Departamento de Estadística, Investigación Operativa Aplicadas y Calidad

Universidad Politécnica de Valencia

APLICACIÓN DE REGRESIÓN CONFORMAL PARA LA MONITORIZACIÓN DE AEROGENERADORES

"Máster Universitario en Ingeniería de Análisis de Datos, Mejora de Procesos y Toma de Decisiones"

Trabajo Fin de Máster

Autor: Isis Caterina Rosario Custodio

Director (a): Ana Isabel Sánchez Galdón

Convocatoria: septiembre 2017

Curso: 2016-2017

RESUMEN

En España, el sector eólico se ha consolidado como una de las principales fuentes de energía renovable con una cobertura de la demanda del 19.4%. Uno de los objetivos del sector eólico es garantizar la máxima disponibilidad de los parques eólicos sin aumentar los costes de mantenimiento. Una forma práctica de conseguir este objetivo es la aplicación de técnicas de monitorización efectivas. En este contexto, el presente trabajo tiene por objetivo la aplicación de la regresión conformal para la monitorización de aerogeneradores a partir de la información registrada en el Sistema de Supervisión, Control y Adquisición de Datos (SCADA). Las discrepancias entre los datos observados y los predichos mediante el modelo podrían ser indicativos de la presencia de fallos o degradaciones en los componentes del sistema. Para evaluar dicha discrepancia se utiliza una medida de no conformidad la cual se obtiene utilizando como algoritmo subyacente Máquina de Soporte Vectorial (SVM). Finalmente, se presenta un caso de aplicación a aerogeneradores de un parque eólico situado en España.

Palabras claves: Energía eólica, Máquinas de Soporte Vectorial, regresión conformal.

RESUM

A Espanya, el sector eòlic s'ha consolidat com una de les principals fonts d'energia renovable amb una cobertura de la demanda del 19.4%. Un dels objectius del sector eòlic és garantir la màxima disponibilitat dels parcs eòlics sense augmentar els costos de manteniment. Una forma pràctica d'aconseguir aquest objectiu és l'aplicació de tècniques de monitoratge efectives. En aquest context, el present treball té per objectiu l'aplicació de la regressió conformal per al monitoratge d'aerogeneradors a partir de la informació registrada en el Sistema de Supervisió, Control i Adquisició de Dades (SCADA). Les discrepàncies entre les dades observades i els predits mitjançant el model podrien ser indicatius de la presència de fallades o degradacions en els components del sistema. Per a avaluar aquesta discrepància s'utilitza una mesura de no conformitat la qual s'obté utilitzant com a algorisme subjacent màquina de suport vectorial (SVM). Finalment, es presenta un cas d'aplicació a aerogeneradors d'un parc eòlic situat a Espanya.

Paraules claus: Energia eòlica, Màquines de Suport Vectorial, regressió conformal.

ABSTRACT

In Spain, the wind energy sector has consolidated as one of the main sources of renewable energy with a coverage of demand of 19.4%. One of the objectives of the wind power sector is to ensure maximum availability of wind farms without increasing maintenance costs. A practical way to achieve this goal is to apply effective monitoring techniques. In this context, the present work aims to apply the conformal regression for the monitoring of wind turbines based on the information recorded in the Supervisory Control and Data Acquisition System (SCADA). The discrepancies between the observed and predicted data using the model could be indicative of the presence of faults or degradations in the components of the system. In order to evaluate this discrepancy a non-conformance measure is used which is obtained using the Support Vector Machine (SVM) as the underlying algorithm. Finally, a case of application to wind turbines of a wind farm located in Spain is presented.

Keywords: Wind energy, Support Vector Machines, conformal regression.

AGRADECIMIENTOS

Agradezco a Dios por darme vida, salud y sabiduría para ver materializado este proyecto, por guiar mis pasos siendo luz en el camino, cuidándome y dándome la fortaleza para continuar y encarar las adversidades con entereza.

Al Ministerio de Educación Superior, Ciencia y Tecnología (MESCyT) y al Gobierno de la República Dominicana por la financiación recibida a través del "Programa de Especialidades, Maestrías y Doctorados en Instituciones de Educación Superior extranjeras" orientadas al desarrollo nacional y a la promoción de la competitividad y la innovación en los sectores productivos y de servicios del país, en el curso 2016-2017.

A mi tutora Ana Isabel Sánchez Galdón, realizar este trabajo de la mano de usted ha sido una experiencia enriquecedora, por su valioso apoyo, orientación y dedicación para la realización de este proyecto, gracias por invertir su tiempo y compartir sus conocimientos con toda disposición.

A mi familia, especialmente a mis padres, que siempre han estado a mi lado brindándome su apoyo para mi desarrollo y crecimiento personal y profesional.

A mi novio, Darwin Leonardo, quien ha estado a mi lado no solo en esta etapa tan importante de mi vida, sino en todo momento ofreciéndome lo mejor y buscando lo mejor para mi persona.

A mis compañeros de piso, por brindarme su respaldo y colaboración en todo el trayecto de esta nueva etapa de mi vida.

CONTENIDO

ÍNDICE DE FIGURAS

ÍNDICE DE TABLAS

CAPÍTULO 1

1. INTRODUCCIÓN

Los aerogeneradores producen electricidad aprovechando la energía natural del viento para impulsar un generador. El viento es una fuente de energía limpia, sostenible que nunca se agota, y la transformación de su energía cinética en energía eléctrica no produce emisiones.

Los aerogeneradores[1] son la evolución natural de los molinos de viento y hoy en día son aparatos de alta tecnología. La mayoría de las turbinas generan electricidad desde que el viento logra una velocidad de entre 3 y 4 m/s, alcanzando su máxima potencia con una velocidad del viento de 15 m/s. Con el objetivo de prevenir daños los aerogeneradores se desconectan cuando hay tormentas con vientos que soplan a velocidades medias superiores a 25 m/s durante un intervalo temporal de 10 minutos.

Como afirma Erich Hau (Hau, 2000), en 1988, la idea de utilizar la energía del viento para generar electricidad seguía siendo el sueño de pocos entusiastas y no fue tomada muy seriamente por la industria eléctrica establecida. Hoy en día, la generación de energía a partir de la energía eólica ha ganado su lugar legítimo en el espectro de suministro de electricidad. En casi todos los países, cada vez se utilizan más aerogeneradores para complementar las centrales eléctricas convencionales y su número está aumentando en todo el mundo.

En España, el sector eólico se ha destacado por su desarrollo industrial y la potencia instalada desde la última década del siglo XX consolidándose como uno de los principales líderes mundiales en esta energía renovable. Concretamente, España se sitúa como el quinto país del mundo por potencia eólica instalada después de China, Estados Unidos, Alemania e India. La potencia instalada en el año 2016 era de (22988+38) MW siendo la tercera tecnología del sistema eléctrico español con una cobertura de la demanda del 19.4%.

Actualmente, un objetivo de la industria eólica es garantizar la máxima disponibilidad de los aerogeneradores sin aumentar los costes de mantenimiento. Una forma práctica de conseguir este objetivo es la aplicación de técnicas de monitorización efectivas.

[1] http://eoliccat.net/la-tecnologia/principios-de-la-energia-eolica/como-funciona-un-aerogenerador/?lang=es

En este contexto, el objetivo fundamental del presente trabajo se centra en la monitorización de la tendencia del funcionamiento de aerogeneradores mediante el seguimiento de parámetros obtenidos del Sistema de Supervisión, Control y Adquisición de Datos (SCADA) que permitan anticiparse al fallo de los equipos mediante la detección e identificación de desviaciones en el comportamiento del aerogenerador que podrían indicar la degradación del mismo permitiendo una mejor gestión de los recursos destinados al mantenimiento.

En este estudio se evaluará la calidad de los resultados a través del uso de metodologías flexibles como la técnica de Predictores Conformales (Conformal Prediction, CP). De acuerdo con Glenn Shafer and Vladimir Vovk, la predicción conformal puede usarse con cualquier método de predicción de puntos para clasificación o regresión, incluyendo Máquinas de Soporte Vectorial, Árboles de Decisión, Redes Neuronales y Predicción Bayesiana. Partiendo del método de predicción de puntos, construimos una medida de no conformidad, que mide cuán inusual es un ejemplo con respecto a ejemplos anteriores, y el algoritmo conformal convierte esta medida de no conformidad en regiones de predicción. Esta teoría permite la asociación de medidas de fiabilidad a las predicciones efectuadas por los algoritmos utilizados en el aprendizaje automático.

Los Predictores Conformales permiten además determinar las medidas de no conformidad utilizando algoritmos subyacentes como las Máquinas de Soporte Vectorial (Support Vector Machine, SVM), la cual es la que utilizaremos en este estudio.

1.1 Motivación

Los aerogeneradores son máquinas empleadas para transformar la energía eólica en energía mecánica. La energía eólica es empleada para generar electricidad, lo habitual para generar la electricidad es instalar diversos aerogeneradores juntos, que forman un parque eólico, y de esta forma aprovechar mejor los recursos de viento del lugar.

Un aspecto fundamental de la explotación de parques eólicos es poder operar un monitoreo eficiente y efectivo del sistema, realizar mantenimiento o reemplazar piezas cuando sea necesario. El exceso de tiempo de inactividad de los aerogeneradores debido a fallas puede incrementar costos operativos que afectan tanto los ingresos anuales como los costos para el consumidor. Lo ideal sería que el operador se diera cuenta lo antes posible de cualquier deterioro o fallas del equipo y de esta forma poder planificar con anticipación las acciones que deben emplearse.

Por lo que en este trabajo se aplica la técnica de Predictores Conformal para la monitorización de aerogeneradores la cual nos permitirá caracterizar su funcionamiento.

1.2 Estructura de la memoria

El presente documento está compuesto por cinco capítulos, los cuales se encuentran organizados de la siguiente manera:

- **Capítulo 1:** en este capítulo se presenta de forma resumida el contexto en el que se desarrolla el presente trabajo destacando la importancia de la energía eólica en el mercado eléctrico español el objetivo del presente trabajo.

- **Capítulo 2:** en este capítulo se presenta una revisión del estado del arte donde se describen las diferentes herramientas que se han presentado en la literatura para la monitorización del estado de aerogeneradores.

- **Capítulo 3:** en este capítulo se presentan los elementos teóricos de la metodología a utilizar para resolver el problema planteado, se muestran las características más importantes de la Predicción Conformal (CP) y las Máquinas de Soporte Vectorial (SVM).

- **Capítulo 4:** se presenta el caso de aplicación de los predictores conformales para la monitorización de los aerogeneradores y los resultados obtenidos.

- **Capítulo 5:** en este capítulo se presentan las conclusiones obtenidas a través de la aplicación de la metodología empleada.

  Finalmente se presentan las referencias bibliográficas y de consultas utilizadas para su realización.

