

Predicción de Series Temporales Económicas: aplicación a los Precios de la Energía en el Mercado Eléctrico Español

Alicia Troncoso Lora¹, Jesús Riquelme Santos¹, José Riquelme Santos²,
José Luis Martínez Ramos¹ y Antonio Gómez Expósito¹

¹ Departamento de Ingeniería Eléctrica, Universidad de Sevilla, Sevilla, España

² Departamento de Lenguajes y Sistemas, Universidad de Sevilla, Sevilla, España
ali@esi.us.es jesus@us.es riquelme@lsi.us.es {camel, age}@us.es

Resumen En este artículo se describe un método local basado en un algoritmo de clasificación de los k vecinos más cercanos para la predicción de series temporales económicas. Se analiza teóricamente cuándo el error cometido con este método es mínimo. Por último, se analizan algunas variantes del método propuesto basadas en la elección de distintas distancias y se comparan los resultados obtenidos de su aplicación a la serie temporal formada por los precios de la energía en el Mercado Eléctrico Español.

1 Introducción

En los últimos quince o veinte años las empresas que habían sido hasta entonces propiedad del Estado están tendiendo progresivamente a la privatización con el objeto de que, al eliminar el sector público en régimen de monopolio, se genere competencia en el mercado, lo que lleva consigo un efecto beneficioso sobre los consumidores, tanto en calidad de servicio como en precios. El Sector Eléctrico Español desde 1998 ha puesto en marcha un proceso de reestructuración siguiendo criterios de liberalización y competencia, al igual que un gran número de países en el mundo como Noruega, Suecia, Finlandia, Portugal, etc. De esta manera, en el sistema Español surgen dos nuevas figuras que hacen posible el funcionamiento de las nuevas reglas que rigen el Mercado [1]: el operador del mercado (OM), encargado de la determinación de los precios de la energía cada hora, y el operador del sistema (OS), encargado de la operación del sistema eléctrico.

En un mercado liberalizado, las herramientas de predicción son muy importantes, ya que en base a una predicción de precios, los agentes que participan en el mercado, como las compañías de generación, elaboran sus estrategias de ofertas para maximizar los beneficios que obtienen por la venta de energía en el mercado [2,3].

En el pasado, se han usado técnicas para predecir la demanda en mercados centralizados basadas tanto en Modelos de Series Temporales [4] como en Inteligencia Artificial [5,6]. Sin embargo, debido a la reciente liberalización del

mercado eléctrico, las técnicas de predicción aplicadas a los precios de la energía son bastantes recientes y poco usuales en la literatura actual, siendo un campo de investigación actualmente en alza. Hoy día, los métodos más usados para la predicción de los precios de la energía en un mercado competitivo son los distintos tipos de Redes Neuronales Artificiales [7,8,9], siendo en este contexto la primera vez que se aplica el método que se presenta en este trabajo.

En general, las series económicas presentan las características siguientes:

1. Generalmente las series temporales disponibles son cortas, excepto las series de alta frecuencia. La serie temporal formada por los precios de la energía es de alta frecuencia puesto que la frecuencia de los precios es de una hora.
2. Los valores de la serie temporal son generalmente difíciles de medir con precisión, ya que hay factores imprevisibles que responden a mecanismos de mercado. Debido a esto, son series con un alto porcentaje de valores inusuales y con un comportamiento caótico.
3. No son estacionarias, es decir no tienen media y varianza constante.
4. Son series no lineales con componentes estocásticos.

En este artículo se describe un método local basado en un algoritmo de clasificación de los k vecinos más cercanos para la predicción de series temporales económicas. Se analiza teóricamente cuándo el error cometido con este método es mínimo. Por último, se analizan algunas variantes del método propuesto basadas en la elección de distintas distancias y se comparan los resultados obtenidos de su aplicación a la serie temporal formada por los precios de la energía en el Mercado Eléctrico Español.

2 Definición del Problema

Las ofertas de venta y compra de energía para el día siguiente se presentan por parte de los agentes del mercado al OM en el mercado diario. El OM incluye estas ofertas en un procedimiento de casación, obteniendo así el precio de la energía para el horizonte diario de programación correspondiente al día siguiente. De esta manera, el problema para un agente del mercado consiste en predecir los precios de la energía en el corto plazo, es decir, las veinticuatro horas del día siguiente y realizar las mejores ofertas en consecuencia.

