UNIVERSIDAD DE SANTIAGO DE CHILE

FACULTAD DE CIENCIAS

DEPARTAMENTO DE FÍSICA

ESTUDIO DEL COMPORTAMIENTO DE LAS VARIABLES FÍSICAS EN INSTRUMENTOS DEL FRONT END DEL OBSERVATORIO ALMA PARA CREACIÓN DE MODELO PREDICTIVO DE FALLAS

FLOR CANDIA C.

UNIVERSIDAD DE SANTIAGO DE CHILE

FACULTAD DE CIENCIAS

DEPARTAMENTO DE FÍSICA

ESTUDIO DEL COMPORTAMIENTO DE LAS VARIABLES FÍSICAS EN INSTRUMENTOS DEL FRONT END DEL OBSERVATORIO ALMA PARA CREACIÓN DE MODELO PREDICTIVO DE FALLAS

FLOR CANDIA C.

Profesor Guía: Dra. Marina Stepánova

TESIS PARA OPTAR AL TÍTULO DE INGENIERO FÍSICO

SANTIAGO DE CHILE

ABRIL 2015

UNIVERSIDAD DE SANTIAGO DE CHILE

FACULTAD DE CIENCIAS

DEPARTAMENTO DE FÍSICA

ESTUDIO DEL COMPORTAMIENTO DE LAS VARIABLES FÍSICAS EN INSTRUMENTOS DEL FRONT END DEL OBSERVATORIO ALMA PARA CREACIÓN DE MODELO PREDICTIVO DE FALLAS

FLOR CANDIA C.

Profesor Guía : Dra. Marina Stepánova

Co-Guía en ALMA : Jaime Guarda

Profesores Comisión : Dr. Ricardo Finger

Profesores Comisión : Dr. Raúl Labbé

TESIS PARA OPTAR AL TÍTULO DE INGENIERO FÍSICO

SANTIAGO DE CHILE

ABRIL 2015

“ESTUDIO DEL COMPORTAMIENTO DE LAS VARIABLES FÍSICAS EN INSTRUMENTOS DEL FRONT END DEL OBSERVATORIO ALMA PARA CREACIÓN DE MODELO PREDICTIVO DE FALLAS”

Trabajo de Graduación presentado a la Facultad de Ciencia, en cumplimiento parcial de los requerimientos exigidos para optar al título de Ingeniero Físico.

UNIVERSIDAD DE SANTIAGO DE CHILE

SANTIAGO DE CHILE

ABRIL 2015

“ESTUDIO DEL COMPORTAMIENTO DE LAS VARIABLES FÍSICAS EN INSTRUMENTOS DEL FRONT END DEL OBSERVATORIO ALMA PARA CREACIÓN DE MODELO PREDICTIVO DE FALLAS”

FLOR CANDIA C.

Este trabajo de graduación fue preparado bajo la supervisión de la profesora guía Dra. Marina Stepánova del Departamento de Física de la Universidad de Santiago de Chile y aprobado por los miembros de la comisión calificadora del candidato.

……
Dr. Ricardo Finger

……
Dr. Raúl Labbé

……
Jaime Guarda
Co-Guía en ALMA

……
Dra. Marina Stepánova
Profesor Guía

……
Dra. Yolanda Vargas H.
Directora Departamento de Física

AGRADECIMIENTOS

Mi más profundo agradecimiento es a Dios, mi Padre Celestial, por guiarme en este camino, ceñirme de fuerza en cada momento y determinar a las personas que me ayudarían y apoyarían a cumplir mi meta.

Agradezco también a mi familia, amigos y compañeros, y en forma particular a quienes han sido clave en la última etapa de mi carrera. Especialmente a Edwin Zárate, por motivarme a buscar oportunidades en donde no creí que las encontraría. A Alejandro Peredo por ser el nexo a emprender mi proyecto de titulación. A Alfredo Plata por todo su apoyo, disposición y ayuda a encaminarme en todo el ámbito profesional.

Un enorme agradecimiento a Jorge Ibsen, no sólo por haber sido mi patrocinador en el Observatorio ALMA, sino además por la gran persona que es, por creer en mí y por todo su impulso a que concrete mis propósitos y objetivos. Asimismo a Jaime Guarda, mi Co-Guía de Tesis, a quien le estoy muy agradecida por su disposición en enseñarme gran parte de la labor que realizan en ALMA y su apoyo en todo este proceso, que sin duda ha sido fundamental. Como también agradezco al equipo de Ingeniería y Computación, quienes siempre tuvieron una gran voluntad por ayudarme y atender todas mis dudas.

Además, agradezco a la profesora Marina Stepánova, quien me guió en el trabajo de titulación y a su vez el respaldo brindado aún antes de ingresar a la universidad. Al profesor Ricardo Finger y Raúl Labbé por su buena disposición y amabilidad. Al profesor y coach Omar Matus, por instruirme a desarrollar habilidades en el entorno personal, académico y profesional.

Finalmente, agradezco a todas aquellas personas maravillosas que he conocido a lo largo de toda mi formación académica, desde la básica hasta la superior, por su pequeño o gran aporte en mi vida, pues son ellos y ellas una gran parte de mi inspiración a seguir cumpliendo mis sueños día a día.

TABLA DE CONTENIDOS

ÍNDICE DE TABLAS

ÍNDICE DE ILUSTRACIONES

ABREVIACIONES Y TÉRMINOS UTILIZADOS

ALMA:	*Atacama Large Millimeter/submillimeter Array*
ACD:	*Amplitude Calibration Device*
ADC:	*ALMA Department of Computing*
ADE/AMG/FE:	*ALMA Department of Engineering/ Array Maintenance Group/ Front End*
AOS:	*Array Operations Site*
CCA:	*Cold Cartdrige Assemby*
CPDS:	*Cartdrige Power Distribution System*
EDFA	*Erbium Dopped Fiber Amplifier*
FEMC:	*Front End Monitor & Control*
FEPS:	*Front End Power Supply*
Frequency LO:	*Frecuency Local Oscillator*
IF Switch:	*Intermediate Frequency Switch*
IFTP:	*Intermediate Frequency Total Power*
KDD:	*Knowledge Discovery in Database*
LPR:	*Local oscillator Photonic Receiver*
LNA:	*Low Noise Amplifier*
PMC:	*PhotoMixer Current*

RBF:	*Radial Basis Function*
SIS:	*Superconductor Isolator Superconductor*
SVM:	*Support Vector Machine*
WCA:	*Warm Cartridge Assembly*
WEKA:	*Waikato Enviroment for Knowledge Analysis*
WVR:	*Water Vapor Radiometer*

RESUMEN

El *Atacama Large Millimeter/submillimeter Array* (ALMA) es un Observatorio radioastronómico de gran complejidad. Su sensibilidad y poder de resolución lo convierte en el más importante de todo el mundo. Compuesto por 66 antenas de 12 y 7 metros de diámetro, operando a frecuencias milimétricas y submilimétricas, puede alcanzar zonas muy distantes del universo, superando en resolución a cualquier radiotelescopio existente hasta el momento. Asimismo, su ubicación a 5000 metros sobre el nivel del mar, en el norte de Chile, es óptima para la obtención de las señales, ya que a esa altura la atmósfera es casi transparente y de muy baja humedad para estas señales. Las antenas están compuestas por receptores que cubren 10 bandas de frecuencias y procesadores digitales de señales, entre otros dispositivos.

Conocer el estado y el comportamiento de los receptores y sus componentes es de suma importancia, tanto para el área de Ciencia como de Ingeniería. Debido a eso, se realizan constantemente monitoreos para programar de forma oportuna mantenciones correctivas.

En esta Tesis se presenta la aplicación del proceso *Knowledge Discovery in Database (KDD)* al desempeño de los componentes del instrumento *Front End,* encargado de medir la radiación en el rango de radiofrecuencias. Se desarrolla la creación de cuatro modelos (uno de clasificación y tres de regresión), cuyo objetivo principal es detectar y predecir fallas del instrumento. Para demostrar esta técnica se escogió el dispositivo electrónico *Warm Cartridge Assembly* (WCA) de la Banda 3 del *Front End.*

El estudio se basa en los análisis de los datos históricos de un WCA, tales como potencia, frecuencia y corriente, con el fin de obtener un conjunto representativo para la predicción de datos. Para procesar toda esta información, el modelo desarrollado considera métodos estadísticos y herramientas de inteligencia computacional, logrando así una predicción de fallas para el dispositivo WCA.

El modelo de clasificación está basado en el Teorema de Bayes de probabilidad condicional y en su entrenamiento logró etiquetar correctamente un 86,96% de las instancias, dadas las clases de falla (F) y normal (N).

La predicción de datos de las variables que componen el dispositivo WCA se adaptaron a los parámetros individuales C y γ, requeridos por los tres modelos de regresión. Para la variable *Intermediate Frequency Total Power* (IFTP) se obtuvo el par de $C = 2^{-3}$ y $\gamma = 2^{-13}$, para *Photo Mixer Current* (PMC) los valores son $C = 2^{-5}$ y $\gamma = 2^{-15}$ y para *Local oscilator Photonic Receiver* (LPR) le corresponden $C = 2^{-4}$ y $\gamma = 2^{-14}$, logrando así el entrenamiento de modelos predictivos mediante las herramientas de Máquinas de Vector soporte para regresión (SVR).

Finalmente, los modelos creados en el software WEKA pueden ser empleados para la detección temprana de fallas del dispositivo WCA estimando su comportamiento, pero el estado final es determinado por los Ingenieros de ALMA mediante pruebas de laboratorio.

Palabras claves: ALMA, Front End, WEKA, KDD, SVR, Clasificación bayesiana, predicción y regresión.

Capítulo 1

Introducción

1.1Áreas de interés.

El presente estudio se ha desarrollado para el Observatorio ALMA, siendo éste el radiotelescopio más potente a nivel global y ubicado a 5000 metros de altura, en el Llano de Chajnantor, a 50 km de San Pedro de Atacama. Su administración se compone de diferentes áreas, las concernientes para esta Tesis son el Departamento de Ingeniería de ALMA, en el Grupo de Instrumentos del equipo de Front End (ADE/AMG/FE), responsable de los detectores de las señales astronómicas y el Departamento de Computación de ALMA (ADC) encargado de mantener el sistema informático, que incorpora los equipos de Tecnologías de la Información, Software y al Grupo de Operaciones de Archivo[1].

1.2 Descripción del problema

La pérdida de condición operativa en instrumentos electrónicos obliga a realizar una mantención correctiva, la cual se lleva a cabo por el Departamento de Ingeniería. Los componentes del equipo de *Front End* son relativamente nuevos y fabricados en el extranjero, por lo que un fallo abrupto ocasionaría pérdidas de información, tiempo e incluso una menor disponibilidad de antenas.

En ocasiones el origen de la falla se desconoce y su reparación puede tomar extensiones de tiempo difíciles de prever. Dependiendo de la magnitud del desperfecto, el equipo de ingenieros especialistas determina si el *Front End* de la antena debe o no debe desinstalarse.

Las razones de por qué ocurren las fallas son numerosas y requieren de tiempo para descubrirlas y prevenirlas. Mantener el rendimiento y perdurabilidad de los componentes electrónicos es clave para la óptima obtención de señales del espacio.

Si se contara con una evaluación previa de los dispositivos se contribuye a implementar mantenciones preventivas. Esto permitiría reducir significativamente los gastos de la mantención y garantizar la entrega de los datos para la investigación científica a los investigadores tanto nacionales como internacionales.

1.3 Objetivos de la Tesis

1.3.1 Objetivo general

Construir un modelo de análisis y clasificación de datos centrado en el dispositivo WCA de la Banda 3 del instrumento *Front End* de una antena del Observatorio *Atacama Large Millimeter/submillimeter Array,* teniendo como propósito predecir fallas para poder programar mantenciones preventivas.

1.3.2 Objetivos específicos

- Obtener los registros de monitoreo y definir el estado actual del dispositivo.

- Seleccionar los datos correspondientes de las variables que caracterizan el dispositivo WCA y que se utilizarán para proceder al entrenamiento, validación y pruebas de los modelos de regresión y de clasificación.

- Determinar qué variables son relevantes para efectos de indicar fallas en el dispositivo.

- Etiquetar y clasificar los datos para uso de modelo clasificador.

- Adaptar datos de monitoreo para elaboración de modelos de regresión y de clasificación.

- Crear modelo clasificador que indique si los datos de predicción corresponden o no a una falla.

- Generar los modelos predictivos.

- Evaluar los modelos predictivos para cada componente del dispositivo y así corroborar la eficacia del modelo a base de datos de entrenamiento.

- Utilizar el modelo clasificador con datos obtenidos por el modelo de predicción para su validación.

1.4 Estructura de la Tesis

El presente trabajo queda organizado de la siguiente manera:

El Capítulo 2 da a conocer los detalles de la instrumentación del Equipo de *Front End* del Observatorio ALMA y una breve reseña del mismo. En el Capítulo 3 se describen las herramientas predictivas, empleadas para el desarrollo de la Tesis y su metodología. El Capítulo 4 presenta los procedimientos y resultados de todas las etapas para el desarrollo de los modelos: clasificador y predictivos.

El Capítulo 5 contiene las conclusiones y recomendaciones para los futuros trabajos relacionados con el desarrollo de nuevos modelos predictivos.

Capítulo 2

Componentes del Equipo Front End de Atacama Large Millimeter/submillimeter Array

2.1 ALMA

Atacama Large Millimeter/submillimeter Array (ALMA) es un complejo astronómico revolucionario, el más importante a nivel mundial en radioastronomía. Es desarrollado en consorcio entre Europa, Norteamérica y Asia del Este y la República de Chile.

ALMA está compuesto por 66 antenas de alta precisión, las cuales funcionan juntas como si fuera sólo una antena virtual de hasta 16 km de diámetro, esta técnica es llamada Interferometría. Al estar dotados de una alta tecnología, los reflectores son capaces de captar longitudes de onda en el rango de las milimétricas y submilimétricas, entre el infrarrojo y las microondas. ALMA es capaz de observar el Universo frío con gran resolución y estudiar zonas hasta ahora ocultas para la ciencia, detectando el polvo, gas molecular, galaxias, formación de estrellas y sistemas planetarios, hasta la radiación proveniente del Big Bang.

Las antenas se encuentran instaladas en el *Array Operations Site* (AOS), lugar correspondiente al Llano de Chajnantor, a 5000 metros de altitud en el Desierto de Atacama en Chile. Son operadas desde el edificio técnico en el centro de operaciones de ALMA, *Operations Support Facility* (OSF), a 2900 metros de altura (Figura 2.1).

La ubicación se debe principalmente a razones científicas y sus características de altitud y sequedad. En este lugar la cobertura de nubes es casi inexistente y la corriente fría de Humboldt junto con el anticiclón del Pacífico mantienen el clima seco del Desierto, complementándose con la uniformidad de la planicie, lo cual es perfecto para que las antenas capten casi sin distorsión la radiación del espacio[2].

