



Universidad Nacional Experimental del Táchira
Vicerrectorado Académico
Decanato de Docencia
Departamento de Ingeniería Electrónica
Trabajo de Aplicación Profesional
Pasantías Profesionales

Sistema de simulación y reconocimiento de patrones de falla en el método de extracción de crudo por Bombeo ElectroSumergible (BES) con máquinas de aprendizaje

Autor: Duarte Montes, Angel Leonardo
Cédula de Identidad V-19.134.002
Correo electrónico: angel.duarte@unet.edu.ve
Teléfono: 0416-0484772
Tutor Académico: López Prato, José Hermes
Correo electrónico: jlopez@unet.edu.ve

San Cristóbal Julio de 2017



Universidad Nacional Experimental del Táchira

Vicerrectorado Académico

Decanato de Docencia

Departamento de Ingeniería Electrónica

Trabajo de Aplicación Profesional

Pasantías Profesionales

Sistema de simulación y reconocimiento de patrones de falla en el método de extracción de crudo por Bombeo Electrosumergible (BES) con máquinas de aprendizaje

Autor: Duarte Montes, Angel Leonardo

Cédula de Identidad V-19.134.002

Correo electrónico: angel.duarte@unet.edu.ve

Teléfono: 0416-0484772

Tutor Externo: Camargo Contreras, Edgar Alexander

Correo electrónico: camargoea@pdvsa.com

San Cristóbal Julio de 2017



PDVSA AIT Corporativo
Mérida, Edo Mérida

Una vez revisado se da el visto bueno al
presente trabajo realizado por el
Br. Angel Leonardo Duarte Montes

A handwritten signature in black ink, appearing to read "Dr. Edgar Camargo".

Dr. Edgar Camargo

Tutor Externo



Distrito Socialista Tecnológico
A.I.T. • Mérida

Mérida, Julio 2017

REPÚBLICA BOLIVARIANA
DE VENEZUELA



UNIVERSIDAD NACIONAL
EXPERIMENTAL DEL TÁCHIRA

DIE.TAP/597

ACTA TRABAJO DE APLICACIÓN PROFESIONAL

Quienes suscriben, miembros del Jurado Evaluador del Trabajo de Aplicación Profesional, designados por el Consejo de Departamento de la Carrera de Ingeniería Electrónica, ciudadanos:

JOSÉ LÓPEZ

C.I. V- 5.688.017

TUTOR ACADÉMICO

FANNY RODRÍGUEZ

C.I. V- 9.220.000

JURADO PRINCIPAL

NELSON DURÁN

C.I. V-11.020.219

JURADO PRINCIPAL

Reunidos en la sede de la Universidad Nacional Experimental del Táchira, ubicada en la ciudad de San Cristóbal, estado Táchira, siendo la hora y fecha acordada para la Defensa del Trabajo de Aplicación Profesional, modalidad Pasantía Profesional, Titulado: "Sistema de simulación y reconocimiento de Patrones de Fallas en el Método de Extracción de Crudo por Bombeo ElectroSumergible (BES) con maquinas de aprendizaje", presentado por el(la) Bachiller **ANGEL LEONARDO DUARTE MONTES**, C.I V-19.134.002, para optar al Título de **INGENIERO ELECTRÓNICO**, luego de consolidar las evaluaciones de acuerdo con lo establecido en las Normas para el Trabajo de Aplicación Profesional, declaran que la nota definitiva obtenida es de NUEVE (9 puntos en la escala de 1 a 9 y por lo tanto dan por **APROBADA** dicha Pasantía Profesional.

En constancia se firma la Presente Acta, el día 08 del mes de noviembre del 2017.

Ing. Fanny Rodríguez
(1) Jurado Principal
C. I. V-9.220.000



Ing. Nelson Durán
(2) Jurado Principal
C. I. V-11.020.219

Edificio Administrativo: Avda. Universidad - Paramillo, Teléfono: 530422. Apartado: 436 - San Cristóbal, Táchira
Oficina de Enlace: Parque Central Edificio Catuche. Nivel de Oficinas 1 – Oficinas 111 – 112
Teléfono: 577-12-16 – Fax: 577-00-80 Caracas – Distrito Federal, Venezuela



Distrito Socialista
Tecnológico
AIT Mérida

Autorización para publicación de versión digital del Trabajo de Aplicación Profesional (TAP)

Recibido por:

Hora:

Fecha:

Dirección
Biblioteca UNET
Su Despacho.

San Cristóbal, 25/07/2017

El suscrito, Br Angel Leonardo Duarte Montes titular de la C.I.N V-19.134.002
domiciliado en Veracruz-Santa Ana pasaje 17 vía la victoria casa 3-111
Municipio Córdoba, Táchira, Venezuela.

Con el fin de optar al título de Ingeniero Electrónico he presentado ante la Biblioteca de la Universidad Nacional Experimental del Táchira el Trabajo de Aplicación Profesional Titulado: Sistema de simulación y reconocimiento de patrones de falla en el método de extracción de crudo por bombeo Electrosumergible (BES) con maquinas de aprendizaje

Por medio del presente, autorizo por tiempo indefinido a la Biblioteca UNET, para publicar en forma electrónica y bajo libre acceso el documento digital de mi Trabajo, con el fin de que este pueda ser consultado públicamente vía Internet. Toda persona que consulte ya sea el formato impreso o el formato digital podrá copiar partes del texto, **citando siempre la fuente**, es decir el autor o autores, título del trabajo y año.

Esta autorización no implica renunciar a la facultad que tengo de publicar total o parcialmente la obra en otras instancias.

La UNET no será responsable de ninguna reclamación que pudiere surgir de terceros que invoquen autoría de la obra que presento.

La publicación electrónica será:

A Texto completo

Inmediata

Parcial
(Especifique Capítulos no Publicables en
Obs)

Después de 1 año
(Para casos donde se requiera
patentar)

Atentamente;

Angel 2 Duarte M.
C.I.:19134002

VBo.:

Edgar Camayo
Tutor Externo
(Sello)

Observaciones: _____

Universidad Nacional Experimental del Táchira
Vicerrectorado Académico
Decanato de Docencia
Departamento de Ingeniería Electrónica
Trabajo de Aplicación Profesional

Sistema de simulación y reconocimiento de patrones de falla en el método de extracción de crudo por Bombeo Electrosumergible con máquinas de aprendizaje

Autor: Duarte Montes, Angel Leonardo
Tutor Académico: López Prato, José Hermes
Tutor Externo: Camargo Contreras, Edgar Alexander

RESUMEN

Una de las desventajas que presenta el levantamiento artificial de crudo por Bombeo Electrosumergible, es el alto costo de sus equipos, haciendo la detección de fallas y tendencias indeseadas una necesidad. Esto nos ha motivado a presentar un sistema computacional que permita, en primer lugar, simular trece escenarios: un comportamiento normal y doce fallas, identificadas a través de la lectura de cartas amperimétricas o registros diarios de corriente. En la simulación se usan como referencia cartas reales, de las que se extraen comportamientos que permiten emular de manera aleatoria situaciones parecidas. En segundo lugar, el sistema identifica y clasifica patrones de fallas aplicando técnicas de máquinas de aprendizaje, específicamente Máquinas de Soporte Vectorial con kernels gaussianos. Para alcanzar este objetivo se requiere un proceso de entrenamiento, acondicionamiento y validación. El acondicionamiento se realiza a través de curvas de aprendizaje (número de cartas usadas vs porcentaje de aciertos) y variación de parámetros propios de la máquina: el parámetro C (compensa problemas de subajuste y sobreajuste) y el gamma (ajusta condiciones de similitud). La validación se hace a través de 4 métodos indicadores de acierto/error en problemas de clasificación (validación simple, cruzada, uno afuera y bootstrap) y mediante gráficas ROC. Se concluye que el uso de máquinas de soporte vectorial en problemas no lineales como el expuesto brinda un desempeño muy satisfactorio.

Palabras clave: Máquinas de aprendizaje, reconocimiento de patrones, bombeo electro-sumergible, fallas, computación.

Universidad Nacional Experimental del Táchira
Vicerrectorado Académico
Decanato de Docencia
Departamento de Ingeniería Electrónica
Trabajo de Aplicación Profesional

Simulation system and failure pattern recognition in the method of crude oil extraction by Electrosubmersible Pumping with learning machines

Author: Duarte Montes, Angel Leonardo
Academic Tutor: López Prato, José Hermes
External Tutor: Camargo Contreras, Edgar Alexander

ABSTRACT

One of the disadvantages of the artificial lifting of crude oil by Pumping Electrosubmersible, is the high cost of its equipment, making the detection of failures a necessity. This has motivated us to present a computational system that allows us, in the first place, simulate thirteen scenarios: a normal behavior and twelve faults, identified through the reading of amperimetric charts or daily current registers. In the simulation, real letters are used as a reference, The behaviors are extracted of that and allow emullate randomly similary situations. Secondly, the system identifies and classifies fault patterns using learning machine techniques, specifically Vector Support Machines with Gaussian kernels. To achieve this goal, a training, conditioning and validation process is required. Conditioning is carried out through learning curves (number of cards used vs. percentage of hits) and variation of machine parameters: parameter C (compensation of under- and over-adjustment problems) and gamma (similarity conditions). Validation is done through four methods of success/error indicators in classification problems (simple validation, cross validation, one outside and bootstrap) and through ROC graphs. It is concluded that the use of vector support machines in non-linear problems such as this one provides very satisfactory performance.

Keywords: Learning machines, pattern recognition, electro-submersible pumping, faults, computing

Dedicatoria

Cuando se deja el alma en cada respiro, en cada experiencia, en cada aventura vendrán días en que despertaras en medio de la nada, en la oscuridad, sin saber de donde vienes ni a donde vas. En esos momentos aparecen las encrucijadas que han de llevarte a tu destino. En ese lugar es donde escucharás la voz de Dios. Cuando sientas que no puedes continuar escucharas el eco distante de alguien que dice :”ya estas acá, la unica opcion es seguir caminando” . Eso es todo lo que necesitas.

A mi madre, quien me enseño a caminar, a hablar; a leer. Que me enseñó que nunca es tarde para emprender la lucha, que me hizo entender que aquel que nace con una llama en el corazon en algun momento debe permitirle arder,esta es una llama que jamás se apaga.

“Mi madre tuvo serio problemas al criarme, pero creo que lo disfrutó” Mark Twain

A mi padre, quien siempre alentó la creatividad y el descubrimiento. Aquel que me enseño que la rectitud, la compasion y la entrega hacia los demas deja una satisfaccion que no puede comprar el dinero. Que mas importante que aprender a golpear es aprender a recibir los golpes.

“Cuando yo tenia catorce años, mi padre era tan ignorante que no podia soportarle, pero, cuando cumplí los veintiuno me parecia increible lo mucho que mi padre habia aprendido en siete años”
Mark Twain

A mis hermanos: Annie y Leonel. Compañeros de vida, de aventuras, de juegos. Que triste es la vida cuando no se tiene a nadie con quien meterse en problemas, que esplendida es la vida cuando los platos rotos se pagan entre tres.

A mi abuela Maba, que siempre ha estado allí defendiéndome, aun cuando no lo merecía, cuidandome, salvandome de los problemas. A mi abuela Juanita, que me inculco el amor a Dios, el contribuir a la comunidad, el tender la mano.

Agradecimientos.

- Al creador: muchos dicen que te llamas Dios y estás en todos lados. No les creí. En mi búsqueda te encontré, y hasta el sol de hoy no sé cómo nombrarte.
- A mis padres quienes siempre antepusieron mis necesidades a las suyas. Quienes jamás dejaron de creer en mí a pesar de las adversidades
- A mis hermanos, quienes han estado allí cuando más los necesito.
- A María José, quien ha sido un baluarte e inspiración en momentos oscuros, sus torpes pasos y caídas sin perder la sonrisa y fe me hicieron entender y aceptar los fracasos.
- A cada uno de los compañeros con quienes tuve el gusto de compartir, aquellos que se ganaron mi profundo respeto y cariño, recuerdo el nombre de cada uno de ustedes con alegría.
- A la Universidad Experimental del Táchira por abrirme las puertas, por la formación dada, que me ha llevado a ser una persona con muchas cosas que aportar al país y al mundo.
- A Todo el personal del DST PDVSA AIT que de manera atenta, respetuosa y profesional me dieron la oportunidad de trabajar a su lado. Al Doctor Edgar Camargo que, con su actitud proactiva, optimista, de trabajo, dedicación y constancia sirvieron de motivación y estímulo.
- Al profesor José López por sus valiosas recomendaciones, consejos y ánimos. Por incentivar ese deseo de investigar, descubrir y aprender.