1.3 Objetivos

El objetivo principal de este trabajo se centra en la aplicación de regresión conformal para la monitorización de aerogeneradores, el cual se realizará a través del algoritmo de la Máquina de Soporte Vectorial. Dentro de los objetivos específicos, tenemos los siguientes:

- Obtener modelos de regresión mediante la técnica de las Máquinas de Soporte Vectorial (SVM) que permitan predecir diferentes variables que caracterizan el comportamiento de un aerogenerador (por ejemplo, la potencia, temperatura devanados, etc.) en condiciones de operación normales.
- Obtener medidas de no conformidad con el método de Predictor Conformal, utilizando como algoritmo subyacente SVM.
- Monitorizar el comportamiento del aerogenerador mediante la detección de anomalías en las medidas de no conformidad.

CAPÍTULO 2

2. ESTADO DEL ARTE

En la literatura se han propuesto diferentes aproximaciones para la monitorización de la condición de aerogeneradores. Entre estas aproximaciones destacan las enfocadas al uso de datos obtenidos de sistemas SCADA. Los parámetros típicamente registrados por los sistemas SCADA se pueden clasificar en: a) parámetros ambientales, b) características eléctricas, c) temperaturas de componentes y d) variables de control. En general, estos parámetros se registran cada 10 minutos y suelen corresponder, dependiendo del parámetro, a valores medios, máximos o mínimos en este periodo.

(Tautz-Weinert and Watson, 2017) presentan una revisión exhaustiva de las diferentes aproximaciones publicadas en la literatura enfocadas al uso de datos procedentes de sistemas SCADA para la monitorización del estado de aerogeneradores. Los autores clasifican estas aproximaciones en: a) análisis de tendencias, b) clustering, c) modelos de comportamiento normal, d) modelos de daño y e) evaluación de alarmas y sistemas expertos.

El presente trabajo se enmarca dentro de los denominados modelos de comportamiento normal. Las aproximaciones basadas en modelos de comportamiento normal usan la idea de detectar anomalías a partir de modelos obtenidos mediante una muestra de entrenamiento obtenida bajo condiciones normales de operación del aerogenerador. El residuo, $\varepsilon_i = |\hat{y}_i - y_i|$, se utiliza como un indicador de posible desviación (fallo) desde la operación normal.

Dentro de los modelos de comportamiento normal las aproximaciones más simples están basadas en modelos de regresión lineal. En este contexto, (Garlick et al., 2009) usan un modelo de ARX lineal para detectar fallos en los rodamientos del generador. El modelo se ajusta con datos procedentes de SCADA de 12 turbinas correspondientes a un periodo de 3 años. En la misma línea, (Cross and Ma, 2015) investigan diferentes aproximaciones utilizando diferentes modelos ARX para la caracterización de las temperaturas de la multiplicadora y los devanados del generador a partir de la velocidad del viento y la potencia activa utilizando datos SCADA correspondientes a 26 turbinas y 16 meses de operación.

(Schlechtingen and Santos, 2010) utilizan una aproximación basada en la Full Signal ReConstruction (FSRC) para la predicción de la temperatura de los rodamientos del generador utilizando como variables independientes la potencia del generador, la temperatura de la

nacelle y la velocidad del eje. Los autores disponen de un histórico de datos de SCADA correspondientes a 10 aerogeneradores de 2 MW de potencia y un periodo de 14 meses. Para la detección del fallo proponen el uso de la media diaria de los residuos.

(Marton et al., 2013) combinan Análisis de Componentes Principales (PCA) y Mínimos Cuadrados Parciales (PLS) para caracterizar las relaciones más importantes entre variables para la predicción del estado del generador.

Otra de las técnicas extendidas en la literatura para la monitorización de aerogeneradores son las Redes Neuronales (RRNN). Las redes neuronales permiten determinar relaciones no lineales entre observaciones. En este contexto, (Garcia et al., 2006) desarrollan un sistema inteligente para mantenimiento predictivo llamado SIMAP basado en redes neuronales. (Zaher et al., 2009) implementan una red neuronal para la predicción de las temperaturas de la multiplicadora y el aceite de refrigeración utilizando datos de SCADA correspondientes a 2 años.

(Brandão & Carvalho, 2015) aplican una aproximación FSRC-RN para la detección de fallos en la multiplicadora y el generador en un parque eólico portugués con 13 turbinas de 2 MW de potencia y un parque de USA de 69 turbinas con una potencia de 1.5 MW.

(Schlechtingen & Santos, 2010) comparan un modelo lineal con dos modelos de RRNN en un estudio donde se utilizan datos de 14 meses desde 10 aerogeneradores. El modelo usa la temperatura del estator del generador, la temperatura de la nacelle, la potencia de salida y la velocidad del generador para predecir la temperatura de los cojinetes del generador.

Otras aproximaciones basadas en RRNN han sido presentadas por (Kusiak & Verma, 2012; Zhang & Wang, 2014; Li et al., 2014) entre otros.

Otra área de investigación se centra en el uso de lógica fuzzy, en este contexto (Schlechtingen et al., 2013) proponen un sistema de inferencia adaptativo neuro-fuzzy (ANFIS) para un modelo de comportamiento normal de aerogeneradores obteniendo modelos para 45 variables de interés tales, como por ejemplo, velocidad del rotor, temperatura en los rodamientos del generador o la potencia. En esta investigación se utilizaron datos de SCADA registrados durante un periodo de 3 años y correspondientes a 18 aerogeneradores.

CAPÍTULO 3

3. METODOLOGÍA

Como se había mencionado anteriormente, en este capítulo se presentarán los conceptos más importantes sobre la metodología a utilizar. La estructura del capítulo la dividimos en tres partes: la primera parte se explican los conceptos de los Predictores Conformales, la segunda parte se habla de las Máquinas de Soporte Vectorial y en la tercera se aborda sobre el concepto del software utilizado.

3.1 Predictores Conformal

La principal herramienta en el aprendizaje automático es la evaluación de la fiabilidad de las predicciones individuales, para determinar el espacio de aplicabilidad de un modelo predictivo, ya sea en el entorno de la regresión o clasificación.

Los predictores conformales (CP, abreviatura del inglés, Conformal Predictor) fueron desarrollados por Vovk, Gammerman y Shafer. La predicción conformal viene siendo utilizada con la finalidad de determinar la incertidumbre asociada a las predicciones proporcionadas por las técnicas tradicionales de clasificación y regresión, en el marco del aprendizaje supervisado, complementando las referidas predicciones con medidas de su fiabilidad (Makili, 2014).

La predicción conformal es una técnica de algoritmos independientes que utiliza la experiencia pasada para determinar niveles precisos de confianza en las nuevas predicciones, funciona con cualquier método predictivo como Máquinas de Soporte Vectorial (SVM, Support Vector Machines en inglés), *K* Vecinos más Cercano (Nearest Neighbor), Bosques Aleatorios (RF, Random Forests en inglés), que genera regiones de confianza para las predicciones individuales en el caso de la regresión y p-valores para las categorías en el ajuste de clasificación.

El método de predicción conformal produce predicciones establecidas que son automáticamente válidas en el sentido de que su probabilidad de cobertura incondicional es igual o superior a un nivel de confianza preestablecido. La idea de la predicción conformal es probar las dos etiquetas diferentes, 0 y 1, para el objeto de prueba, y para cualquiera de las

etiquetas postuladas para probar la suposición de aleatoriedad, verificando cuán bien el ejemplo de prueba se ajusta al conjunto de entrenamiento (Vovk, 2012).

3.1.1 Transducción e Inducción

La diferencia entre la transducción y la inducción que se aplican a los problemas de predicción fue formulada por Vapnik (Pérez, 2011). En la *Figura 3.1* se puede observar un esquema relativo al proceso de predicción.

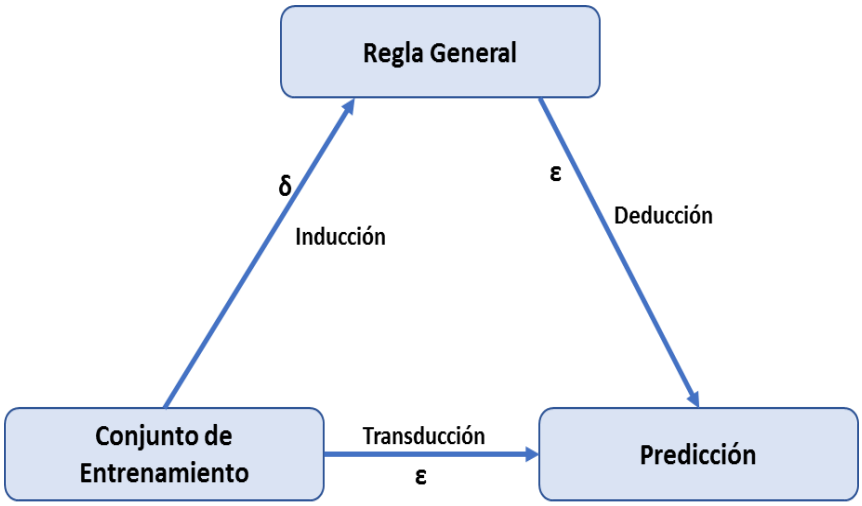

Figura 3.1. Predicción transductiva e inductiva

En la predicción inductiva los ejemplos se conducen con una regla la cual podríamos llamar de predicción o decisión, un modelo o una teoría. En este paso inductivo, cuando se muestra un nuevo ejemplo, se obtiene una predicción más general que en el paso deductivo. En la predicción transductiva se toma una entrada directa, de los ejemplos anteriores se pasa directamente a la predicción de un nuevo ejemplo (Pérez, 2011).

En el razonamiento del aprendizaje inductivo se realizan dos pasos, los cuales son inductivo y deductivo. En el primero, se hace el razonamiento en dirección del caso particular al general, estableciéndose una regla de predicción, a partir del conjunto de datos disponibles. En el segundo paso, al estar un nuevo ejemplo disponible, a partir de la regla general se deriva una predicción, en éste se razona en dirección del caso general al particular.

La transducción es una forma de inferencia en la que el razonamiento se realiza de particular hacia particular. Al emplear este tipo de inferencia, para cada objeto nuevo la predicción de las etiquetas se construye directamente, a partir de los ejemplos observados anteriormente, sin la construcción de una regla general. En este caso se implementa el principio que establece que al solucionar un problema se debe evitar resolver, como paso intermedio, un problema más general, siempre que la cantidad de información disponible sea limitada (Makili, 2014).

Dentro de las diferencias entre el aprendizaje transductivo e inductivo podemos destacar las siguientes:

La transducción, es más sencillo que la inducción. El objetivo de la inferencia inductiva es la estimación de un modelo o una función para todos los valores posibles del espacio de entradas, mientras que para la transducción el problema es más sencillo, ya que la inferencia se aplica de forma que se estiman los valores del modelo o la función sólo en un conjunto discreto de entradas (Cherkassky & Mulier, 2007).