Para predecir los precios horarios del día siguiente se usa una ventana formada por precios horarios de días pasados de la base de datos histórica disponible. La longitud de la ventana es un parámetro que hay que determinar de manera óptima mediante el análisis de los precios correspondientes a los días anteriores que más influyen en los precios del día siguiente. Los valores que forman la ventana pueden ser representados como puntos en un espacio m -dimensional, \mathbb{R}^m , donde m es la longitud de la ventana.

Para generar predicciones hay dos tipos de aproximaciones:

1. **Predicción directa:** la ventana para predecir los veinticuatro precios horarios del día siguiente es fija. De esta manera la ventana se desplaza veinticuatro horas cada vez. Así se tiene una asociación entre los precios pasados como puntos de \mathbb{R}^m y los precios futuros como puntos de \mathbb{R}^{24} .

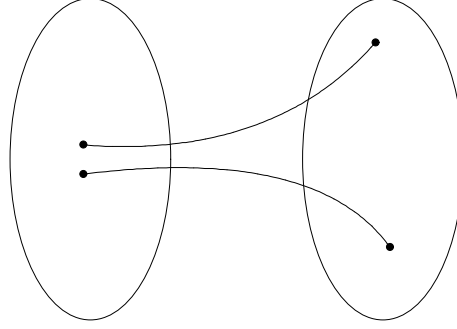


Figura 1. Falsos vecinos.

2. **Predicción iterada:** la ventana se usa para predecir el precio de una hora y la predicción obtenida forma parte de una nueva ventana para predecir la hora siguiente, desplazando así la ventana una hora cada vez. De esta manera la asociación se tiene entre precios pasados como puntos de \mathbb{R}^m y precios futuros como puntos de \mathbb{R} . Con este último tipo de predicción los errores se van acumulando en el vector de entrada que forma la ventana, sobre todo en las últimas horas del horizonte de predicción.

Para determinar de una manera óptima la longitud de la ventana m se usa el **Método del Falso Vecino más Cercano** (FNN, False Nearest Neighbour) [10]. Este método consiste en calcular la longitud de la ventana que hace el número de falsos vecinos más cercanos mínimo. Dos puntos de \mathbb{R}^m , que representan precios de la energía en el pasado, son falsos vecinos si la distancia entre sus correspondientes precios futuros (puntos de \mathbb{R}^{24} o \mathbb{R} según el tipo de predicción que se haga) crece como se muestra en la Figura 1.

Este radio de crecimiento entre precios pasados y futuros se calcula como sigue:

1. Predicción directa:

$$R_i = \frac{d(P_{i+1}, P_{j+1})}{d(P_i, P_j)} \quad (1)$$

2. Predicción iterada:

$$R_i = \frac{|P_{i+1} - P_{j+1}|}{d(P_i, P_j)} \quad (2)$$

donde $d(\cdot, \cdot)$ es una distancia entre puntos de \mathbb{R}^{24} . Si R_i es mayor que un umbral dado, el punto P_j es un falso vecino más cercano de P_i . Se trata entonces de elegir la longitud de la ventana suficientemente grande para que el número de falsos

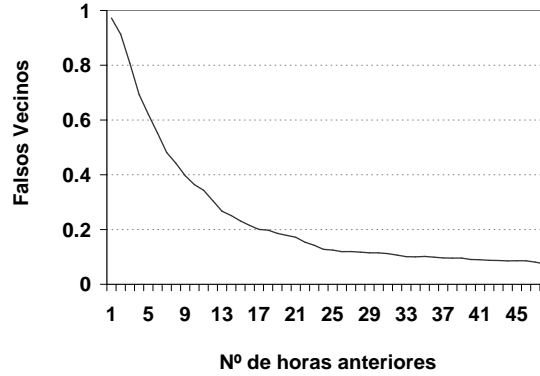


Figura 2. Fracción de falsos vecinos en el periodo Marzo-Agosto 2001 mediante predicción iterada.

vecinos más cercanos sea suficientemente pequeño. La Figura 2 muestra cómo disminuye el número de falsos precios vecinos cuando aumenta el número de precios de horas anteriores usados para predecir la hora siguiente. Como se observa, veinticuatro horas anteriores son suficientes para predecir la hora siguiente, ya que el número de falsos vecinos no disminuye de una manera considerable con el uso de más horas.