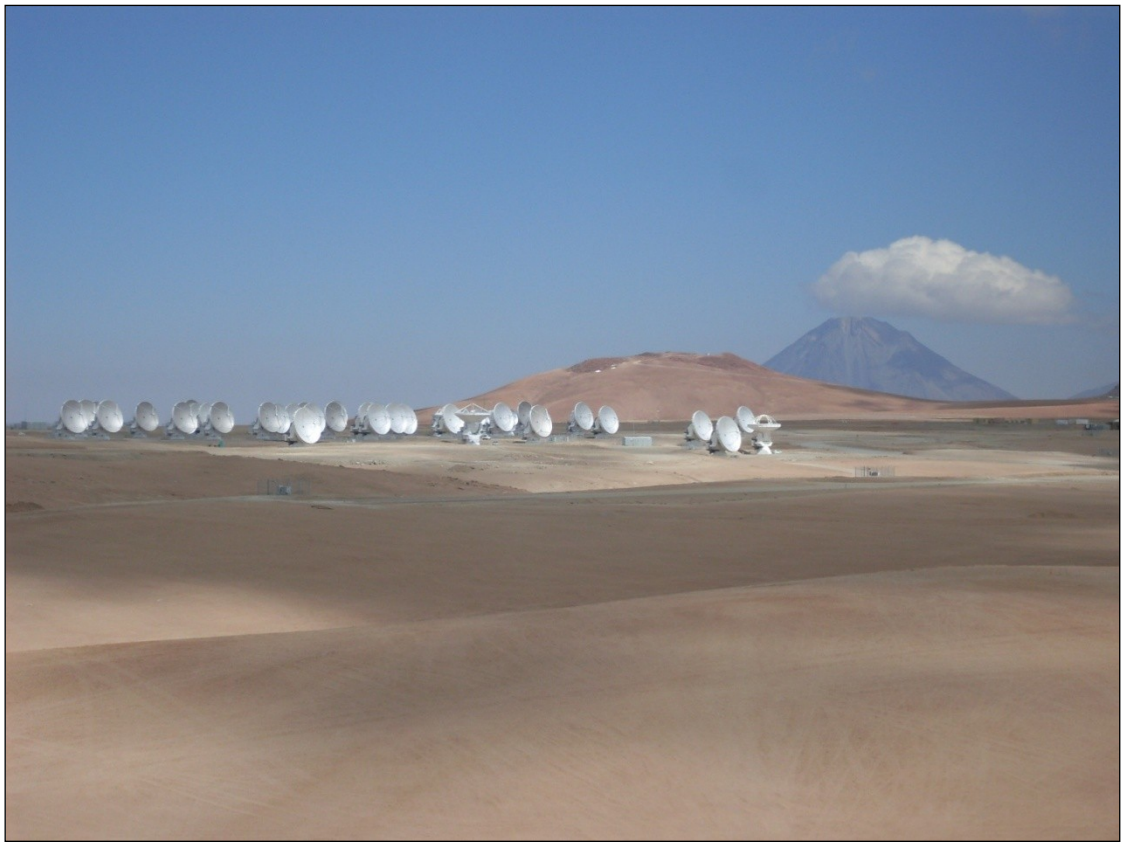


Figura 2.1.Conjunto de antenas en el Llano de Chajnantor, 5000 m sobre el nivel del mar (Fotografía de F. Candia).

2.2 ¿Qué es el Front End?

El *Front End* de ALMA es un sistema diseñado para recepcionar y convertir señales de radio en diez frecuencias diferentes[2].

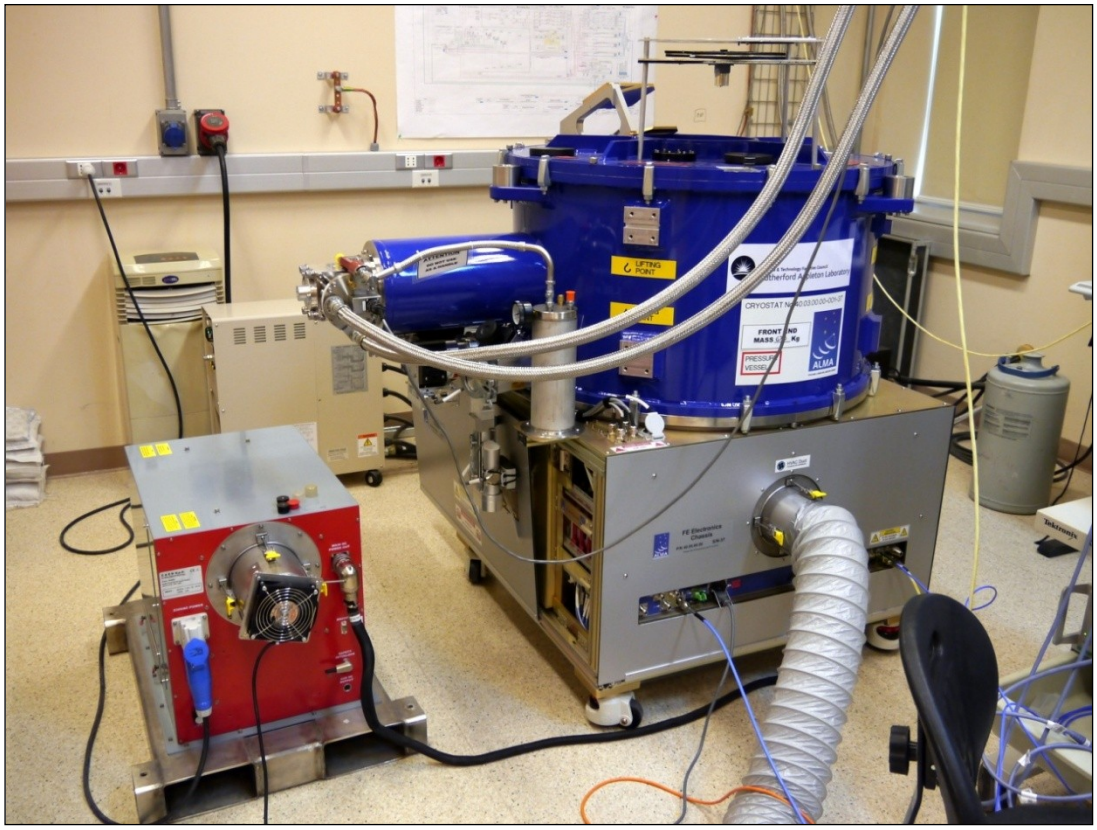


Figura 2.2: Laboratorio de Equipo Front End. El cilindro azul es el criostato, el cual contiene en su interior las 10 bandas de frecuencias.

Contiene diversos componentes, como el sistema óptico que está formado por espejos, bocinas de alimentación y rejillas de polarización. El elemento más visible del Front End es el *criostato* y tiene forma de un cilindro. La temperatura de operación requerida es extremadamente baja, llegando a ser de 4 K (-269 C), lo cual es necesario para que sus componentes superconductores funcionen de manera óptima en el SIS (Figura 2.2).

El criostato alberga los receptores de las diez bandas de radiofrecuencias. Estos están montados en *cartridges* y convierten la onda electromagnética en señales eléctricas, para enviarlas a otro sistema denominado llamado *Back End,* donde son digitalizadas y sincronizadas, para luego ser enviadas mediante fibra óptica al computador central denominado Correlacionador (ver figura 2.3)[3].

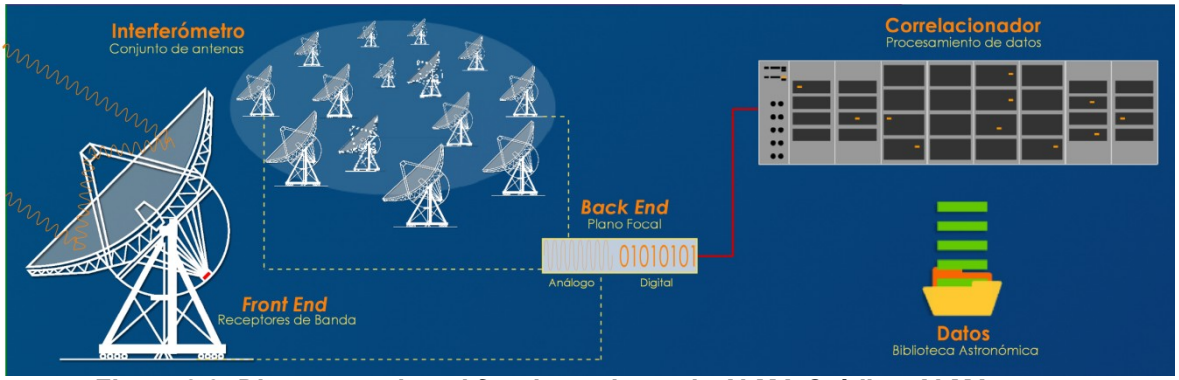


Figura 2.3: Diagrama sobre el funcionamiento de ALMA.Crédito: ALMA (ESO/NAOJ/NRAO).

2.2.1 Bandas de los receptores de ALMA.

La Tabla 2.1 detalla los rangos de frecuencias correspondientes a cada banda.

Tabla 2.1: Frecuencias de las 10 Bandas de los receptores de ALMA.

Banda	Rango de frecuencia GHz
1	35-50
2	67-90
3	84-116
4	125-163
5	163-211
6	211-275
7	275-373
8	385-500
9	602-720
10	787-950

Un ejemplo de *cartridge* es el que se ilustra en la figura 2.3. Este instrumento tiene tres secciones separadas por discos. La sección superior opera a una temperatura de 4K, la intermedia a 15 K y la inferior a 110 K.

Figura 2.4. Ejemplo de cartridge, Banda 7. Crédito: ALMA (ESO/NAOJ/NRAO).

2.2.2 Dispositivos electrónicos y sus parámetros

- **Receptores criogénicos de señal, *Cold Cartdrige Assemby* (CCA):** Este dispositivo contiene el SIS mixer y amplificadores de bajo ruido (LNA). El SIS mixer mezcla la señal recibida con la señal del oscilador local (LO) para bajar la frecuencia por método heterodino y obtiene una señal manejable por circuitos electrónicos en el rango de 4-8 GHz y 4-12 GHz, llamada frecuencia intermedia (IF). Para el *Mixer* se determina su curva característica (Corriente v/s Voltaje) la cual debe ser no lineal y dado que los LNAs son amplificadores, sólo se observa su estado de consumo.

- **Generadores de oscilador local, *Warm Cartgrige Assembly* (WCA):** Dispositivo asociado al CCA y una de sus funciones es generar la señal LO, siendo corregido por dos señales ópticas provenientes del Oscilador Local Central (CLO). Estas señales ópticas ingresan al *Photomixer*, de lo que resulta una señal eléctrica con la diferencia de frecuencia de las dos

señales ópticas. Lo que se verifica es la cantidad de corriente que se genera para una cierta potencia fotónica.

- **Receptores fotónicos, *LO Photonic Receiver* (LPR):** Dispositivo que contiene un circuito de amplificación óptico (EDFA), controlado por voltaje y un conmutador óptico para enviar los láseres provenientes del CLO al *Photomixer.* Se monitorea el estado del amplificador verificando la potencia óptica de salida versus el voltaje de polarización del EDFA.

- **Conmutadores de frecuencia intermedia, *Intermediate Frequency Switchs* (IF Switch):** El *Front End* tiene 4 salidas hacia *el Back End* análogo (2 polarizaciones y 2 bandas laterales por polarización). Internamente hay un conmutador de RF (radiofrecuencia) que selecciona de qué banda toma la señal. Trabaja con las señales convertidas y amplificadas.

- **Radiómetro de vapor de agua, *Water Vapor Radiometer* (WVR):** Es un receptor de RF, que detecta la frecuencia de emisión espectral del vapor de agua atmosférico cercano a 182-183 GHz y cuenta con su propio sistema de calibración integrado.

- **Dispositivos de calibración de receptores, *Amplitude Calibration Device* (ACD):** Es un brazo robótico con movimiento en dos ejes, que se posiciona sobre cada una de las 10 bandas del *Front End* y el WVR, dos cuerpos negros con temperaturas conocidas y distintas para calibrar las potencias medidas y posteriormente convertidas de potencia a temperatura para señales observadas en el espacio. Posee además, un filtro para observaciones solares. Se monitorea el tiempo de posicionamiento en las bandas, que generalmente debe ser menor a un cierto límite, de modo que se verifique su deterioro.

- **Fuentes de alimentación, *Front End Power Supply* (FEPS):** Son fuentes de poder para alimentar al *Front End* con los distintos voltajes requeridos. Se deben monitorear que los voltajes estén dentro de los valores aceptables.

- **Monitor y Control Front End, *Front End Monitor & Control (*FEMC):** Es la unidad que recibe los comandos de control del software de ALMA y procesa los requerimientos para enviar los comandos al módulo específico. Además, contiene los archivos de configuración básica del hardware instalado.

- **Sistema de distribución, Cartdrige *Power Distribution System* (CPDS):** Corresponde a un sistema de regulación y filtrado del voltaje, distribuido a los módulos de CCA y WCA de las 10 bandas. Se deben monitorear que las tensiones sean adecuadas.

2.3 Fallas en dispositivos

La gran mayoría de instrumentos electrónicos van perdiendo funcionalidad a través del tiempo. Diversas razones contribuyen a su deterioro, ya sea el uso excesivo, una inadecuada manipulación, etc., pero en este caso en particular, los dispositivos del equipo *Front End* sufren tanto degradación como fallas abruptas de los componentes. En consecuencia se detiene el funcionamiento de una parte del equipo y en el peor de los casos, se procede a desinstalarlo de la antena. Las reparaciones son realizadas en el laboratorio, de acuerdo al desperfecto encontrado.

A continuación se describen algunas de las principales fallas detectadas por los Ingenieros de ALMA.

Dispositivo	Falla
CCA	• *Mixers:* degradación de la curva característica corriente-voltaje, corto-circuito o circuito abierto del SIS. • *Cold amplifiers*: degradación o falla completa de la ganancia. • *Bias Module*: error en los voltajes aplicados a los diferentes dispositivos del CCA o error en los valores leídos.
WCA	• *IF Warm Amplifiers*: degradación de la ganancia o falla completa (no generan señal de salida). • *Photomixer:* degradación o falla completa de la generación de RF a partir de señal óptica. • *PLL block*: falla en cualquiera de las etapas internas del PLL, oscilador controlado por voltaje, multiplicador de frecuencia. • *LO amplifiers*: degradación o falla completa en los amplificadores que amplifican la salida del PLL para generar la señal de LO.
WVR	• Lecturas de temperatura incorrectas. • Lecturas de temperaturas con variabilidad mayor que la especificada. • Consumo de corriente del sistema de calibración mayor a lo normal.
ACD	• Lecturas de temperaturas de carga ambiente y carga caliente incorrectas. • Lecturas de temperaturas de carga ambiente y carga caliente, cambian más rápido que lo especificado. • Posiciones de calibración sobre las bandas no coinciden con lo especificado
FEPS	• Temperaturas de operación fuera de lo normal. • Voltajes de operación fuera de lo normal.

Capítulo 3

Marco teórico y Metodología

3.1 Predicción de datos

El uso de herramientas estadísticas y técnicas de análisis predictivos son de gran utilidad en diversos ámbitos tanto en el rubro comercial como en la ingeniería.

Las predicciones de datos requieren de estudios meticulosos de eventos pasados, por lo que es importante basarse en la probabilidad de ocurrencias, considerando que su comportamiento no depende del transcurso del tiempo. Al observar registros y monitoreos, se tiene un punto de partida para obtener una tendencia de comportamientos a corto y largo plazo. La idea es favorecer la minimización de mantenciones correctivas y potenciar las preventivas y predictivas.

Entre las herramientas de predicción más utilizadas, se encuentran:

-Filtro Kalman, definido como una estimación secuencial y estadísticamente óptima para sistemas dinámicos. Su mayor ventaja es la fácil adaptación a cualquier cambio en las observaciones. Por ejemplo, se puede usar para fines meteorológicos[4].

-Modelo Autorregresivo de Media Móvil (ARMA), es un modelo que representa un proceso aleatorio de variables en el tiempo. Se ha utilizado para la predicción a corto plazo de la velocidad del viento, dada su naturaleza aleatoria de las condiciones meteorológicas [5].

-Distribución de Weibull, corresponde a un modelo estadístico aplicado a problemas de confiabilidad, ya sea en tiempos de falla y duración de un componente o equipo. Representa la probabilidad de fallo del instrumento dentro un cierto tiempo predeterminado[6].

-Regresión Lineal Múltiple, es útil cuando se ajustan datos experimentales, en donde la variable que se está analizando es función de otras[7]. A modo de ejemplo, se tiene un análisis de regresión para correlacionar datos en catálogos de sismicidad[8].

-Redes Neuronales Artificiales (RNA), consiste en la creación de un modelo inspirado en la biología del cerebro humano, para clasificación y predicción de datos. Entre sus aplicaciones se destacan el Procesado de imágenes en el área de Medicina, Robótica y Control[9]. Asimismo, casos de predicciones meteorológicas como el modelo de RNA aplicado en cinco estaciones de monitoreo en la ciudad de Santiago de Chile[10]. A su vez, un estudio realizado de la predicción en la evolución de la demanda horaria de energía eléctrica en [11].

-Support Vector Machine (SVM), es una herramienta que ha tomado fuerza en el último tiempo, debido a su desarrollo en la técnica para clasificación y regresión, llegando a obtener mejores resultados que las redes neuronales, de acuerdo al error del modelo[12]. A modo de ejemplo, se tiene su aplicación en el movimiento del IBEX-35 (Índice Bursátil Español), de la compra y venta en el mercado, concluyendo este estudio que se producen resultados de predicción más aceptables en comparación con las RNA[13].