Algún método inductivo establece una regla de predicción la cual puede ser implementada a varios objetos nuevos, tiene como objetivo que al aplicar la regla se obtenga una alta probabilidad de realizar predicciones con elevada precisión. Esto incluye el dominio de dos parámetros, los cuales son la precisión deseada para la regla (ε) y la probabilidad de alcanzar esa precisión (δ); mientras que los métodos transductivo por poseer un objetivo más modesto, implican el control de un parámetro llamado la probabilidad del error a ser tolerado (ε) al aplicar el método (Makili, 2014).

El modo transductivo implementa la inferencia directa a partir de datos de entrenamiento disponibles para predecir los valores de salida en puntos dados, sin un paso intermedio de estimar una regla general (función). Otra diferencia importante es que en la transducción las predicciones se hacen conjuntamente para todos los valores de entrada (de interés), mientras que durante el razonamiento inductivo la predicción (deducción) se realiza independientemente para cada valor de entrada (de interés) (Cherkassky & Mulier, 2007).

3.1.2 Predictores conformales

En este apartado se define el concepto de una medida de No conformidad, la cual nos permite medir qué tan diferente es un nuevo ejemplo de los ejemplos anteriores.

Cuando usamos este método, predecimos que un nuevo objeto tendrá una etiqueta que lo hace similar a los antiguos ejemplos de alguna manera especificada, y usamos el grado en que el tipo especificado de similitud se mantiene dentro de los ejemplos antiguos para estimar la confianza en la predicción. Los predictores conformales son, en otras palabras, "predictores de confianza".

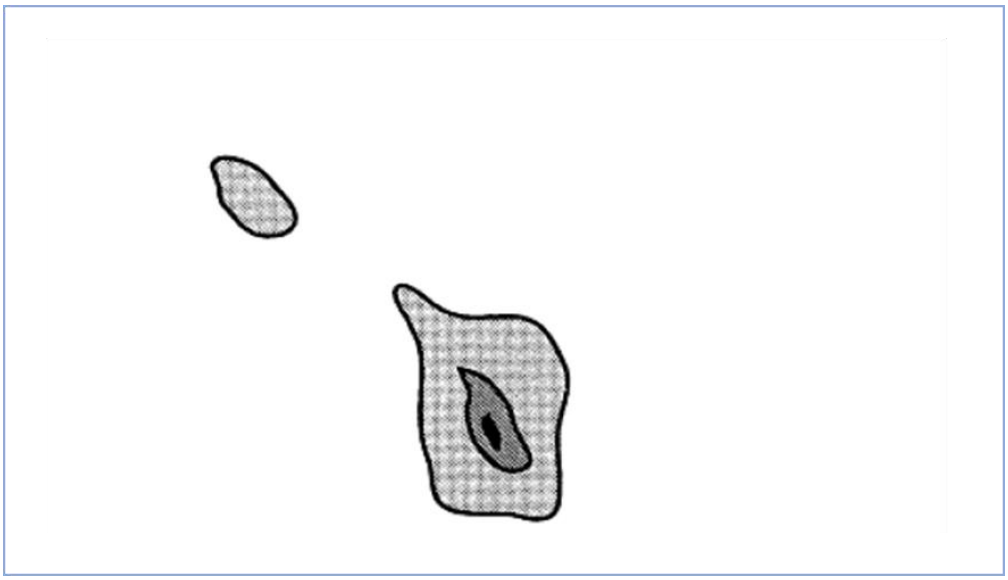

Figura 3.2. Ejemplo de una familia de conjuntos anidados en la predicción

Se puede observar en la *Figura 3.2* el ejemplo de una familia de conjuntos anidados en la predicción, donde se muestra en negro la zona con menor confianza, la confianza media en gris obscuro y muy confiados en gris claro (Vovk, Gammerman, et al., 2005).

Los dos indicadores principales de la calidad de los predictores establecidos y los predictores de confianza son los que llamamos validez (cuán confiables son) y eficiencia (cuán informativos son).

Considerando los algoritmos de predicción que emiten un conjunto de elementos de $\boldsymbol{Z}$ como su predicción; dicho conjunto se denomina conjunto de predicción. La declaración implícita en un conjunto de predicción es el que contiene el ejemplo de prueba Z_{l+1}, y el conjunto de

predicción se considera erróneo si y sólo si no contiene Z_{l+1} (Balasubramanian, Ho, et al., 2014).

Un predictor de conjunto es una función Γ que mapea cualquier secuencia $(z_1, \dots, z_l) \in Z^l$ a un conjunto $\Gamma(z_1, \dots, z_l) \subseteq \boldsymbol{Z}$ y satisface la siguiente condición de mensurabilidad:

$$\{(z_1, \dots, z_{l+1}) \qquad z_{l+1} \in \Gamma(z_1, \dots, z_l)\} \tag{3.1}$$

es mensurable en Z^{l+1}.

Con frecuencia se considera las familias anidadas de predictores establecidos dependiendo de un parámetro $\epsilon \in [0,1]$, el cual llamamos *nivel de significación*, reflejando la confiabilidad requerida de la predicción. La parametrización de la fiabilidad será tal que los valores más pequeños de ϵ corresponden a una mayor fiabilidad.

Un predictor de confianza es una familia $(\Gamma^\epsilon : \quad \epsilon \in [0,1])$ de predictores establecidos que están anidados en el siguiente sentido:

cuando $0 \leq \epsilon_1 \leq \epsilon_2 \leq 1$,

$$\Gamma^{\epsilon_1}(z_1, \dots, z_l) \supseteq \Gamma^{\epsilon_2}(z_1, \dots, z_l) \tag{3.2}$$

Para aplicar Predictor Conformal a un algoritmo tradicional se tiene que desarrollar una *medida de no conformidad* (en inglés *nonconformity measure*) basada en ese algoritmo. Esta medida marca la diferencia de un nuevo ejemplo de un conjunto (conjunto múltiple) de ejemplos anteriores. Las medidas de no conformidad se construyen utilizando como base el algoritmo tradicional al que se está aplicando CP, denominado algoritmo subyacente del resultante Predictor Conformal. En efecto, las medidas de no conformidad miden el grado en que el nuevo ejemplo no está de acuerdo con la relación atributo-etiqueta de los ejemplos anteriores, de acuerdo con el algoritmo subyacente de la CP. Hay que señalar que para cada algoritmo tradicional se pueden construir muchas medidas diferentes de no conformidad, y cada una de esas medidas define un CP diferente. Esta diferencia, no afecta la validez de los resultados producidos por los CPs, solo afecta su eficiencia (Papadopoulos, Vovk, et al., 2011).

Dado $n \in \mathbb{N}$, $donde\ \mathbb{N} = \{1,2,\dots\}$ es el conjunto de números naturales. Una medida n de no conformidad es una función mensurable A que asigna a cada secuencia $(z_1, \dots z_n)$ de n ejemplos una secuencia $(\alpha_1, \dots, \alpha_2)$ de n números reales los cuales son equivariante con respecto a las permutaciones: para cualquier permutación $\pi\ de\ \{1, \dots, n\}$,

$$(\alpha_1, \dots, \alpha_n) = A(z_1, \dots, z_n) \Longrightarrow \left(\alpha_{\pi(1)}, \dots, \alpha_{\pi(n)}\right) = A\left(z_{\pi(1)}, \dots, z_{\pi(n)}\right) \qquad (3.3)$$

Dado $n = l + 1$. El predictor conformal determinado por A como una medida de no conformidad es definido por:

$$\Gamma^{\epsilon}(z_1, \dots, z_l) := \{z \mid p^z > \epsilon\}, \qquad (3.4)$$

donde para cada $z \in \boldsymbol{Z}$ el p-valor correspondiente p^z es definido por:

$$p^z := \frac{|\{i=1,\dots,l+1 \mid \alpha_i^z \geq \alpha_{l+1}^z\}|}{l+1} \qquad (3.5)$$

y la secuencia correspondiente de los valores de no conformidad es definida por:

$$(\alpha_1^z, \dots, \alpha_{l+1}^z) := A(z_1, \dots, z_l, z) \qquad (3.6)$$

Similarmente, el predictor conformal determinado por A como una medida de conformidad es definido por (3.4) – (3.6) con $\alpha_1^z \geq \alpha_{l+1}^z$ en (3.5) remplazado por $\alpha_1^z \leq \alpha_{l+1}^z$ (para este caso, la ecuación (3.6) se denominan como valores de conformidad).

Es fácil observar que el conjunto de predicción de la ecuación (3.4) producido por el predictor conformal Γdepende de ϵ a través de:

$$[\epsilon]_l := \frac{\lfloor \epsilon(l+1) \rfloor}{l+1}$$

Observación: $\Gamma^{\epsilon 1} = \Gamma^{\epsilon 2}$ cuando $\epsilon_1 \, y \, \epsilon_2$ son l -equivalente, en el sentido $[\epsilon_1]_l = [\epsilon_2]_l$. Teniendo en cuenta que $[\epsilon]_l$ es el valor más pequeño que es l-equivalente para ϵ.

3.2 Máquinas de Soporte Vectorial

Las Máquinas de Soporte Vectorial (SVMs por su nombre en inglés Support Vector Machines) son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik que pueden ser aplicados a problemas de regresión o clasificación de datos. Las SVMs inicialmente fueron introducidas para clasificar clases de objetos linealmente separables, éstas actualmente se utilizan para resolver otros tipos de problemas como de regresión, agrupamiento, multiclasificación. Las Máquinas de Soporte Vectorial buscan un hiperplano que separa de forma óptima los puntos de una clase de la otra. En la actualidad un gran número de investigadores han utilizado las SVMs en aplicaciones como son identificador de firmas, reconocimiento de objetos, recuperación de información, detección de rostros, categorización de textos, entre otras utilidades.

Las SVMs poseen la capacidad de construir regiones de decisión no lineales de una forma discriminativa a través de la introducción de una función núcleo (Goddard, Silva et al., 2000).

Según Christopher J.C. Burgues (Burges, 1998) el problema que propulsó el desarrollo inicial de las Máquinas de Soporte Vectorial se presenta en varios aspectos, como son: la compensación de la varianza de sesgo, el control de capacidad. En términos generales, para una tarea determinada de aprendizaje, con una cantidad dada de datos de entrenamiento finitos, se obtendrá el mejor rendimiento de generalización si se alcanza el equilibrio adecuado entre la precisión obtenida en ese conjunto de entrenamiento particular y la "capacidad" de la máquina, es decir, la capacidad de la máquina para aprender cualquier conjunto de entrenamiento sin error.

(Betancourt, 2005) afirma que la teoría de la SVM está basada en la idea de minimización de riesgo estructural. Las SVM han mostrado tener un gran desempeño y han sido introducidas como herramientas poderosas para resolver problemas de clasificación. Una SVM primero mapea los puntos de entrada a un espacio de características de una dimensión mayor (i.e: si los puntos de entrada están en $\Re^2$ entonces son mapeados por la SVM a $\Re^3$) y encuentra un hiperplano que los separe y maximice el margen *m* entre las clases en este espacio, como se aprecia en la *Figura 3.3*.