3 Descripción del Método

En esta sección se describe un método local de predicción de series temporales basado en un algoritmo de clasificación de los k vecinos más cercanos (kNN, k Nearest Neighbour) [11].

Esta predicción consiste en estimar los precios de la energía de mañana como una media de los precios de la energía de los días siguientes a los k días con precios similares al día de hoy. Esta media está ponderada con unos pesos que representan la importancia de los días con precios similares según su mayor o menor parecido. Al precio de la energía del día siguiente al día en el que los precios fueron los más parecidos a los de hoy le corresponde un peso mayor que al precio de la energía del día siguiente al día en el que los precios fueron los menos parecidos a los de hoy. En el caso de un vecino la idea geométrica se muestra en la Figura 3. Los precios de hoy y los precios a estimar de mañana están representados con una circunferencia. Los puntos negros son los precios vecinos más cercanos a los precios de hoy, siendo el día r aquel cuyos precios son los más parecidos. Entonces parece lógico estimar los precios de mañana con los precios que hubo el día siguiente al día r , es decir los precios del día $r + 1$.

Este método usa para predecir básicamente una media local ponderada con unos pesos. Para determinar los precios que forman parte de esta media basta

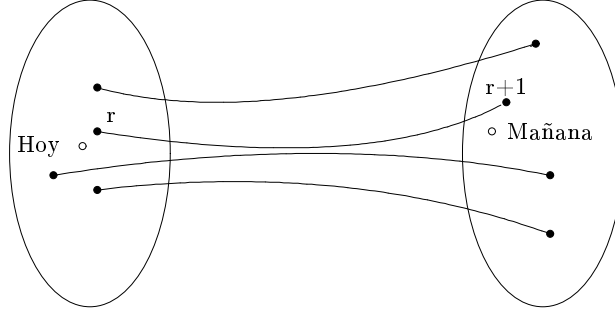


Figura 3. Aproximación local.

con calcular los k precios más parecidos a los precios de hoy. Para ello hay que elegir una distancia que mida la similitud entre los precios de dos días cualesquiera de una manera adecuada a las características específicas de la serie que se esté tratando de predecir. Tradicionalmente, la distancia más común ha sido la distancia Euclídea. Sin embargo, muchas aplicaciones del mundo real requieren una distancia un poco más flexible a la hora de modelar las características de la serie, como la distancia definida por:

$$d_w^2(q, z) = \sum_{i=1}^{24} w_i \cdot (q_i - z_i)^2 \quad \text{donde } w_i \in [0, 1] \quad (3)$$

En el caso de los precios de la energía como serie temporal, los pesos w_i representan la importancia de cada hora, ya que no todas las horas tienen la misma influencia sobre los precios del día siguiente.

El uso de diferentes distancias puede dar lugar a mejores aproximaciones en el algoritmo de búsqueda de los k vecinos más cercanos.

De esta manera, la predicción de los precios de la energía del día $d + 1$ viene dada por:

$$\hat{P}_{d+1} = \frac{1}{\alpha_1 + \dots + \alpha_k} \sum_{l=1}^k \alpha_l \cdot P_{v_l+1} \quad (4)$$

donde v_l son los k días con precios más parecidos a los precios del día d , ordenados desde el más cercano al más lejano y donde los pesos vienen dados por

$$\alpha_l = \frac{d(P_d, P_{v_k}) - d(P_d, P_{v_l})}{d(P_d, P_{v_k}) - d(P_d, P_{v_1})} \quad (5)$$

El peso α_l es igual a cero cuando l corresponde al día considerado más lejano y uno cuando el día considerado es el más cercano.

4 Acotación del Error

En esta sección se hace un breve análisis de cual es el mínimo error que se comete en la predicción de una serie temporal con el método local descrito en la sección anterior. Primero se analiza el caso de un solo vecino y finalmente se concluye para el caso más general de k vecinos más cercanos.

