. Mitchell, *Web Scraping com Python*, 2nd ed. São Paulo, 2019.
- [28] H. Hahn, S. Meyer-Nieberg, and S. Pickl, “Electric load forecasting methods: Tools for decision making,” *Eur. J. Oper. Res.*, vol. 199, no. 3, pp. 902–907, 2009, doi: 10.1016/j.ejor.2009.01.062.