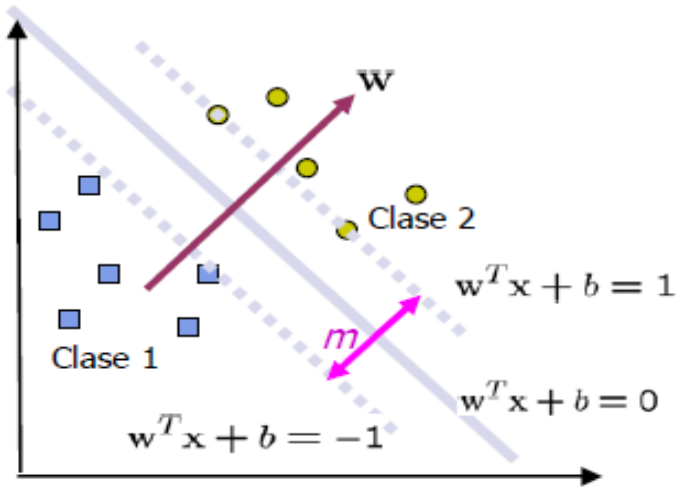

Figura 3.3. Frontera de decisión

Los datos utilizados para hallar la frontera de decisión (hiperplano, el cual divide ambas clases) se les denomina *vectores de entrenamiento* o *de aprendizaje*. La Máquina de Soporte Vectorial encuentra el hiperplano óptimo utilizando el producto punto con funciones en el espacio de características, los cuales son llamadas kernels. La solución del hiperplano óptimo puede ser escrita como la combinación de unos pocos puntos de entrada llamados *vectores de soporte*.

A través de unos datos de entrada $\mathbf{x_i}$, las Máquinas de Soporte Vectorial proporcionarán su clase de acuerdo con la regla de clasificación $f(\mathbf{x_i}) = \text{signo}(h(\mathbf{x_i}))$.

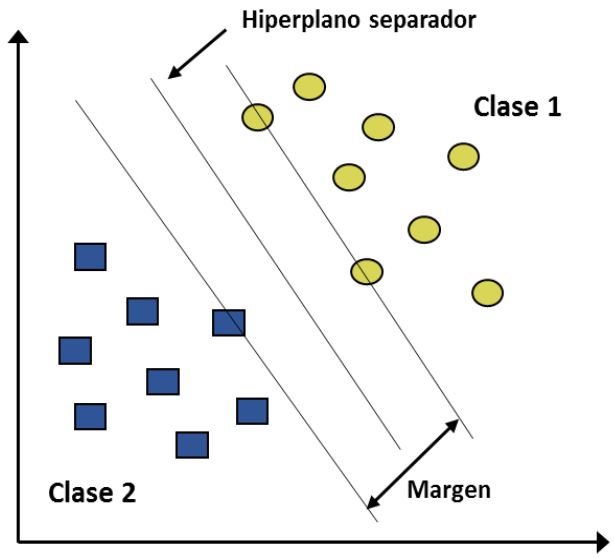

Figura 3.4. Separación de un conjunto de datos

Tipos de Kernel	Expresión
Polinómica	$K(x,x') = (x^T.x' + c)^d \qquad c \in \mathbb{R}, \quad d \in \mathbb{N}$
Gaussiana	$K(x,x') = e^{\left(\frac{-\|\|x-x'\|\|^2}{2\sigma^2}\right)} \qquad \sigma > 0$
Sigmoidal	$K(x,x') = tanh(s(x^T.x') + r) \quad s,r \in \mathbb{R}$

Tabla 3.1. Tipos de kernels utilizados en SVM

3.1.3 Máquinas de Soporte Vectorial para regresión

Las máquinas de Soporte Vectorial además de utilizarse para clasificación se pueden adaptar para resolver problemas de regresión mediante la introducción de una función de pérdida alternativa, éstas son llamadas por el acrónimo SVR (por sus siglas en inglés Support Vector Regression).

La función de pérdida debe modificarse para incluir una medida de distancia (Gunn, 2010). En la *Figura 3.5* se muestran cuatro funciones de pérdidas.

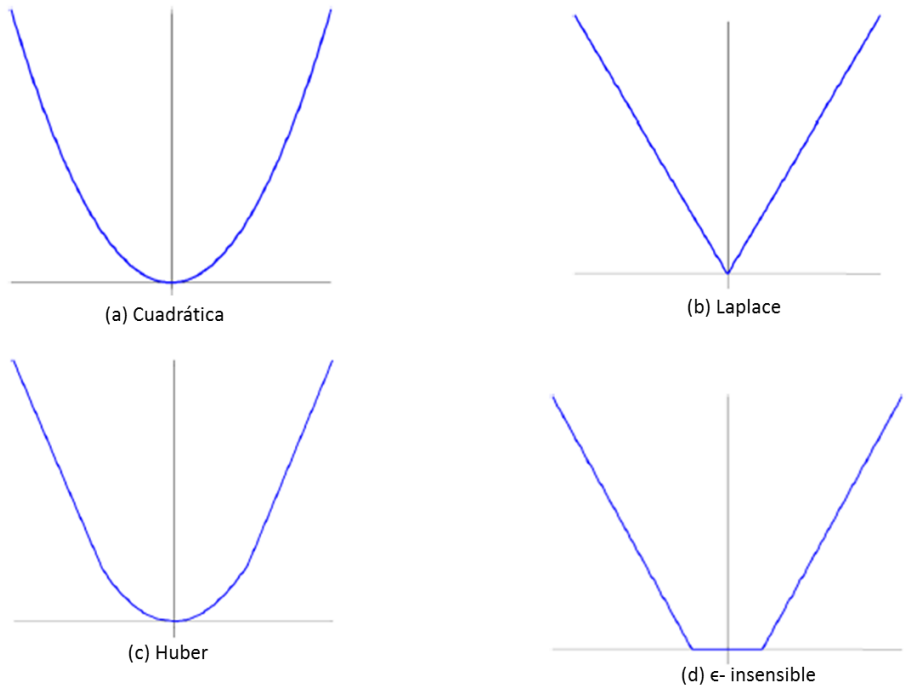

Figura 3.5. Funciones de Pérdidas

La función de pérdida en la *Figura 3.5 (a)* pertenece al criterio de error de mínimos cuadrados convencional. La *Figura 3.5 (b)* muestra la función de pérdida laplaciana, la cual es menos sensible a los datos atípicos que la función de pérdida cuadrática. Huber recomendó la función de pérdida de la *Figura 3.5 (c)* como una función robusta que tiene propiedades óptimas cuando la distribución subyacente es desconocida. Esas tres funciones no producirán dispersión en los vectores soporte. La función de pérdida presentada en la *Figura 3.5 (d)* fue propuesta por Vapnik como un acercamiento a la función de pérdida de Huber que permite obtener un escaso conjunto de vectores soporte.

- **Regresión lineal**

Dado un problema de aproximación de un conjunto de datos:

$$D = \{(x_1, y_1), \dots, x_n y_n)\}, \qquad donde \quad x_i \in \mathbb{R}^d, y_i \in \mathbb{R} \tag{3.7}$$

Proporcionado los datos, queremos hallar una función lineal de la forma:

$$f(x) = \langle w, x \rangle + b \tag{3.8}$$

La función de regresión óptima está dada por el mínimo de la funcionalidad:

$$\Phi(w, \xi) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i^- + \xi_i^+) \tag{3.9}$$

Donde C es un valor pre-especificado, y ξ_i^-, ξ_i^+ Son variables de holgura que representan las menores limitaciones en las salidas del sistema.

- **Función de pérdida ϵ-insensible**

Se utiliza esta función para admitir un nivel de ruido en los ejemplos de entrenamiento y así poder debilitar la condición de error entre el valor predicho por la función y el valor real. La función ϵ-insensible, $\mathbf{L_\epsilon}$, se encuentra caracterizada por ser una función lineal con una zona insensible, en la que el error es nulo, ver *Figura 3.6*, en la cual se muestra la relación entre las variables ξ_i^-, ξ_i^+ y la función de pérdida, $\mathbf{L_\epsilon}$.

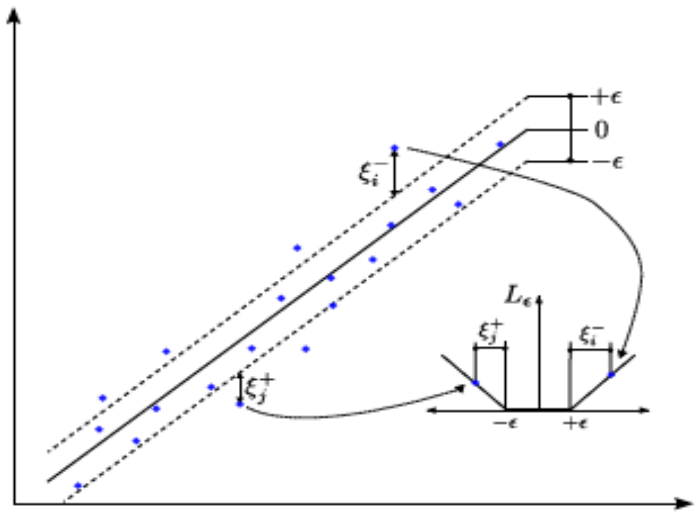

Figura 3.6. SVM con margen blando

Usando la función de pérdida:

$$L_\epsilon(y) = \begin{cases} 0 \\ |f(x) - y| - \epsilon \end{cases} \qquad si\ |f(x) - y| < \epsilon \qquad (3.10)$$

La solución está dada por:

$$maxW\ (\alpha, \alpha^*) = \max(\alpha, \alpha^*) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)\langle x_i, x_j\rangle + \sum_{i=1}^{n}\alpha_i\,(y_i - \epsilon) - \alpha_i^*(y_i + \epsilon) \qquad (3.11)$$

$$\text{s.a.} \qquad 0 \le \alpha_i\alpha_i^* \le C, \qquad i = 1, \dots, n \qquad (3.12)$$
$$\sum_{i=1}^{n}(\alpha_i\alpha_i^*) = 0$$

Resolviendo la ecuación (3.11) con las restricciones de la ecuación (3.12) se determinan los multiplicadores de Lagrange, α, α^*, la función de regresión viene dada por la ecuación (3.8),

$$\text{donde:} \qquad \bar{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i \qquad (3.13)$$
$$\bar{b} = -\frac{1}{2}\langle \bar{w}, (x_r + x_s)\rangle$$

Las condiciones KKT las cuales son satisfecha por la solución:

$$\bar{\alpha}_i\bar{\alpha}_i^* = 0, \qquad i = 1, \dots, n \qquad (3.14)$$

Los vectores soporte son aquellos puntos donde uno de los multiplicadores de Lagrange es mayor que cero. Cuando ϵ = 0, obtenemos la función de pérdida L1 y el problema de optimización es simplificado:

$$min\,\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i\beta_j\langle x_i, x_j\rangle - \sum_{i=1}^{n}\beta_i y_i \qquad (3.15)$$

s.a. $-C \leq \beta_i \leq C \qquad i = 1, \ldots., n$ (3.16)

