

Análisis exploratorio y reducción de dimensionalidad de datos de energía eléctrica de una subestación eléctrica industrial

Eliel Alexis Fonseca-Aguilar

Instituto Politécnico Nacional, Escuela Superior de Ingeniería Mecánica y Eléctrica
efonsecaa1700@alumno.ipn.mx

<https://orcid.org/0009-0004-1890-5331>

Claudia Lizbeth Martínez-González

Instituto Politécnico Nacional, Escuela Superior de Ingeniería Mecánica y Eléctrica
clmartinezg@ipn.mx

<https://orcid.org/0000-0002-1307-0847>

Jaime Robles-García

Instituto Politécnico Nacional, Escuela Superior de Ingeniería Mecánica y Eléctrica
jarobles@ipn.mx

Resumen— La energía eléctrica es un servicio esencial, pero su creciente demanda y uso ineficiente plantean desafíos. La calidad de la energía garantiza estabilidad y pureza en el suministro. En este trabajo se realizó un análisis exploratorio univariante y bivariante, así como la reducción de dimensionalidad de datos con PCA (Principal Component Analysis) como primera etapa para evaluar la calidad energética en instalaciones industriales. Los resultados muestran la alta relación entre los distintos tipos de datos eléctricos, y también la probable reducción de componentes principales, para realizar un análisis completo. Si bien el voltaje del sistema se mantiene estable, las corrientes presentan un desbalance notable, especialmente en una de las fases, lo que señala la necesidad de tomar medidas correctivas para asegurar el cumplimiento de estándares.

Palabras Clave — Análisis de componentes principales (PCA), Aprendizaje de máquina, Calidad de la energía eléctrica, Clasificación, Sistemas eléctricos industriales.

Abstract. - *Electrical energy is an essential service, but its growing demand and inefficient use pose significant challenges. Energy quality ensures stability and purity in the supply. To address this, data science is employed to analyze and improve energy quality in industrial facilities. The data science methodology was applied, extending to data exploration. The results reveal a strong relationship between different types of electrical data and the potential reduction of principal components for a comprehensive analysis. The findings indicate that while the system voltage remains stable, the currents show a notable imbalance, particularly in one of the phases, highlighting the need for corrective measures to ensure compliance with standards.*

Keywords — *Principal Component Analysis (PCA), Machine Learning, Electrical Energy Quality, Classification, Industrial Electrical Systems.*

I. INTRODUCCIÓN

Actualmente, la energía eléctrica se considera un servicio básico para todos los seres humanos, desde el hogar, hasta las grandes industrias. El rápido crecimiento demográfico requiere más recursos y servicios, como la energía eléctrica, ocasionando que su demanda sea mayor y la utilización frecuentemente no tiene un uso óptimo [1].

La calidad de la energía se refiere a la estabilidad y pureza de la energía eléctrica suministrada en un sistema, asegurando que los equipos funcionen de manera eficiente y sin perturbaciones. Incluye factores como la tensión, frecuencia, distorsión armónica y fluctuaciones que pueden afectar el rendimiento de los dispositivos conectados. Una buena calidad de energía minimiza pérdidas, reduce fallos y prolonga la vida útil de los equipos [2].

La penalización por la falta de cumplimiento de la normativa vigente en cuanto a la calidad de la energía eléctrica en México ha ido en incremento, tanto para las empresas que suministran, como para los usuarios que consumen energía eléctrica. En particular los usuarios que consumen más de 100 kilowatts bimestrales corresponden a plazas comerciales, hospitales, parques, centros de esparcimiento, estadios, auditorios, así como centros de producción o fábricas, ensambladoras, madereras, mineras, empresas de producción refresquera, plásticos y cualquier giro que requiera una instalación eléctrica industrial.

En la actualidad para uso industrial se utiliza equipo basado en electrónica de potencia que puede causar disturbios a los sistemas eléctricos, lo cual produce inestabilidad [3].