Índice General

Capítulo I.....	1
Introducción.....	1
Identificación de la empresa	1
Gerencia de AIT	2
Misión.....	2
Visión	3
Objetivos.....	3
Identificación del área de trabajo.....	3
Descripción de la situación actual	5
Identificación del problema	5
Proyecto y oportunidad	6
Objetivo general.....	7
Objetivos específicos.....	7
Justificación e importancia	7
Alcance.....	8
Capítulo II	9
Fundamentos teóricos	9
Antecedentes.....	9
Bases Teóricas	10
Proceso de producción de un pozo petrolero	10
Sistema de Bombeo Electrosumergible (BES)	13
Equipos de Superficie.....	14
Equipos de Subsuelo	15
Ventajas y desventajas.....	18
Cartas Amperimétricas en el sistema de levantamiento por bombeo electrosumergible BES	19
Carta de operación normal.....	19
Carta amperimétrica con picos de corriente	20
Carta amperimétrica gas libre en la bomba	21
Carta amperimétrica Bloqueo por gas en la bomba	21
Carta Amperimétrica en baja carga.....	22
Carta Amperimétrica por sobrecarga	23

Carta Amperimétrica bajo nivel de fluido	23
. Carta Amperimétrica excesivos intentos de arranque	24
Carta Amperimétricas Emulsiones o cargas en superficie	25
Carta Amperimétrica sólidos en la bomba	26
Carta Amperimétrica descarga de fluido de control	26
Máquinas de Aprendizaje	27
Tipos de aprendizaje.....	28
Aprendizaje Supervisado.....	28
Aprendizaje no Supervisado	28
Predicción	28
Minería de datos.....	29
Descubrimiento de Conocimiento en Bases de Datos.....	29
Máquinas de Soporte Vectorial SVM	30
Solución de problemas no lineales con SVM, kernels	35
Tipos de Evaluación	37
Validación simple	37
Validación cruzada.....	37
Validación uno afuera	38
Bootstrap:.....	38
Curvas ROC (Receiver Operating Characteristics)	39
Capítulo III	41
Marco metodológico	41
Tipo y nivel de investigación.....	41
Diseño de la investigación.....	42
Fases de desarrollo del proyecto	43
Fase I Búsqueda de información.....	43
Fase II: Diseñar y elaborar el sistema computacional de simulación y reconocimiento de patrones de falla en procesos de levantamiento artificial BES	43
Fase III: Realizar pruebas y análisis.....	44
Fase IV Documentar.....	44
Técnicas e instrumentos de recolección de datos	44
Capítulo IV	49
Desarrollo y Resultados del plan de trabajo	49
Etapa I: Investigar, estudiar y comprender el método de levantamiento artificial	

por bombeo electrosumergible (BES).....	49
Etapa II: Recolectar, aprender, clasificar y delimitar las alternativas que proponen las máquinas de aprendizaje para resolver problemas de reconocimiento de patrones, aplicar un algoritmo de aprendizaje que, teóricamente y basado en experiencias previas se ajuste a las condiciones descritas	50
Etapa III: Diseñar y elaborar el sistema computacional de simulación y reconocimiento de patrones de fallas en procesos de levantamiento artificial BES	52
a) Simulador.....	54
b) Reconocimiento de patrones.....	62
Etapa IV: Realizar pruebas del sistema computacional para reconocimiento de fallas	66
Etapa V: Desarrollar escenarios de prueba para la evaluación y análisis estadístico del sistema	66
Validación Simple:.....	68
Validación cruzada:.....	69
Validación Bootstrap:.....	70
Análisis ROC:.....	72
Etapa VI: Elaborar la documentación relacionada con la instalación y uso del sistema	73
Capítulo V	74
Discusión y análisis de resultados	74
Conclusiones.....	77
Recomendaciones	79
Referencias Bibliográficas	80
Anexo: Manual del programa.....	83

Índice de Tablas

Tabla 1. Matriz de confusión para la validación simple	69
Tabla 2. Porcentaje de aciertos en validación cruzada	69
Tabla 3. Matriz de confusión en validación cruzada	70
Tabla 4. Porcentaje de aciertos en validación bootstrap	71
Tabla 5. Matriz de Confusión pesimista en validación bootstrap.....	71
Tabla 6. Matriz de Confusión optimista en validación bootstrap.....	72

Índice de Figuras.

Figura 1. Proceso de producción en un pozo petrolero.....	12
Figura 2. Levantamiento por bombeo electrosumergible (BES).....	13
Figura 3. Banco de transformación-reducción.....	14
Figura 4. Variador de frecuencia.	15
Figura 5. Banco transformador-elevador.	15
Figura 6. Caja de venteo.....	15
Figura 7. Cabezal	16
Figura 8. Bomba Electrosumergible.	16
Figura 9. Separador de gas.	17
Figura 10. Sello	17
Figura 11. Motor.....	17
Figura 12. Unión de descarga.....	17
Figura 13. Carta Amperimétrica operación normal.	20
Figura 14. Carta Amperimétrica Picos de corriente..	20
Figura 15. Carta Amperimétrica Gas libre en la bomba	21
Figura 16. Carta Amperimétrica bloqueo por gas.	22
Figura 17. Carta Amperimétrica en baja carga.....	22
Figura 18. Carta Amperimétrica en sobrecarga. k	23
Figura 19. Carta Amperimétrica bajo nivel de fluido con gas y sin gas	24
Figura 20. Carta Amperimétrica excesivos intentos de arranque.....	24

Figura 21. Carta Amperimétrica excesivos ciclos de operación	25
Figura 22. Carta Amperimétrica Emulsiones o cargas en superficie.....	26
Figura 23. Carta Amperimétrica sólidos en la bomba.	26
Figura 24. Carta Amperimétrica descarga de fluido de control.	27
Figura 25. Fases del proceso KDD.....	30
Figura 26. Esquema de una máquina de aprendizaje a partir de ejemplos	31
Figura 27. Barrera de decisión de un problema de clasificación lineal.	34
Figura 28. Transformación no lineal.....	35
Figura 29. Esquema de una máquina de soporte vectorial.....	36
Figura 30.Curvas ROC.....	40
Figura 31. Diagrama del sistema de simulación y reconocimiento.	52
Figura 32. Diagrama del sistema de simulación 1.....	53
Figura 33. Diagrama del sistema de simulación 2.....	53
Figura 34. Diagrama del sistema de Reconocimiento de patrones.....	54
Figura 35. Carta operación Normal simulada	55
Figura 36. Carta con picos de corriente simulada	55
Figura 37. Carta apagado por gas en bomba simulada	56
Figura 38. Carta gas libre en la bomba simulada	56
Figura 39. Carta corriente en baja carga simulada	57
Figura 40. Carta corriente en sobre carga simulada	57
Figura 41. Carta descarga de fluido simulada	58
Figura 42. Carta bajo nivel de fluido con gas simulada.....	58
Figura 43. Carta bajo nivel de fluido sin gas simulada.....	59
Figura 44. Carta número excesivo de arranques simulada.....	59
Figura 45. Carta excesivos ciclos de operación simulada.....	60
Figura 46. Carta cargas o emulsiones en superficie simulada.....	60
Figura 47. Carta presencia de sólidos en la bomba simulada.....	61
Figura 48. Curva de aprendizaje para MSV	67
Figura 49. Grafica C vs Gamma	68
Figura 50. Graficas ROC por clase	72
Figura 51. Grafica ROC promedio	73

Capítulo I

Introducción

Identificación de la empresa

Petróleos de Venezuela, S.A. es la corporación estatal de la República Bolivariana de Venezuela que se encarga de la exploración, producción, manufactura, transporte y mercadeo de los hidrocarburos de manera eficiente, rentable, segura, transparente y comprometida con la protección ambiental. Es la empresa insigne nacional con incidencia directa en la economía, en las políticas públicas, financieras y sociales. Es la quinta empresa petrolera más importante del planeta según estudio realizado en el 2014 por Petroleum Intelligence Weekly (PIW), órgano informativo especializado que desde hace 50 años analiza el mercado global de la industria del crudo y el gas.

La empresa está formada por cuatro bloques de trabajo:

- Exploración y Producción: se encarga de la evaluación, exploración, certificación y perforación de yacimientos de petróleo. Siendo el primer eslabón de la cadena, cubre además la perforación y construcción de los pozos petroleros
- Refinación: tiene como responsabilidad mejorar y obtener los productos y derivados del petróleo a través de plantas de procesamiento y refinerías.
- Distribución y comercialización: su función es colocar los productos obtenidos (crudos) en los diferentes mercados internacionales, y (derivados) en mercados nacionales e internacionales
- Gas: Con unas reservas probadas por 147 billones de pies cúbicos, Venezuela es una de las potencias mundiales del sector gaseoso.

Aparte de estos cuatro bloques, Petróleos de Venezuela S.A necesita cubrir áreas de investigación, desarrollo social, tecnología, mercadeo entre otras, propias de una empresa de envergadura. Por ello delega funciones y objetivos en distintas gerencias, una de ellas es la gerencia de Automatización, Informática y Telecomunicaciones (AIT).

Gerencia de AIT

La gerencia de Automatización, Informática y Telecomunicaciones (AIT) se orienta a la generación, adaptación y aplicación de tecnologías que permitan aumentar la productividad, disminuir costos operacionales, mejorar la calidad de los productos y reducir el impacto ambiental de los diferentes proyectos que se realizan dentro del plan de negocios. Nace en Mérida, el 10 de noviembre de 2006, como:

“una necesidad de la industria petrolera para apalancar la soberanía tecnológica en materia de Automatización, Informática y Telecomunicaciones”.

Misión

Ser la organización que rige, provee y mantiene los servicios y soluciones integrales en tecnologías de automatización, información y comunicaciones de la corporación, contribuir a mantener la continuidad operativa y ejecutar planes, innovar y actuar como agente de transformación en PDVSA y en la sociedad venezolana con corresponsabilidad social, económica y ambiental; potenciar un ecosistema tecnológico que impulse los poderes creadores del pueblo, el conocimiento libre, el desarrollo endógeno sustentable y la economía social productiva para lograr la soberanía tecnológica.

Visión

Tener soberanía plena tecnológica que permita generar soluciones en el área de Automatización, Informática y Telecomunicaciones para el sector energético que le sirva, además, para agregar valor social.

Objetivos

- Minimizar los efectos adversos en el ambiente de producción de AIT en la implantación de cambios a la plataforma, mediante la utilización de normas y procedimientos establecidos en la organización.
- Garantizar una plataforma de automatización, tecnología de información y comunicaciones, única, integrada y coherente, que asegure el desempeño eficiente de las actividades medulares de la corporación y apalancue la tecnología en el estado y la nación.
- Apalancar con tecnología de AIT el desarrollo eficiente de las actividades de los distintos negocios de la cadena de valor.
- Habilitar la implantación de los sistemas corporativos que soportan la transparencia y rendición de cuenta de la corporación.
- Impulsar el desarrollo y estabilización de un ecosistema tecnológico que provea productos y servicios a la corporación y que genere a la vez nuevas fuentes de riqueza para la nación.

Identificación del área de trabajo

El trabajo de aplicación profesional (TAP) se desarrolló en las

instalaciones del Distrito Social Tecnológico AIT PDVSA MÉRIDA, estado Mérida, ciudad de Mérida, municipio Libertador. Dicha infraestructura es un anexo de la sede del Instituto Nacional de Capacitación y Educación Socialista (INCES) Eliecer Otaiza y se ubica en la segunda planta. Tiene un área aproximada de 150 metros cuadrados, se divide en 4 secciones:

- la oficina de la Gerencia Corporativa de prevención y control de perdidas PCP donde laboran unas tres personas (por turno) encargadas de la seguridad y control de ingreso del personal.
- Set de cubículos del Área del personal del área territorial conformada por unas ocho personas.
- Escritorios de Recursos humanos y labores administrativas en donde ejercen funciones de cuatro a seis personas
- Escritorios de Desarrollo de proyectos, donde hacen vida un personal de cinco a siete trabajadores (sin contar los pasantes). En esta área es donde se realiza el TAP, el escritorio asignado está equipado con un computador portátil y un monitor anexo.
- Sala de exposiciones y/o reuniones de trabajo

La sede del INCES Eliécer Otaiza se encuentra en la avenida los Próceres, frente a la funeraria la inmaculada.

El Distrito Social Tecnológico AIT PDVSA Mérida es una dependencia de la gerencia de Automatización, Informática y Telecomunicaciones (AIT), la organización es concebida como una red productiva para la generación de soluciones tecnológicas y sociales innovadoras, de calidad y especializadas en las áreas de automatización, informática y telecomunicaciones, que garanticen la

independencia tecnológica y el desarrollo endógeno sostenible, para la industria petrolera, gasífera y el sector energético nacional.

Descripción de la situación actual

Al ser el sistema de levantamiento artificial de crudo por Bombeo Electrosuminergible un método ampliamente aplicado en pozos y macollas, Petróleos de Venezuela S.A cuenta con estudios previos de temas como: modelados, estimación de variables, supervisión y control. Dichos estudios a nivel teórico han demostrado ser eficientes y acordes a lo esperado, sin embargo, muchos de ellos no han sido aplicados en campo. Como consecuencia, la responsabilidad a nivel operativo sigue recayendo en gran medida a la experiencia y pericia del personal, situación indeseada ya que el objetivo de toda organización es contar con métodos y procedimientos que permitan a cualquier trabajador debidamente capacitado y entrenado sustituir ausencias y faltas. Esta situación se torna en un problema ante el éxodo de personal en los últimos años.

Cuando situaciones como las descritas ocurren los niveles de producción, el tiempo de vida y la confiabilidad de los equipos involucrados en el proceso varían negativamente, generando pérdidas monetarias y riesgos en el personal que pudieron ser evitados o aminorados de haber contado con las herramientas adecuadas aplicadas en situaciones puntuales.

Identificación del problema

En la detección de fallas en el sistema BES suelen aplicarse procesos metódicos basados en manuales, experiencias previas y situaciones. Un ejemplo de esto es la metodología de análisis causa raíz (ACR) donde, al seguir una serie de pasos y deducciones se identifica y clasifica una posible anomalía. Aunque resultan ser efectivos volvemos al problema de necesitar un experto que aplique

dicha metodología. Es importante acotar que en este tipo de métodos no suelen considerarse las variables calculadas, medidas o sensadas con el suficiente peso o importancia, es decir, son procesos descriptivos. Para cambiar la situación se propone una alternativa: el uso de mecanismos y sistemas que ejecuten la misma tarea recurriendo a técnicas estadísticas y datos. Esta alternativa se apoya en casos precisos y ciertos deducidos por expertos, con anterioridad y en distintos escenarios.

Proyecto y oportunidad

Al ser el método BES ampliamente aplicado en la industria petrolera, se cuenta con información y casos avalados por expertos que relacionan los tipos de falla con el comportamiento de la corriente a través de cartas amperimétricas (registros de corriente por 24 horas). Esto les ha permitido generar patrones que sirvan para comparar visualmente la carta e identificar a qué tipo de falla corresponde.

Entonces, nace una oportunidad: Si en la actualidad existen algoritmos estadísticos (puntualmente, máquinas de aprendizaje) que permiten, a través de una serie de patrones y casos generar modelos para realizar tareas de reconocimiento tan complejas como identificar rostros, imágenes, paisajes, desglosar cadenas genéticas, realizar navegación autónoma y muchas más. ¿Por qué no aplicarlas en situaciones mucho más sencillas como es el reconocer una falla en el sistema BES? ¿Por qué no realizar un simulador que, de manera aleatoria genere escenarios de fallas?

De esta manera podrían estudiarse y plantearse distintas situaciones, además, abre la puerta al análisis de una gran cantidad de cartas en corto tiempo (ventaja si se desea conocer como se ha comportado el pozo hasta la fecha y estimar cómo debería comportarse en el futuro) sin interferir en la operación del

sistema de levantamiento.