3.1.1 Criterios de selección del modelo

Un modelo predictivo se construye en base a ecuaciones lineales o no lineales. Enlaza tanto valores pasados como futuros junto con sus variables, adaptándose a los datos para su entrenamiento con el objetivo minimizar errores al momento de predecir.

Para la elección del modelo predictivo a utilizar se toman en cuenta los resultados de las aplicaciones descritas previamente y de forma paralela la descripción del problema, de acuerdo a la información proporcionada por el Equipo de *Front End* de ALMA.

La recolección de los datos históricos de las variables físicas a predecir es otro punto a considerar. Tratándose de información cuantitativa, el análisis estadístico es de gran utilidad en la selección de variables que mejor describen el comportamiento de un equipo, ayudando a explorar cuándo se está en presencia de datos *outliers,* es decir, datos fuera de los rangos aceptables.

Finalmente, es relevante considerar que la implementación del modelo debe ser aplicada en una plataforma de software libre para su uso en ALMA.

3.1.2 Propuesta de solución

Como solución se plantea un modelo que logre predecir el estado del dispositivo WCA de la Banda 3. Para ello, se basa en los registros de datos de los monitoreos, que serán utilizados para el entrenamiento de dicho modelo, llegando a estimar y clasificar los valores obtenidos de las variables.

Si el estado del dispositivo corresponde a una falla o no, se determina en conjunto con el análisis realizado a los datos, con el propósito de evaluar la probabilidad de fallo. El Equipo de Front End es quién toma las medidas

necesarias de corrección en caso de deterioro, minimizando las demoras y gastos en reparación.

3.2 Análisis estadístico

3.2.1 Rango intercuartílico (RI): es la diferencia entre el tercer y primer cuartil de los datos. Donde cuartil (Q) corresponde a los valores que dividen la variable en cuatro grupos, con igual número de observaciones (25%, 50% y 75%). El RI permite obtener una idea de la dispersión de los datos, cuanto mayor es el rango más dispersos están en el conjunto[14].

$$Rango_{IQ} = Q_3 - Q_1 \quad (1)$$

Por otro lado, se determinan los límites inferior (L_i) y superior (L_s) que ayudan a identificar los valores *outliers* (atípicos), siendo éstos los datos que están fuera del intervalo (L_i,L_s). Dichos límites se obtienen de la siguiente forma: [15]

$$L_i = Q_1 - 1.5 \cdot Rango_{IQ}$$

$$L_s = Q_3 + 1.5 \cdot Rango_{IQ} \quad (2)$$

3.2.2 Correlación de Pearson: es una medida de relación entre dos o más variables, determinando su grado de asociación lineal al tomar valores entre -1 y 1. La asociación es fuerte si el valor es cercano a 1 o -1 y débil si es casi cero[16].

3.2.3 Normalización de datos: consiste en transformar los valores dentro de la base de datos para mejorar la precisión, eficiencia y tiempos computacionales en su procesamiento.

La normalización de los datos es necesaria para adecuarlos a los problemas de clasificación, debido a que muchas veces, éstos no están definidos en las mismas escalas numéricas.

Se utiliza el método de normalización ***Min-Max*** (MM) para transformar los valores a un rango entre cero y uno y viene dado por:

$$s'_{ij} = \frac{s_{ij} - min_j}{max_j - min_j} \quad (3)$$

Donde, s'_{ij} es el dato i transformado del conjunto de datos j; s_{ij} es el dato i original del conjunto de datos; min_j corresponde al valor mínimo del conjunto de datos j y max_j el valor máximo del conjunto de datos j[17].

3.3 Modelo de Clasificación

3.3.1 ¿Qué es clasificación?

La clasificación es un proceso que consiste en separar un conjunto de datos distribuyéndolos para entrenamiento y pruebas, teniendo como propósito la construcción de un modelo.

Cada instancia en el conjunto de entrenamiento contiene un "valor objetivo", es decir, las etiquetas de clase y varios atributos correspondientes a las características o variables observadas[18]. Por esta razón, la clasificación se enmarca en un aprendizaje supervisado.

En la creación del modelo se utilizan los datos clasificados para el entrenamiento, mientras que con los datos de pruebas se corrobora si su resultado es aceptable. De esta forma es posible clasificar nuevos conjuntos de datos, en donde su etiqueta es desconocida[19].

3.3.2 Clasificador bayesiano

Es un clasificador probabilístico basado en el teorema de Bayes descrito como[20]:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \tag{4}$$

Donde $P(h)$ es la probabilidad de la hipótesis h sin ninguna observación, es decir, la probabilidad a priori de h. $P(D)$ corresponde a la probabilidad a priori del conjunto de entrenamiento D. $P(D|h)$ es la probabilidad de observar los datos D, dado que se tiene la hipótesis h, se le denomina también *verosimilitud.* $P(h|D)$ es la probabilidad a posteriori de h, cuando se ha observado D.

El teorema busca la hipótesis h más probable, dado los datos observados D y un conocimiento inicial sobre las probabilidades a priori de h, proporcionando de forma directa el cálculo de dichas probabilidades.

Al buscar la hipótesis más probable para el clasificador bayesiano se debe partir por tener un conjunto H. En donde se intenta encontrar una hipótesis $h \epsilon H$ que recibe el nombre de *hipótesis máximum a posteriori* o MAP. Lo que significa, que se clasifican las instancias como las que tienen máxima probabilidad a posteriori, por lo tanto, se tiene que:

$$h_{MAP} = argmax_{h \in H} P(h|D)$$

$$= argmax_{h \in H} \frac{P(D|h)P(h)}{P(D)}$$

$$= argmax_{h \in H} P(D|h)P(h) \tag{5}$$

Donde $argmax$ es una notación que expresa el argumento máximo, es decir, el valor máximo en $P(h|D)$. Además, se ha eliminado $P(D)$ por ser independiente de h .

Para el modelo clasificador se considera el método *Naïve Bayes*, más conocido como *modelo bayesiano ingenuo.* Se utiliza éste por tratarse de un modelo simple y efectivo. Los atributos que describen a los ejemplos son independientes entre sí, conocido el valor de la variable "clase"[21].

En cada ejemplo x se describe que todas las probabilidades de los atributos son $(a_1, a_2, \dots, a_n)$, es decir, los a_i corresponden a los valores de los datos con los que se clasificarán.

Si al tomar un conjunto finito V con v_j cada una de las clases que se quiere clasificar, entonces, como la clasificación viene dada por el valor de máxima probabilidad a posteriori (5) , se tiene por tanto[22]:

$$V_{MAP} = argmax_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j) \quad (6)$$

Considerando que los atributos son independientes entre sí con respecto al conjunto "objetivo", entonces:

$$P(a_1, a_2, \dots, a_n | v_j) = \prod_i P(a_i | v_j)$$

Por lo tanto, la aproximación para el clasificador bayesiano ingenuo es:

$$V_{NB} = argmax_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad (7)$$

Así, V_{NB} expresa la salida del valor "objetivo" por el clasificador bayesiano ingenuo. Por lo tanto, este método lleva una etapa de aprendizaje donde los términos $P(v_j)$ y $P(a_i|v_j)$ son estimados en base a los datos de entrenamiento.

3.4 Modelo de Regresión

3.4.1 Máquinas de Vector Soporte

La teoría de Máquinas de Vector de Soporte (*Support Vector Machine,* SVM) tiene como propósito producir un modelo de clasificación. Las SVM forman parte de los clasificadores lineales, ya que inducen separadores lineales o hiperplanos en espacio de características de alta dimensionalidad[23]. Estas Máquinas mapean el conjunto de puntos de entrada a un espacio de características de una dimensión mayor, encontrando un hiperplano que los separe y maximice el margen entre las clases.

Matemáticamente, si para una función $f: R^N \rightarrow \{\pm 1\}$ se tiene un conjunto de entrenamiento, consistente en l muestras $(x_1, y_1), \ldots, (x_l, y_l) \in R^N \rightarrow \{\pm 1\}$, es posible que pueda ser separado por un hiperplano dado por:

$$D(x) = (w \cdot x) + b \qquad w \in R^N, b \in R \tag{8}$$

Este hiperplano (separador) satisface la restricción:

$$y_i[(w \cdot x_i) + b] \geq 1 \qquad i = 1, \ldots, l; \; y_i \epsilon \{\pm 1\} \tag{9}$$

Entre todos los hiperplanos que separan los datos se busca encontrar uno que sea *óptimo,* que otorgue el máximo margen de separación de las clases. Por lo cual, se considera como un problema de optimización y para encontrar dicho hiperplano, se toma en cuenta (9) junto con:

$$\tau(w) = \frac{1}{2}\|w\|^2 \tag{10}$$

Donde se define el *margen* τ correspondiente a la distancia mínima del hiperplano separador al punto de dato más cercano.

Si se tiene el punto x en el hiperplano separador con

$$D(x) = (w \cdot x) + b = 0 \tag{11}$$

y un punto x' fuera del hiperplano dado por

$$D(x') = (w \cdot x') + b \tag{12}$$

Al sustraer (12) menos (11) se obtiene: $D(x') = w \cdot (x' - x)$, por lo tanto, la distancia del punto x' al hiperplano separador es:

$$\frac{D(x')}{\|w\|} \tag{13}$$

Dado que si existe un τ para todos los patrones se cumple:

$$\frac{y_k D(x_k)}{\|w\|} \geq \tau \ , \quad k = 1, \dots, n \tag{14}$$

Pero, para hallar dicho hiperplano también es necesario encontrar el w que maximiza el margen τ. Puesto que el número de soluciones es grande, se fija la escala siguiente para limitarlas:

$$\tau \|w\| = 1 \tag{15}$$

Entonces el margen máximo es equivalente a minimizar la norma de w. Esto se debe a que τ se relaciona directamente con la generalización del hiperplano separador. Los puntos que se encuentran sobre el margen, son aquellos para los cuales (9) es igual a 1 y reciben el nombre de *vectores de soporte.*

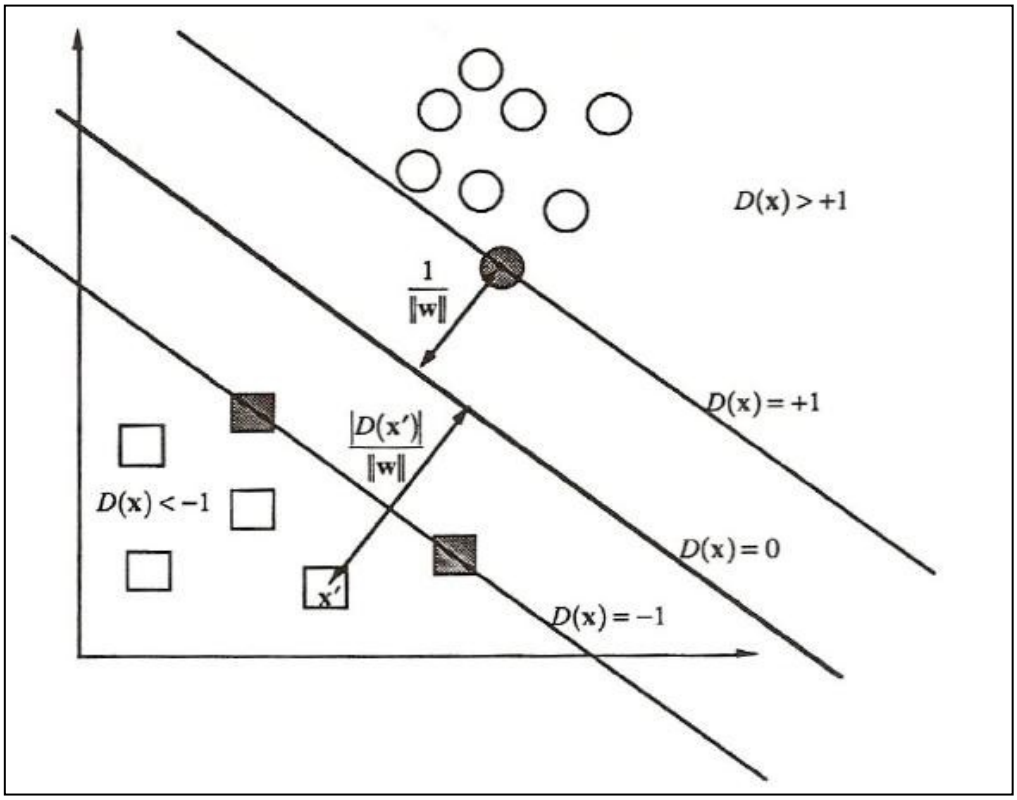


Figura 3.1 El hiperplano óptimo está definido por los puntos x, en donde D(x)=0. Pero para un x' la distancia entre el hiperplano es $D(x')/\|w\|$ ***y para un vector soporte entre su distancia que define el margen y el hiperplano óptimo es*** $1/\|w\|$.

De lo anterior y junto con la figura 3.1 se infiere, que al tratarse de un problema de optimización hay casos donde los datos no pueden ser separados linealmente a través de un hiperplano óptimo. Para ello, la transformación de los datos de un espacio lineal a otro de mayor dimensión se realiza mediante las *funciones núcleo o kernel.*

Por definición, *kernel* es un producto interno en el espacio de características, tiene su equivalente en el espacio de entrada de una transformación no lineal y se describe como:

$$K(x,x') = < \Phi(x), \Phi(x') > \quad (16)$$

Donde K es una función simétrica positiva y $< \Phi(x), \Phi(x') >$ corresponde al producto escalar.

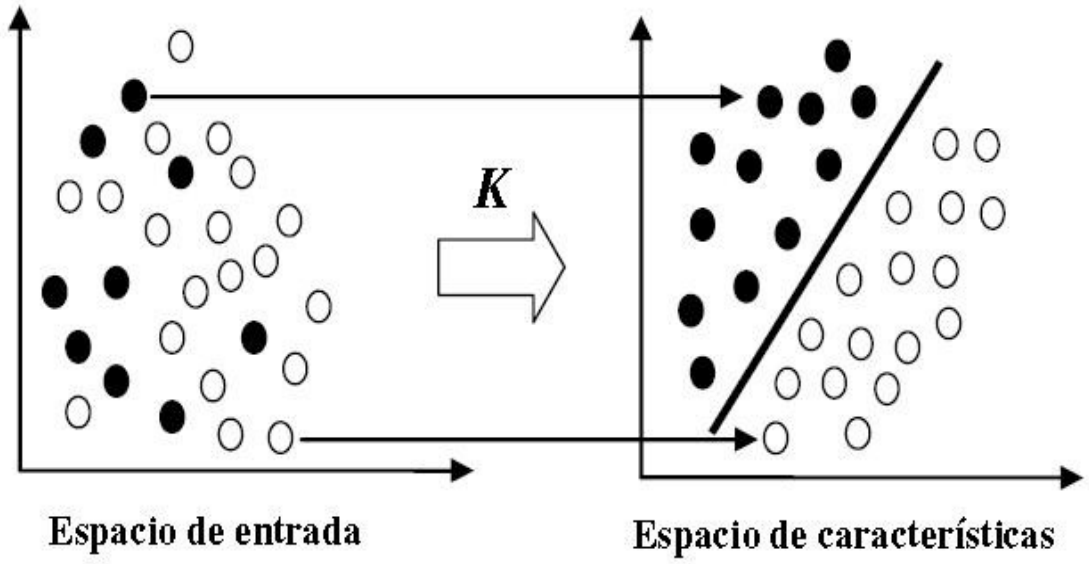


Figura 3.2 Representación de la separación de datos mediante función kernel.

La figura 3.2 ilustra la separación y el traslado de los datos al espacio de características mediante la función kernel.

3.4.2 Máquinas de Vector Soporte para regresión (SVR)

El algoritmo de Máquinas de Vector soporte para regresión trata de construir una función lineal en el espacio de características, de manera que los puntos de entrenamiento se encuentren a una distancia $\varepsilon > 0$.