La predicción para los precios horarios de la energía de mañana, P_{d+1} , se estima con los precios del día siguiente al día v en el que hubo los precios más parecidos a los precios horarios de la energía de hoy, P_d . Es decir,

$$\hat{P}_{d+1} = P_{v+1} \text{ donde } v \text{ es tal que } d(P_d, P_v) \text{ es mínima} \quad (6)$$

El error mínimo se cometerá cuando los precios horarios de la energía de mañana sean estimados con los precios más parecidos a éstos, es decir

$$\hat{P}_{d+1} = P_r \text{ donde } r \text{ es tal que } d(P_{d+1}, P_r) \text{ es mínima} \quad (7)$$

Entonces el error es mínimo cuando $P_{v+1} = P_r$, o lo que es igual $v + 1 = r$. Es decir, el día siguiente al día vecino de hoy es el vecino más cercano de mañana.

En el caso de k vecinos más cercanos, razonando de manera análoga, el error es mínimo cuando:

$$\frac{1}{\alpha_1 + \dots + \alpha_k} \sum_{l=1}^k \alpha_l \cdot P_{v_l+1} = P_r \quad (8)$$

Es decir, el error es mínimo cuando la media ponderada de los precios de los días siguientes a los k precios vecinos más cercanos de los precios de hoy es proporcional al precio vecino más cercano de los precios de mañana. La constante de proporcionalidad es $\alpha_1 + \dots + \alpha_k$, donde α_l vienen dados por (5).

Obsérvese que para predecir de forma óptima los precios horarios de la energía de mañana se necesita calcular P_r minimizando la distancia a los precios que se quieren predecir, luego este análisis es útil para saber si el error que hemos cometido después de haber predicho localmente con este método se desvía mucho del error óptimo que se pudiera haber obtenido.

Véase también que cuando la predicción por este método se aleja de ser la óptima, es decir, $d(P_{v+1}, P_{d+1})$ es suficientemente grande, P_d y P_v son falsos vecinos. Además la diferencia entre el error obtenido y el error óptimo da una idea intuitiva de la divergencia media que sufren las trayectorias que son vecinas en el pasado pero no en el futuro, como se muestra en la Figura 1.

El Error Absoluto Medio (EAM) al predecir de manera óptima viene dado por:

$$EAM = \frac{1}{24} \sum_{h=1}^{24} |\hat{P}_{d+1,h} - P_{d+1,h}| = \frac{1}{24} \sum_{h=1}^{24} |P_{r,h} - P_{d+1,h}| \quad (9)$$

De (7) se deduce que el EAM es mínimo si se elige como distancia la norma 1, puesto que r es tal que $d(P_{d+1}, P_r)$ es mínima. Es decir, si la distancia elegida

para el cálculo del vecino más cercano es la norma 1, al predecir de manera óptima el error absoluto medio será menor que el obtenido con cualquier otra distancia. Lo mismo ocurre con el error cuadrático medio (ECM) si se elige la distancia Euclídea o con el error relativo medio (ERM) si se elige como distancia entre dos puntos x e y :

$$d(x, y) = \frac{|x - y|}{x} \quad (10)$$

5 Resultados

El método de predicción basado en un algoritmo de clasificación kNN ha sido aplicado para obtener la predicción de los precios de la energía en el Mercado Eléctrico Español. El periodo Marzo-Agosto 2001 se ha usado como conjunto de entrenamiento para analizar de una manera visual el comportamiento y las propiedades de la serie, así como para determinar el número óptimo de vecinos. El periodo Septiembre-Octubre 2001 se ha usado como conjunto test para validar el método propuesto. Para evitar que los errores de la predicción se acumulen en las últimas horas del día siguiente se ha elegido un esquema de predicción directa.

El precio medio de la serie es 2.79 Céntimos de Euros por kWh y la desviación estándar respecto de la media 1.14, lo cual es un primer indicador de la dispersión que presentan los precios de la energía, debido principalmente a que la serie viene dada por el funcionamiento del Mercado Eléctrico Español en régimen de oligopolio.

La Figura 4 muestra la distribución de los precios durante Marzo-Agosto 2001. Nótese que el 21% de los precios tienen valores mayores que 4 céntimos

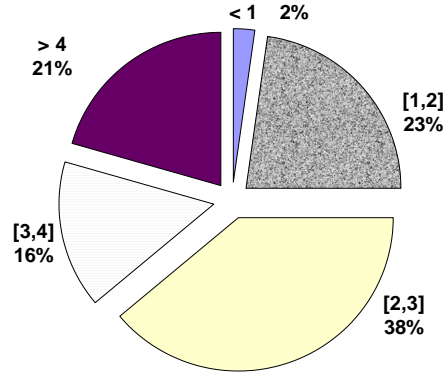


Figura 4. Distribución de los precios en los meses Marzo-Agosto 2001.