Se ha desarrollado investigación con ciencia de datos para analizar la calidad de la energía en instalaciones eléctricas industriales, con el fin de identificar disturbios, particularmente con el uso de aprendizaje automático, lo que permitiría disminuir el tiempo y mejorar la precisión del método convencional. En la Tabla 1 se presenta algunos trabajos relacionados.

Tabla 1.- Trabajos relacionados con el análisis de la calidad de la energía con aprendizaje automático

Ref	Problema	Tipo de estudio / datos	Muestra	Métodos	Precisión
[5]	Disturbios en la calidad de la energía eléctrica	Detección y clasificación de disturbios en la calidad de la energía eléctrica.	52,000 valores de tensión	SVMs	95%
[6]	Perturbaciones en la calidad de la energía eléctrica.	Clasificación de las perturbaciones en la calidad de la energía eléctrica de una sola etapa.	85,000 valores de tensión	LDA RF KNN SVM NB	96 %
[7]	Perturbaciones en la calidad de la energía eléctrica.	Clasificación automática de las perturbaciones de la calidad de la energía eléctrica.	147,000 valores de tensión	RNA: LSTM GRU SAE ResNet50	94 %
[8]	Disturbios en la calidad de la energía eléctrica.	Identificación de perturbaciones en la calidad de la energía eléctrica.	115,000 valores de tensión	RNN CNN	95 %

SVM: Máquinas de soporte vectorial (Support Vector Machines); KNN: k-vecinos más cercanos (k-Near neighbors); RF: Bosque aleatorio (Random Forest); LDA: Análisis discriminante lineal (Linear Discriminant Analysis); NB: Clasificador Bayesiano (Naive Bayes); LSTM: Redes neuronales de memoria de corto-largo plazo (Long short-term memory); GRU: Unidades recurrentes cerradas (Gated Recurrent Unit); SAE: Autocodificador (Space Autoencoder); RNN: Red neuronal recurrente (Recurrent Neural Network); ResNet50: Red convolucional con 50 capas; CNN: Red neuronal convolucional (Convolutional Neural Network).

Se analizaron estudios recientes sobre la detección y clasificación de perturbaciones en la calidad de la energía eléctrica utilizando técnicas de aprendizaje automático. Li et al. (2020) [5] y Zhao et al. (2024) [8] abordaron disturbios con métodos como SVMs, KNN, RNN y CNN, logrando precisiones del 95%. Singh et al. (2020) [6] y Wang et al. (2020) [7] se enfocaron en la clasificación de perturbaciones, empleando algoritmos como LDA, RF, LSTM y ResNet50, con resultados del 96% y 94%, respectivamente. Los trabajos futuros incluyen mejorar la velocidad de los modelos, comparar métodos e incorporar datos híbridos los cuales son una combinación entre mediciones física y datos sintéticos.

La primera fase en un problema de ciencia de datos es la exploración de los datos, que consiste en la visualización, el análisis univariante y bivariante, así como la reducción de dimensionalidad. En este trabajo se llevó a cabo esta fase, el análisis exploratorio y la reducción de dimensionalidad de datos eléctricos para el análisis de la calidad de la energía eléctrica en una subestación industrial.

II. METODOLOGÍA/DESARROLLO

Para llevar a cabo la investigación, se requiere analizar el impacto que tiene la calidad de la energía eléctrica en el consumidor industrial para conocer el contexto de la problemática, realizar el análisis de los parámetros para la evaluación de la calidad en el consumo de la energía eléctrica.

La metodología de la ciencia de datos (Figura 1) sigue un proceso estructurado que incluye la definición del problema, la

recopilación y limpieza de datos, el análisis exploratorio, la selección y aplicación de modelos, la evaluación de resultados y la implementación de soluciones. Este enfoque sistemático permite transformar datos en valiosa información para la toma de decisiones [4].