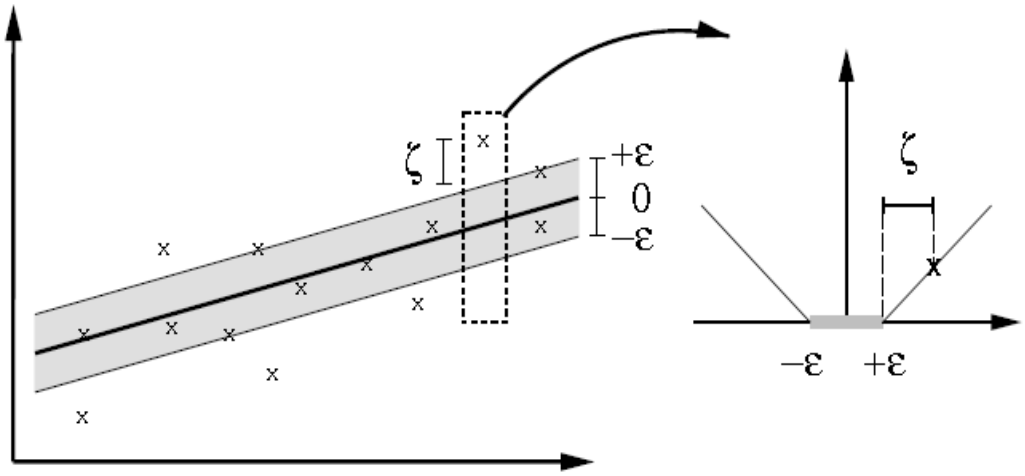


Figura 3.3 Ajuste del margen blando de Máquina de Vector Soporte.

La figura 3.3 muestra la variable de holgura ζ, con lo que es posible redefinir la condición del hiperplano como:

$$y_i[(w \cdot x_i) + b] \geq 1 - \zeta_i \quad \text{con } i = 1, \ldots, l \tag{17}$$

La razón de introducir ζ se debe a que existen datos de entrada erróneos, *outliers* o ruido en los datos de entrenamiento, lo que podría afectar negativamente a los resultados. Por lo que se crea un margen blando que pueda tolerar el ruido. En la figura 3.3 los puntos que están fuera del área sombreada con tamaño ε, contribuyen en el error.

Por otro lado, para permitir esta flexibilidad se incluye una constante C, que determina la holgura del margen blando y es obtenida en la etapa de aprendizaje.

De acuerdo a esto, se requiere la solución del siguiente problema de optimización:

$$\boldsymbol{Minimizar}: \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\zeta_i + \zeta'_i) \tag{18}$$

Sujeto a:

$$y_i - \langle w, x_i\rangle - b \leq \varepsilon + \zeta_i$$

$$\langle w, x_i\rangle + b - y_i \leq \varepsilon + \zeta'_i$$

$$\zeta_i, \zeta'_i \geq 0$$

La resolución de lo anterior es mediante cálculos de los multiplicadores de *Lagrange*, desarrollando el siguiente problema a maximizar[24]:

$$\boldsymbol{Maximizar}: -\frac{1}{2}\sum_{i,j=1}^{m}(\alpha_i - \alpha'_i)(\alpha_j - \alpha'_j)\langle x_i, x_j\rangle$$

$$-\varepsilon\sum_{i=1}^{m}(\alpha_i + \alpha'_i) + \sum_{i=1}^{m} y_i(\alpha_i - \alpha'_i) \tag{19}$$

Sujeto a:

$$\sum_{i=1}^{m}(\alpha_i - \alpha'_i) = 0$$

$$\alpha_i, \alpha'_i \epsilon\, [0, C]$$

Por lo tanto, es posible encontrar el vector w descrito como una combinación lineal dada por:

$$w = \sum_{i=1}^{m}(\alpha_i - \alpha'_i)x_i \tag{20}$$

Finalmente, la ecuación se expresa de la siguiente forma:

$$f(x) = \sum_{i=1}^{m}(\alpha_i - \alpha'_i)\langle x_i, x\rangle + b \quad (21)$$

Para la implementación de la Máquina de Vector Soporte para Regresión se ha utilizado la función núcleo de *Radial Basis Function* (RBF), ya que acuerdo con la referencia [18] este núcleo mapea muestras de forma no lineal, en un espacio dimensional superior y posee solo un parámetro γ, tal como se describe en la siguiente ecuación:

$$K(x, y) = e^{-\gamma\|x-y\|^2} \quad (22)$$

En la formulación del modelo de predicción de datos se escogen los mejores valores para parámetros de C y γ, con el propósito de caracterizar a todo el conjunto de entrenamiento. Para ello, se utilizan las técnicas de cross-*validation* y *grid-search*, que ayudan evaluar e identificar unos buenos pares de parámetros para una predicción más exacta. Estos parámetros se obtienen mediante las ecuaciones siguientes:

$$C = 2^x \quad (23)$$

$$\gamma = 2^x \quad (24)$$

3.5 Indicadores de evaluación

Los algoritmos utilizados en la creación de los modelos se evalúan mediante cuatro índices para conocer su confiabilidad. Estos índices sirven para comparar la precisión entre métodos de predicción, seleccionando los menores.

En ellos, h representa el modelo de regresión para la función objetivo f y D el set de datos. Estos indicadores son[25].

1-*Mean squared error (MSE):*

$$MSE = \frac{1}{n}\sum_{x \epsilon D}(h(x) - f(x))^2$$

2-*Mean absolute percentage error (MAPE)*:

$$MAPE = \left[\frac{1}{n}\sum_{x \epsilon D}\frac{|h(x) - f(x)|}{f(x)}\right] * 100$$

3-*Root Mean Squared Error (RMSE),* su resultado se presenta en unidades originales:

$$RMSE = \sqrt{\frac{1}{n}\sum_{x \epsilon D}(h(x) - f(x))^2}$$

4-*Mean absolute error (MAE):*

$$MAE = \frac{1}{n}\sum_{x \epsilon D}|h(x) - f(x)|$$

3.6 Metodología

Para el estudio se utiliza una metodología llamada *Knowledge Discovery in Databases* (KDD)[26]**.** La cual es un proceso que ahonda en la información de repositorios de datos para extraer conocimiento útil, con el objetivo de descubrir tendencias o patrones[27].

A continuación se describen las etapas del proceso KDD, que abarca desde la obtención de datos hasta la aplicación del conocimiento[28].

1. Integración y recopilación: se determina la procedencia de los datos y el tipo de información relevante a utilizar para el análisis. En este caso, la fuente son los registros de monitoreos hechos por el equipo *Front End* para el dispositivo WCA de la Banda 3.

2. Selección y preprocesamiento: en esta etapa se preparan los datos obtenidos transformándolos a un mismo formato, con el propósito de obtener un óptimo procesamiento de ellos. Asimismo, se realiza una selección de datos para eliminar inconsistencias, dejando sólo los datos y variables más relevantes para el análisis de clasificación y predicción[29].

3. Minería de datos: tiene por finalidad integrar los métodos de aprendizaje y estadísticas, para obtener los modelos que serán aplicados. El objetivo es extraer conocimiento, describir tendencias y predecir comportamientos.

4. Evaluación e interpretación: se evalúan los resultados obtenidos al emplear los modelos de clasificación y regresión, mediante datos de pruebas y validación.

5. Difusión y uso: esta fase posterior a la validación, comprende el uso que se le dará a los modelos, tanto para el de clasificación como el de predicción. Se utiliza el conocimiento extraído, ayudando a la toma de decisiones, en cuanto a posibles fallas o no del dispositivo estudiado (WCA)[30].

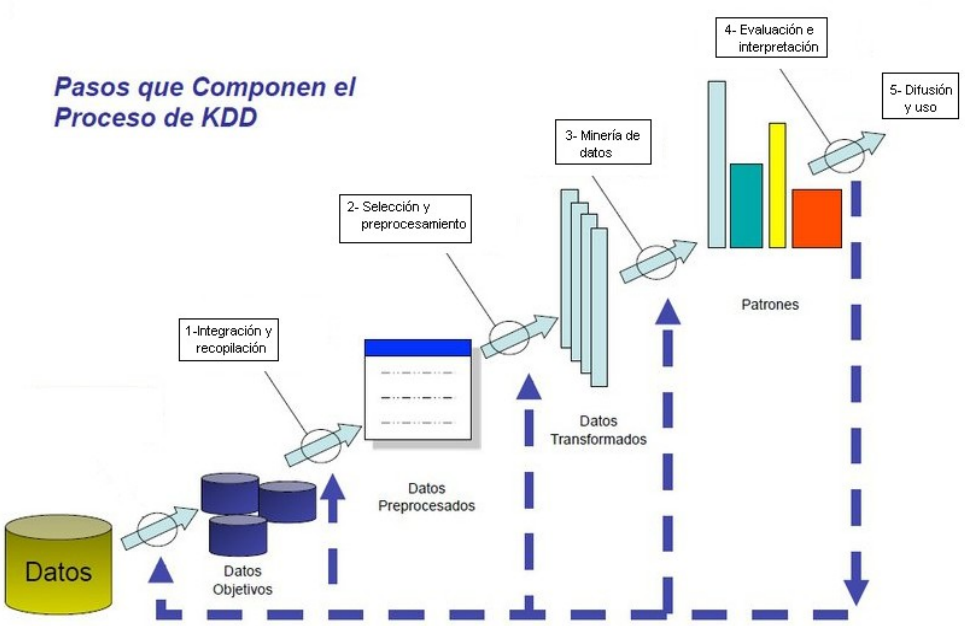


***Figura 3.4 Diagrama de proceso Knowledge Discovery in Databases* (KDD).**

El proceso KDD se emplea utilizando el Software WEKA versión 3.7, que es el acrónimo *Waikato Enviroment for Knowledge Analysis*. Fue desarrollado en la Universidad de Waikato, en Nueva Zelanda. Escrito en Java como software de código abierto, bajo los términos de la GNU GPL (*General Public Licence*).Contiene una colección de algoritmos de aprendizaje automático y minería de datos, los cuales se pueden aplicar directamente a conjuntos de datos, dadas sus herramientas para preprocesamiento, reglas de asociación, clasificación, regresión, entre otras, mediante interfaces[31] [32].

Capítulo 4

Análisis y resultados de Knowledge Discovery in Databases (KDD)

4.1 Integración y recopilación de datos del dispositivo WCA

En la primera etapa de KDD se han adquirido los registros de datos, consistentes en los valores de las variables del dispositivo WCA proporcionado por el Equipo de *Front End*. Los Ingenieros a cargo analizan y crean repositorios de datos en formato Excel, siendo ellos quienes determinan el estado operativo del instrumento. En base a dicho análisis se clasifican los valores monitoreados, etiquetándose 26 de ellos como normales (N) y 7 como fallas (F) para un total de 33 registros.

La base de datos del WCA utilizada contiene ocho conjuntos de datos para las variables *Intermediate Frequency Total Power* (IFTP) [V], *Photo Mixer Current* (PMC) [mA] y *Local oscilator Photonic Receiver* (LPR) [mW]. Asimismo, se cuenta con la variable independiente *Frequency LO* en GHz, la cual corresponde al rango de frecuencias en las que trabajan las variables anteriores, operadas en la Banda 3.

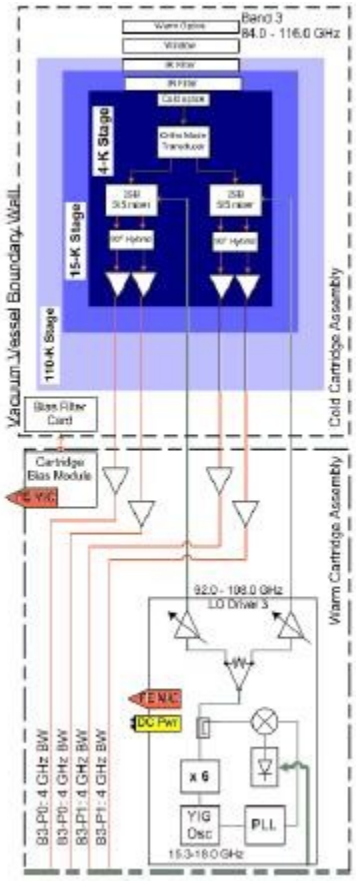


Figura 4.1 Diagrama esquemático de la Banda 3.

4.2 Selección y preprocesamiento de las variables.

Esta fase contribuye con la depuración de los datos y la determinación del conjunto con el cual se crearán los modelos.

Rango Intercuartílico (RI)

Utilizando la herramienta de Rango Intercuartílico se ha establecido la dispersión de los datos, su calidad y la posible presencia de *outliers*. Los *data sets* son evaluados con sus propios valores mínimos, máximos y cuartiles (Q1, Q2 y Q3)

La siguiente tabla resume el cálculo estadístico.

Tabla 4.1: Resultados de aplicar el Rango intercuartílico para los atributos del dispositivo WCA.

Atributos / Valores	IFTP [V]	PMC [mA]	LPR [mW]
Min	0.01	0.01	0.01
Max	4.90203857	2.44842594	5.47973396
Q1	2.45040894	1.22421297	2.73986698
Q2=Media	4.88861084	2.44842594	5.47973396
Q3	4.89578247	2.44842594	5.47973396
RI	2.44537354	1.22421297	2.73986698
Li	-1.21765137	-0.61210648	-1.36993349
Ls	8.56384277	4.28474539	9.58953444
Outliers	No hay	No hay	No hay

La Tabla 4.1 ilustra los resultados del RI, indicando que no se han encontrado datos *outliers*, por lo que el conjunto es fidedigno y no distorsiona al resto de los registros.

Correlación de Pearson

La relación lineal entre los atributos del dispositivo WCA se detalla en la Tabla 4.2 a continuación.

Tabla 4.2 Resultados de la asociación lineal entre los atributos.

Atributos	IFTP [V]	PMC[mA]	LPR[mW]
IFTP	1	0.999990568	0.999991069
PMC[mA]		1	0.999999325
LPR[mW]			1

De la Tabla 4.2 se infiere que los tres atributos tienen una alta asociación lineal dados sus valores muy cercanos a 1, por lo que estas variables son esenciales como conjunto en sí y no se debe eliminar ninguna.

Normalización de datos

Esta etapa muestra el resultado de normalizar los datos transformándolos a una escala común mediante (3), como fase previa al entrenamiento de los modelos de clasificación y predicción.

Un ejemplo de datos normalizados se presenta en la Tabla 4.3.

Tabla 4.3 Registro de datos normalizados con sus respectivas clases de un mismo WCA.

Nor. IFTP[V]	Nor. PMC[mA]	Nor. LPR[mW]	Etiqueta
0.99878355	1	1	N
0.99887712	1	1	N
0.99953213	1	1	N
1	1	1	N
0.99884593	1	1	N
0.99937618	1	1	N
0.99872117	1	1	N
0.99775424	1	1	N
0.99850283	1	1	N
0.99975047	1	1	N
0.99744233	1	1	N
0.99794139	1	1	N
0.74909128	0.748974748	0.74954294	N
0.49791744	0.497949497	0.49908588	N
0.24767152	0.246924245	0.24862882	F
0	0	0	F
0	0	0	F
0	0	0	F
0	0	0	F
0	0	0	F
0.24696973	0.246924245	0.24862882	F
0.49885317	0.497949497	0.49908588	N
0.74757072	0.748974748	0.74954294	N
0.99794139	1	1	N
0.99946975	1	1	N
0.99787901	1	1	N
0.99681851	1	1	N
0.99766067	1	1	N
0.99725519	1	1	N
0.99157841	1	1	N
0.99606993	1	1	N
0.99606993	1	1	N
0.99401132	1	1	N

La Tabla 4.3 detalla el método utilizado, clasificando los datos normalizados en el rango entre 0 y 1, dejando el valor mínimo y máximo con estos valores respectivamente. Asimismo, las etiquetas para los datos son dadas según sea su clase, si es de carácter Normal (N) o Falla (F).

Filtros para relevancia de atributos

La importancia de cada atributo se determina por ranking mediante los filtros *Information Gain Attribute Eval* y *Gain Ratio Attribute Eval,* La razón de este procedimiento es simplificar el modelo de clasificación y costo computacional, dejando sólo los atributos más relevantes.

La Tabla 4.4 muestra los resultados aplicados a los tres atributos que describen el dispositivo WCA, obtenidos mediante el software WEKA, los detalles se encuentran en el Apéndice A.

Tabla 4.4 Resumen de ranking para filtros Information Gain y Gain Ratio.

Ranking	Variables
1	LPR[mW]
2	PMC[mA]
3	IFTP[V]

El orden de relevancia viene dado considerando el primer atributo más importante que el siguiente. Al ser una cantidad pequeña de variables se determina no eliminar ninguna, por lo tanto, para las etapas siguientes se han de considerar siempre las tres.