$\sum_{i=1}^{n} \beta_i = 0$

La función de regresión viene dada por la ecuación (3.8), donde:

$\bar{w} = \sum_{i=1}^{n} \beta_i x_i$ (3.17)

$\bar{b} = -\frac{1}{2}\langle \bar{w}, (x_r + x_s) \rangle$

- **Función de pérdida cuadrática**

Empleando la función de pérdida:

$L_{cuadrática}(f(x) - y) = (f(x) - y)^2$ (3.18)

La solución está dada por:

$maxW(\alpha, \alpha^*) = max -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_{i-}\alpha_i^*)\left(\alpha_j - \alpha_j^*\right)\langle x_i, x_j \rangle + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i - \frac{1}{2C}\sum_{i=1}^{n}(\alpha_i^2 + (\alpha_i^*)^2$ (3.19)

La optimización correspondiente puede ser simplificada empleando las condiciones Karush-Kuhn-Tucker (KKT) de la ecuación (3.14) y por tanto esto implica $\beta_i^* = |\beta_i|$. El problema de optimización resultante es:

$min\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i\beta_j\langle x_i x_j \rangle - \sum_{i=1}^{n}\beta_i y_i + \frac{1}{2C}\sum_{i=1}^{n}\beta_i^2$ (3.20)

s.a. $\sum_{i=1}^{n}\beta_i = 0$ (3.21)

La función de regresión está dada por las ecuaciones (3.8) y (3.17)

- **Función de pérdida Huber**

Empleando la función de pérdida:

$$L_{huber}(f(x)-y)=\begin{cases}\frac{1}{2}(f(x)-y)^2 & de\ otra\ forma \quad |f(x)-y|<\mu \\ \mu|f(x)-y|-\frac{\mu^2}{2}\end{cases} \tag{3.22}$$

La solución viene dada por:

$$maxW(\alpha,\alpha^*)=max-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle x_i,x_j\rangle+\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)y_i-\frac{1}{2C}\sum_{i=1}^{n}(\alpha_i^2+(\alpha_i^*)^2)\mu \tag{3.23}$$

Como resultado del problema de optimización se tiene:

$$min\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i\beta_j\langle x_i,x_j\rangle-\sum_{i=1}^{n}\beta_i y_i+\frac{1}{2C}\sum_{i=1}^{n}\beta_i^2\mu \tag{3.24}$$

s.a. $$-C\le\beta_i\le C, \qquad i=1,\dots,n \tag{3.25}$$

La función de regresión está dada por las ecuaciones (3.8) y (3.21).

- **Regresión no lineal**

De manera semejante a los problemas de clasificación, habitualmente se requiere un modelo no lineal para modelar adecuadamente los datos. De la misma manera que el acercamiento de los Vectores Soporte para Clasificación (SVC) no lineal, un mapeo no linear puede utilizarse para mapear los datos en un espacio de característica dimensional alta en la cual se realiza la regresión lineal. Nuevamente se aplica el enfoque del núcleo para abordar el curso de la dimensionalidad. Usando la función de pérdida ϵ-insensible, la solución de Soporte Vectorial para Regresión (SVR) no lineal viene dada por:

$$\max W\ (\alpha,\alpha^*)=max\sum_{i=1}^{n}\alpha_i^*(y_i-\epsilon)-\alpha_i(y_i+\epsilon)-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(x_{i,}x_j) \tag{3.26}$$

s.a. $0 \leq \alpha_i, \alpha_i^* \leq C \qquad i = 1, \dots, n$ (3.27)

$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$

Resolviendo la ecuación (3.26) con las restricciones de la ecuación (3.27) se determinan los multiplicadores de Lagrange, $\alpha_i - \alpha_i^*$, y la función de regresión viene dada por:

$$f(x) = \sum_{SVs}(\bar{\alpha}_i, \bar{\alpha}_i^*)K(x_i, x) + b \tag{3.28}$$

Donde

$$\langle \bar{w}, x \rangle = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x_j) \tag{3.29}$$
$$\bar{b} = -\frac{1}{2}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(K(x_i, x_r) + K(x_i, x_s)$$

Al igual que con los Vectores Soporte para Clasificación, la restricción de igualdad puede ser eliminada si el Kernel contiene un término de sesgo, *b* siendo acomodado dentro de la función Kernel y la función de regresión dada por:

$$f(x) = \sum_{i=1}^{n}(\bar{\alpha}_i - \bar{\alpha}_i^*)K(x_i, x) \tag{3.30}$$

Para las demás funciones de pérdida vistas anteriormente en la regresión lineal, los criterios de optimización se obtienen de manera similar reemplazando el producto de punto por una función de núcleo.

La función de pérdida insensible es atractiva porque a diferencia de las funciones de coste cuadrático y Huber, donde todos los puntos de datos serán vectores de soporte, la solución de soporte vectorial puede ser escasa. (Gunn, 2010).

3.3 Software R

El Software utilizado es R Studio, versión 3.3.3.

R es un entorno y lenguaje de programación con una orientación al análisis estadístico. En los últimos años se ha convertido en uno de los lenguajes más utilizados por la comunidad estadística en investigaciones siendo muy popular en el campo de minería de datos. Una de sus ventajas es la amplia gama de herramientas estadísticas que proporciona, las cuales incluyen análisis de datos y generación de gráficos.

R es un lenguaje de programación interpretado, de distribución libre, que se mantiene en un ambiente para el cómputo estadístico y gráfico. El término ambiente pretende caracterizarlo como un sistema totalmente planificado y coherente, en lugar de una acumulación gradual de herramientas específicas y poco flexibles (Santana & Farfán, 2014).

Respecto a las librerías concretas utilizadas en el siguiente trabajo, se ha utilizado el paquete "Conformal" (Cortes, 2016) para la obtención del modelo de regresión conformal el cual utiliza "Caret" (Khun, 2017) en la implementación de máquinas de vector soporte. Por último, el paquete utilizado para la detección de anomalías a partir de las medidas de no conformidad ha sido "Structchange".

CAPÍTULO 4

4 APLICACIÓN DE REGRESIÓN CONFORMAL PARA LA MONITORIZACIÓN DE AEROGENERADORES Y RESULTADOS

4.1 Aerogenerador y sus componentes

Un aerogenerador[2] es un dispositivo que convierte la energía cinética del viento en energía eléctrica. Las palas de un aerogenerador giran entre 13 y 20 revoluciones por minuto, según su tecnología, a una velocidad constante o bien a velocidad variable, donde la velocidad del rotor varía en función de la velocidad del viento para alcanzar una mayor eficiencia.

Los componentes de un aerogenerador se muestran en la *Figura 4.1.*

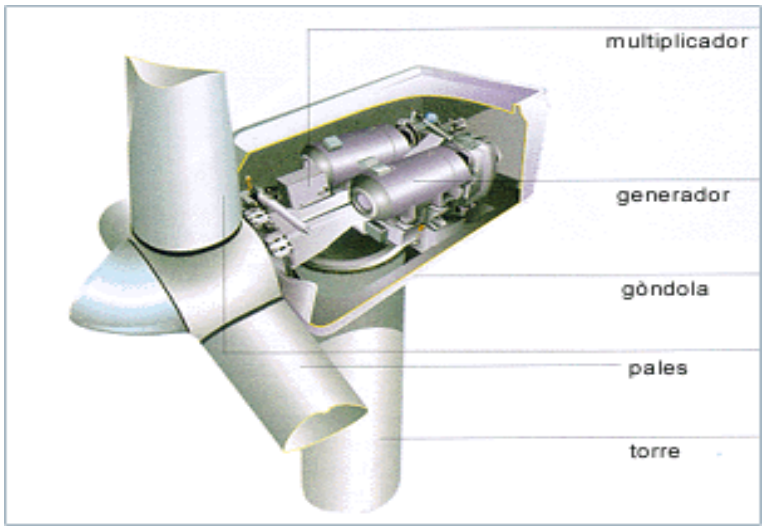

Figura 4.1. Componentes de un aerogenerador

[2] https://www.acciona.com/es/energias-renovables/energia-eolica/aerogeneradores/

- **Multiplicador:** es la transmisión que aumenta la velocidad de giro del eje para el funcionamiento del motor.

- **Generador:** convierte la energía mecánica producida por el rotor en energía eléctrica. El rotor es el conjunto que está formado por las palas y el eje al que van unidas.

- **Góndolas:** caja que protege la multiplicadora, el generador eléctrico y los sistemas de control, orientación y freno. Ésta se sitúa en la parte superior de la torre.

- **Palas:** son las aspas que giran, éstas son unas de las partes más importantes, debido a que se encargan de recoger la energía del viento. Normalmente se fabrican con una mezcla de fibra de vidrio y resina, y son tan firmes como las alas de un avión.

- **Torre:** es la estructura que soporta el peso de la góndola y mantiene elevadas las palas de la turbina del suelo.

4.2 Descripción de los datos

Los datos utilizados en el trabajo se han obtenido de un aerogenerador instalado en un parque eólico español y corresponden a la información registrada por un sistema SCADA en el periodo 2009-2012. Los valores son registrados cada 10 minutos y corresponde a valores medios, máximos o mínimos dependiendo del tipo de variable.

El aerogenerador analizado tiene una potencia de 2 MW (2000 kW) y una tensión de 690 V AC. Las palas del rotor tienen una longitud de 39 m y el rotor un diámetro de 80 m.

La mayoría de los componentes del sistema se monitorizan mediante un conjunto de sensores registrándose datos atmosféricos (velocidad del viento, temperatura ambiente, etc.), temperatura y presión de diferentes equipos (temperatura del aceite del grupo hidráulico, temperatura multiplicadora, etc.) e información relativa a la potencia (potencia total de salida, potencia reactiva, etc.).

Como paso previo a la realización de análisis los datos han sido preprocesados de la siguiente manera:

- Se eliminaron los registros en los cuales faltaban valores de una o varias de las variables analizadas.
- Se eliminaron variables para las cuales no se disponía del total del histórico o para las que se había modificado el criterio de registro de la información.

- Se eliminaron los registros que corresponden a periodos de tiempo en los cuales el aerogenerador se encontraba en mantenimiento correctivo o preventivo, así como las dos horas posteriores a la puesta en funcionamiento del aerogenerador tras la realización de la actividad de mantenimiento.

Tras el preprocesado se obtuvo un conjunto de datos constituido por 3000 observaciones y 23 variables. Todos estos registros reflejan estados de funcionamiento que reflejan un rendimiento del 100%. Para el entrenamiento de los modelos en condiciones normales de operación se utilizaron los datos correspondientes al año 2009 considerándose únicamente los registros que reflejan estados de funcionamiento con un rendimiento del 100%. En la *Tabla 4.1* se muestra una descripción de las variables disponibles.