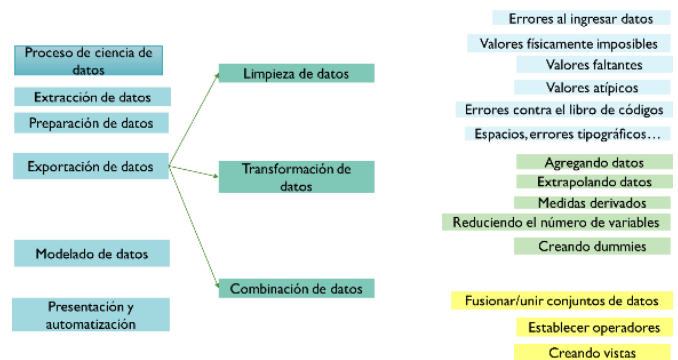


Figura 1.- Metodología del proceso de ciencia de datos (DSP)
 Fuente: Extraído de Cielan (2016)

En la Figura 1 se muestran las fases del proceso de ciencia de datos, de los cuales se realizó la extracción de los datos, la limpieza, la preparación de los datos, revisando si existen valores atípicos, faltantes, etc. Y por último se continuó este análisis hasta la transformación de los datos, reduciendo el número de variables.

Se utilizaron datos generados de la subestación eléctrica industrial que corresponde a la plaza comercial Las Palmas. Esta plaza comercial está integrada por tiendas de lujo, boutiques de moda, restaurantes gourmet y algunos servicios de alto estándar. Además, se destacan sus espacios amplios, con jardines y áreas de descanso que crean una atmósfera relajada y exclusiva.

Las mediciones de los datos eléctricos se obtuvieron con un analizador de redes Enetics LM5400 el cual es el encargado de recopilar la información correspondiente a los parámetros de calidad de la energía eléctrica. Asimismo, se utilizó el software Power Scape, el cual se encarga de extraer la información que el analizador de redes obtuvo durante las mediciones.

El dataset final se compone por 48 parámetros, conformados por tensión, corriente, potencia y frecuencia, THD, factor de potencia y corriente armónica; Contiene 8640 instancias. La Tabla 2 muestra el diccionario de datos que muestra los parámetros que fueron recabados en la medición a la subestación eléctrica. Para realizar el análisis completo de la calidad de la energía es necesario tomar en cuenta todos los parámetros, en este análisis exploratorio se tomaron los valores de tensión y corriente de las fases A, B y C.

Tabla 2.- Diccionario de datos de las mediciones eléctricas.

Lugar de medición	Valor de medición	Unidad de medida	Valor de medición	Lugar de medición	Unidad de medida
Fase A	A-N Min	V	Fase B	A-N Min	V
	A-N Prom	V		A-N Prom	V
	A-N Max	V		A-N Max	V
	A-N Min	A		A-N Min	A
	A-N Prom	A		A-N Prom	A
	A-N Max	A		A-N Max	A
	Potencia real	W		Potencia real	W
	Potencia reactiva	VAR		Potencia reactiva	VAR
	Potencia aparente	VA		Potencia aparente	VA
	Factor de potencia	-		Factor de potencia	-
	THD	%		THD	%
Corriente armónica	A	Corriente armónica	A		
Fase C	A-N Min	V	Neutro	N-G Min	V
	A-N Prom	V		N-G Prom	V
	A-N Max	V		N-G Max	V
	A-N Min	A	Tri Fásico	Potencia aparente	VA
	A-N Prom	A		Factor de potencia	-
	A-N Max	A		Potencia activa	W
	Potencia real	W		Potencia reactiva	VAR

Potencia reactiva	VAR	Frecuencia mínima	Hz
Potencia aparente	VA	Frecuencia promedio	Hz
Factor de potencia	-	Frecuencia máxima	Hz
THD	%		
Corriente armónica	A		

A-N: Fase A a Neutro, B-N: Fase B a Neutro, C-N: Fase C a Neutro, N-G: Neutro a Tierra (Ground), THD: Distorsión armónica total (Total Harmonic Distortion).