4.2.1 Entrenamiento Modelo bayesiano ingenuo

El conjunto total de datos contiene 33 instancias (registros por variable) y es dividido en 70%, 20% y 10%, para datos de entrenamiento, prueba y validación respectivamente, seleccionados de forma aleatoria. Los detalles de estos resultados se encuentran en el Apéndice B.

Una vez entrenado, el clasificador bayesiano ingenuo entrega los valores porcentuales de clasificación, correspondiendo para el subconjunto del 70% con 23 instancias valores de:

- Porcentaje de instancias clasificadas correctamente: 86.96%
- Porcentaje de instancias clasificadas incorrectamente: 13.04%

De acuerdo a estos resultados, existe una gran proporción de confianza en cómo serán clasificados los datos de predicción.

La evaluación del modelo previo se ha realizado también con los datos de prueba y validación, siendo éstos de 6 y 4 instancias respectivamente. Al tratarse de una cantidad tan pequeña de datos en ambos casos, el modelo clasifica los datos al 100%.

4.3 Aplicación de la técnica de Minería de datos para modelo de regresión

El proceso de minería abordó el desarrollo del modelo de regresión basado en Máquinas de Vector Soporte (SVR). De acuerdo a la teoría de aprendizaje estadístico, se construyen tres modelos predictivos, uno para cada atributo en particular, dados sus respectivos parámetros C y γ. Una vez obtenidos los valores de predicción, se procede a clasificarlos mediante el modelo bayesiano,

el cual indicará la clase de los datos (Normal o Falla), entregando una estimación del estado del dispositivo.

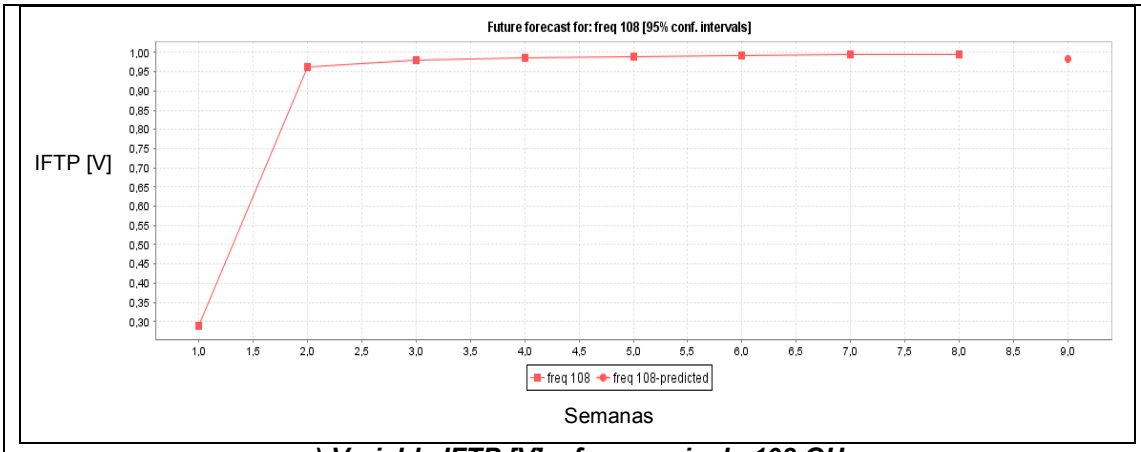
Búsqueda de C y γ para dispositivo WCA

El par de parámetros C y γ no se conoce de antemano, por lo que se realiza una búsqueda exhaustiva mediante las ecuaciones (23) y (24) y en dos iteraciones para cada uno. El criterio usado para la determinación del par se basa en los resultados de las iteraciones con la mejor precisión dada para los datos de predicción, utilizando el más viable para el entrenamiento del modelo[18].

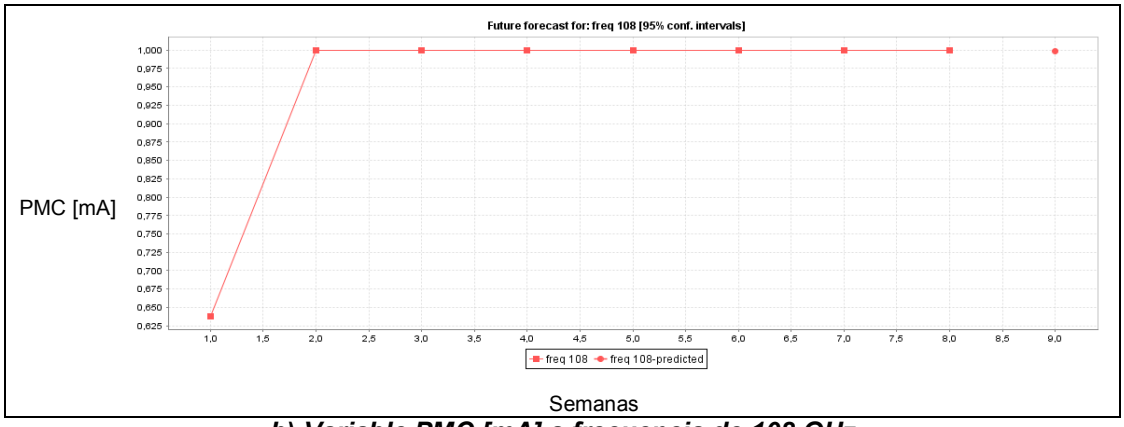
La primera iteración contiene valores altos y espaciados en un intervalo de 3 hasta completar 10 pruebas, identificando un rango específico que contenga mejores resultados y así, realizar la segunda iteración con valores más finos (igualmente en un intervalo de 0.5). Las tablas con los resultados de las iteraciones se detallan en el Apéndice C.

El parámetro épsilon (ε) usado en (18) tiene un valor de 0.001 de acuerdo a indicaciones en WEKA.

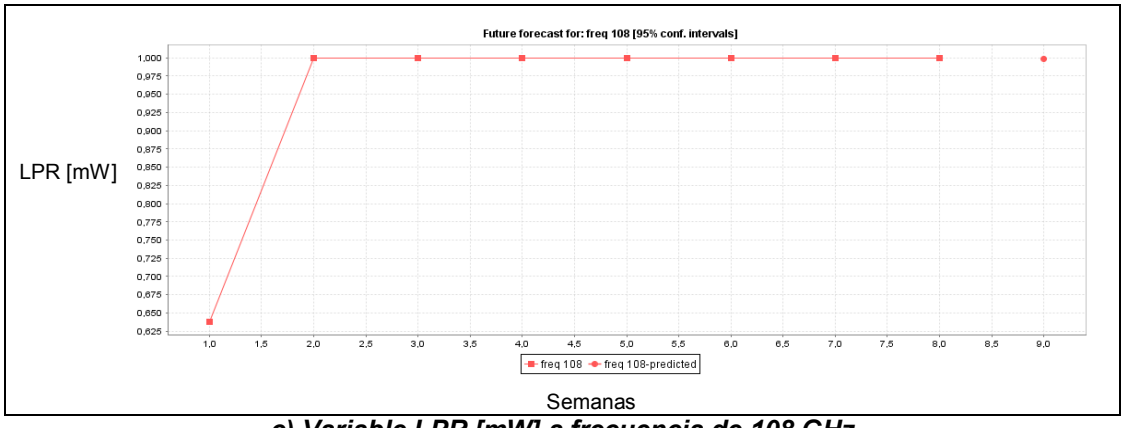
La figura 4.1 contiene el gráfico obtenido al realizar la predicción de dato por dato de cada una de las variables. A modo de ejemplo, se presentan los resultados correspondientes a la variable *Frequency LO* en el valor de 108 GHz. Los ocho primeros datos pertenecen a los registros históricos de los *data sets*, por tanto el noveno corresponde a la predicción.



a) Variable IFTP [V] a frecuencia de 108 GHz.



b) Variable PMC [mA] a frecuencia de 108 GHz.



c) Variable LPR [mW] a frecuencia de 108 GHz.

Figura 4.2. Ilustración de tendencia futura en un paso hacia delante para cada variable del dispositivo WCA.

A continuación, se detallan los resultados para cada variable con su respectiva curva, considerando que los *data sets* registrados se obtienen de observaciones monitoreadas cada dos semanas aproximadamente. Además, se ha procedido a desnormalizar los datos para ilustración de sus comportamientos.

Variable IFTP [V]

La primera iteración para el parámetro C se inicia en el intervalo entre 2^{-5} y 2^{22}, de igual forma para el parámetro γ pero con un rango entre 2^{-15} hasta 2^{12}. Los diez pares de valores (ilustrados en el Apéndice C) son ocupados para entrenar el modelo de SVR y de acuerdo a las predicciones obtenidas, tratar de estimar el mejor par que entregará los valores de predicción.

En base a los resultados detallados en el Apéndice C se ha elegido un nuevo rango, para realizar la iteración 2. En el caso de C se tiene valores entre $2^{-5.5}$ y 2^{-1}, en cuanto para γ entre $2^{-15.5}$ hasta 2^{-11}.

De la segunda iteración se obtiene que el mejor par de parámetros con los que finalmente se entrenará el modelo para la variable IFTP es: $C = 2^{-3}$ y $\gamma = 2^{-13}$.

La curva obtenida del comportamiento histórico más el set de datos de predicción se observa en la figura 4.2 que refleja la continua tendencia de la variable IFTP [V] a valores bajos para las frecuencias entre 99 GHz y 102 GHz. Tal comportamiento se observa a partir del monitoreo del *data set 2*.

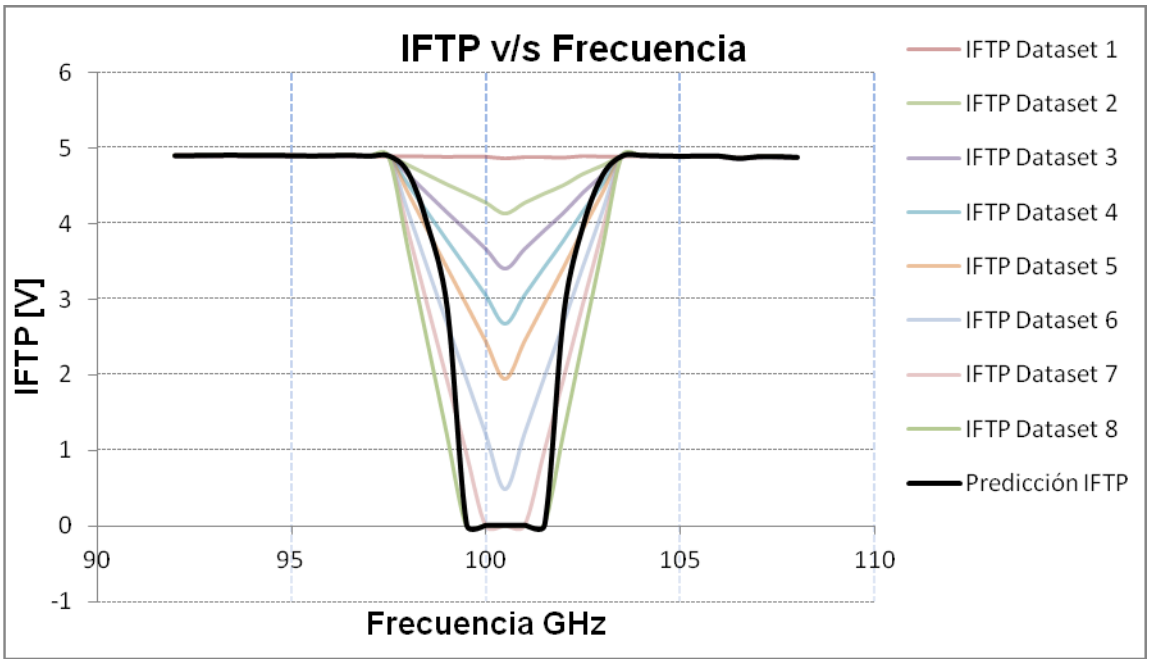


Figura 4.3. Comportamiento de variable IFTP, ilustrando los data sets monitoreados y el de predicción.

Variable PMC [mA]

El procedimiento para la variable PMC es idéntico al de la IFTP, exceptuando los valores dados para la segunda iteración, por lo que para la primera se tiene para C entre 2^{-5} y 2^{22} y para γ 2^{-15} y 2^{12}. Ahondando para la segunda iteración se encuentra para C valores de $2^{-2.5}$ y 2^{-7}, en el caso de γ $2^{-12.5}$ y 2^{-17}.

Dado estos resultados se determina que el par de parámetros más adecuado para el entrenamiento corresponde a $C = 2^{-5}$ y $\gamma = 2^{-15}$.

El comportamiento de esta variable se observa en la figura 4.3, la cual tiende a decaer obteniendo sus valores más bajos en las frecuencias 99.5 GHz y 102.5 GHz.

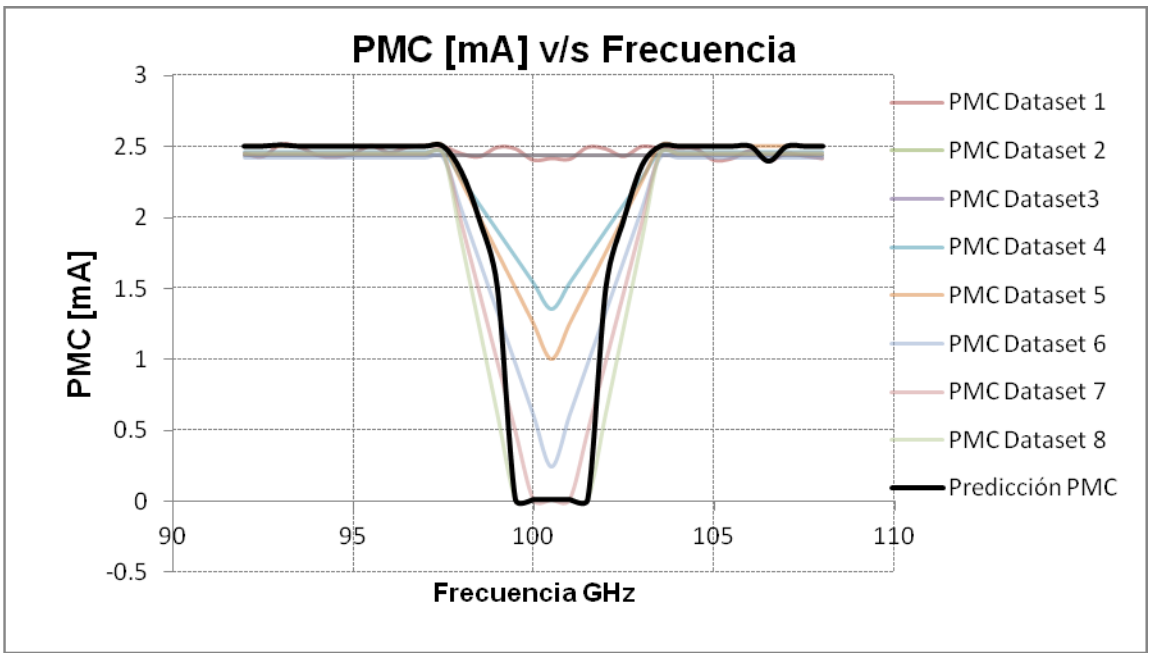


Figura 4.4. Comportamiento de variable PMC, ilustrando los data sets monitoreados y el de predicción.

Variable LPR [mW]

El método de las iteraciones se repite para esta variable, en donde la primera tiene un intervalo que varía entre 2^{-5} y 2^{22} para C y para γ 2^{-15} hasta 2^{12}. En el caso de la segunda iteración, el C está entre $2^{-6.5}$ y 2^{-2} y γ entre $2^{-16.5}$ y 2^{-12}. Por lo tanto, el par de parámetros definitivos para la predicción en este modelo es: $C = 2^{-4}$ y $\gamma = 2^{-14}$.