Objetivo general

Realizar un Sistema computacional para simulación y reconocimiento de Patrones de Fallas en el proceso de Levantamiento artificial de crudo por Bombeo Electrosumergible (BES) usando herramientas de máquinas de aprendizaje (Machine Learning)

Objetivos específicos

1. Conocer en qué consiste el Proceso de Levantamiento de Crudo por Bombeo Electrosumergible.
2. Conocer Técnicas de Máquinas de Aprendizaje.
3. Desarrollar el sistema computacional para simulación y reconocimiento de Patrones de fallas en un sistema BES.
4. Realizar pruebas del Sistema computacional para reconocimiento de fallas
5. Desarrollar escenarios de prueba para la evaluación y análisis estadístico del sistema
6. Elaborar la documentación relacionada con la instalación y uso del sistema

Justificación e importancia

El sistema simula escenarios y situaciones de fallas a conveniencia sin interferir en la operación del pozo, es abierto al permitir modificar condiciones de simulación de acuerdo a nuevas consideraciones, experiencias y opiniones. Dichas tareas son útiles debido a que las características de operación del pozo se ven reflejadas en la corriente, lo que da valor a esta fuente de datos.

Considerando que la herramienta permite realizar estudios, análisis y planteamientos, así como identificar patrones al generar un conjunto de datos ilimitados aleatorios pero controlados. Teniendo en cuenta que los componentes del sistema BES son costosos, que su reparación y mantenimiento requiere personal calificado se hace evidente el aporte que el sistema computacional brindaría en el reconocimiento de fallas.

Alcance

El proyecto permite generar situaciones de operación normal y doce escenarios de falla clasificados, además estima a qué tipo de falla corresponden los valores de cada carta. Como el modelo predictor se entrena con data simulada se cuenta con cantidad ilimitada de información, en la realidad no se tiene esta ventaja por lo que es importante tenerlo en cuenta. Debido a no tener acceso a datos históricos no fue posible realizar pruebas con valores reales. Sin embargo, y de acuerdo a las características propias de la máquina de soporte vectorial y la comparación entre lo simulado (carta generada) y lo esperado (carta real) es muy buena, se espera un comportamiento aceptable al estimar y/o entrenar con valores reales. El sistema aplica al levantamiento artificial BES

Capítulo II

Fundamentos teóricos

Antecedentes

Guillén M. Paredes J. y Camacho O. (2009) presentaron el estudio titulado “*Un enfoque para la detección y diagnóstico de fallas en la instrumentación de un proceso usando reconocimiento de patrones en el dominio wavelet*” en donde, a través de un análisis espectral de frecuencia de señal, transformadas wavelet y algoritmos de reconocimiento de patrones y clasificación (distancia Euclíadiana y distancia entre vecinos más cercanos) lograron, en un tanque de mezclado, clasificar de manera exitosa 5000 señales (patrones) con fallas en 4 clases e índices de error aceptables (entre 8% y 5%). Lo interesante del estudio es que ofrece una alternativa: la transformada de wavelet para procesar datos y extraer características de entrada del modelo con la idea de mejorar el acierto de los algoritmos clásicos mencionados.

Zumoffen D. (2008) en su tesis doctoral “*Desarrollo de sistemas de diagnóstico de fallas integrado al diseño de control tolerante a fallas en procesos químicos*” aborda el diagnóstico y correcto manejo de situaciones anormales mediante distintos tipos de soluciones aplicadas a una sola unidad de proceso. Usa herramientas como la transformada de wavelet discreta, identificación de sistemas, análisis de componentes principales, sistemas de lógica difusa y redes neuronales artificiales. Una de sus premisas es obtener diseños que sean independientes de factores como: dimensión, complejidad, operatividad, tipo de proceso y estrategias de control. El sistema de diagnóstico fue aplicado en: una planta de tratamiento de aguas residuales con cuatro tipos de fallas clasificadas y en una planta de pulpa y papel donde el reconocimiento de patrones de falla y clasificación se manejó en un universo de 14 clases o fallas definidas.

Cardozo N. y Pabón L. (2009) en su trabajo de grado “*análisis de falla de pozos de Bombeo Electrosumergible en el campo cantagallo*” muestran la aplicación de la metodología de análisis causa a raíz como medio para lograr la reducción de fallas en el Bombeo Electrosumergible. Para ello recurre al uso de datos de producción, reportes de reparación de bombas y demás información propia del sistema.

Leal R. Camargo E. Aguilar J. Ríos A. Travé-MassuyèsL (2015) en su trabajo “*Análisis de Diagnosticabilidad y Localización de Sensores en un Pozo de Extracción de Petróleo por Inyección de Gas*” realizaron un análisis de fallas con modelos matemáticos de referencia. Dichos modelos comparan el comportamiento del sistema y, cuando ocurre una falla en la comprobación produce un resultado distinto a cero o residuo. Puntualmente usaron el modelo de relaciones de redundancia analíticas que, combinado con algoritmos genéticos multi-objetivos determinan la detectabilidad y aislabilidad de cinco fallas en el sistema de extracción de crudo por inyección de gas. Cuando no se cumplen las propiedades de detectabilidad y aislabilidad ubica dichos de sensores.

Bases Teóricas

Proceso de producción de un pozo petrolero

Camargo, Aguilar, Ríos y Rivas (2009), reseñan que:

El proceso de producción en un pozo de petróleo se inicia desde el instante en que los fluidos comienzan a moverse desde el radio externo de drenaje en el yacimiento y termina en el separador de producción en la estación de flujo.

El sistema de producción está compuesto por cuatro módulos principales: yacimiento, completación, pozo y línea de flujo superficial.

Las distintas etapas de desplazamiento del crudo son:

Transporte en el yacimiento: el fluido se desplaza desde el yacimiento donde se encuentra a una presión estática promedio, pasa a través del medio poroso hasta llegar a la arena que se encuentra a distinta presión, el fluido va perdiendo energía a medida que el medio sea de baja capacidad de flujo.

Transporte en las perforaciones: Los fluidos aportados por el yacimiento atraviesan la completación. Al atravesar la completación los fluidos entran al fondo del pozo con una nueva presión

Transporte en el pozo: Los fluidos ascienden hasta el cabezal a través de la tubería de producción venciendo la fuerza de gravedad y la fricción en las paredes internas de la tubería.

Transporte en la línea de flujo superficial: Al salir del pozo ocurre una caída brusca de presión a la descarga del reductor la presión es la de la línea de flujo, luego atraviesa la línea de flujo superficial llegando al separador.

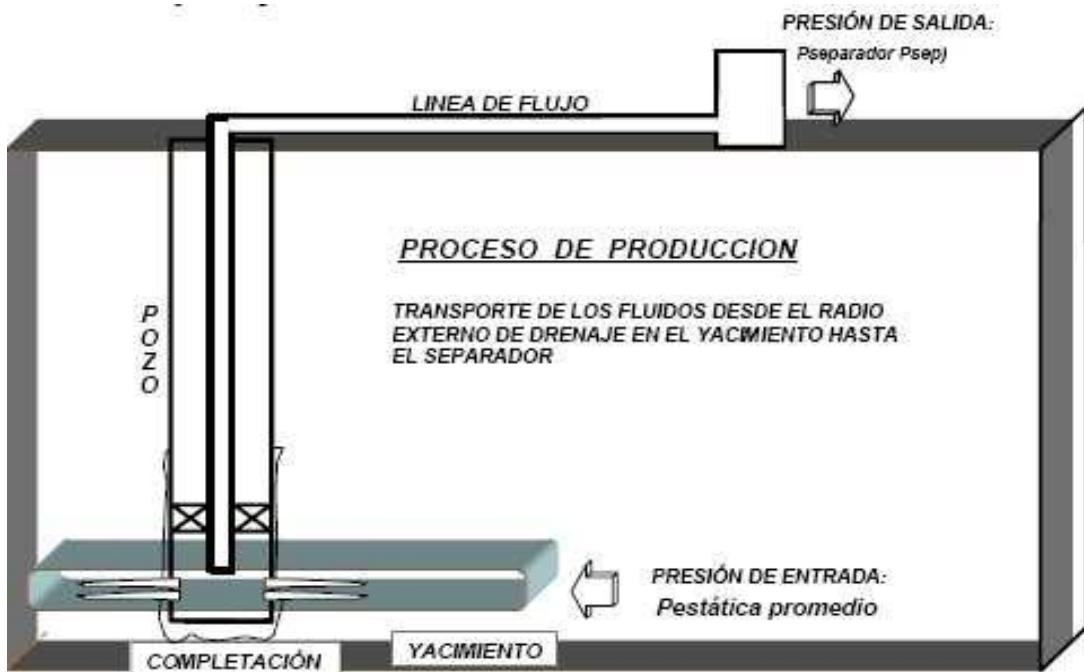


Figura 1. Proceso de producción en un pozo petrolero. Fuente: “Supervisión inteligente en la producción industrial”. Camargo E

Cuando la energía natural de un yacimiento es suficiente para impulsar los fluidos desde un punto del yacimiento hasta el fondo del pozo, y de allí hasta la superficie, se dice que el pozo produce por flujo natural. A medida que la energía del yacimiento declina, la producción del pozo disminuye hasta el punto en que no puede producir por sí solo, debiéndose adoptar una manera de disminuir la presión del fondo del pozo y la vez de transportar los fluidos hasta la superficie, incrementando de esta forma el aporte de fluidos de la formación al pozo. Esto implica, el uso de un sistema que permita proporcionar energía de manera artificial al pozo, a estos sistemas se les conoce como Sistemas de Levantamiento Artificial. (Carrillo J. 2010)

Los métodos de levantamiento artificial de mayor aplicación en la industria petrolera son: el levantamiento artificial por gas (LAG), bombeo mecánico (BMC) por cabilla de succión, bombeo electrosumergible (BES), bombeo de cavidad progresiva (BCP) y bombeo hidráulico (BHR y BHJ).

Sistema de Bombeo Electrosumergible (BES)

Los elementos primordiales del sistema BES son el motor y la Bomba. El motor se encuentra en el fondo del pozo y recibe energía eléctrica a través de cables de alta resistencia que bajan por el casing. El motor al girar transforma su energía eléctrica en energía cinética, esta energía se transfiere a la bomba a través del eje que la une al motor. La bomba es de tipo centrífuga por lo que a través de impulsores (las bombas son multietapas en el sistema BES) entra en contacto con el fluido y le transfiere energía cinética. La combinación impulsor-difusor transforma esta energía en potencial, originando una diferencia de presión en el tubing que hace que el fluido ascienda hasta la superficie.

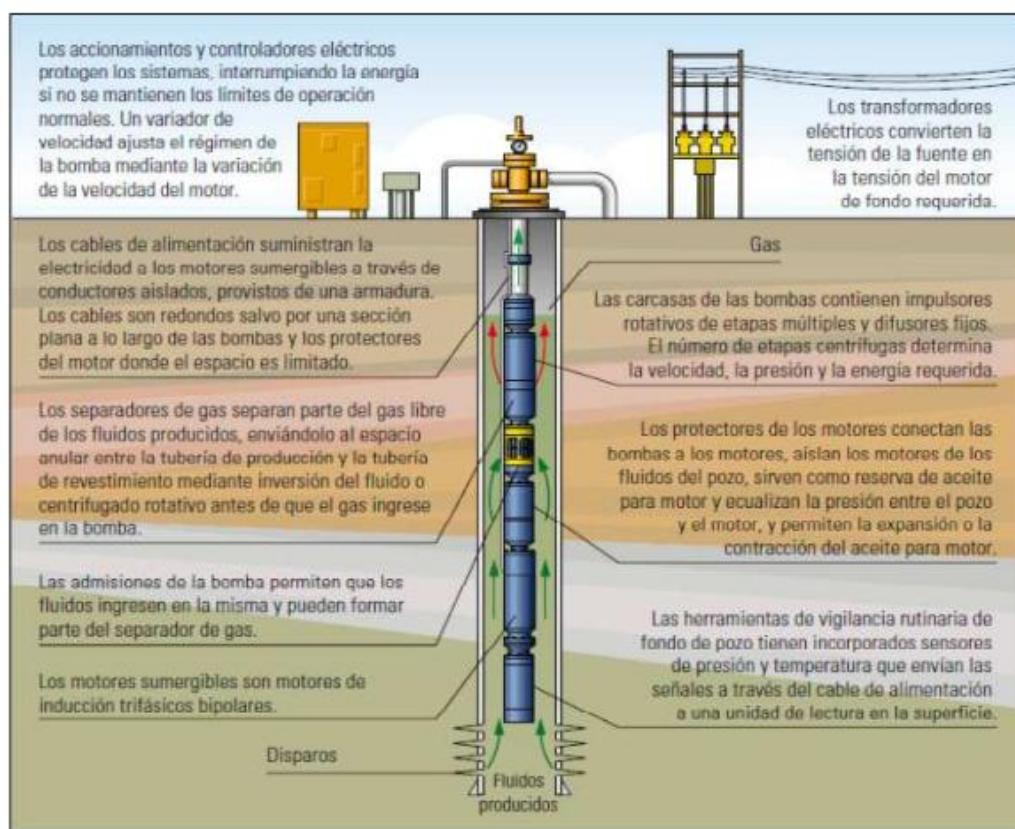


Figura 2. Levantamiento por bombeo electrosumergible (BES) Fuente: Schlumberger REDA

Los equipos que conforman el sistema por bombeo electrosumergible se dividen en equipos de superficie y equipos de subsuelo.

Equipos de Superficie

Los equipos de superficie están formados por:

Una fuente de alimentación con voltajes del orden de los kv. Un banco de transformación-reducción (figura 3) que lleva el voltaje de kv a v y suministra la energía requerida por el variador de frecuencia (figura 4). El variador de frecuencia se encarga de controlar la velocidad del motor y provee el voltaje trifásico variable necesario al banco transformador-elevador (figura 5). Este banco adecúa la tensión al valor nominal del motor que depende del fabricante, número de etapas y número de motores.

Una caja de venteo (figura 6) que sirve como punto de conexión entre el transformador-elevador y el cable de potencia, ayuda a liberar en la atmósfera cualquier gas presente en la armadura de protección del cable; también se usa como punto de acceso para pruebas y medidas. Un registrador o sensor de amperaje que se encarga de monitorear la corriente de los cables de potencia que llegan al motor, se puede sensar una sola fase, o las tres, esta última condición permite detectar con facilidad desbalances eléctricos.



Figura 3. Banco de transformación-reducción. Fuente: “análisis del desempeño de dos sistemas de levantamiento artificial, bombeo electrosumergible y bombeo mecánico VSH2 utilizados en el campo caño limón” Machete E.



Figura 4. Variador de frecuencia. Fuente: “ESP Oil International training group” ESP OIL Engineering Consultants.



Figura 5. Banco transformador-elevador. Fuente: “análisis del desempeño de dos sistemas de levantamiento artificial, bombeo electrosumergible y bombeo mecánico VSH2 utilizados en el campo caño limón” Machete E.