El análisis exploratorio consistió en el análisis estadístico descriptivo, el análisis univariante y bivariante, y el análisis de componentes principales. Para el análisis univariante se obtuvo la distribución de probabilidad; en el análisis bivariante se realizaron gráficos de dispersión y la matriz de correlación en un mapa de calor.

El análisis de componentes principales (PCA, por sus siglas en inglés) es una técnica estadística utilizada para reducir la dimensionalidad de un conjunto de datos, transformando las variables originales en un nuevo conjunto de variables no correlacionadas llamadas componentes principales. Estos componentes capturan la mayor variabilidad de los datos, permitiendo simplificar su interpretación y visualización sin perder información significativa [9].

El análisis se realizó en Python, con las bibliotecas numpy, pandas, matplotlib y seaborn.

III. RESULTADOS Y DISCUSIÓN

En la Tabla 3 se muestra el resumen estadístico descriptivo con las medidas de tendencia central. V1, V2 y V3 se refieren a las tensiones de las fases A, B y C, mientras que I1, I2 e I3 se refieren a las corrientes de las fases A, B y C.

Tabla 3.- Resumen estadístico descriptivo

	V1	V2	V3	I1	I2	I3
Instancias	8756.00	8756.00	8756.00	8756.00	8756.00	8756.00
Media	130.46	131.74	132.11	57.76	65.71	105.77
Desviación estándar	2.35	2.03	1.98	7.09	11.69	37.06
Min	122.32	125.46	125.96	44.46	44.78	45.16
25%	128.62	130.14	130.56	52.00	56.72	62.72
50%	130.72	132.04	132.44	57.88	66.28	133.64
75%	132.38	133.40	133.70	63.10	74.23	137.68
Max	135.06	135.70	135.98	82.46	102.48	142.82

El análisis de tensión (V1, V2, V3) muestra un sistema estable y balanceado, con medias cercanas a 130-132 y desviaciones estándar bajas (1.98 a 2.35). Los valores mínimos y máximos están dentro de un rango estrecho (122.32 a 135.98), lo que confirma que no hay variaciones extremas. Los percentiles indican una distribución simétrica, sugiriendo que

los datos no están sesgados. Esto refleja un comportamiento saludable en el suministro de tensión.

Por otro lado, el análisis de corriente (I1, I2, I3) revela un desequilibrio significativo, especialmente en la fase I3, con una media de 105.77 y una desviación estándar alta (37.06). Los rangos son amplios (45.16 a 142.82), indicando picos de corriente o cargas variables.

Este desbalance podría deberse a cargas mal distribuidas, armónicos o fallas en equipos, por lo que se recomienda investigar y corregir estas anomalías para mejorar la eficiencia del sistema.

En las Figuras 3 y 4 se muestra el análisis univariante mediante la distribución de probabilidad de las tensiones V1, V2 y V3, y de las corrientes de fase I1, I2 e I3.