No. Variables	Variables	Definición
1	time	Tiempo en que se toma cada muestra (mints).
2	tgbxoil	Temperatura del aceite de la multiplicadora (°C).
3	tgbxhss	Temperatura de la multiplicadora (°C).
4	tgenDE	Temperatura de los cojinetes del generador en el LA (°C).
5	tgenNDE	Temperatura de los cojinetes del generador en el LOA (°C).
6	tgenslr	Temperatura del disipador de calor (°C).
7	tgenMAX	Temperatura máxima en los devanados del generador (°C).
8	thydr	Temperatura del grupo hidráulico (°C).
9	phydr	Presión en el grupo hidráulico (mbar).
10	tnac	Temperatura en el interior de la góndola (°C).
11	tamb	Temperatura en el exterior de la góndola, medido por el anemómetro(°C).
12	txfmrMAX	Temperatura máxima en uno de los bobinados del transformador (°C).
13	wmsh	Velocidad de rotación del eje principal, (velocidad del rotor) (rpm).
14	maxwgen	Máxima velocidad del generador.
15	angbl	Ángulo de las palas o del sistema de Pitch (°).
16	velwind	Velocidad de viento media en el anemómetro (m/s).
17	angwind	Dirección del viento respecto de la góndola (°).
18	vgrid	Tensión de la red (Voltio).
19	pwrea	Potencia reactiva en el generador (KVA).
20	pwest	Potencia del estator del generador (kw).
21	pwrot	Potencia del rotor del generador (kw).
22	pwtot	Potencia del generador (kw).
23	prod	Producción total acumulada de la máquina (kwh).
24	state	Estado en el que se encuentra el aerogenerador durante los 10 minutos.

Tabla 4.1. Descripción de las variables

En el conjunto de datos disponibles correspondiente al aerogenerador analizado se observó que en octubre del año 2010 hay una sustitución del generador que es uno de los equipos de mayor coste. En este contexto, el objetivo del caso de aplicación se centra en analizar la aplicabilidad de la regresión conformal para detectar con antelación la degradación previa a dicho fallo con el objetivo de planificar de forma adecuada la sustitución del mismo minimizando las pérdidas económicas derivadas de la no disponibilidad del equipo durante un periodo prolongado. Para la detección de dicha degradación se han obtenido modelos para las siguientes variables:

- Potencia del generador (pwtot).
- Temperatura de los cojinetes del generador en el LOA (tgenNDE).

4.3 Construcción de modelos con SVM

Para la construcción de los modelos se han utilizado los datos correspondientes al comportamiento del aerogenerador en condiciones normales de operación correspondientes al año 2009 los cuales se han dividido en un 75% para entrenamiento y el 25% restante para validación, esta división se realizó a través del método *Hold Out*.

El método *hold-out* separa el conjunto de datos disponibles en dos subconjuntos, uno para entrenar el modelo y el otro para realizar la validación del mismo. Se crea el modelo con los datos de entrenamiento. Con el modelo creado se generan datos de salida los cuales se comparan con el conjunto de datos reservados para realizar la validación (Pérez-Planells, Delegido, et al., 2015).

En el contexto de la regresión conformal los datos del conjunto de entrenamiento se utilizan para definir como de diferente es un nuevo dato con respecto a los datos utilizados en la fase de entrenamiento. La cuantificación de la medida de no conformidad de un nuevo dato con respecto al conjunto de entrenamiento puede ser realizada mediante diferentes métricas. En este caso concreto se ha utilizado una medida estándar la cual viene dada por:

$$\boldsymbol{\alpha} = \frac{|\boldsymbol{y} - \hat{y}|}{\hat{\boldsymbol{\rho}}} \qquad (4.1)$$

siendo α el valor de la medida de no conformidad, y el valor observado de la variable independiente, $\hat{y}$ el valor predicho y $\hat{\rho}$ el error predicho el cual se evalúa mediante un modelo

de error. Tanto el modelo para la predicción de $\hat{y}$ como para la predicción del error $\hat{\rho}$ han sido obtenidos mediante SVR.

Los intervalos de confianza asociados a las predicciones se estiman a partir del modelo de predicción de la variable dependiente y , y del modelo de error.

4.3.1 Potencia (pwtot)

En la *Figura 4.2* se muestran los inputs incluidos en el modelo de regresión para la modelización de la potencia (pwtot). Dichas variables han sido seleccionadas de un estudio previo.

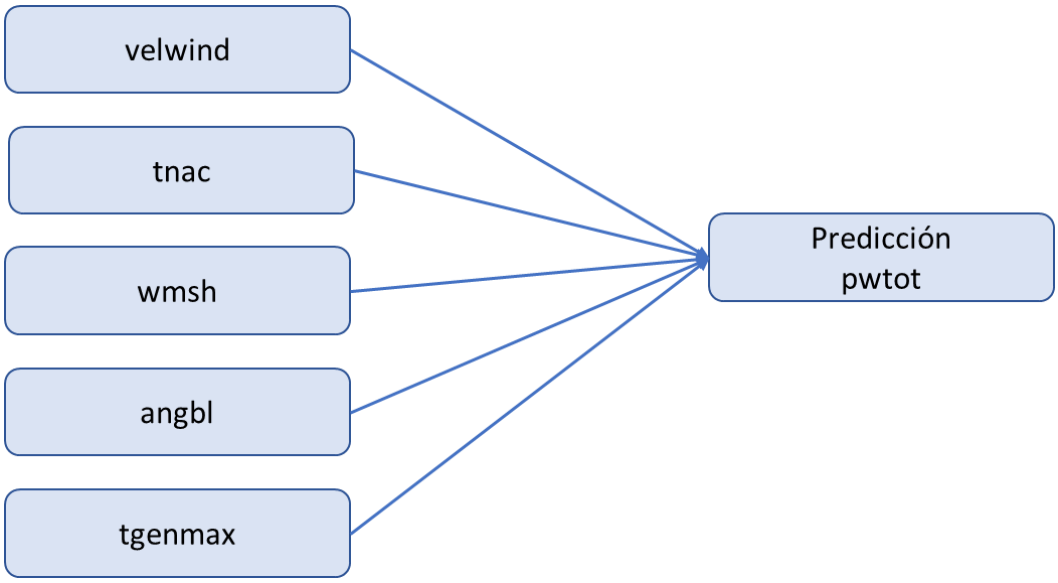

Figura 4.2. Variables incluidas en el modelo de predicción pwtot

Para la construcción del modelo SVR se utilizó la función núcleo o Kernel *radial.* Los parámetros de coste (C) y σ se determinaron mediante validación cruzada sobre el conjunto de entrenamiento seleccionando los valores de los parámetros para los cuales se obtuvo el menor valor del RMSE (Root Mean Squared Error). Los valores finales obtenidos de los parámetros son σ=0.015625 y C = 100.

En la *Tabla 4.2* se muestran los valores obtenidos del RMSE y del coeficiente de determinación para el modelo obtenido para la variable *pwtot*.

Parámetro	Valor
RMSE	90.58
R^2	0.98

Tabla 4.2. Medidas de bondad de ajuste del modelo

La *Figura 4.3* muestra los valores observados vs a los predichos de la variable respuesta *pwtot* en el conjunto de test. Asimismo, se representan los intervalos de confianza.

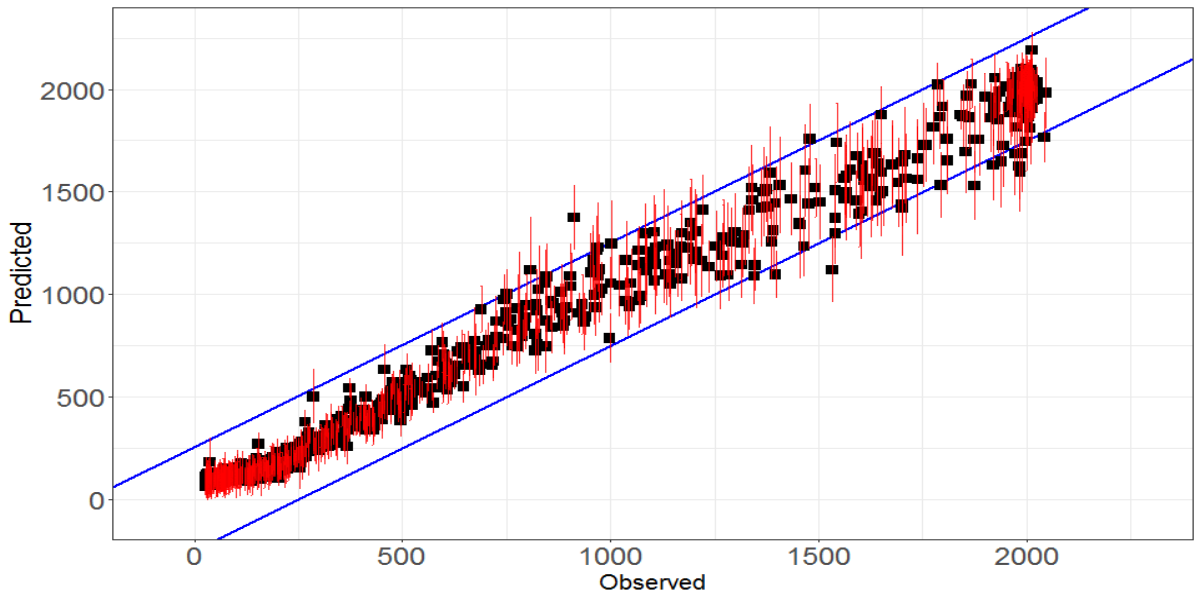

Figura 4.3. Observados vs Predichos

En la *Figura 4.4* se presenta el histograma del error de predicción de la variable pwtot (potencia del generador) utilizando un modelo SVR.

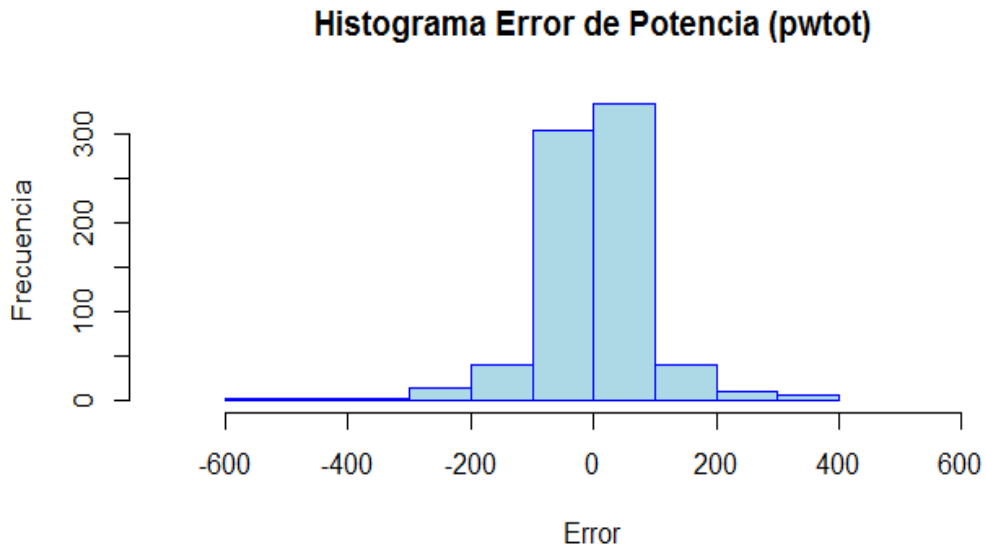

Figura 4.4. Histograma error de predicción (pwtot)

Una vez obtenido el modelo de predicción conformal que permite obtener la medida de no conformidad (MNC), ver ecuación (4.1), la misma se utiliza para la detección de comportamientos anómalos previos a la sustitución del generador en octubre de 2010.