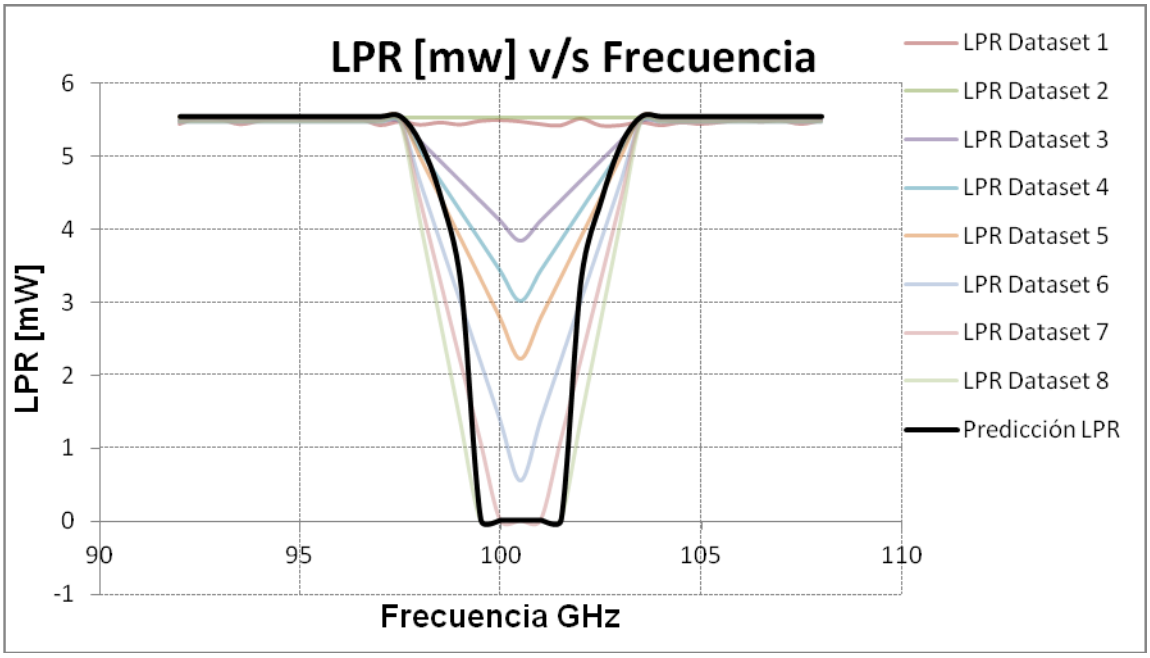


Figura 4.5. Comportamiento de variable LPR, ilustrando los data sets monitoreados y el de predicción.

La figura 4.4 detalla el comportamiento que ha tenido en el tiempo la variable LPR de los ocho monitoreos realizados y la predicción. A partir del *data set 3*, comienza a disminuir sus valores en el rango de frecuencias entre 99 GHz y 102.5 GHz.

Cabe destacar, que la interpretación final del estado y comportamiento del dispositivo es determinado por los análisis que realizan los Ingenieros de ALMA. Para que ocurra un desperfecto total del dispositivo WCA deben fallar las tres variables que lo componen (IFTP, PMC y LPR), siendo posible que continúe su funcionamiento hasta con dos de ellas con tendencia al deterioro.

Indicadores de evaluación

La Tabla 4.5 muestra un promedio de los índices con que se evalúan los modelos al momento de realizar cada predicción. Los resultados obtenidos indican las medidas del rendimiento al que los modelos se ajustan.

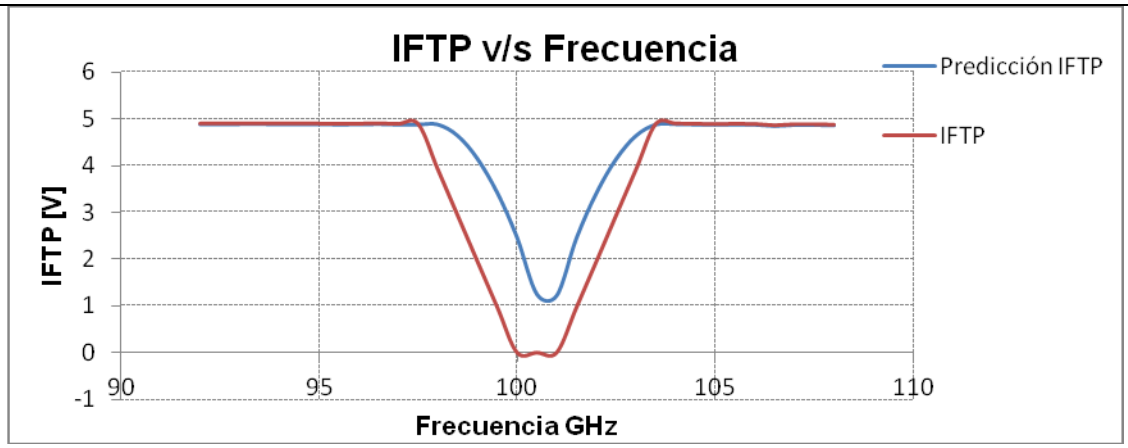
Tabla 4.5. Resultados de índices de evaluación del modelo SVR para predicción de datos (normalizados). A menor valor mayor es la precisión.

Índices	IFTP $C = 2^{-3}$, $\gamma = 2^{-13}$	PMC $C = 2^{-5}$, $\gamma = 2^{-15}$	LPR $C = 2^{-4}$, $\gamma = 2^{-14}$
MSE	0.038841935	0.037793939	0.007365625
MAPE	17.24850645	17.879225	14.71402903
RMSE	0.064854839	0.067824242	0.03585
MAE	0.058003226	0.063225	0.030065625

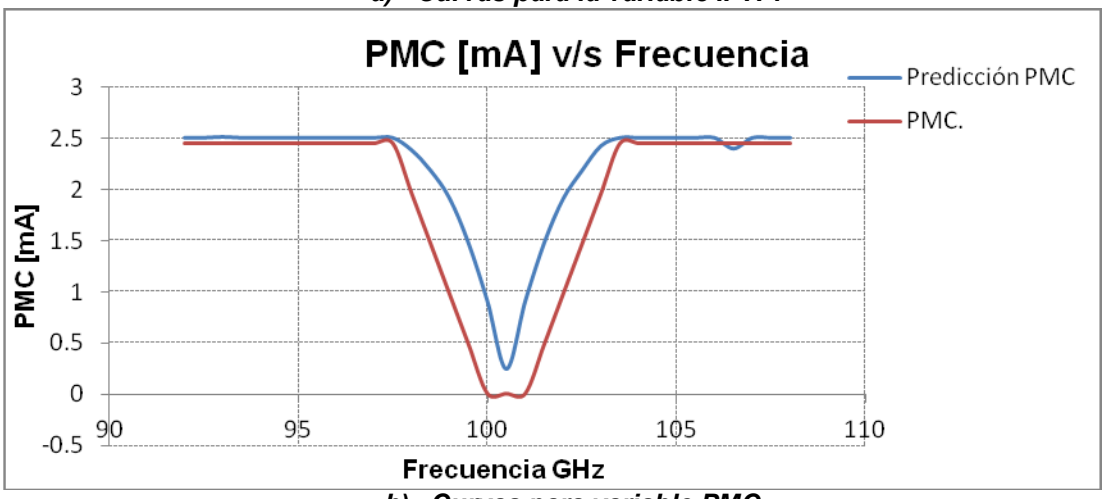
En base al entrenamiento con los pares de parámetros C y γ y la predicción de datos, los índices obtenidos son bajos en la evaluación de los algoritmos, lo cual indica la precisión de cada modelo para cada una de las variables que describen el dispositivo WCA.

Predicción data set 7

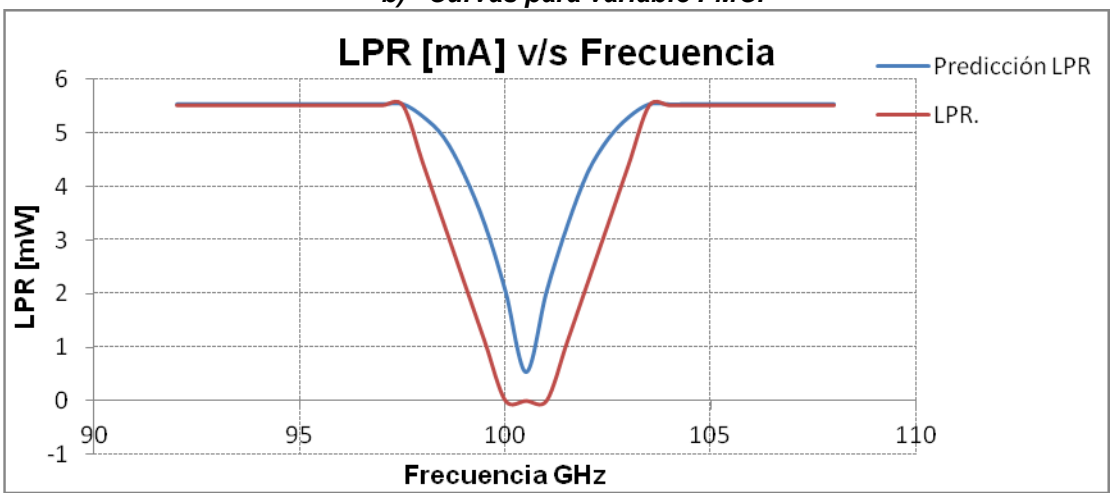
El modelo de predicción creado con SVR anteriormente se ha utilizado para predecir un nuevo data set, a modo de corroboración de sus parámetros C y γ encontrados para el dispositivo WCA. Para ello, se ha tomado nuevamente los registros monitoreados, pero sólo hasta el *data set* seis, de modo que la predicción resultante sea la número siete y se compare con los datos originales de los registros, tal como lo ilustra la figura 4.5.



a) Curvas para la variable IFTP.



b) Curvas para variable PMC.



c) Curvas para variable LPR.

Figura 4.6. Imágenes comparativas de las curvas entre registros.

De las curvas presentadas en la figura 4.5 se observa que hay una tendencia a presentar valores bajos para frecuencia que bordean los 100 GHz, tanto en los datos originales como los de predicción, por lo tanto, se corrobora que los pares de parámetros (C y γ) para realizar predicciones en el dispositivo WCA son los más indicados para el modelo.

La Tabla 4.6 muestra los resultados obtenidos de los índices evaluadores por variable.

Tabla 4.6. Resultados de índices de evaluación del modelo SVR para predicción de datos (normalizados). A menor valor mayor es la precisión.

Índices	IFTP $C = 2^{-3}$, $\gamma = 2^{-13}$	PMC $C = 2^{-5}$, $\gamma = 2^{-15}$	LPR $C = 2^{-4}$, $\gamma = 2^{-14}$
MSE	0	0	0
MAPE	0.1022125	0.032286667	0.10116333
RMSE	0.00094375	0.00031613	0.00081613
MAE	0.00087813	0.03149375	0.00081613

Los resultados de los índices entregados son bajos al evaluar los algoritmos, indicando la precisión de cada modelo para cada una de las variables del dispositivo WCA.

4.4 Evaluación e interpretación de modelos: Clasificación y regresión

En esta etapa del proceso KDD son evaluados los datos resultantes obtenidos por los modelos de predicción ingresándolos al modelo clasificador entrenado en la sección 4.2.

Las clases entregadas por el modelo bayesiano ingenuo han sido comparadas con respecto al último *data set* monitoreado por el Equipo de *Front End*. La Tabla 4.7 ilustra ambos resultados.

Tabla 4.7. Comparación de clasificación entre los valores de predicción obtenidos y última muestra de monitoreo.

Instancias	Datos de predicción	Data set 8	Resultado
1	N	N	✓
2	N	N	✓
3	N	N	✓
4	N	N	✓
5	N	N	✓
6	N	N	✓
7	N	N	✓
8	N	N	✓
9	N	N	✓
10	N	N	✓
11	N	N	✓
12	N	N	✓
13	N	N	✓
14	N	N	✓
15	N	F	✗
16	F	F	✓
17	F	F	✓
18	F	F	✓
19	F	F	✓
20	F	F	✓
21	F	F	✓
22	N	N	✓
23	N	N	✓
24	N	N	✓
25	N	N	✓
26	N	N	✓
27	N	N	✓
28	N	N	✓
29	N	N	✓
30	N	N	✓
31	N	N	✓
32	N	N	✓
33	N	N	✓

Tabla 4.8. Comparación de datos clasificados.

Instancias	Data set 7 Original	Data set 7 Predicción	Resultado
1	N	N	✓
2	N	N	✓
3	N	N	✓
4	N	N	✓
5	N	N	✓
6	N	N	✓
7	N	N	✓
8	N	N	✓
9	N	N	✓
10	N	N	✓
11	N	N	✓
12	N	N	✓
13	N	N	✓
14	N	N	✓
15	F	N	✗
16	F	N	✗
17	F	F	✓
18	F	F	✓
19	F	F	✓
20	F	F	✓
21	F	F	✓
22	N	N	✓
23	N	N	✓
24	N	N	✓
25	N	N	✓
26	N	N	✓
27	N	N	✓
28	N	N	✓
29	N	N	✓
30	N	N	✓
31	N	N	✓
32	N	N	✓
33	N	N	✓

De los resultados del modelo bayesiano ingenuo se observa que es capaz de clasificar las instancias de forma bastante cercana, con respecto al último monitoreo. La comparación de los resultados con el *data set* siete, se encuentran en la Tabla 4.8 en paralelo a la clasificación con los valores originales.

4.5-Difusión y uso

Posteriormente a la creación y evaluación de los modelos tipo clasificador y de predicción exclusivos para el dispositivo WCA, su uso e interpretación de resultados ayudan a la toma de decisiones de posibles fallas, para que los Ingenieros lo ocupen como parte del análisis del estado operacional del instrumento.

Capítulo 5

Conclusiones y Trabajo futuro

5.1 Conclusión

El estudio de fallas en el dispositivo *Warm Cartgrige Assembly* (WCA) del equipo Front End de una antena de ALMA, se desarrolló bajo la metodología del proceso *Knowledge Discovery in Databases* (KDD). Se logró crear y probar un modelo de clasificación y tres de predicción, basados en los repositorios de datos del dispositivo indicado. El uso de estos modelos contribuirá a minimizar las mantenciones correctivas, dando lugar a las preventivas.

Las herramientas estadísticas aplicadas como el Rango intercuartílico, Correlación de Pearson, Normalización de datos, favorecieron la creación de los modelos, facilitando su tratamiento en el proceso KDD, principalmente en la etapa de selección y preprocesamiento, donde comienza la creación de modelos, mediante el Software de Inteligencia computacional, WEKA.

El clasificador bayesiano ingenuo entregó las etiquetas de clase de los datos, pudiendo ser normal (N) o de falla (F). El modelo entrenado obtuvo una clasificación correcta de 86,96% y de un 13.04% de clasificación incorrecta.

La predicción de datos se basó en Máquinas de Vector Soporte para regresión (SVR). Se construyeron tres modelos, uno para cada una de las variables que componen el dispositivo WCA. Posteriormente para dichas variables se buscaron tres pares de parámetros y se eligieron los que mejor se ajustaran a las predicciones. Estos pares fueron encontrados por medio de la técnica *grid-*

search, obteniéndose para IFTP el par de $C = 2^{-3}$ y $\gamma = 2^{-13}$, en el caso de PMC el par de $C = 2^{-5}$ y $\gamma = 2^{-15}$ y el par de LPR $C = 2^{-4}$ y $\gamma = 2^{-14}$.

La base de datos utilizada cuenta con ocho *data sets* por variables, por lo que al implementar los modelos de clasificación y de regresión se obtuvo la predicción de un *data set* nueve. Para corroborar los modelos, se realizó una segunda predicción, donde se tomó de la base de datos conjunto completo hasta el *data set* seis, con el objetivo de predecir el siete y compararlo con el original. En ambos casos se logró obtener que los datos clasificados y por ende predichos, son consistentes con las muestras de datos originales, por lo tanto, los modelos funcionan para el dispositivo en específico, el WCA.

El dispositivo WCA es sólo uno más entre los muchos que componen el criostato del Equipo *Front End*, entre ellos están el *Cold Cartdrige Assemby* (CCA), *Water Vapor Radiometer (WVR), Amplitude Calibration Device* (ACD). Cada instrumento contiene variables que los caracterizan y a la vez se enlazan con los demás. Como cualquier equipo y al igual que el WCA, sufren fallas abruptas y/o deterioros paulatinos.

Como Observatorio, ALMA debe mantener las antenas operativas para las investigaciones de los astrónomos, por lo que es de suma importancia tener un análisis preventivo de su estado y así lograr en forma óptima sus ambiciosos y lejanos descubrimientos del Universo frío.