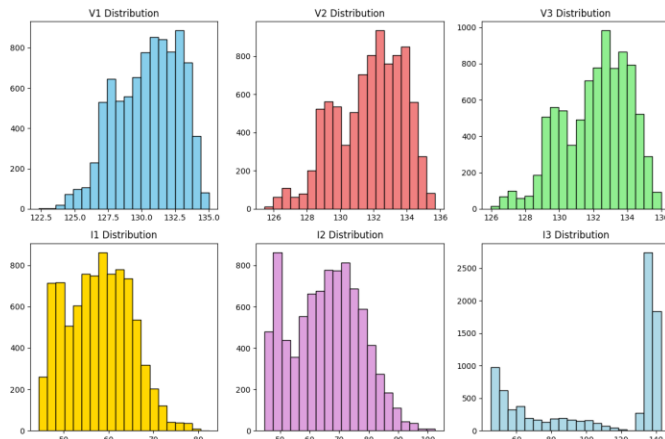


Figura 3.- Tensiones y corrientes en cada fase

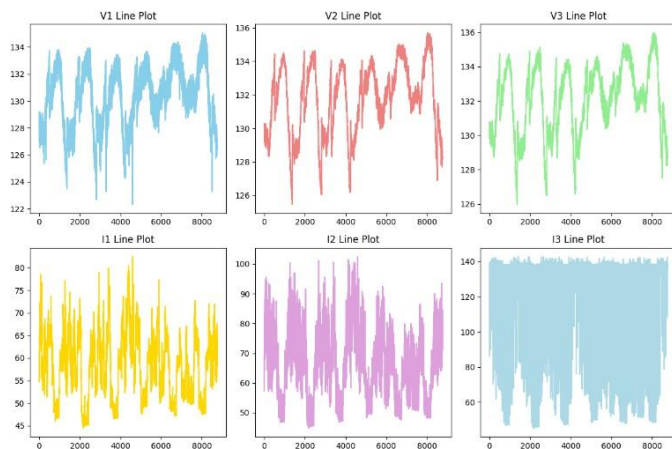


Figura 4.- Tensiones y corrientes en cada fase

A partir del análisis de las Figuras 3 y 4 se puede obtener la siguiente información:

Tensiones (V1, V2, V3): Estables, balanceadas y dentro de rangos aceptables.

Corrientes (I1, I2, I3): Desequilibradas, especialmente en I3, lo que requiere atención para corregir posibles problemas de carga o armónicos.

Las tres fases de tensión son estables y balanceadas, con medias cercanas y desviaciones estándar bajas. Esto sugiere un suministro de energía eléctrica saludable.

Las corrientes presentan un desequilibrio notable, especialmente en I3, que muestra una media alta y una gran variabilidad. Esto podría deberse a cargas desbalanceadas, armónicos o fallas en equipos.

La Figura 5 presenta el análisis bivariante entre las variables V1 y V2 que son la tensión promedio en la fase A y B.

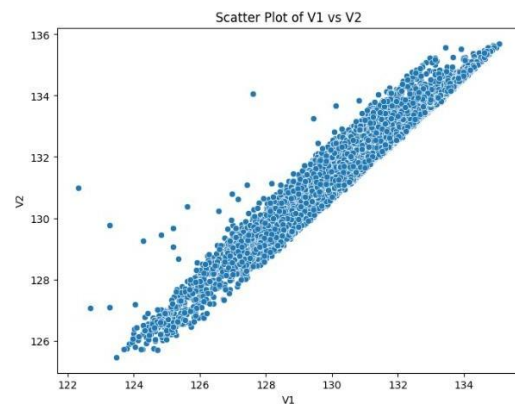


Figura 5.- Análisis bivariante entre V1 y V2

La Figura 5 muestra la relación entre V1 y V2. Los puntos en el gráfico siguen una línea recta o una tendencia clara, esto indica que existe una correlación entre V1 y V2. Una correlación positiva (pendiente ascendente) sugiere que cuando V1 aumenta, V2 también lo hace.

En cuanto a la estabilidad del sistema, los puntos se agrupan cerca de una línea diagonal ($V1 \approx V2$), esto sugiere que las tensiones en ambas fases son similares y estables.

Respecto a los valores atípicos (outliers), algunos puntos se alejan significativamente del grupo principal podrían indicar anomalías en el sistema, como picos de tensión o fallas momentáneas.

La Figura 6 presenta un mapa de calor con la correlación entre las variables.