Para detectar posibles cambios en el comportamiento de la medida de no conformidad se ha utilizado un test- F basado en el test de Chow (Chow, 1960). La *Figura 4.5* muestra los resultados obtenidos. Como el estadístico F supera el límite, existe evidencia de un cambio estructural (α=0.05) en el comportamiento de MNC. Dicho cambio se observa en agosto de 2010 permitiendo anticiparse a la ocurrencia del fallo con, aproximadamente, un mes y medio antes de su ocurrencia.

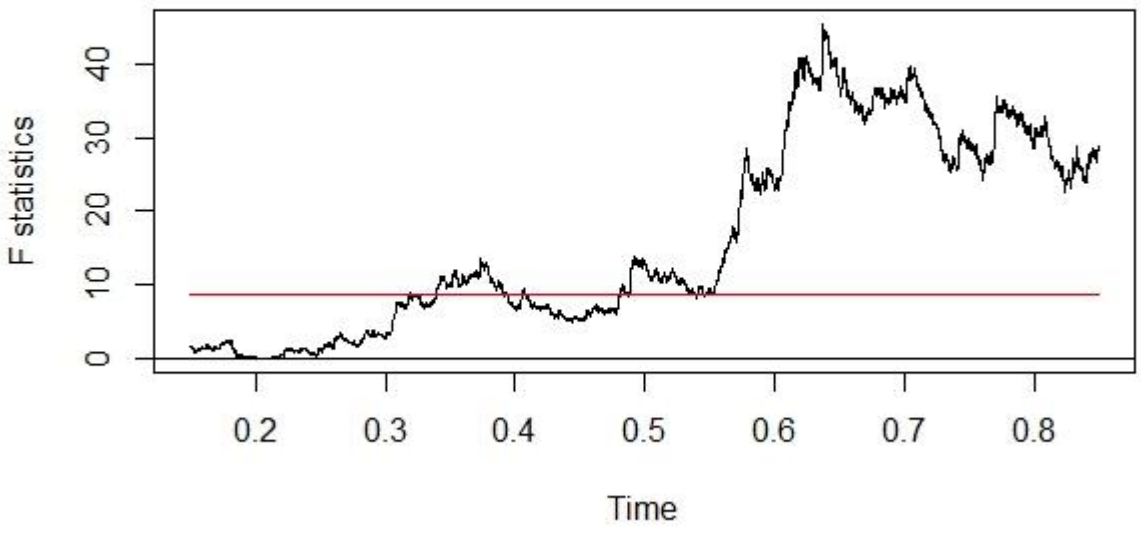

Figura 4.5. Estadístico F para la MNC de pwtot

4.3.2 Temperatura de los cojinetes del generador en el LOA (tgenNDE)

La segunda variable analizada ha sido la temperatura de los cojinetes del generador en el LOA (tgenNDE). En la *Figura 4.6* se muestran las variables incluidas en la predicción de la variable tgenNDE.

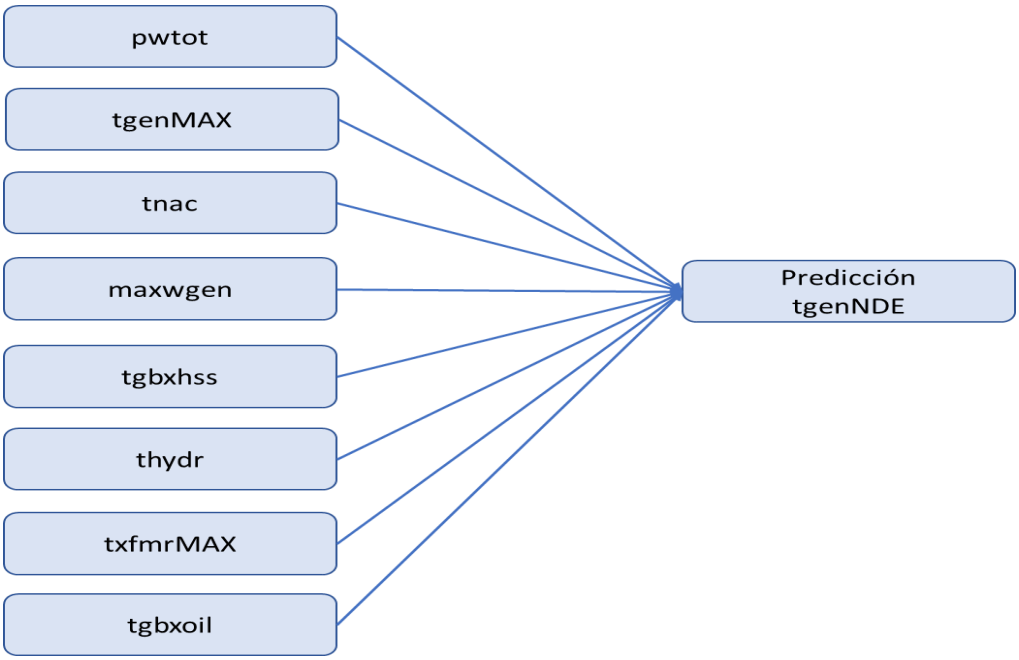

Figura 4.6. Variables incluidas en el modelo de predicción tgenNDE

Al igual que en el modelo presentado en la sección anterior, se utilizó la función núcleo radial para la construcción del modelo SVR obteniéndose los valores de los parámetros C y σ mediante validación cruzada sobre el conjunto de entrenamiento. Los valores de C y σ para los cuales se obtuvo el menor RMSE son σ=0.015 y C = 100.

En la *Tabla 4.3* se muestran los valores obtenidos del RMSE y del coeficiente de determinación para el modelo obtenido para la variable *tgenNDE*.

Parámetro	Valor
RMSE	1.956
R^2	0.913

Tabla 4.3. Medidas de bondad de ajuste del modelo

En la *Figura 4.7* se muestran las predicciones obtenidas de la variable respuesta tgenNDE a partir del conjunto de test junto con los intervalos de confianza obtenidos por regresión conformal.

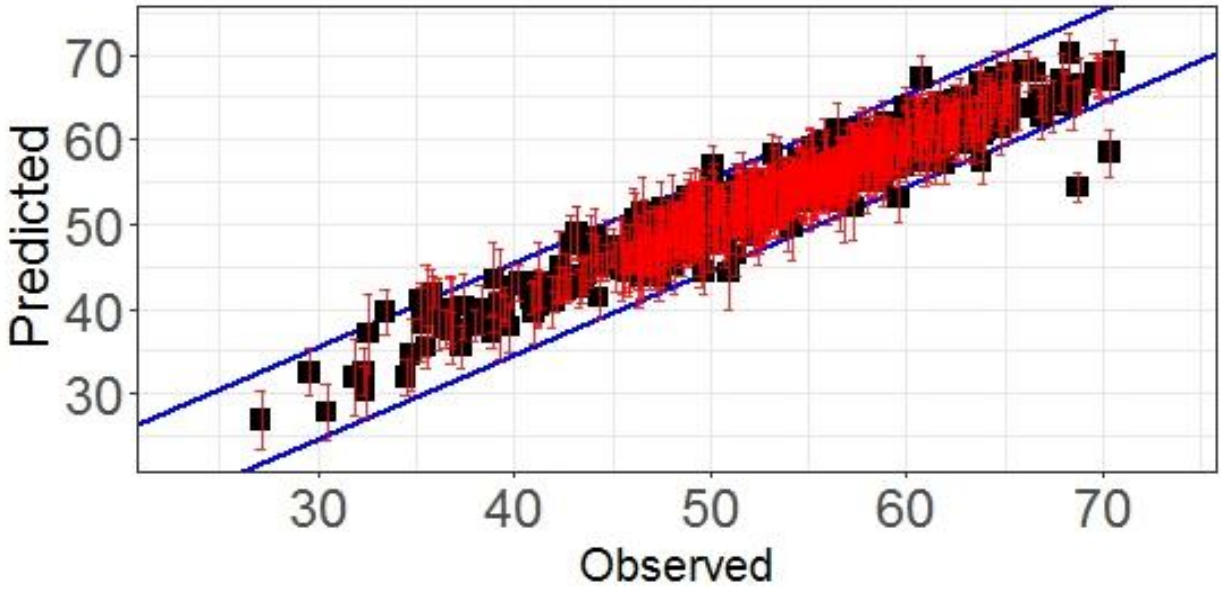

Figura 4.7. Observados vs Predichos

En la *Figura 4.8* se presenta el histograma del error de predicción de la variable tgenNDE (temperatura de los cojinetes del generador en el LOA) utilizando un modelo SVR.

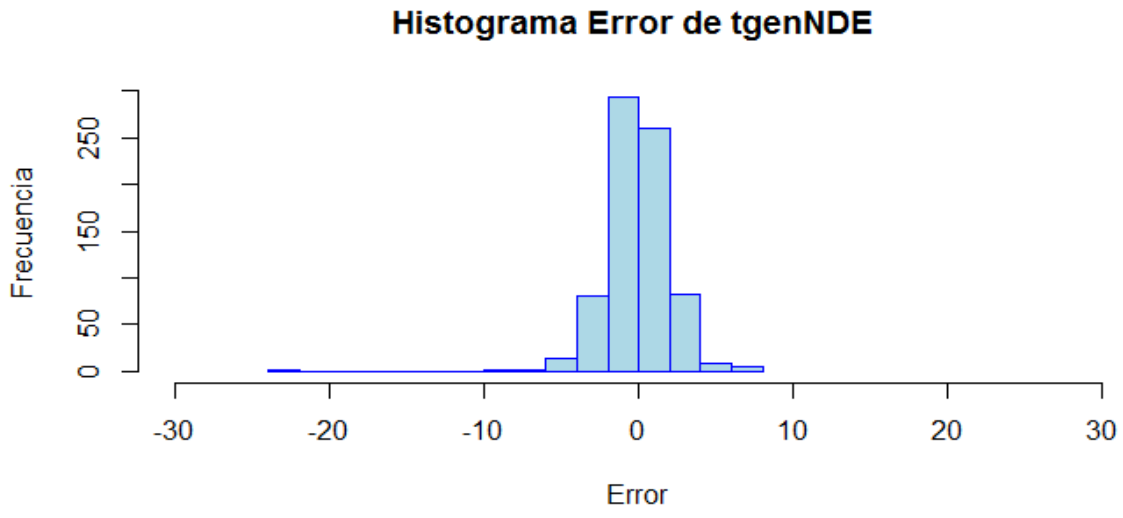

Figura 4.8. Histograma error de predicción (tgenNDE)

Una vez obtenido el modelo de predicción conformal y las medidas de no conformidad bajo comportamiento normal del aerogenerador ambas se utilizan para analizar el comportamiento del aerogenerador antes del fallo del generador en octubre de 2010 y detectar cambios en la medida de no conformidad que indiquen señales de degradación previas a la ocurrencia de dicho fallo.