5.2 Trabajo futuro

Para próximas investigaciones o proyectos se proponen los siguientes puntos:

- Construir una aplicación que incorpore los registros de monitoreo entregando los datos etiquetados según sea su clase.
- Generar una herramienta que facilite la tarea de análisis estadístico en etapa de preprocesamiento de datos, para disminuir tiempo en desarrollo.
- Crear otros modelos de clasificación a modo de comparar los resultados de predicción.
- Realizar una aplicación que automatice la búsqueda de los parámetros C y γ.
- Utilizar otras técnicas, como Redes Neuronales para la clasificación y predicción de datos.
- Investigar la forma de adaptar los modelos al software de uso cotidiano en ALMA.

Bibliografía

[1] Sitio web: http://www.almaobservatory.org/es/sobre-alma/la-gente/administracion- de-alma/247-jorge-ibsen

[2] Sitio web: http://www.almaobservatory.org

[3] ALMA Memo 455, Cartridge Test Cryostats for ALMA Front End.

[4] Applications of Kalman filters based on non-linear functions to numerical weather predictions. G. Galanis, P. Louka, P. Katsafados, I. Pytharoulis and G. Kallos. 2006.

[5] Short term scheduling in a wind/diesel autonomus energy System. G.C. Contaxis, J. Kabouris. Members IEEE. 1991

[6] Metodologías y criterios de mantenibilidad aplicados a la organizacion y planificacion del proceso de mantenimiento de equipo electrónico de impresión. Juan F. Catalán Gudiel, Universidad de San Carlos de Guatemala, Facultad de Ingeniería, 2007.

[7] Métodos numéricos para Ingenieros. Con aplicaciones en computadoras personales. Ph.D. Steven C. Chapra, Ph.D. Raymond P. Canale. Editorial McGraw-Hill.

[8] Análisis de regresión lineal para correlacionar datos del valor *b* en catálogos de sismicidad, obtenidos con dos técnicas. Ernesto Guadalupe López Briceño. 2011.

[9] Redes Neuronales artificiales y sus aplicaciones. Xabier Basogain Olabe. Escuela Superior de Ingeniería de Bilbao, EHU. 2011.

[10] An integrated neural network model for PM10 forecasting. Patricio Pérez, Jorge Reyes. Universidad de Santiago de Chile. 2006.

[11] Predicción de la demanda eléctrica horaria mediante redes neuronales artificiales. Carlos Mallo G. Departamento de Economía Cuantitativa. Universidad de Oviedo. 2006.

[12] Time series prediction using support vector machines. Juan D. Velásquez, Yris Olaya, Carlos J. Franco. Revista chilena de ingeniería. 2009.

[13] Predicción del índice IBEX-35 aplicando Máquinas de Soporte Vectorial y Redes Neuronales. Rosillo R., Dunis CL., De la Fuente D., Pino R. 2012.

[14] Introducción a la estadística descriptiva para economistas- Joaquín Alegre Martín, Magdalena Cladera Munar, Universitat de les Balears. Palma, 2002.

[15] Diagrama de cajas y bigotes. Área de Matemáticas. Estadística y Probabilidad I. Universidad Nacional Autónoma de México.

[16] Herramientas estadísticas. Anexo 4. Instituto Nacional de economía y cambio climático (INECC).

[17] Análisis del procesamiento de los datos de entrada para un localizador de fallas en sistemas de distribución. Mg. Walter Julián Gil González, Dr. Juan José Mora Flórez, Dra. Sandra Milena Pérez Londoño. Grupo de Investigación en Calidad de Energía Eléctrica y Estabilidad de la Universidad Tecnológica de Pereira (UTP), Colombia.

[18] A Practical Guide to Support Vector Classification. Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin. Department of Computer Science. National Taiwan University, Taipei. 2010.

[19] Software para clasificación/predicción de datos. Jorge Enrique Rodríguez Rodríguez. Magíster en Ingeniería de Sistemas. Docente investigador de la Universidad Distrital Francisco José de Caldas. 2007.

[20] Métodos bayesianos. Carlos J. Alonso González. Grupo de Sistemas Inteligentes. Departamento de Informática. Universidad de Valladolid.

[21] Técnicas de clasificación en el entorno de WEKA para la determinación de cultivos de regadío (cítricos) en Librilla, Murcia (Se España). J.C. González, M.Castellón y M. J. Castejón. 2009.

[22] Aprendizaje bayesiano. Oscar J. Prieto Izquierdo. Machine Learning, Capítulo 6, Tom M. Mitchell, McGraw-Hill International Editions.

[23] Máquina de vectores de soporte. Inteligencia artificial. Capítulo IV. Gerardo Colmenares.

[24] A tutorial on Support Vector Regression. Alex J. Smola and Bernhard Schölkopf,2003.

[25] ¿Cómo medir la precisión y efectividad de los pronósticos?. Centro Ejecutivo de Logistica S. C.(Celogis). Instrutor: Ing. Tomás Gálvez Martínez, MBA y PhD.

[26] Minería de datos. Universidad Nacional del Nordeste. Departamento de Informática. Mg. David L. La Red Martínez, Ramón D. E.Lezcano.

[27] Introducción a la Minería de y al aprendizaje automático. Carlos A. González. Departamento de Informática, Universidad de Valladolid; Juan J. Rodríguez D. Departamento de Ingeniería Civil, Universidad de Burgos. Grupo de Sistemas Inteligentes.

[28] Sitio Web: http://www.webmining.cl/2011/01/proceso-de-extraccion-de-conocimiento/#B1

[29] Introducción al Data Mining. Fernando Berzal. DECSAI, Departamento de Ciencias de la Computación e I.A. Universidad de Granada.

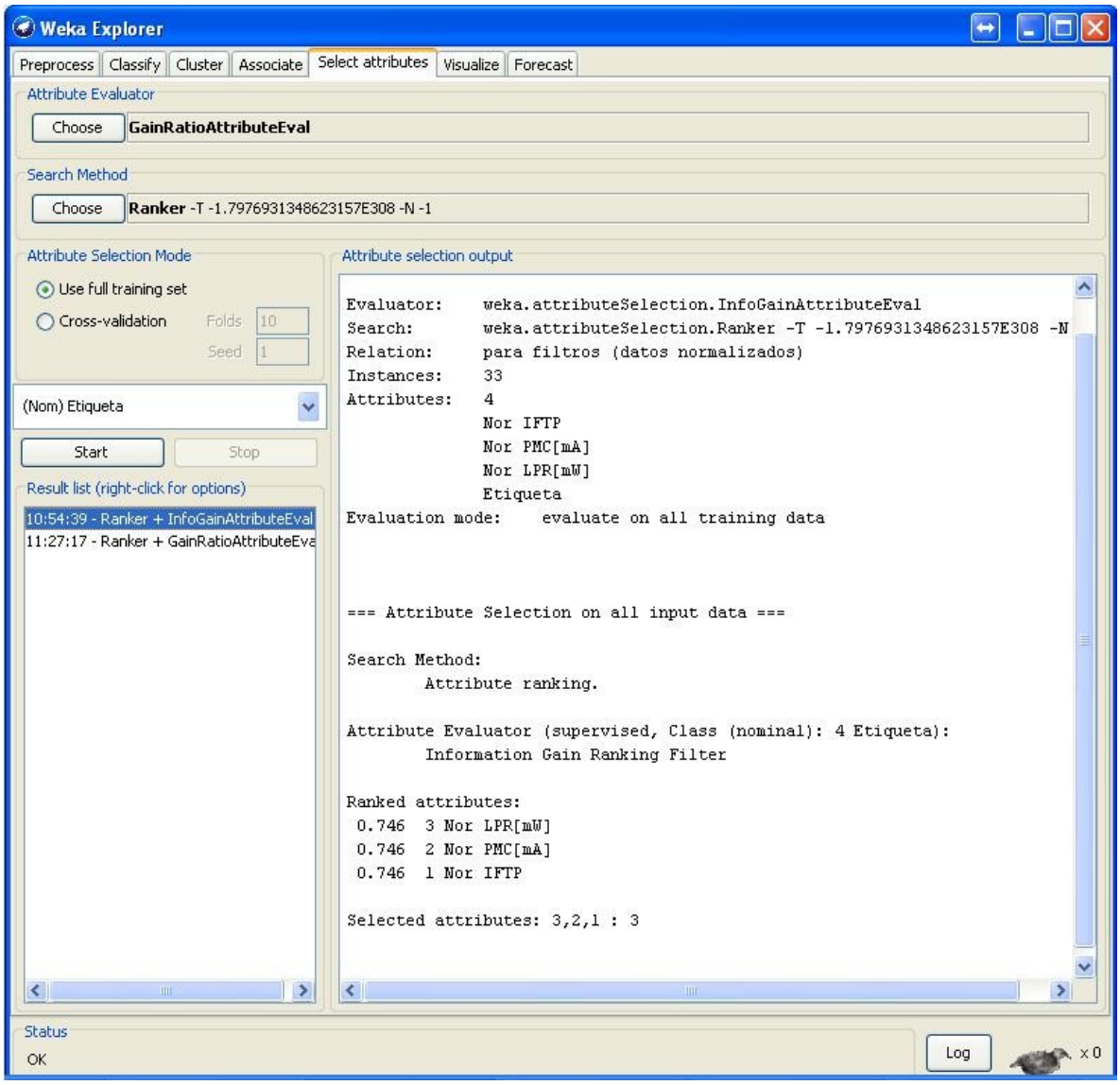
[30] Extracción de conocimiento en base de datos astronómicas. Miguel A. Montero N. Sevilla, 2009.

[31] Manual de WEKA. Diego García Morate.

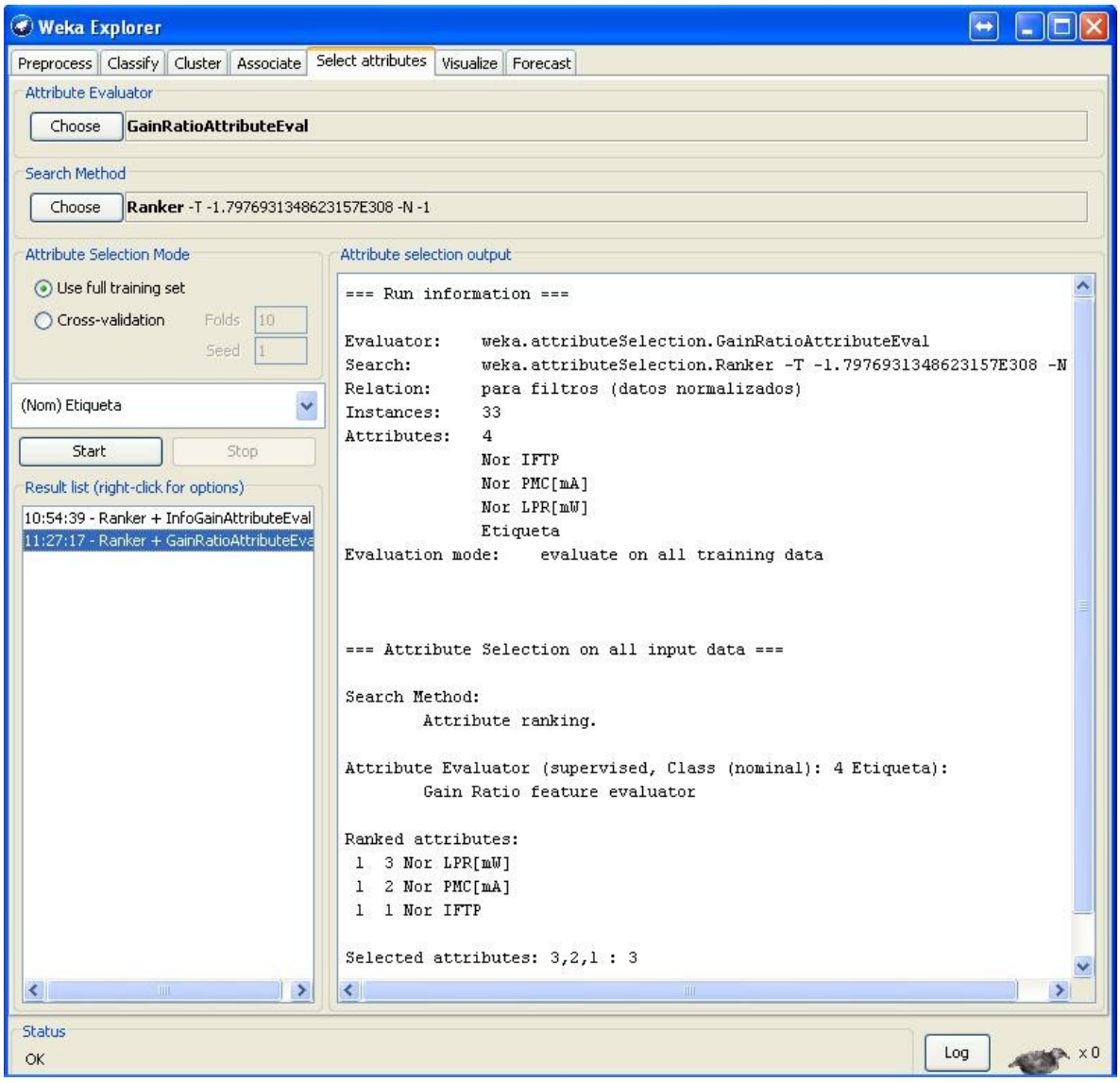
[32] Sitio web: http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka

APENDICE A. FILTROS PARA RANKING DE VARIABLES

A.1 Filtro*: Information Gain Ranking Filter*



A.2 Filtro*: Gain Ratio feature evaluator*



APENDICE B. EVALUACIÓN DE MODELO CLASIFICADOR BAYESIANO INGENUO

En las siguientes figuras, se detallan los resultados del entrenamiento para el modelo clasificador del dispositivo WCA, con datos del 70%, 20% y 10%.