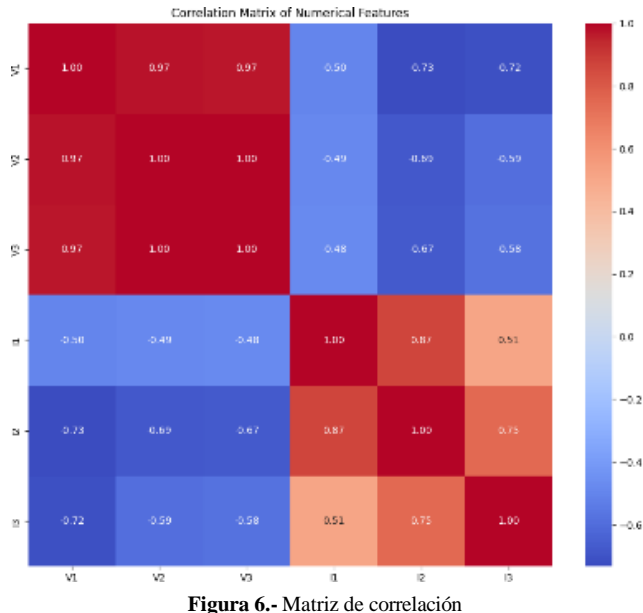


Figura 6.- Matriz de correlación

Esta matriz de correlación proporciona una visión general de cómo se relacionan entre sí las diferentes variables del conjunto de datos: tres tensiones (V1, V2, V3) y tres corrientes (I1, I2, I3). Cada celda de la matriz representa el coeficiente de correlación entre dos variables, y el color indica la fuerza y dirección de esa relación.

En cuanto a la correlación entre las tensiones, las tensiones V1, V2 y V3 están fuertemente correlacionadas entre sí ya que obtienen los valores de tensión de la misma fuente.

En cuanto a la relación inversa entre tensiones y corrientes, las tensiones (V1, V2, V3) tienen una correlación negativa con las corrientes (I1, I2, I3). Esto es consistente con la Ley de Ohm, que establece que la corriente es inversamente proporcional a la resistencia para un voltaje constante.

En la Figura 7 se muestra la varianza acumulativa explicada. Esta gráfica muestra cómo la varianza total de un conjunto de datos es capturada por un número creciente de componentes principales (PCs) obtenidos a través del Análisis de Componentes Principales (PCA). En otras palabras, indica qué porcentaje de la información original se puede conservar al reducir la dimensionalidad de los datos.

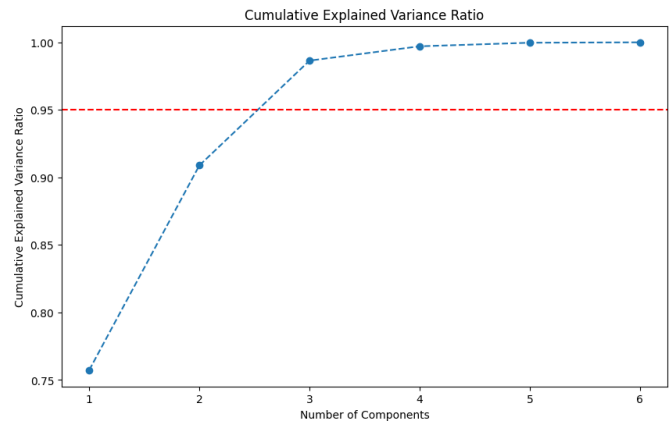


Figura 7.- Varianza acumulativa explicada

De acuerdo con la gráfica, se podría considerar que 2 o 3 componentes principales serían suficientes para representar los datos, ya que a partir de ese punto la ganancia en términos de varianza explicada es marginal.

La Figura 8 muestra la matriz de componentes principales.

Tabla 3.- Matriz de componentes principales.

	PC1	PC2	PC3	PC4
V1	0.45	0.3	-0.05	0.03
V2	0.43	0.37	0.19	0.04
V3	0.43	0.38	0.21	0.07
I1	-0.34	0.64	-0.45	-0.52
I2	-0.42	0.42	-0.03	0.8
I3	-0.37	0.19	0.85	-0.27

La matriz muestra cómo las variables originales (V1, V2, V3, I1, I2, I3) contribuyen a cada componente (PC1, PC2, PC3, PC4).