En la *Figura 4.9* se muestra la evolución del estadístico F asociado a la medida de no conformidad, desde junio de 2010 se observa como en un primer periodo el estadístico F supera el valor establecido. Posteriormente, se observa una tendencia ascendente en el comportamiento del estadístico y aproximadamente 100 días antes de la sustitución se supera el límite antes de los 100 días.

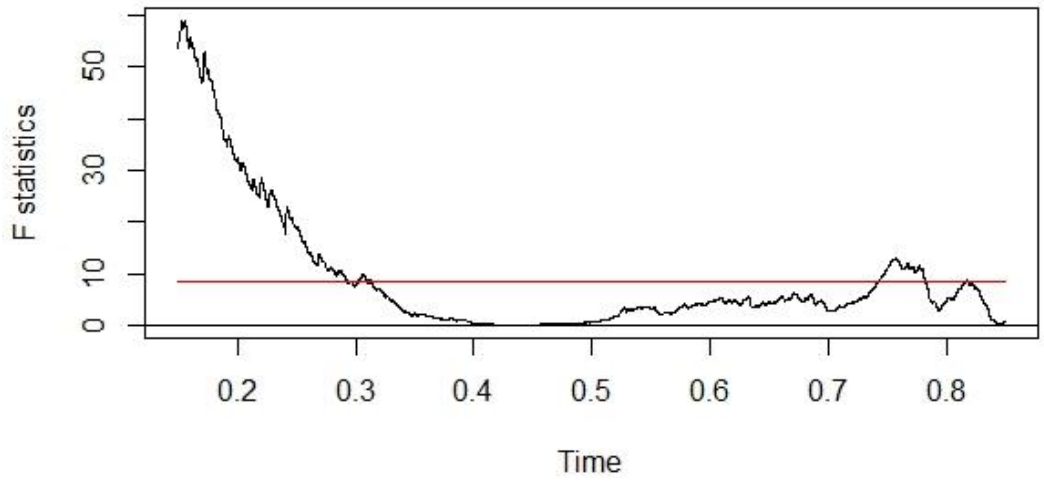

Figura 4.9. Estadístico F para la MNC de tgenNDE

CAPÍTULO 5

5 CONCLUSIONES

La monitorización de aerogeneradores para la detección de comportamientos anómalos previos a la aparición de fallos es un objetivo de la industria eólica con la finalidad de garantizar la máxima disponibilidad de los aerogeneradores sin aumentar los costes de mantenimiento.

En la literatura se han propuesto diferentes metodologías para la monitorización del estado de degradación de los componentes a partir del uso de los datos registrados en los sistemas SCADA. En este contexto, en el presente trabajo se ha analizado la aplicabilidad de la regresión conformal, utilizando como algoritmo subyacente SVR, para el desarrollo de modelos de comportamiento normal de variables operacionales de un aerogenerador y la posterior monitorización de la medida de no conformidad para la detección de comportamientos anómalos que indiquen la aparición de síntomas de degradación del equipo. El modelo de predicción ha sido construido utilizando el histórico operacional disponible en la base de datos SCADA de un aerogenerador perteneciente a un parque eólico español.

Los resultados obtenidos muestran la aplicabilidad de la metodología para la monitorización del comportamiento de aerogeneradores. Respecto a los resultados alcanzados con el algoritmo SVR los valores obtenidos del coeficiente de determinación han sido de 0.98 y de 0.91 para los modelos de *pwtot* y de *tgenNDE*, respectivamente. En ambos casos la selección de los parámetros asociados a SVR se realizó mediante validación cruzada seleccionando los valores que minimizaban el RMSE. La monitorización de la medida de no conformidad obtenida mediante regresión conformal mostró la aparición de cambios en la serie temporal previos al fallo.

Como trabajo futuro, se plantean diferentes líneas, como son:

- La aplicación de los modelos desarrollados a otros generadores de características similares con el objetivo de verificar la validez del modelo tanto en su potencia para detectar comportamientos anómalos como la probabilidad de falsas alarmas.
- Estudiar otras técnicas alternativas para la monitorización de la medida de no conformidad.
- El uso de otros algoritmos subyacentes como, por ejemplo, Vecinos Cercanos (K-NN) y comparación de los resultados obtenidos con SVR.

BIBLIOGRAFÍA

Alonso, E. G. (2007). Aplicación de las máquinas de soporte vectorial para el reconocimiento de matrículas, (March), 1–109. Retrieved from http://www.iit.upcomillas.es/pfc/resumenes/467f930a0aa05.pdf

Balasubramanian, V. N., Ho, S.-S., & Vovk, V. (2014). *Conformal Prediction for Reliable Machine Learning: Theory, Adaptations and Application*. https://doi.org/http://dx.doi.org/10.1016/B978-0-12-398537-8.00015-8

Betancour, G. (2005). Las máquinas de soporte vectorial (SVMs). *Scientia Et Technica*, (27), 67–72.

Brandão, R.F.M., Carvalho, J.A.B.: 'Intelligent system for fault detection in wind turbines gearbox'. PowerTech Eindhoven 2015, 2015

Burges, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.*, *2*(2), 121–167. https://doi.org/10.1023/A:1009715923555

Carmona Suárez, E. J. (2014). Tutorial sobre Máquinas de Vectores Soporte (SVM), 1–25.

Cherkassky, V., & Mulier, F. (2007). *Learning From Data. Concepts, Theory, and Methods* (Second Edi). New Jersey, USA: John Wiley & Sons, Inc.

Cristianini, N., Shawe-Taylor J., (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press.

Cortés, I. (2016). Conformal Predictors for Regression and Classification. Package Conformal.

Cross, P., Ma, X.: 'Model-based and fuzzy logic approaches to condition monitoring of operational wind turbines', Int. J. Autom. Comput., 2015, 12, (1), pp. 25–34.

Garcia, M.C., Sanz-Bobi, M.A., del Pico, J.: 'SIMAP: intelligent system for predictive maintenance application to the health condition monitoring of a windturbine gearbox', Comput. Ind., 2006, 57, (6), pp. 552–568

Garlick, W.G., Dixon, R., Watson, S.J.: 'A model-based approach to wind turbine condition monitoring using SCADA data'. 20th Int. Conf. System Engineering, 2009.

Gasch, R., Twele, J., (2012). "Wind Power Plants: Fundamentals, Design, Construction and Operation", 2nd ed., New York: Springer.

Goddard, J., Gerardo, S., Silva, D. L. C., P, B. R., & Angel, M. (2000). Un Algoritmo Para El Entrenamiento De Máquinas De Vector Soporte Para Regresión, *7*, 107–116.

Gunn, S. R. (2010). Support vector machines for classification and regression. *The Analyst*, *135*(2), 230–267. https://doi.org/10.1039/b918972f

Hau, E., (2000) "Wind Turbines: Fundamentals, Technologies, Applications, Economics", 3rd ed., New York: Springer.

Khun, M. (2017). Classification and Regression Training. Package Caret.

Kusiak, A., Verma, A.: 'Analyzing bearing faults in wind turbines: a data mining approach', Renew. Energy, 2012, 48, pp. 110–116.

Li, J., Lei, X., Li, H., et al.: 'Normal behavior models for the condition assessment of wind turbine generator systems', Electr. Power Compon. Syst., 2014, 42, (11), pp. 1201–1212.

Makili, L. E. (2014). SISTEMAS DE CLASIFICACIÓN AUTOMÁTICOS CON CONFIANZA Y CREDIBILIDAD EN FUSIÓN TERMONUCLEAR.

Marton I., Sanchez A., Carlos S., Martorell S. (2013). Application of data driven methods for condition monitoring maintenance. Chemical Engineering Transactions, 33, 301-306 DOI: 10.3303/CET1333051.Pérez-Planells, L., Delegido, J., Rivera-Caicedo, J. P., & Verrelst, J. (2015). Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos. Revista de Teledeteccion, 2015(44), 55–65. https://doi.org/10.4995/raet.2015.4153

Papadopoulos, H., Vovk, V., & Gammerman, A. (2011). Regression conformal prediction with nearest neighbours. *Journal of Artificial Intelligence Research*, *40*, 815–840. https://doi.org/10.1613/jair.3198

Pérez-Planells, L., Delegido, J., Rivera-Caicedo, J. P., & Verrelst, J. (2015). Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos. *Revista de Teledeteccion*, *2015*(44), 55–65. https://doi.org/10.4995/raet.2015.4153

Pérez, N. V. R. (2011). INFORMÁ TICA " Aplicación de Predictores Conformales a Señales de Fusión, *0*.

Santana, J. S., & Farfán, E. M. (2014). El Arte de programar en R, 182.

Schlechtingen, M., Santos, I.F.: 'Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection', Mech. Syst. Signal Processing, 2010, 25, pp. 1849–1875

Schlechtingen, M., Santos, I.F., Achiche, S.: 'Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: system description', Appl. Soft Comput., 2013, 13, (1), pp. 259–270.

Shafer, G., & Vovk, V., (2008). *A tutorial on conformal prediction*, Journal of Machine Learning Research, 9, 371–421.

Tautz-Weinert, J. and Watson, S.J., 2017. Using SCADA data for wind turbine condition monitoring - a review. IET Renewable Power Generation, 11 (4), pp.382-394

Vapnik, V.N., (2000). "The Nature of Stastical Learning Theory", 2nd ed., Springer, New York.

Vovk, V. (2012). Cross-conformal predictors, 1–10. Retrieved from http://arxiv.org/abs/1208.0806

Vovk, V., Gammerman, A., & Shafer, G. (2005). “Algorithmic learning in a random world.” Retrieved from http://www.alrw.net/

Zaher, A., McArthur, S.D.J., Infield, D.G., et al.: ‘Online wind turbine fault detection through automated SCADA data analysis’, Wind Energy, 2009, 12, (6), pp. 574–593.

Zhang, Z.-Y., Wang, K.-S.: ‘Wind turbine fault detection based on SCADA data analysis using ANN’, Adv. Manuf., 2014, 2, (1), pp. 70–78.