Figura 6. Caja de venteo. Fuente: “análisis del desempeño de dos sistemas de levantamiento artificial, bombeo electrosumergible y bombeo mecánico VSH2 utilizados en el campo caño limón” Machete E.

Equipos de Subsuelo

Están conformados por:

El Cabezal del pozo (figura 7) cuya función es monitorear y controlar altas presiones, está constituido por el conjunto de válvulas y otros accesorios que permiten conectar el cable de superficie y el cable de potencia. Tubería de producción que es por donde circula el fluido durante el proceso de levantamiento. La bomba electrosumergible (figura 8) que es del tipo centrífuga

multietapas, cada etapa tiene un impulsor (dinámico) y un difusor (estático), el número de etapas determina la carga total generada y la potencia requerida, el tamaño de etapa usada determina el volumen del fluido que va a producirse. El separador de gas (figura 9) tiene por finalidad eliminar la mayor cantidad de gas en solución contenida en el fluido, pueden ser de flujo inverso o rotativo.

El sello (figura 10) se localiza entre el motor y la bomba, está diseñado para igualar la presión del fluido del motor y la presión externa del fluido del pozo a la profundidad de colocación del apparejo. El motor electrosumergible (figura 11) trifásico, de inducción tipo jaula de ardilla, tres fases conectadas al bobinado del motor que termina en conexión estrella, pueden montarse en tandem de dos o tres motores para duplicar o triplicar la potencia La unión de descarga (figura 12) es la conexión entre todo el sistema BES con la tubería de producción.

El sensor de fondo suele ser opcional, se instala en la base del motor, monitorea la temperatura del motor y la presión de entrada a la bomba.



Figura 7. Cabezal. Fuente: “Electrical Submersible Pumping (ESP) Systems” Amao M

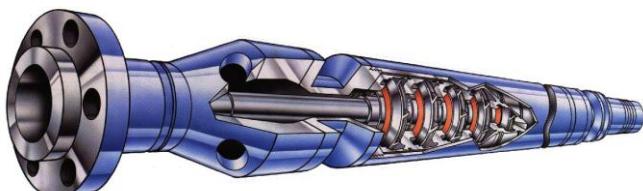


Figura 8. Bomba Electrosuminergible. Fuente: “Electrical Submersible Pumping (ESP) Systems” Amao M



Figura 9. Separador de gas. Fuente: “Electrical Submersible Pumping (ESP) Systems” Amao M

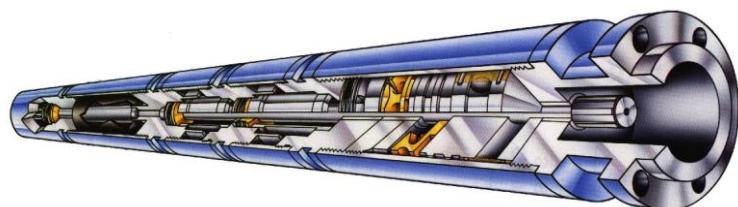


Figura 10. Sello Fuente: “Electrical Submersible Pumping (ESP) Systems” Amao M

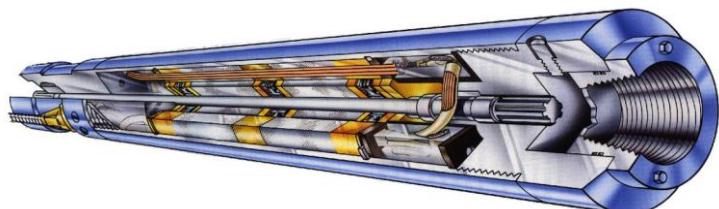


Figura 11. Motor Fuente: “Electrical Submersible Pumping (ESP) Systems” Amao M



Figura 12. Unión de descarga. Fuente: “análisis del desempeño de dos sistemas de levantamiento artificial, bombeo electrosumergible y bombeo mecánico VSH2 utilizados en el campo caño limón” Machete E

Ventajas y desventajas

El sistema de levantamiento artificial por bombeo electrosumergible comparado con cualquier proceso industrial también posee ventajas, desventajas y limitaciones.

Conforme a lo expuesto por Vásquez (2011), las ventajas del sistema BES son:

- Puede levantar altos volúmenes de fluidos.
- Puede ser usado a baja presión de fondo.
- Maneja altos cortes de agua (aplicable en costa afuera).
- Su vida útil puede ser muy larga.
- Trabaja bien en pozos desviados.
- No causa destrucciones en ambientes urbanos.
- Fácil aplicación de tratamientos contra la corrosión y formaciones de escamas.
- No tiene casi instalaciones de superficie a excepción de un control de velocidad del motor.
- La motorización es eléctrica exclusivamente y el motor se encuentra en la bomba misma al fondo del pozo.
- Su tecnología es la más aplicada y cara, pero son preferidas en caso de tener que elevar grandes caudales.

De acuerdo con Arcentales y Caizapanta (2011), la principal desventaja del sistema BES es que se requiere una inversión inicial muy alta, sumándole una baja rentabilidad en pozos con baja producción y ciertas limitaciones como:

- La temperatura del cable debe ser conocida y revisada, si no se produce deterioro de los cables.
- La potencia disponible debe suplir los Hp requeridos por la bomba. Altos contenidos de gas pueden disminuir la capacidad del sistema.
- Alto contenido de sólido causa fallas prematuras en el sistema.

- Diseño complejo.
- Carece de flexibilidad en la tasa de producción.
- Usualmente se requiere de un tiempo mayor que en los otros métodos para corregir fallas. Esto se debe a que la unidad total se encuentra en el subsuelo.

Cartas Amperimétricas en el sistema de levantamiento por bombeo electrosumergible BES

La carta amperimétrica es una herramienta valiosa que permite realizar un monitoreo al pozo, ya que registra los valores sensados en la corriente del motor. Como cualquier situación inesperada en el sistema se refleja en esta corriente, un correcto análisis permite determinar anomalías. Debido a esto se pueden tomar correctivos y estudiar el comportamiento del pozo y/o los elementos que lo conforman a lo largo del tiempo. Usualmente las cartas son generadas cada 24 horas, en algunas situaciones se generan semanalmente; nos centraremos en el primer caso.

Antiguamente la carta se generaba en hojas de papel que debían ser cambiadas, hoy en día estos valores son tomados de manera digital. Sin embargo el comportamiento de la corriente sigue siendo el mismo. Por ello el llevar estos valores de corriente a coordenadas polares nos arroja una representación idéntica, propicia a la interpretación.

Carta de operación normal

Bajo condiciones normales debería trazarse una curva suave y simétrica con un valor cercano a la corriente nominal, se espera un pico de corriente en el arranque

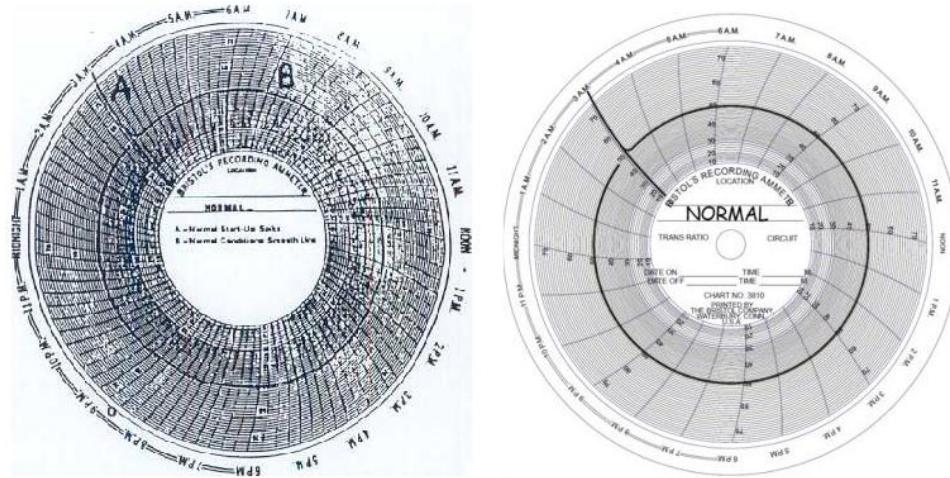


Figura 13. Carta Amperimétricas operación normal. Fuente: Electrical Submersible Pump HandBook

Carta amperimétrica con picos de corriente

Picos que perturban las condiciones normales de operación, la causa más común son cargas periódicas del sistema que pueden ser consecuencia de descargas eléctricas, fluctuaciones en la fuente primaria o encendido de otros elementos conectados a la misma fuente.

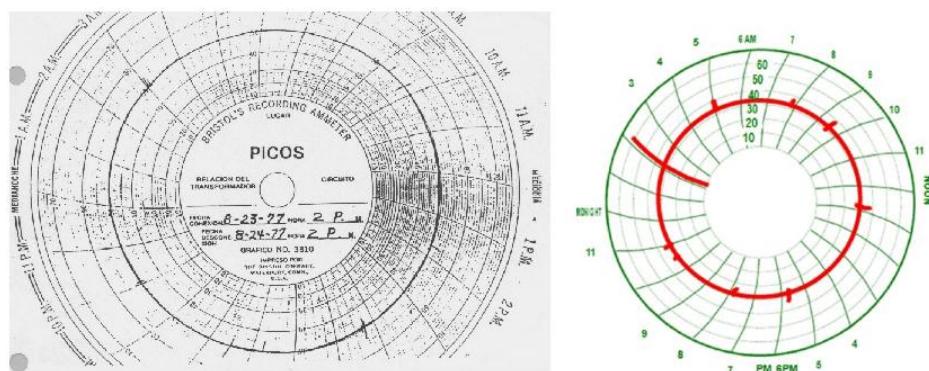


Figura 14. Carta Amperimétricas Picos de corriente. Fuente: “ESP Oil International training group” ESP OIL Engineering Consultants.

Carta amperimétrica gas libre en la bomba

El valor de la corriente se encuentra cercano nominal, pero con variaciones erráticas que no alcanzan a superar los valores de corriente máximo y mínimo necesarios para apagar la bomba. La presencia de gas puede deberse a la presencia de fluidos emulsionados, para solventarlo se sugiere descender aún más la bomba en el pozo, o colocar un separador de gas (en caso de no tenerlo)

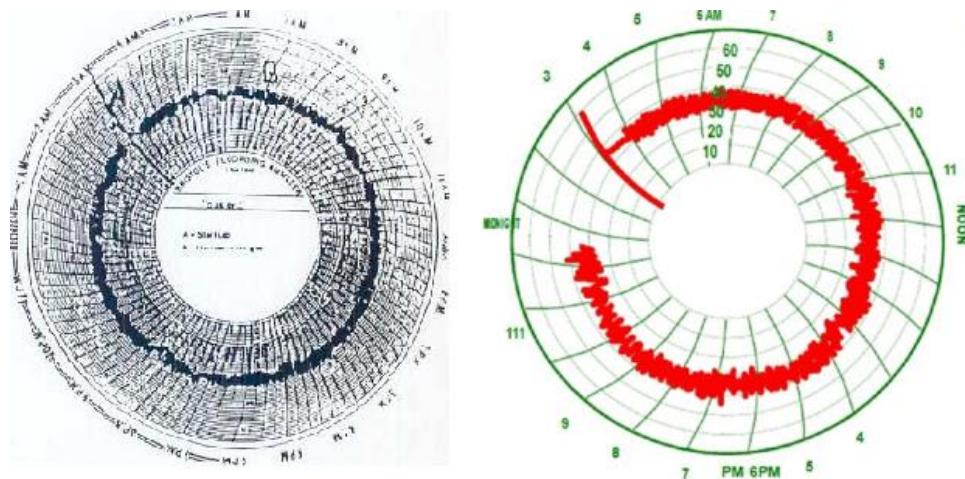


Figura 15. Carta Amperimétrica Gas libre en la bomba Fuente: Electrical Submersible Pump HandBook

Carta amperimétrica Bloqueo por gas en la bomba

La bomba arranca y el nivel de corriente se estabiliza por determinado tiempo, luego la presencia del gas hace que la corriente varíe erráticamente superando los límites permitidos. Como consecuencia la bomba se apaga, las posibles causas son: disminución del nivel del fluido o formación de emulsiones, se puede descender o rediseñar la bomba

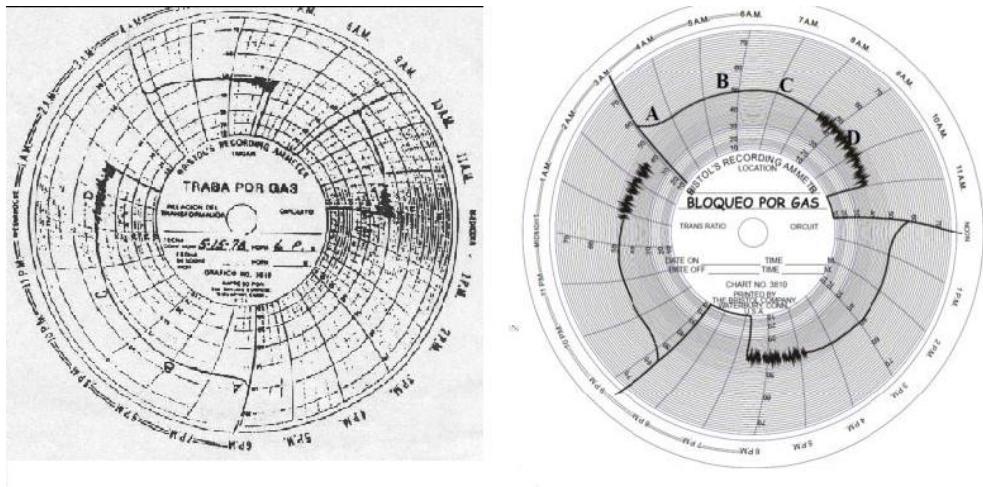


Figura 16. Carta Amperimétrica bloqueo por gas. Fuente: “ESP Oil International training group” ESP OIL Engineering Consultants.

Carta Amperimétrica en baja carga

La unidad bombea fluido del pozo hasta que no hay fluido para producir, pero el motor continúa funcionando sin carga hasta que se genera suficiente calor para apagar o quemar el motor. Puede deberse a una unidad sobredimensionada o a la elección de la alarma de baja carga ajustada a un valor muy bajo.