En el primer componente principal PC1, las variables V1, V2 y V3 tienen contribuciones positivas (alrededor de 0.43-0.45), mientras que I1, I2 e I3 tienen contribuciones negativas (alrededor de -0.34 a -0.42), lo que sugiere que PC1 contrasta las variables V con las I. En el segundo componente principal PC2, las variables I1 e I2 dominan con contribuciones altas (0.64 y 0.42), mientras que en PC3 e PC4, I3 e I2 tienen contribuciones muy altas (0.85 y 0.80, respectivamente), indicando que estos componentes están fuertemente influenciados por estas variables.

En general, las variables V1, V2 y V3 parecen estar altamente correlacionadas, ya que tienen contribuciones similares en PC1 y PC2. Por otro lado, las variables I1, I2 e I3 muestran

contribuciones más variadas, capturando diferentes aspectos de la variabilidad.

IV. CONCLUSIONES

Los resultados muestran que las tensiones (V1, V2, V3) son estables y balanceadas, con medias cercanas a 130-132 V y desviaciones estándar bajas, lo que indica un suministro eléctrico saludable. Sin embargo, las corrientes (I1, I2, I3) presentan un desequilibrio significativo, especialmente en la fase I3, con una alta variabilidad y valores extremos, lo que sugiere posibles problemas como cargas desbalanceadas o presencia de armónicos.

El análisis de correlación y componentes principales (PCA) confirma la fuerte relación entre las tensiones y su contraste con las corrientes, destacando la importancia de corregir estas anomalías para mejorar la eficiencia del sistema.

En conclusión, el estudio demuestra la utilidad del análisis exploratorio de datos y técnicas como el PCA para identificar patrones y problemas en la calidad de la energía eléctrica.

Los hallazgos sugieren que, aunque el sistema de tensión es estable, el desbalance en las corrientes requiere atención para garantizar el cumplimiento de normas como la IEC 61000.

Este trabajo sienta las bases para continuar el análisis para incorporar modelos de aprendizaje automático para la detección y clasificación de perturbaciones, optimizando así la gestión de la calidad de la energía en este entorno industrial.

REFERENCIAS

- [1] Enríquez Harper, Gilberto. *El ABC de la calidad de la energía eléctrica*. 265 2012.
- [2] IEEE Standards Association. *IEEE Std 1159-2019: IEEE Recommended Practice for Monitoring Electric Power Quality*. IEEE 2019.
- [3] Bollen, M. H. J., & Gu, I. Y. H. *Signal processing of power quality disturbances*. Wiley-IEEE 2006.
- [4] Cielen, D., Meysman, A., & Ali, M. *Introducing data science: Big data, machine learning, and more, using Python tools*. Manning Publications 2016.
- [5] J. Li, Z. T. Q. T. and J. S. Detection and classification of power quality disturbances using double resolution s-transform and dagsvms. *IEEE Transactions on Instrumentation and Measurement*, 65, 2302–2312 2020.
- [6] G. Singh, Y. P. and A. K. D. Classification of power quality disturbances using linear discriminant analysis. *Applied Soft Computing*, 138, 1–16 2023.
- [7] S. Wang and H. Chen. “A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network.” *Applied energy*, 235, 1126–1140 2019.
- [8] Zhao, F., Liu, G., Chen, X., & Wang, Y. Power Quality Disturbance Identification Basing on Adaptive Kalman Filter and Multi-Scale Channel Attention Fusion Convolutional Network. *Energy Engineering*, 121(7), 1865–1882 2024.
<https://doi.org/10.32604/EE.2024.048209>
- [9] Jolliffe, I. T. *Principal Component Analysis* (2nd ed.). Springer 2002.