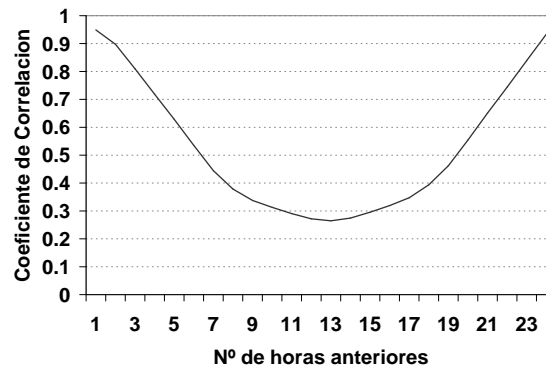


Figura 5. Coeficiente de correlación.

por kWh y el 2% valores menores que 1 céntimo por kWh. Estos precios son considerados inusuales, y su alto porcentaje dificulta obtener una buena predicción.

La Figura 5 muestra el coeficiente de correlación de los precios de la energía en una hora y los precios correspondientes a las veinticuatro horas anteriores. Como se puede ver, las horas pasadas más influyentes en la hora actual son justo la hora uno y la veinticuatro, que coincide con la hora anterior y con la misma hora que la actual pero del día anterior, respectivamente.

La Figura 6 muestra la influencia del número de vecinos usados sobre los distintos tipos de errores, EAM, ECM y ERM, de la predicción de los precios de la energía en el periodo de test considerado usando la distancia definida

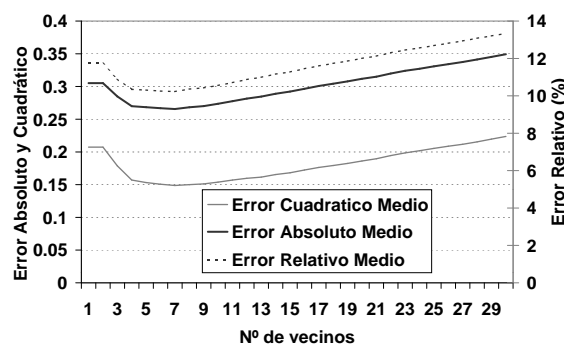


Figura 6. Número óptimo de vecinos.

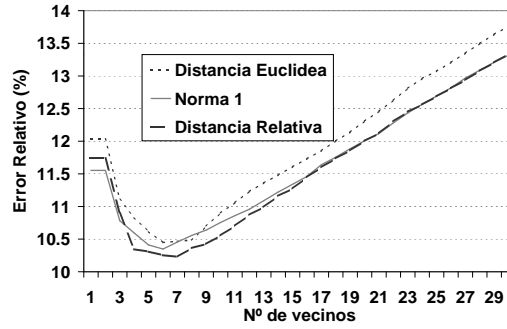


Figura 7. Influencia de la distancia en el tipo de error.

por (10). Como se puede observar, los distintos tipos de errores son mínimos cuando el número de vecinos es 7. Se ha hecho el mismo estudio con la distancia Euclídea y la norma 1 habiendo obtenido el mismo número óptimo de vecinos. Es decir, el número óptimo de vecinos para la predicción mediante esta técnica es independiente de la distancia usada.

La Figura 7 muestra la influencia de la distancia usada para determinar los precios vecinos, mediante el algoritmo kNN, en el Error Relativo Medio. Generalmente, este error, al usar la distancia relativa definida por (10), es menor que con otra distancia cualquiera, aunque no siempre como se puede observar cuando el número de vecinos es menor o igual que 3.

La Tabla 1 muestra los valores de los distintos tipos de errores, obtenidos mediante la predicción óptima de los precios de la energía descrita en la sección anterior, según la distancia usada para el cálculo de un único vecino. Nótese que, el ECM mínimo se obtiene cuando se usa la distancia Euclídea, el EAM mínimo cuando se usa la norma 1 y el ERM cuando se usa la distancia relativa (10), como se demostró teóricamente en la sección anterior.