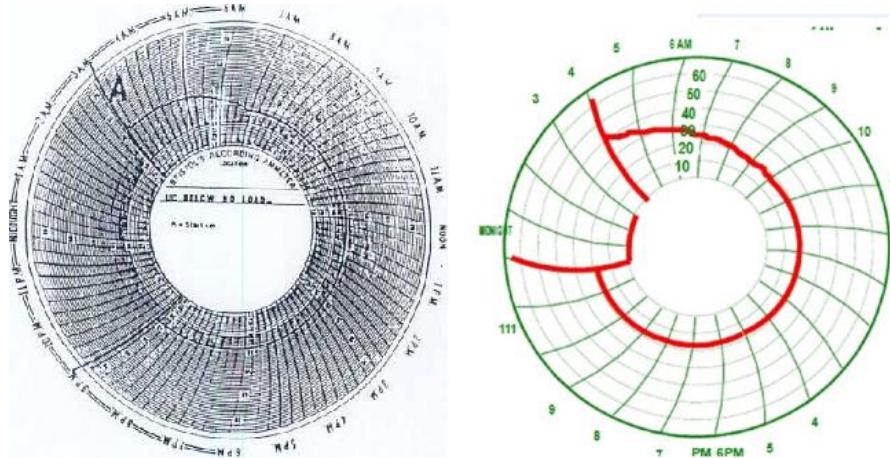


Figura 17. Carta Amperimétrica en baja carga. Fuente: “Electrical Submersible Pump HandBook

Carta Amperimétrica por sobrecarga

Se debe a un incremento en la viscosidad del fluido, a la formación de arena, problemas eléctricos o mecánicos en el equipo de subsuelo o de superficie. Hasta que la causa de sobrecarga no sea determinada no se debe intentar arrancar el equipo

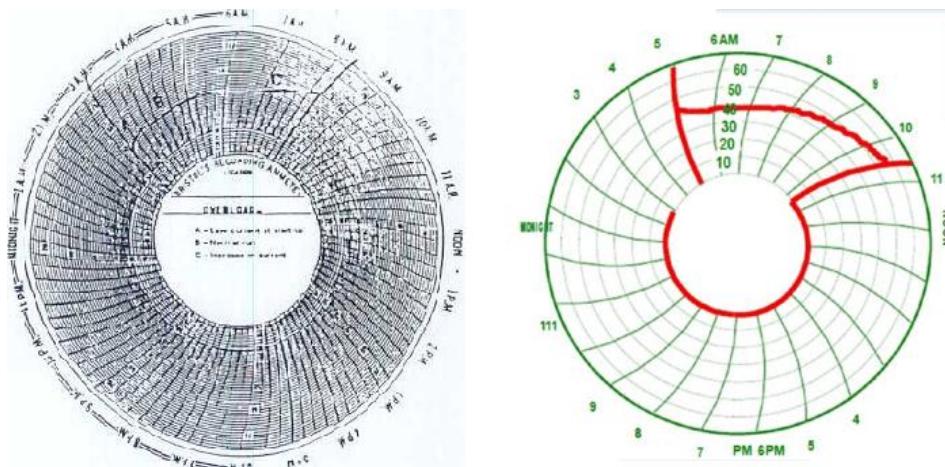


Figura 18. Carta Amperimétrica en sobrecarga. Fuente: "Electrical Submersible Pump HandBook"

Carta Amperimétrica bajo nivel de fluido

Con gas: La bomba opera a una rata mayor al aporte del pozo, se reduce el nivel del fluido y el gas comienza a separarse llegando a un bloqueo por gas. Las posibles soluciones son: recircular fluido, profundizar la bomba, rediseñar la bomba u operar el equipo a menor frecuencia

Sin gas: el nivel del fluido cae y se presenta un descenso en la corriente de operación. El pozo no alcanza a llenar el anular para cargar la bomba, y el intento de arranque es fallido. El sistema es muy grande para la aplicación y debe ser rediseñado o el pozo estimulado

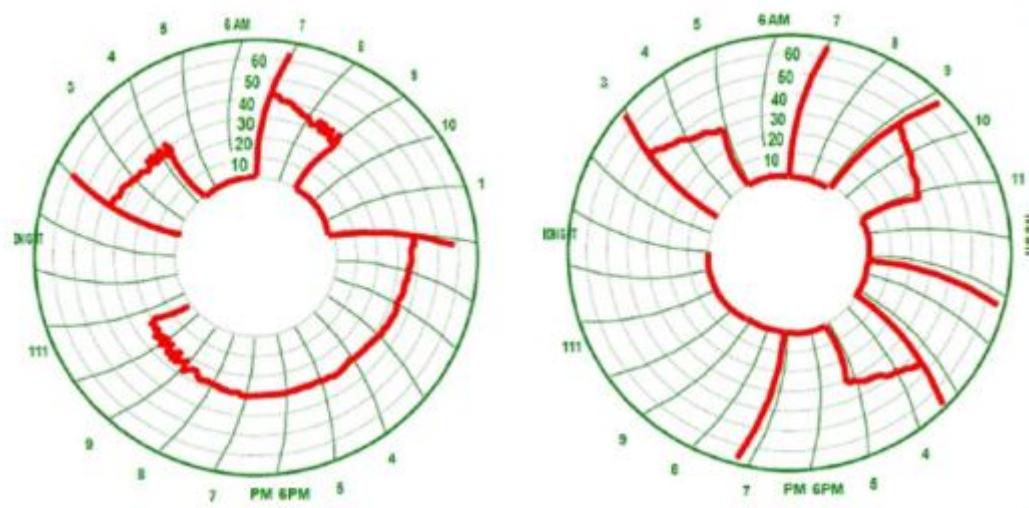


Figura 19. Carta Amperimétrica bajo nivel de fluido con gas, bajo nivel fluido sin gas.
Fuente: "ESP Oil International training group" ESP OIL Engineering Consultants.

. Carta Amperimétrica excesivos intentos de arranque

En general ocurre ante problemas de alimentación, de presentarse no debe seguir intentando arrancar el motor hasta detectar la causa de este comportamiento, se sugiere no realizar más de tres intentos de arranque al día.

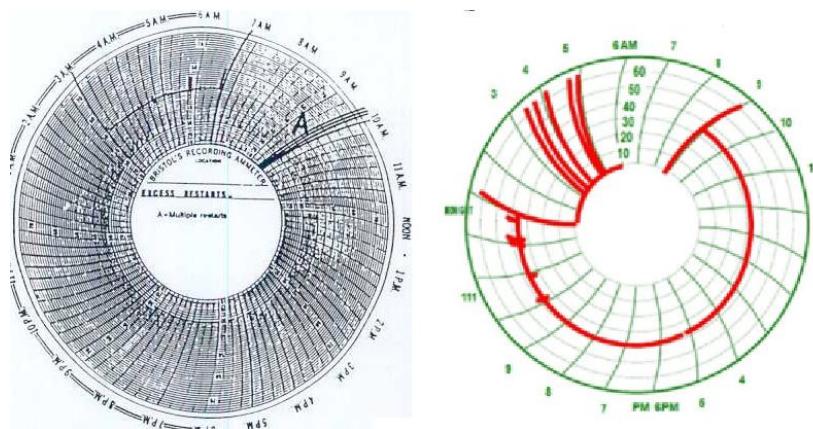


Figura 20. Carta Amperimétrica excesivos intentos de arranque. Fuente: "Electrical Submersible Pump HandBook"

Carta Amperimétrica excesivos ciclos de operación

Puede ocurrir ante la elección de un motor sobredimensionado para el sistema, por un taponamiento de la tubería, válvulas cerradas o fugas en el tubing. Se sugiere tomar el nivel del fluido después del apagado y chequear la presión de la tubería

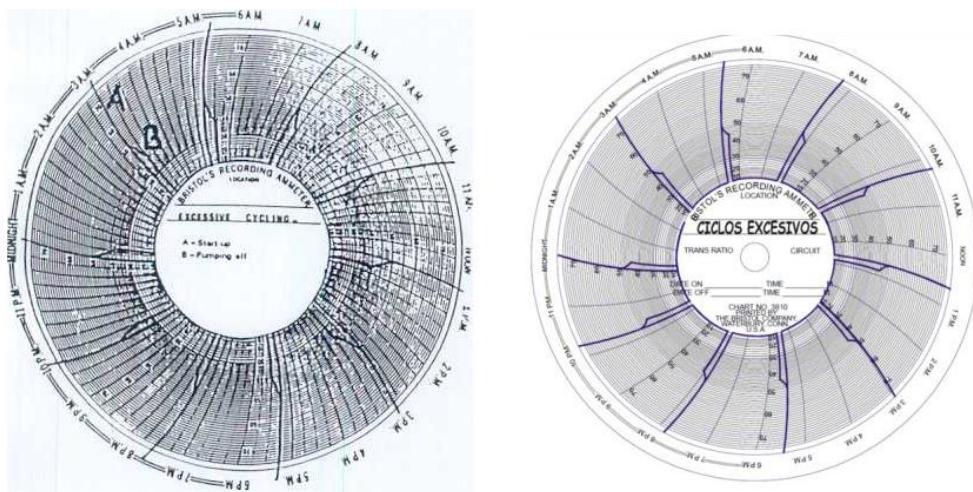


Figura 21. Carta Amperimétricas excesivos ciclos de operación. Fuente: "Electrical Submersible Pump HandBook"

Carta Amperimétricas Emulsiones o cargas en superficie

Formación de emulsiones dentro de la bomba que son desalojadas periódicamente, disminución en el voltaje de superficie debido a equipos conectados en la misma línea que se encienden u apagan cíclicamente, también puede ser resultado de una falla en el sistema de control del generador.

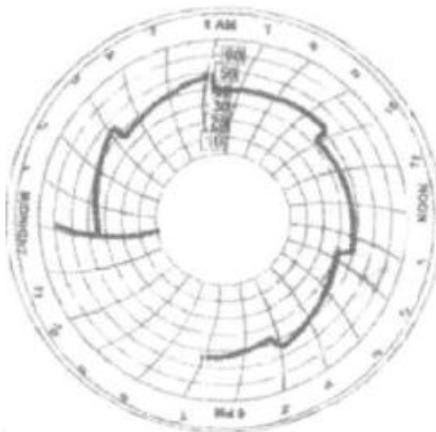


Figura 22. Carta Amperimétricas Emulsiones o cargas en superficie Fuente: “ESP Oil International training group” ESP OIL Engineering Consultants

Carta Amperimétrica sólidos en la bomba

Ocurre cuando partículas sólidas ingresan a la bomba (sedimentos, arena, lodo). El pozo debe ser limpiado siempre para remover materiales extraños antes de arrancar la bomba.

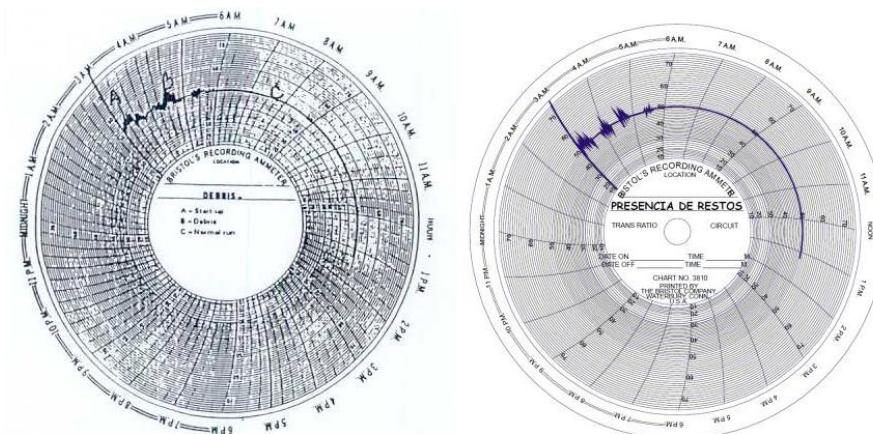


Figura 23. Carta Amperimétricas sólidos en la bomba. Fuente: “Electrical Submersible Pump HandBook

Carta Amperimétrica descarga de fluido de control

Cuando se descarga el fluido de control su peso lleva a un incremento de

corriente mientras es desplazado completamente, se recomienda ajustar el setting de sobrecarga por encima del 115% durante el tiempo de descarga, y arrancar a una frecuencia menor de la operación.

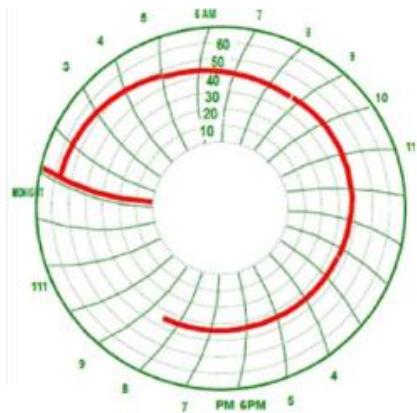


Figura 24. Carta Amperimétrica descarga de fluido de control Fuente: “ESP Oil International training group” ESP OIL Engineering Consultants.

Máquinas de Aprendizaje

Mitchell (1997) dice que el problema del aprendizaje máquina (Learning machine) se reduce a la inferencia de una función a través de un conjunto de datos conocidos (conjunto de entrenamiento), que permita estimar o predecir la salida de una serie de datos de entrada ajenos al conjunto de entrenamiento pero que comparten condiciones o características similares.

Witten, I., Frank, E. y Hall, M (2011) afirman que: “las cosas aprenden cuando cambian su comportamiento en una dirección en que les lleve a mejorar su funcionamiento y desarrollo en el futuro”. Si al comparar el comportamiento presente con el comportamiento pasado se observa una mejoría en resultados y/o procedimientos se puede afirmar que una máquina está “aprendiendo”.

Tipos de aprendizaje

Aprendizaje Supervisado

Es el tipo de aprendizaje que agrupa a aquellos algoritmos que predicen o clasifican salidas según un conjunto de entradas o características propias del problema. Para alcanzar su objetivo se apoyan en conjuntos de datos en donde se conocen valores de entrada y valores de salida. La naturaleza del conjunto de datos debe ser la misma del caso que se desea predecir o clasificar. Algunos de estos algoritmos son: algoritmos genéticos, redes neuronales, árboles de decisión, regresión, máquinas de soporte vectorial.

Aprendizaje no Supervisado

En este tipo de aprendizaje los algoritmos descubren patrones y tendencias en los datos sin tener ningún tipo de conocimiento previo acerca de cuáles son los patrones buscados, ejemplos: Clustering, Análisis de enlace, de frecuencia.

Clasificación

Consiste en examinar las características de un nuevo objeto y asignarle una clase o categoría de acuerdo a un conjunto previamente definido, por ejemplo, dividir un grupo de personas por género (masculino o femenino), altura (bajo, medio y alto), tipo de vivienda (casa, apartamento, residencia, quinta, cabaña).