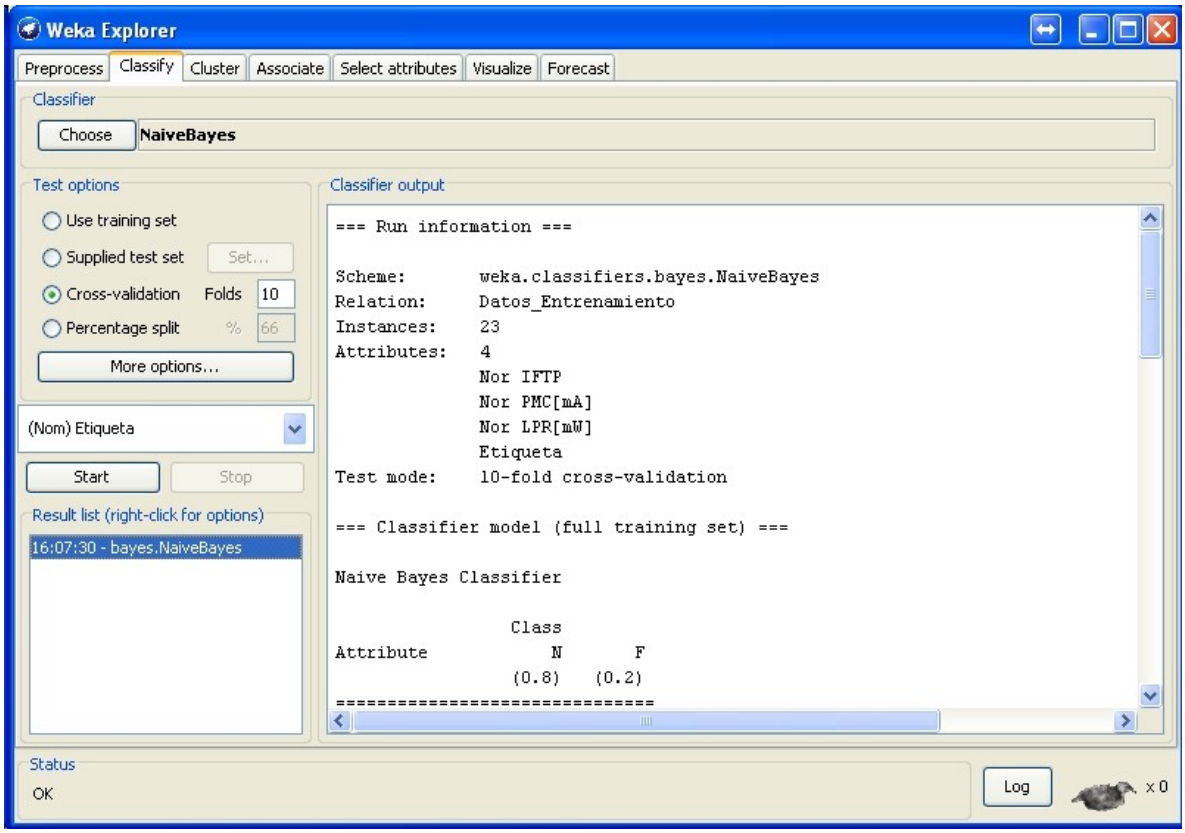


Figura B.1. Entrenamiento modelo bayesiano ingenuo para dispositivo WCA

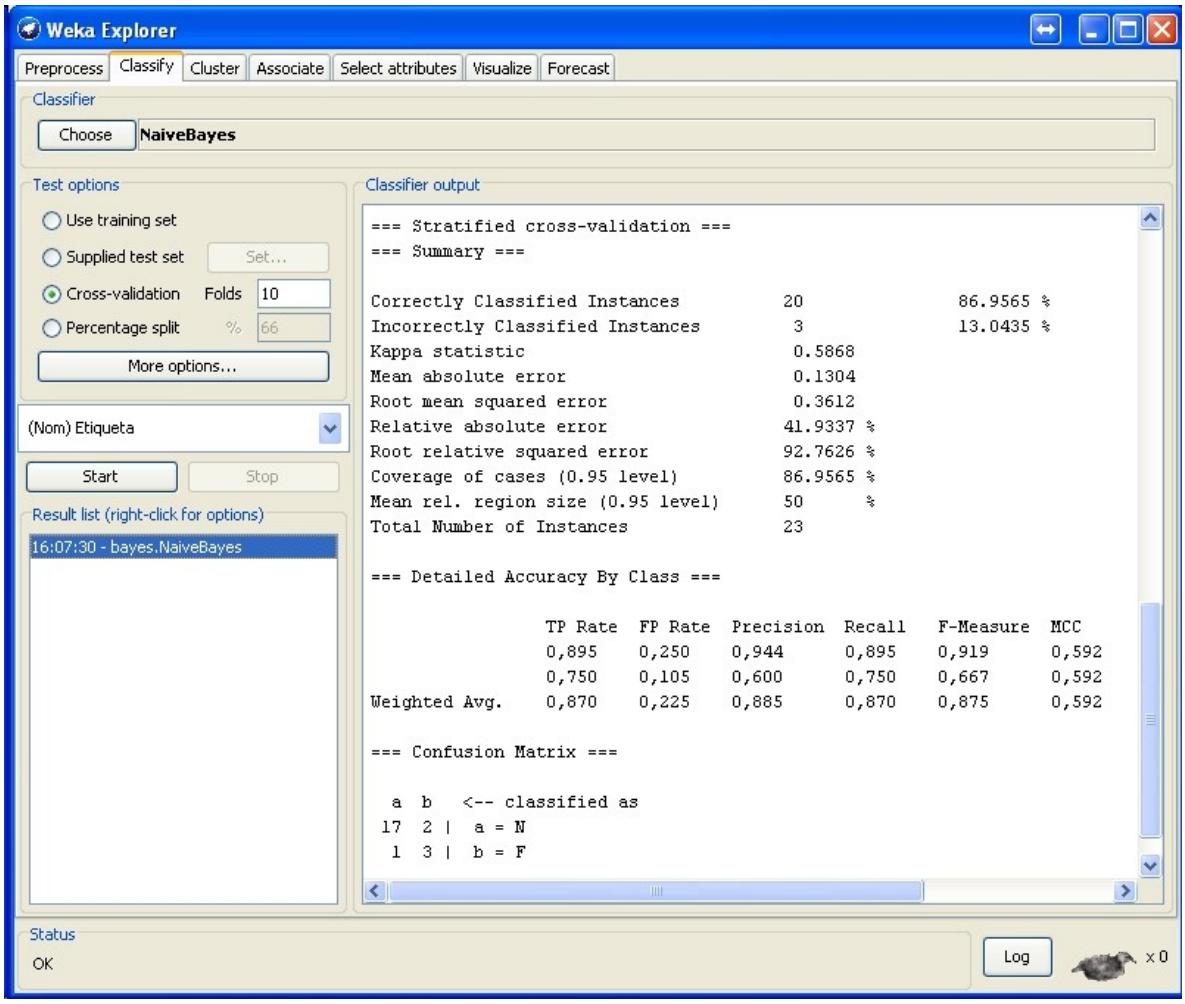


Figura B.2. Continuación a la Figura B.1 del entrenamiento de modelo bayesiano ingenuo. El conjunto de datos es el 70%.

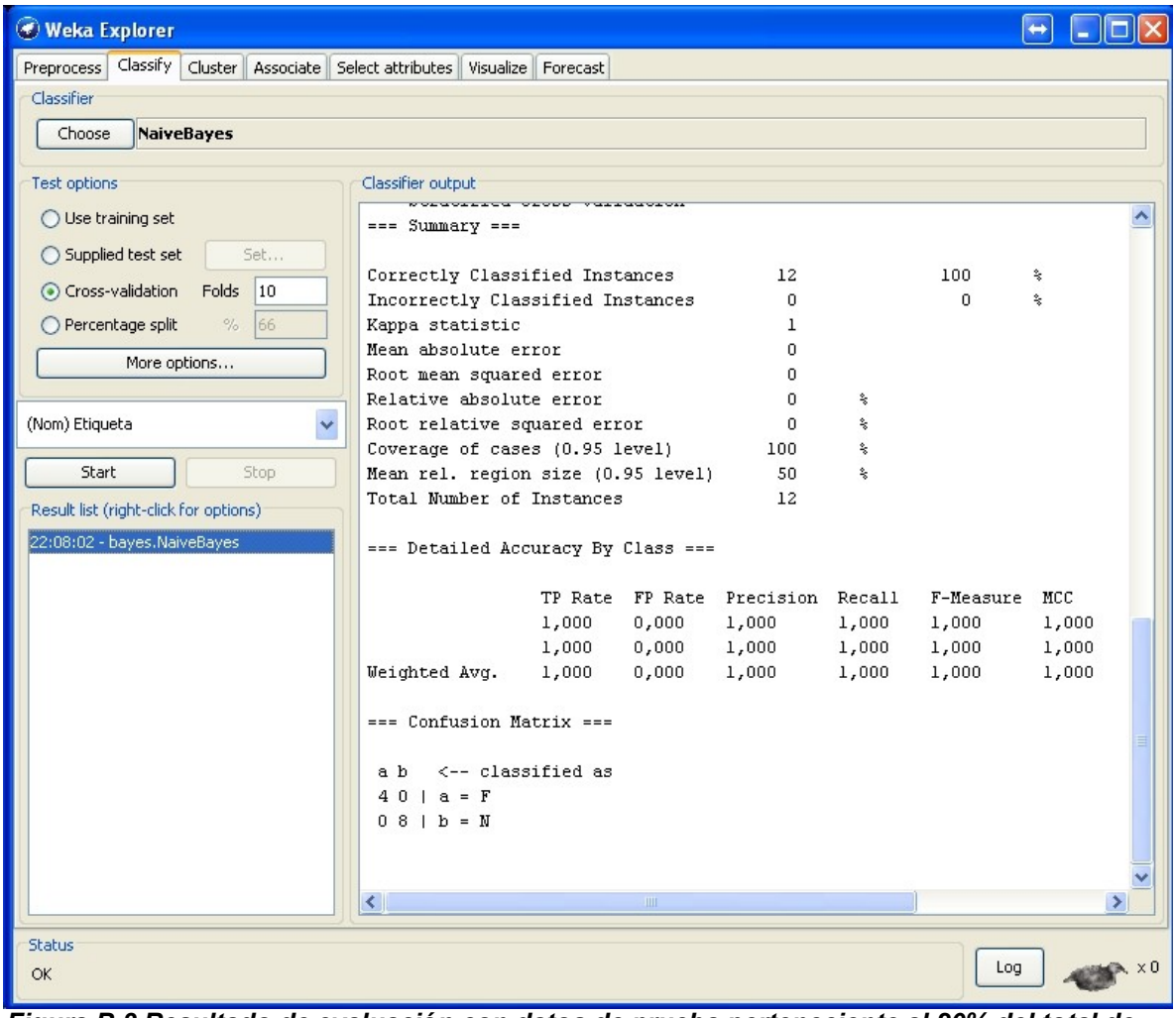


Figura B.3 Resultado de evaluación con datos de prueba perteneciente al 20% del total de instancias.

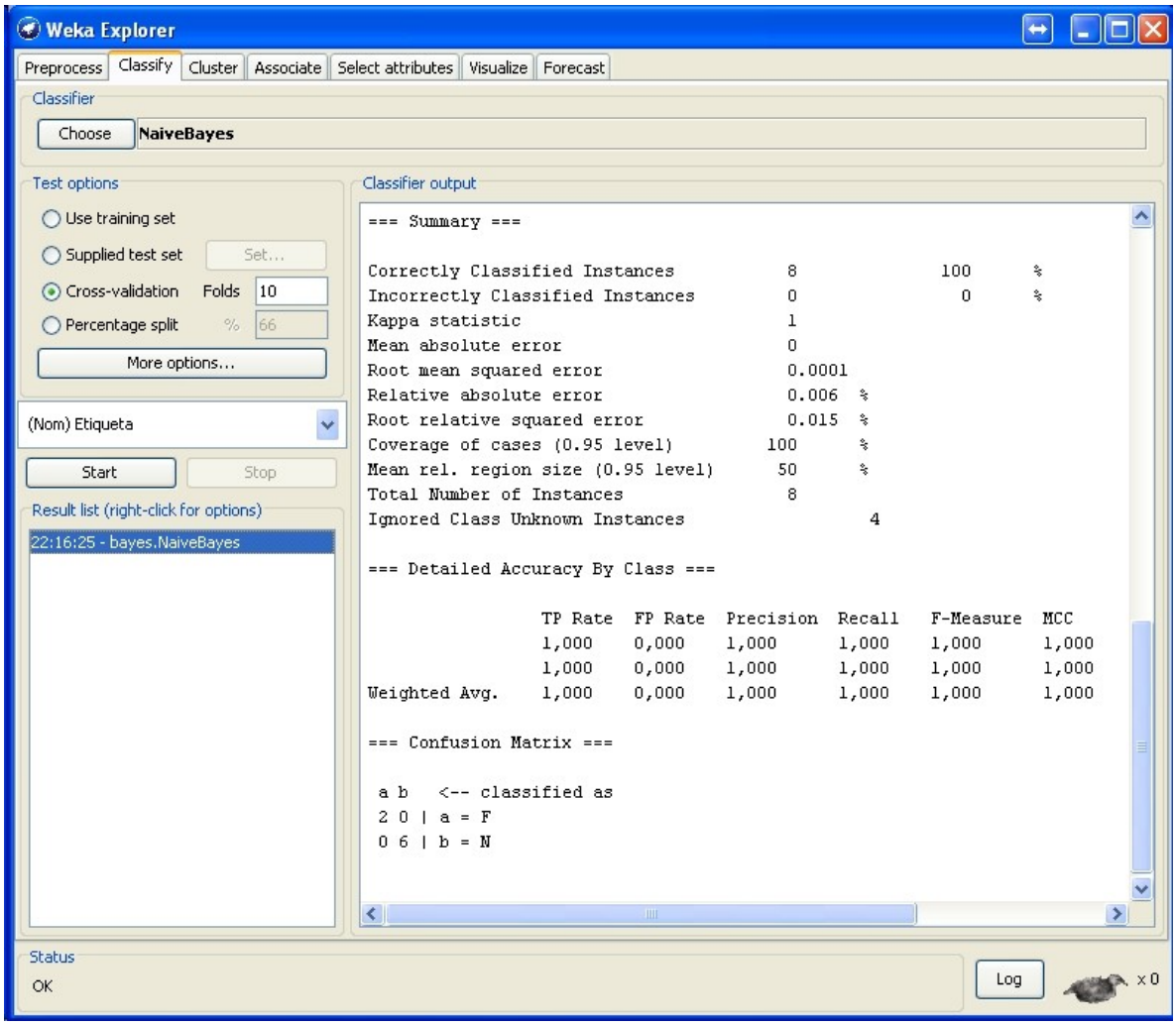


Figura B.4. Resultados para modelo bayesiano ingenuo con datos de evaluación, pertenecientes al 10% del total de instancias.

APENDICE C. TABLAS DE ITERACIONES PARA LA BÚSQUEDA DE PARÁMETROS C Y γ.

Tabla C.1: Iteración uno para atributo IFTP.

Iteración 1	C		γ	
	2^x	Valor	2^x	Valor
1	2^-5	0.03125	2^-15	3.05E-05
2	2^-2	0.25	2^-12	0.000244
3	2^1	2	2^-9	0.001953
4	2^4	16	2^6	0.015625
5	2^7	128	2^-3	0.125
6	2^10	1024	2^0	1
7	2^13	8192	2^3	8
8	2^16	65536	2^6	64
9	2^19	524288	2^9	512
10	2^22	4194304	2^12	4096

Tabla C.2: Iteración dos para atributo IFTP.

Iteración 2	C		γ	
	2^x	Valor	2^x	Valor
1	2^-5.5	0.022097	2^-15.5	2.16E-05
2	2^-5	0.03125	2^-15	3.05E-05
3	2^-4.5	0.044194	2^-14.5	4.32E-05
4	2^-4	0.0625	2^-14	6.10E-05
5	2^-3.5	0.088388	2^-13.5	8.63E-05
6	2^-3	0.125	2^-13	0.000122
7	2^-2.5	0.176777	2^-12.5	0.000173
8	2^-2	0.25	2^-12	0.000244
9	2^-1.5	0.353553	2^-11.5	0.000345
10	2^-1	0.5	2^-11	0.000488

Tabla C.3: Iteración uno para atributo PMC.

Iteración 1	C		γ	
	2^x	Valor	2^x	Valor
1	2^-5	0.03125	2^-15	3.05E-05
2	2^-2	0.25	2^-12	0.000244
3	2^1	2	2^-9	0.001953
4	2^4	16	2^6	0.015625
5	2^7	128	2^-3	0.125
6	2^10	1024	2^0	1
7	2^13	8192	2^3	8
8	2^16	65536	2^6	64
9	2^19	524288	2^9	512
10	2^22	4194304	2^12	4096

Tabla C.4: iteración dos para atributo PMC.

Iteración 2	C		γ	
	2^x	Valor	2^x	Valor
1	2^-2.5	0.176777	2^-12.5	0.000173
2	2^-3	0.125	2^-13	0.000122
3	2^-3.5	0.088388	2^-13.5	8.63E-05
4	2^-4	0.0625	2^-14	6.10E-05
5	2^-4.5	0.044194	2^-14.5	4.32E-05
6	2^-5	0.03125	2^-15	3.05E-05
7	2^-5.5	02^-5.5	2^-15.5	2.16E-05
8	2^-6	0.015625	2^-16	1.53E-05
9	2^-6.5	0.011049	2^-16.5	1.08E-05
10	2^-7	0.007813	2^-17	7.63E-06

Tabla C.5: Iteración uno para atributo LPR.

Iteración 1	C		γ	
	2^x	Valor	2^x	Valor
1	2^-5	0.03125	2^-15	3.05E-05
2	2^-2	0.25	2^-12	0.000244
3	2^1	2	2^-9	0.001953
4	2^4	16	2^6	0.015625
5	2^7	128	2^-3	0.125
6	2^10	1024	2^0	1
7	2^13	8192	2^3	8
8	2^16	65536	2^6	64
9	2^19	524288	2^9	512
10	2^22	4194304	2^12	4096

Tabla C.6: Iteración dos para atributo LPR.

Iteración 2	C		γ	
	2^x	Valor	2^x	Valor
1	2^-6.5	0.011049	2^-16.5	1.08E-05
2	2^-6	0.015625	2^-16	1.53E-05
3	2^-5.5	0.022097	2^-15.5	2.16E-05
4	2^-5	0.03125	2^-15	3.05E-05
5	2^-4.5	0.044194	2^-14.5	4.32E-05
6	2^-4	0.0625	2^-14	6.10E-05
7	2^-3.5	0.088388	2^-13.5	8.63E-05
8	2^-3	0.125	2^-13	0.000122
9	2^-2.5	0.176777	2^-12.5	0.000173
10	2^-2	0.25	2^-12	0.000244