La Tabla 2 presenta los valores de los distintos tipos de errores obtenidos de la aplicación del método descrito en este trabajo a los precios de la energía,

	Predicción óptima		
	EAM	ECM	ERM (%)
Distancia Euclídea	0.226	0.114	5.77
Norma 1	0.216	0.126	5.55
Distancia Relativa	0.218	0.126	5.50

Tabla 1. Influencia de la distancia en la optimización de los distintos tipos de errores.

	Predicción no óptima		
	EAM	ECM	ERM (%)
Distancia Euclídea	0.36	0.25	9.25
Norma 1	0.36	0.25	9.19
Distancia Relativa	0.37	0.26	9.36

Tabla 2. Influencia de la distancia en los errores de la predicción.

según la distancia usada. El número de vecinos usados ha sido el óptimo, 7 en este caso. La diferencia entre los errores obtenidos y los errores óptimos de la Tabla 1 es aproximadamente un 3.5%. Esta desviación del error medio óptimo en el periodo Septiembre-Octubre 2001 es debido principalmente a los precios que son considerados falsos vecinos.

6 Conclusiones

En un mercado liberalizado las herramientas de predicción son muy importantes ya que en base a una predicción de precios los agentes que participan en el mercado elaboran sus estrategias de ofertas. En este artículo se ha descrito un método local basado en un algoritmo de clasificación de los k vecinos más cercanos para la predicción de series temporales económicas. Se ha analizado teóricamente cuándo el error cometido con este método es mínimo. Por último, se han analizado algunas variantes del método propuesto basadas en la elección de distintas distancias y se han comparado los resultados obtenidos de su aplicación a la serie temporal formada por los precios de la energía en el Mercado Eléctrico Español.

Agradecimientos

Los autores agradecen a la DGICYT (PB97-0719, DPI2001-2612, TIC2001-1143-C03-02) la financiación recibida.

Referencias

1. A. Canoyra, C. Illán, A. Landa, J.M. Moreno, J.I. Pérez Arriaga, C. Sallé and C. Solé: The Hierarchical Market Approach to the Economic and Secure Operation of the Spanish Power System. Bulk Power System Dynamic and Control IV, Agosto 24-28, Santorini, Grecia.
2. J.L. Martínez, A. Gómez, J. M. Riquelme, A. Troncoso Lora and A. R. Marulanda Guerra: Influence of ANN-based Market Price Forecasting Uncertainty on Optimal Bidding. Proceedings of the PSCC. Sevilla, España. Junio 2002.
3. A. J. Conejo, J. Contreras, J.M. Arroyo and S. de la Torre: Optimal Response of an Oligopolistic Generating Company to a Competitive Pool-Based Electric Power Market. To appear in IEEE Trans. on Power System.

4. A. D. Papalexopoulos and T. C. Hesterberg: A Regression-Based Approach to Short-Term System Load Forecasting. *IEEE Trans. on Power System*, Vol. 5, pp. 1535-1547. 1990.
5. A. S. Alfuhaid and M. A. El-Sayed: Cascaded Artificial Neural Network for Short-Term Load Forecasting. *IEEE Trans. on Power System*, Vol. 12, pp. 1524-1529. 1997.
6. R. Lamedica et al: A Neural Network Based Technique for Short-Term Forecasting of Anomalous Load Periods. *IEEE Trans. on Power System*, Vol. 11, no 4, Noviembre 1996.
7. B. Ramsay and A. J. Wang: An Electricity Spot-Price Estimator with Particular Reference to Weekends and Public Holidays. *Proceedings of the Universities Power Engineering Conference UPEC'97*, Manchester, Reino Unido, Septiembre 1997.
8. F. Gao, X. Guan, X-R. Cao and A. Papalexopoulos: Forecasting Power Market Clearing Price and Quantity using a Neural Network Method. *Proceedings of the Power Engineering Summer Meeting*, Seattle, USA, Julio 2000.
9. B. R. Szkuta, L. A. Sanabria and T.S. Dillon: Electricity Price Short-Term Forecasting using Artificial Neural Networks. *IEEE Trans. on Power Systems*, vol. 14, no. 3, pp. 851-857, Agosto 1999.
10. M. B. Kennel, R. Brown and H. D. I. Abarbanel: Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review*, A45, 3403, 1992.
11. B.V. Dasarathy (Ed): Nearest neighbour (NN) Norms: NN pattern classification techniques. *IEEE Computer Society Press*, 1991.