Predicción

Es el proceso de inferir resultados en base a información útil conocida. Mediante el uso de un conjunto de datos y herramientas estadísticas propias del machine learning es posible estimar valores de casos que tengan modelos

matemáticos complejos, desconocidos o difíciles de implementar.

Minería de datos

La minería de datos tiene como propósito identificar conocimiento, obtenido a partir de bases de datos que aporten y dirijan hacia una toma de decisión. Es la exploración y análisis de grandes cantidades de datos con el objeto de encontrar patrones y reglas significativas (conocimiento) Aguilar, J. (2010)

Descubrimiento de Conocimiento en Bases de Datos

Del inglés Knowledge Discovery in Databases, puede ser descrito como: El Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles, a partir de los datos. Tiene las siguientes fases

1. *Integración y recopilación de datos:* en esta parte se agrupan (en caso de no estar en la misma área de trabajo) registros, archivos que contengan información, bases de datos. etc.

2. *Selección, limpieza y transformación:* se analiza la calidad de la información recopilada, se excluyen datos que no se ajusten al comportamiento de los demás, así como aquellos que no contengan información (datos perdidos). Si la cantidad de información es muy extensa se selecciona una porción de la misma de forma aleatoria, se pueden redefinir y/o construir nuevas variables que faciliten el proceso, esta parte es particular del problema.

3. *Minería de datos* acá se aplica el modelo a usar, puede ser descriptivo a través de reglas de asociación, de clustering, o agrupamiento, predictivos, de regresión o clasificación supervisada (entre ellos se encuentran las Máquinas de soporte vectorial)

4. *Evaluación e interpretación*: una vez se define el modelo a través de alguno de los algoritmos de la fase 3 viene el proceso de evaluación entre los que tenemos validación simple, validación cruzada (k-fold), validación uno afuera, bootstrap, reglas de asociación, clustering, error cuadrático medio, porcentaje de bien clasificados, matriz de confusión, análisis ROC entre otros.

5. *Difusión y uso*: consiste en distribuir el producto a los posibles usuarios, medir la evolución a lo largo del tiempo. Se recomienda reevaluar, reentrenar y reconstruir el modelo cada cierto tiempo.



Figura 25. Fases del proceso de Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery in Databases) KDD. Fuente: Introducción minería de datos, metodologías y técnicas de minería de datos Aguilar, J.

Máquinas de Soporte Vectorial SVM

La teoría de las SVMs fue desarrollada inicialmente por V. Vapnik a principios de los años 80 y se centra en lo que se conoce como Teoría del Aprendizaje Estadístico. Su razón de ser es buscar, para una tarea de aprendizaje dada, con una cantidad finita de datos, una función que permita llevar a cabo una

generalización que sea resultado de una adecuada relación entre la precisión alcanzada con un particular conjunto de entrenamiento y la capacidad del modelo

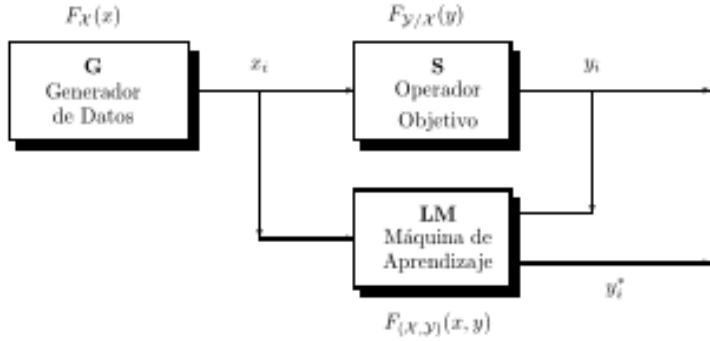


Figura 26. Esquema de una máquina de aprendizaje a partir de ejemplos Fuente: “Modelos de Clasificación basados en Máquinas de vectores soporte” González, L

Para llevar a cabo su objetivo requiere un conjunto de:

Vectores de entrada (inputs) $\{x_i, i = 1, \dots, n\} \subseteq X \subseteq R^d$

Valores de salida (outputs) $\{y_i, i = 1, \dots, n\} \subseteq Y \subseteq R$

Donde los vectores x_i son independientes e idénticamente distribuidos a través de alguna función de distribución desconocida, pero que no debe variar a lo largo del proceso de aprendizaje. Cada uno de estos vectores es la entrada a un operador objetivo que llamaremos S el cual arroja una salida según una función de distribución condicional $F_{Y=x_i}(y)$, por otra parte la máquina de aprendizaje (LM) que en este caso es la SVM recibe el siguiente conjunto de entrenamiento.

$$Z = \{(x_1, y_1), \dots, (x_n, y_n)\} \subseteq X \times Y = Z$$

A partir de él construye una aproximación al operador desconocido. Ya que la función de distribución condicional puede resultar en un problema complicado y al no ser estrictamente necesario llegar a ella el problema se aborda desde una función de costo. Al minimizar dicha función de costo se hallan

parámetros o pesos que definen a la máquina. Una función de costo sirve para dos cosas:

- Calcular los parámetros de la máquina mediante un proceso de minimización
- Calcular qué tanto se ajusta el valor real al valor estimado

Otra cuestión a considerar en el proceso de aprendizaje es el sobreajuste. El sobreajuste es una característica propia del entrenamiento y ocurre cuando la máquina se amolda tan bien al conjunto de datos que, ante valores ajenos al mismo la estimación difiere considerablemente. Para ello se usan términos de regularización. Al agrupar la función de costo con el término de regularización se persiguen dos objetivos: en primer lugar, minimizar la función de costo a un valor bajo, cercano a cero y por otro reducir en lo posible el aporte de cada característica al entrenamiento. El sobreajuste suele ocurrir cuando una característica o grupo de características inciden en demasía sobre el modelo, como es difícil saber cuáles son (un problema puede manejar datos con una cantidad enorme de características de entrada) se opta por reducir el aporte de todas ellas.

$$J(\theta) = \frac{1}{m} \sum_{i=0}^m \text{costo}((h_\theta(x)^{(i)}, y^{(i)}) + \text{término de regularización}$$

Esta es la ecuación general. La función de costo depende del valor estimado por la hipótesis $h_\theta(x)$ a cada ejemplo de entrenamiento y de su valor real conocido, se promedia cada uno de ellos. Cada caso aporta un costo a la función, idealmente el aporte de cada ejemplo debería ser cero, esto equivale a decir que la salida estimada o hipótesis coincide perfectamente con la salida real conocida. El término de regularización busca mantener los valores de los parámetros θ lo más bajo posible. En la Máquina de soporte vectorial la función de costo es:

$$J(\theta) = C * \left[\sum_{i=0}^m y^{(i)} * cost_1(\theta^T * x^{(i)}) + (1 - y^{(i)}) * cost_0(\theta^T * x^{(i)}) \right] + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

Donde C es el parámetro de ajuste, que se encarga de evitar el sobreajuste y juega con características muy importantes: la varianza y la oscilación.

La varianza indica que tan diferente se considera un valor de otro cuando dicho valor cambia, es decir, si se tiene un valor de $x=1$ y el mismo se desplaza un valor de paso $h=0.01$ tendremos un valor de $x_1=1.01$. Si la varianza es baja puede decirse que x es muy similar a x_1 , incluso se puede decir que $x=x_1$, pero si la varianza es alta se dice que x es muy diferente a x_1 . La oscilación se refiere a que tanta seguridad se tiene que un valor catalogado como cero sea cero, y un valor catalogado como uno sea uno. Mientras menor sea la oscilación mayor certeza se tiene que el valor esté adecuadamente catalogado.

La idea es buscar un equilibrio pues cuando la varianza es alta la oscilación suele ser baja y viceversa, si la varianza es baja la oscilación suele ser alta. Aunque se pueda pensar que la primera condición es idónea no es así, porque en este caso es que precisamente ocurre el sobreajuste.

La función de costo planteada responde a problemas de clasificación binarios. Por ello existen dos funciones de costos: cost1 que aplica cuando la salida real es uno, en este caso si el valor estimado es uno o muy cercano a uno el aporte del ejemplo a la función de costo es muy bajo, caso contrario, si el valor real es uno y el valor estimado es cero o muy cercano a cero se penaliza con un costo muy alto. Cost0 aplica cuando la salida real es 0 y tiene el mismo planteamiento que el anterior con las obvias diferencias. El término de regularización trabaja con la distancia vectorial entre los parámetros, la misma debe ser lo más cercana a cero para que el aporte a la función de costo sea mínimo. Esto se consigue con valores de parámetros lo suficientemente pequeños.

La gran ventaja que presentan las máquinas de soporte vectorial con respecto a otras máquinas de aprendizaje es la frontera de decisión, dicha frontera busca un margen amplio entre los valores positivos y negativos que están más cercanos entre sí.

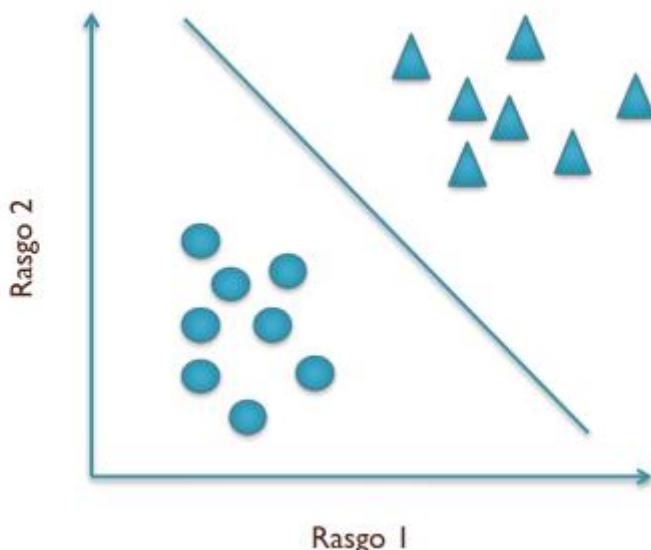


Figura 27. Barrera de decisión de un problema de clasificación binario linealmente separable. Fuente: “Aprendizaje de Máquina y reconocimiento de patrones” Ruiz, S.

Cuando se quiere realizar una clasificación múltiple se usa la clasificación por SVM binaria aplicada a cada clase. Ejemplo: si se quieren clasificar un conjunto de patrones en 4 posibles clases C1, C2, C3 y C4 primero se toma a C1 como clase positiva y al resto como clase negativa, se obtiene un resultado o probabilidad de que este patrón pertenezca a esta clase. Luego se toma C2 como clase positiva y el resto como clase negativa y se obtiene la probabilidad de que el patrón pertenezca a esta clase, esto se repite con cada clase y al final se comparan las probabilidades. En aquella en que el patrón tenga mayor probabilidad es donde se clasifica, este método se llama uno vs todos.

Solución de problemas no lineales con SVM, kernels

Muchas veces los datos no son linealmente separables, es decir, una frontera de decisión lineal no permite dividir ejemplos positivos de negativos. En este caso se transforman las características de los patrones con la idea de llevarlos a una nueva dimensionalidad donde, a través de hiperplanos sea posible realizar la clasificación, esto se conoce como transformación no lineal. Existen muchas maneras de llevarlo a cabo, pero las Máquinas de soporte vectorial usan funciones kernel o núcleos.

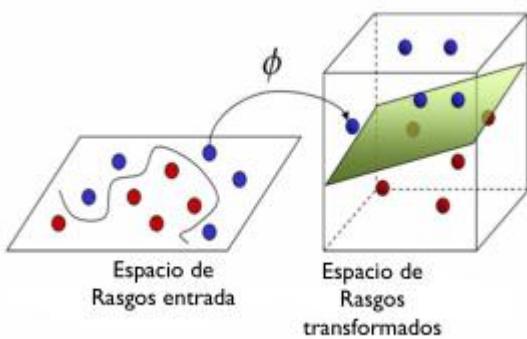


Figura 28. Transformación no lineal, Fuente: “Aprendizaje de Máquina y reconocimiento de patrones” Ruiz, S.

Un kernel es una función que genera nuevas características dependiendo de que tanto se parece un ejemplo de entrenamiento a los demás, también se les llama funciones de similitud. Existe gran variedad de kernels ya que cada uno de ellos tiene sus criterios para evaluar semejanza, existen kernels lineales, kernels polinómicos, kernels gaussianos entre otros. Se trabajará con el kernel Gaussiano.

El Kernel gaussiano está descrito por la siguiente ecuación:

$$f_i = \exp\left(-\frac{\|x - l^i\|^2}{2\sigma^2}\right)$$

Siendo f_i un vector $[f_1 \ f_2 \ f_3 \dots \ f_m]$ el cual será el nuevo conjunto de

características. El problema tendrá $R^{m \times m}$ características nuevas donde m es el número de ejemplos de entrenamiento. Él compara cada ejemplo de entrenamiento con una marca. Las marcas a su vez son los mismos ejemplos, de allí que la matriz de características sea cuadrada. Para el criterio de similitud se usa la distancia entre el patrón y la marca, la varianza en el denominador sirve para el ajuste.

En el caso en que el patrón y la marca sean similares, la función exponencial, al estar elevada a un número pequeño negativo que tiende a cero tratará de acercarse a un valor de uno. Si la distancia entre el patrón y la marca es grande el kernel tratará de acercarse a un valor cero. En conclusión, el kernel dará mayor peso a las características en donde exista alta similitud. La frontera de decisión se crea en base a estas comparaciones y de allí es que es posible que esta pueda ser altamente no lineal.

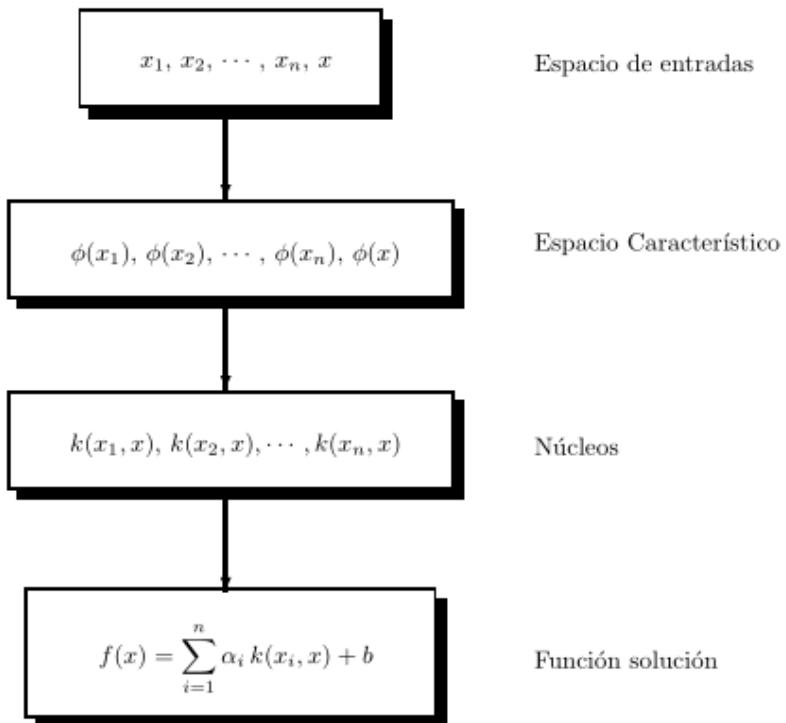


Figura 29. Esquema de una máquina de soporte vectorial Fuente: “Modelos de Clasificación basados en Máquinas de vectores soporte” González, L.

Tipos de Evaluación

Validación simple

Se evalúa que tan bien se comporta el modelo una vez entrenado, para ello se divide el conjunto de datos en dos grupos: un conjunto de entrenamiento y un conjunto de prueba. Una relación típica es tomar un 70% para el entrenamiento y el resto para la evaluación; se compara el valor estimado con el valor real y en base a ello se estima un porcentaje de acierto.

$$\text{porcentaje de acierto} = \frac{\text{numero de aciertos}}{\text{numero de datos evaluados}} * 100$$

$$\text{porcentaje de error} = \frac{\text{numero de desaciertos}}{\text{numero de datos evaluados}} * 100$$

Aunque es un método de validación sencillo y rápido de aplicar a nivel computacional no es muy confiable ya que depende de la manera en que se segmenten los datos para el entrenamiento y validación

Validación cruzada

Suponga que la cantidad de data disponible para el entrenamiento y la evaluación de la máquina es limitada, si se quiere aplicar una evaluación simple se debe separar la data en dos partes. Ahora bien, ¿qué ocurre si el conjunto de valores tomados para el entrenamiento resulta ser no representativo? Puede ocurrir que los datos usados para entrenar omitan ejemplos de cierta clase, entonces dicha clase no estará bien representada y otra clase puede quedar sobre-representada (las clases estarán desproporcionadas). Para evitarlo se desordena el conjunto de datos

antes de dividirlo, esto se conoce como estratificación.

La estratificación reduce la probabilidad que ocurra la situación expuesta pero no es una garantía. Por ello se presenta una nueva manera de evaluar: la validación cruzada. La validación cruzada divide la data en n partes, siendo n_1 =primera partición, n_2 = segunda partición hasta n_n = enésima partición. El set de entrenamiento estará formado por $n-1$ particiones y el set de prueba tendrá una partición. Se realizarán tantos entrenamientos como particiones se tengan. Luego, en cada entrenamiento se toman set de entrenamientos diferentes y el sobrante se usa para validar, se calcula el porcentaje de aciertos y al final se promedian para obtener el porcentaje de la validación cruzada.

Validación uno afuera

Es similar a la validación cruzada, se entrena tantas veces como ejemplos tenga la data, si la data tiene m ejemplos de entrenamiento se usarán $m-1$ ejemplos para entrenar y 1 ejemplo para validar. En cada entrenamiento se cambiará el ejemplo de validación. Una vez calculado el porcentaje de aciertos o de error se promedian para obtener el porcentaje de la validación uno afuera.

Bootstrap:

El método está basado en el procedimiento estadístico de muestreo con reemplazo, en donde se toman aleatoriamente ejemplos del set de datos para conformar el conjunto de entrenamiento, con la particularidad que se puede tomar más de una vez el mismo ejemplo. Witten, I. y Frank, E. (2011) describen una variante llamada 0.632 bootstrap, en ella, un conjunto de datos con n ejemplos o instancias es muestreado con reemplazo para crear un nuevo conjunto de datos con las mismas dimensiones. Debido a que algunos elementos en el segundo conjunto estarán repetidos muchas de las instancias del conjunto original no son

tomadas para el entrenamiento, estas instancias serán usadas para formar un conjunto de prueba.

Al evaluar con el modelo y conjunto de prueba anteriormente descritos se tendrá una estimación pesimista. Se dice que es pesimista ya que el conjunto de entrenamiento tiene alrededor de un 63% de las instancias originales en comparación, por ejemplo, a una validación cruzada con diez particiones, en la que se considera un 90% de ellas. Para compensar se propone combinar dicha estimación con el error de resustitución. El error con restitución se calcula considerando todas las instancias (incluso aquellas que se usaron en el entrenamiento) por ello se dice que es un error optimista. El error total viene determinado por:

$$\text{error} = 0.632 * \text{error}_{\text{instancias prueba}} + 0.368 * \text{error}_{\text{todas instancias}}$$

Se recomienda hacer esta validación varias veces y al final sacar un promedio del error obtenido. Este tipo de validación es muy usado cuando se cuenta con datas pequeñas.

Curvas ROC (Receiver Operating Characteristics)

Estas curvas evalúan el rendimiento del clasificador sin necesidad de tener la distribución de las clases o el error de costos, imprime la rata de aciertos positivos en el eje vertical contra la rata de aciertos negativos en el eje horizontal

TP Rate=rata de aciertos positivos; FP Rate= rata de desaciertos positivos

TP= número de aciertos positivos; FP=número de desaciertos positivos

TN=número de aciertos negativos; FN=número de desaciertos negativos.

$$TP \text{ Rate} = 100 * \frac{TP}{TP+FN} \quad FP \text{ Rate} = 100 * \frac{FP}{FP+TN}$$

Cuando la data está desbalanceada (existen mucho más ejemplo de una clase que de otra) el uso del porcentaje de acierto/error no es una métrica adecuada para evaluar al desempeño del sistema. Suponga el siguiente ejemplo:

Se tiene una máquina de clasificación binaria para identificar si un producto es o no defectuoso, después de entrenar se obtiene un modelo que cataloga todos los ejemplos como no defectuosos, un modelo evidentemente inútil. ¿Qué ocurre si se evalúa el desempeño en un set de prueba con 99 ejemplos no defectuosos y solo un ejemplo defectuoso mediante el porcentaje de acierto/error?

$$\text{porcentaje de acierto} = \frac{99}{100} * 100 = 99\%$$

$$\text{porcentaje de error} = \frac{1}{100} * 100 = 1\%$$

En este caso la métrica me dice que estamos ante un sistema con alto desempeño, aunque la realidad es otra, allí es donde entra en juego la rata de aciertos positivos vs la rata de desaciertos positivos. Al hacer una comparación de estos valores en función del número de ejemplos se obtienen las gráficas ROC.

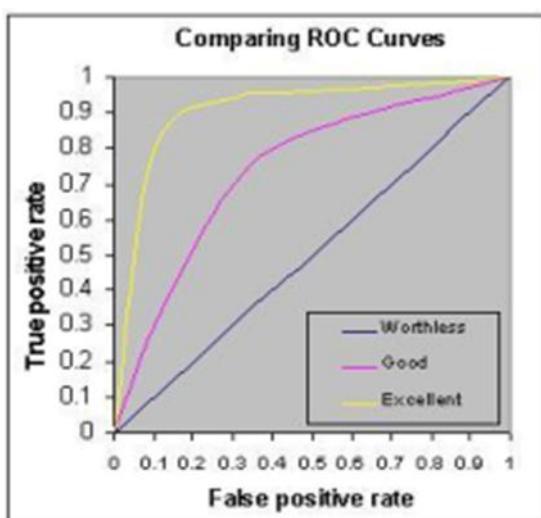


Figura 30.Curvas ROC: “Aprendizaje de Maquina y reconocimiento de patrones” Ruiz, S

Capítulo III

Marco metodológico

Tipo y nivel de investigación

Sabino (2002) define los tipos de investigación según objetivos extrínsecos o externos, es decir, según la utilidad que, fuera del ámbito estrictamente científico, van a tener las conclusiones que saquemos en: puras y aplicadas. Al referirse a una investigación aplicada afirma que:

La investigación aplicada persigue fines directos e inmediatos... Para que ello se produzca es preciso que existan tanto la voluntad como los recursos que pueden llevar las conclusiones teóricas al plano de la vida real.

Como la finalidad del Sistema computacional para simulación y reconocimiento de patrones es ser una herramienta práctica, y útil, que sirva para el estudio, análisis e identificación de fallas en el método de BES. Se dice que cae en el tipo de investigación aplicada

Padrón (2008) describe un nivel de investigación aplicativa como:

Aquella cuyo objetivo central está en proveer tecnologías o esquemas de acción derivados de los conocimientos teóricos construidos dentro de la secuencia de la Línea.

Estas investigaciones carecen, propiamente hablando, de preguntas. Más bien tienden a establecer una relación productiva, ingeniosa y creativa, entre las posibilidades de un modelo teórico, por un lado, y las dificultades o necesidades que se confrontan en el terreno de la práctica por otro lado.

Según el enfoque que presenta el sistema de simulación y reconocimiento

de patrones de falla, es necesario un estudio profundo de temas teóricos relacionados con el sistema de levantamiento artificial de crudo por bombeo electrosumergible, cartas amperimétricas y su relación con anomalías. También es pertinente el estudio herramientas de máquinas de aprendizaje, máquinas de soporte vectorial, métodos de evaluación, validación y prueba, diseños, procedimientos y sintaxis de programación en Python. Con esto se busca desarrollar un software de uso libre que sea una herramienta de apoyo en el análisis e identificación de fallas. La investigación recae en la categoría Aplicativa.

Diseño de la investigación

Palella y Martins (2012), dicen que en el diseño no experimental se observan los hechos tal y como se presentan en su contexto real y en un tiempo determinado o no, para luego analizarlos. Por lo tanto, en este diseño no se construye una situación específica si no que se observan las que existen. El diseño bibliográfico por su parte se fundamenta en la revisión sistemática, rigurosa y profunda de material documental de cualquier clase. Se procura el análisis de los fenómenos o el establecimiento de la relación entre dos o más variables. Cuando opta por este tipo de estudio, el investigador utiliza documentos; los recolecta, selecciona, analiza y presenta resultados coherentes.

El diseño bibliográfico utiliza los procedimientos lógicos y mentales propios de toda investigación: análisis, síntesis, deducción, inducción, entre otros. La investigación sigue un diseño no experimental debido a que se trabaja sobre premisas probadas donde relacionan cada tipo de falla con la forma y el comportamiento que sigue la corriente eléctrica de motor. La investigación se desarrolla siguiendo procesos propios del diseño bibliográfico.

Fases de desarrollo del proyecto

Fase I Búsqueda de información

En esta etapa se investigó, estudió y comprendió el método de levantamiento artificial de crudo por bombeo electrosumergible, sus elementos, sus características, métodos, ventajas, desventajas y áreas de aplicación. Se hizo énfasis en el análisis de cartas ampermétricas y su relación con la identificación de fallas. Esta tarea se llevó a cabo a través de manuales, publicaciones, textos, investigaciones y entrevistas no estructuradas con el tutor industrial.

Se recolecta, clasifica y delimita las alternativas que ofrecen las máquinas de aprendizaje para el problema de reconocimiento de patrones. Puntualmente las máquinas de soporte vectorial aplicadas a problemas de clasificación.

Fase II: Diseñar y elaborar el sistema computacional de simulación y reconocimiento de patrones de falla en procesos de levantamiento artificial BES

Se desarrolló el código fuente en leguaje Python, definiendo los métodos a seguir: primero dividiendo las funciones necesarias por el programa en 6 scripts:

- Principal: en donde son llamadas las funciones
- Menús: contiene las funciones encargadas de mostrar de menús y submenús, recibir valores de selección suministrados por el usuario
- Manage_files: contiene funciones encargadas de cargar datos, crear archivos, mostrar información
- Generate_Ammeter_Charts: contiene las funciones que definen cada uno de los casos de simulación de fallas, identificadas de manera correcta
- Training_Set: tiene los métodos relacionados con el entrenamiento y prueba del sistema
- Load_and_Form_data: brinda métodos para cargar valores, reordenar

matrices, etc.

Una vez se plantea esta estructura se establecen las metas: generar fallas aleatorias controladas, visualizar cartas, predecir fallas a través de una carta amperimétrica usando máquinas de aprendizaje.

.

Fase III: Realizar pruebas y análisis

Se realizaron pruebas para determinar errores de ejecución en el programa y se le dieron su respectiva corrección, además se presentaron distintos escenarios donde, a través de procesos de validación simple, cruzada, uno afuera, bootstrap, curvas ROC y de acierto se evaluó el comportamiento y desempeño del simulador y clasificador de patrones de falla

.

Fase IV Documentar

Se presentan conclusiones, resultados, manuales de usuario, de administrador y documentación del sistema

.

Técnicas e instrumentos de recolección de datos

Arias (1999), define a las técnicas de recolección de datos como "...las distintas formas o maneras de obtener la información., además plantea que el análisis documental:

Consiste en registrar los datos que se van obteniendo en la revisión bibliográfica, en fin, en las diferentes etapas y procesos que se van desarrollando. Entre los beneficios de esta técnica es necesario señalar que constituye un factor de claridad, pues permite recoger con autonomía los diferentes aspectos que se desea estudiar, posibilita la estructuración

ordenada y lógica de las ideas, permite cotejar fácilmente las citas de las referencias consultadas, ahorra tiempo.

Y agrega que “...Los instrumentos son los medios materiales que se emplean para recoger y almacenar la información”

Al ser una investigación del tipo documental la recolección de datos se hizo en base a fuentes bibliográficas, informes, manuales.

PLAN DE TRABAJO DE LA PASANTIA PROFESIONAL

Pág. 1/2

Empresa: PDVSA-DST

Departamento: Gerencia AIT

Dirección / Teléfono/correo: Edificio Sede PDVSA, DST Mérida, 0261-8064787 / 0426-5808582
camargoea@pdvsa.com**Tutor Externo:** Edgar Alexander Camargo Contreras.

Líder de Control de procesos

Pasante: Angel Leonardo Duarte Montes

C. I: V-19.134.002

Teléfono: 0416-0484772

Correo Electrónico: angel.duarte@unet.edu.ve

Departamento: Ingeniería Electrónica

Período de Pasantías: (20-03-2017)/(07-07-2017)

TITULO

Sistema de Simulacion y reconocimiento de Patrones de Fallas en el Método de Extracción de Crudo por Bombeo ElectroSumergible (BES) con maquinas de aprendizaje

DESCRIPCION DEL PROYECTO

En la industria petrolera la extracción de crudo puede darse de dos maneras: de forma natural cuando las condiciones de presión en el pozo permiten que el hidrocarburo salga a la superficie, y de forma artificial cuando se requieren sistemas elaborados de levantamiento siendo los mas usados: levantamiento artificial por gas (LAG), bombeo mecánico (BMC), por cabilla de succión, bombeo ElectroSumergible (BES), bombeo de cavidad progresiva (BCP) y bombeo hidráulico (BHJ).

El sistema BES es un método de levantamiento ampliamente utilizado que está constituido por equipos complejos y de alto costo, debido a esto un mal funcionamiento o falla inesperada puede resultar catastrófica en términos de producción, seguridad del personal y precios; es sumamente deseable contar con alguna herramienta que permita predecir e identificar fallas. El tener un sistema que reconozca fallas puede resultar útil a la hora de realizar correcciones, planificar rutinas de mantenimiento, identificar elementos propensos a mal funcionamiento entre otros.

Teniendo esto en cuenta se propone un sistema computacional para reconocimiento de patrones de falla aplicado al método de levantamiento artificial por Bombeo ElectroSumergible (BES) usando Maquinas de Aprendizaje; Las máquinas de aprendizaje usadas en el reconocimiento de patrones presentan buen desempeño y entre unas de sus aplicaciones están: diagnóstico de procesos, historial de mantenimiento de equipos, predicción de comportamientos y otras en las cuales se manejen datas estructuradas y suficientemente grandes [Mitchell (1997)]. Así, en la industria petrolera existen situaciones en las cuales el uso de máquinas de aprendizaje ha dado buenos resultados, tal es el caso de la separación del crudo del gas natural; este proceso es un requisito previo esencial para la etapa de refinación y el control del mismo es un trabajo difícil. Por tal motivo la British Petroleum ha usado maquinas de aprendizaje para crear reglas y establecer parámetros logrando realizar en 10 minutos lo que expertos humanos suelen hacer en de un día (Ian H. Witten, Frank Eibe, Mark A. Hall, 2011)

En tal sentido, las máquinas de aprendizaje apoyadas en la rápida evolución de la computación que le ha permitido manejar con mayor celeridad y robustez grandes volúmenes de información y procesos de cálculo, así como su comprobada efectividad para realizar predicciones de manera precisa y confiable se presenta como una alternativa idónea para desarrollar un sistema computacional con óptimo desempeño.



PLAN DE TRABAJO DE LA PASANTIA PROFESIONAL

Pág. 2/2

OBJETIVO GENERAL

Realizar un Sistema computacional para simulacion y reconocimiento de Patrones de Fallas en el proceso de Levantamiento artificial de crudo por Bombeo ElectroSumergible (BES) usando herramientas de máquinas de aprendizaje

OBJETIVOS ESPECIFICOS

1. Conocer en qué consiste el Proceso de Levantamiento de Crudo por Bombeo ElectroSumergible.
2. Conocer Técnicas de Maquinas de Aprendizaje.
3. Desarrollar el sistema computacional para simulacion y reconocimiento de Patrones de Fallas en un sistema BES.
4. Realizar pruebas del Sistema computacional para reconocimiento de fallas
5. Desarrollar escenarios de prueba para la evaluación y análisis estadístico del sistema
6. Elaborar la documentación relacionada con la instalación y uso del sistema

ANTECEDENTES

R Leal E. Camargo, J Aguilar, A. Ríos, "Modelo de Diagnosticabilidad para Procesos de Producción Petrolero" 2012.

M. Guillén, J. Paredes, O. Camacho. An approach for fault detection and diagnosis in the instrumentation process using patterns recognition in the wavelet domain 2010

R. Leal, E. Camargo, J. Aguilar, A. Ríos, Travé-MassuyèsL. "Análisis de Diagnosticabilidad y Localización de Sensores en un Pozo de Extracción de Petróleo por Inyección de Gas" 2015

METODOLOGÍA

ETAPA 1: Investigar, estudiar y comprender el metodo de levantamiento artificial de crudo por Bombeo ElectroSumergible (BES)

ETAPA 2: recolectar, aprender, clasificar y delimitar las alternativas que proponen las maquinas de aprendizaje para resolver problemas de reconocimiento de patrones; aplicar un algoritmo de aprendizaje que, teóricamente y basado en experiencias previas se ajuste a las condiciones descritas

ETAPA 3: Diseñar y elaborar el sistema computacional de simulacion y reconocimiento de patrones de fallas en procesos de levantamiento artificial BES; engloba: desarrollo de código fuente en lenguaje Python 3.4 , definir e implementar métodos de visualización de resultados (gráficas, tablas u otros),

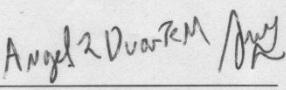
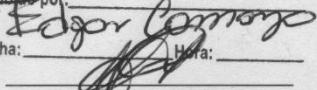
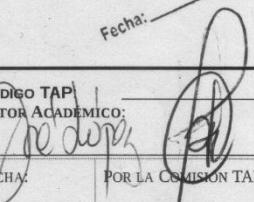
ETAPA 4: Realizar pruebas; en esta etapa se definen los criterios para evaluar el desempeño del sistema computacional (condiciones que me indiquen si funciona como se espera), corregir si es necesario

ETAPA 5: Realizar análisis: presentar distintos escenarios y condiciones, realizar análisis estadístico y de efectividad

ETAPA 6: Documentar: realizar manuales de usuario, de administrador y documentación del sistema

PDVSA
Distrito Socialista
Tecnológico
AIT Mérida

Recibido por:
Fecha:
Hora:

Preparado por: •  [NOMBRE PASANTE /FIRMA]	Aprobado por: Recibido por: Fecha:  Hora: [Nombre Tutor/Firma/Sello Empresa]	CÓDIGO TAP TUTOR ACADÉMICO: FECHA: _____ POR LA COMISIÓN TAP: 
--	---	--



Instituto Socialista
Tecnológico
AIT Mérida

Ali Mérida

Fecha:

Horas:

Programa de Pasantías

Gerencia: Distrito Socialista Tecnológico AIT PDVSA Mérida

Nombre del Pasante: Angel Duarte

Nombre del Tutor: Edgar

Cronograma de Actividades

Nº	ACTIVIDADES	SEMANAS														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Curso de inducción a la empresa	x														
2	Conocer en qué consiste el Sistema de Levantamiento de Crudo por Bombeo ElectroSumergible		x	x	x	x										
3	Conocer Técnicas de Maquinas de Aprendizaje			x	x	x	x									
4	Realizar un Sistema de Simulación y Reconocimiento de Patrones de Fallas del Sistema ElectroSumergible		x	x	x	x	x	x	x	x						
5	Realizar pruebas del Sistema Desarrollador para Pozos por Electrosumergible (validación y análisis)			x	x	x	x	x	x	x						
6	Elaborar la documentación relacionada con la instalación y uso del sistema					x	x	x	x	x						

Tutor Académico

Tutor Industrial

showing rope

200

Avenida Libertador, edición Perdices de Venezuela, Tome Este, La Campaña. Apartado postal 163, Caracas 11. Teléfonos de VENEZUELA, SA.

1960-61
1961-62
1962-63
1963-64
1964-65
1965-66
1966-67
1967-68
1968-69
1969-70
1970-71
1971-72
1972-73
1973-74
1974-75
1975-76
1976-77
1977-78
1978-79
1979-80
1980-81
1981-82
1982-83
1983-84
1984-85
1985-86
1986-87
1987-88
1988-89
1989-90
1990-91
1991-92
1992-93
1993-94
1994-95
1995-96
1996-97
1997-98
1998-99
1999-2000
2000-01
2001-02
2002-03
2003-04
2004-05
2005-06
2006-07
2007-08
2008-09
2009-10
2010-11
2011-12
2012-13
2013-14
2014-15
2015-16
2016-17
2017-18
2018-19
2019-20
2020-21
2021-22

Capítulo IV

Desarrollo y Resultados del plan de trabajo

Para llevar a cabo cada uno de los objetivos presentados en el plan de trabajo se abordó el problema en 6 etapas:

Etapa I: Investigar, estudiar y comprender el método de levantamiento artificial por bombeo electrosumergible (BES)

Se inició con el estudio del proceso de extracción de crudo, y de las fases por las cuales pasa el fluido desde que se extrae del yacimiento hasta que llega a la superficie. Una vez hecho esto se plantean las características que determinan si un pozo es capaz de producir por flujo natural o si necesita de un método de levantamiento artificial. Luego se realiza la descripción de los distintos tipos de levantamiento. Finalmente se profundiza en el método de bombeo electrosumergible.

En el bombeo electrosumergible se hizo una breve reseña sobre su historia, implementación y uso en el territorio nacional, las ventajas y desventajas que presenta en relación a otros métodos y a las características del pozo. Posteriormente se procedió a identificar cada uno de los equipos que conforman el sistema (equipos de superficie y equipos de subsuelo) así como el rol que desempeñan. Se finaliza con un análisis básico de la física involucrada en el proceso de levantamiento BES.

A nivel de detección de fallas se investigó la metodología en que se identifican, diagnostican y clasifican las mismas. Esto con la finalidad de elegir alguna característica o procedimiento que permitiera el uso de máquinas de aprendizaje y reconocimiento de patrones. Luego de este análisis se escoge el estudio de las cartas amperimétricas, ya que describen de manera clara el

comportamiento de la corriente del motor electrosumergible a través de valores sensados exentos de análisis cualitativos (los valores de corriente no dependen de la interpretación). Esto presenta a la carta amperimétricas como un patrón por excelencia.

Una vez se encontraron patrones se establecen relaciones entre dichos patrones de corriente y distintos tipos de falla que les permita ser clasificadas. La clasificación depende de la interpretación de expertos combinada con comportamientos esperados. Se entiende como comportamiento esperado tales condiciones: si a un motor eléctrico se le incrementa su carga física aumenta el torque y la corriente aumenta, si se le disminuye su carga física la corriente disminuye, si pasa los umbrales de corriente el motor se quema o apaga, si la carga física varía la corriente fluctúa, etc.

La etapa 1 fue realizada mediante búsqueda bibliográfica en artículos científicos, manuales de operación, trabajos de pre-grado, post-grado, maestrías, sitios web relacionados con la industria petrolera. Se complementó con asesorías y entrevistas no estructuradas con el tutor industrial, el mismo se encargó de suministrar documentos e investigaciones precedentes realizadas en el DST PDVSA Mérida, así como de brindar su experiencia y conocimiento en el área.

Etapa II: Recoleistar, aprender, clasificar y delimitar las alternativas que proponen las máquinas de aprendizaje para resolver problemas de reconocimiento de patrones, aplicar un algoritmo de aprendizaje que, teóricamente y basado en experiencias previas se ajuste a las condiciones descritas

Se recurre a la investigación bibliográfica, textos especializados en minería de datos, aprendizaje máquinas y análisis estadísticos para el estudio y correcta comprensión de que es una máquina de aprendizaje. De estos textos se

tocaron puntos como: tipos de aprendizaje supervisado, no supervisado, minería de datos, tratamiento de datos, fases a seguir para desarrollar un proyecto. También se comprenden conceptos de clasificación, patrón, conjunto de datos, conjunto de entrenamiento, validación, prueba, acierto, modelos, sobreajuste, ajuste, subajuste entre otras. Estas definiciones han de ser manejadas con soltura a la hora de abordar un problema como el planteado.

Una vez se tiene claro el panorama se estudian y comparan algunos de los modelos que presentan las máquinas de aprendizaje delimitando el problema a:

- Es un problema de aprendizaje supervisado porque se parte de patrones de datos con entradas y salidas conocidas para inferir nuevos patrones de los cuales solo se conocen sus características de entrada.
- Es un problema de clasificación porque sus salidas son valores discretos y bien delimitados, además es de clasificación múltiple porque la salida puede tomar valores que van desde 1 hasta 13
- Debido a que se estudian fallas (comportamientos no esperados) la estructura de los datos de entrada no sigue un comportamiento lineal, tal como se evidencia al revisar la forma de las cartas. Al hacer un análisis se nota que en algunos casos la corriente toma valores muy diferentes en intervalos de tiempo muy cortos, casi instantáneos. También presenta incrementos y decrementos en formas que no se adaptan a un modelo lineal o quasi-lineal.
- En los patrones se encuentran características similares que les permite ser clasificados. Esto significa que un patrón puede ser asociado a una clase; por lo tanto el problema se aborda desde la metodología de reconocimiento de patrones.

- Al ser un problema de aprendizaje supervisado, de clasificación múltiple, no lineal y de reconocimiento de patrones se decide aplicar máquinas de soporte vectorial con kernels gaussianos. Tal elección se debe a sus características y precedentes en la resolución de problemas similares tanto en el reconocimiento de patrones, en problemas de clasificación y en el campo de datos no lineales

En esta etapa el estudio bibliográfico teórico, la investigación de casos similares, los métodos a seguir y las recomendaciones del tutor académico resultaron ser un aporte valioso

Etapa III: Diseñar y elaborar el sistema computacional de simulación y reconocimiento de patrones de fallas en procesos de levantamiento artificial BES

Se desarrolla un esquema general para el programa:

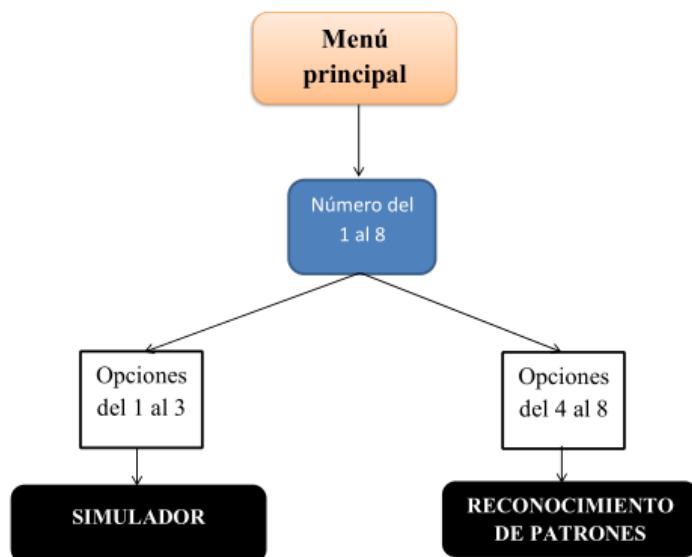


Figura 31. Diagrama del sistema de simulación y reconocimiento de patrones de falla en el sistema BES. Elaboración Propia.

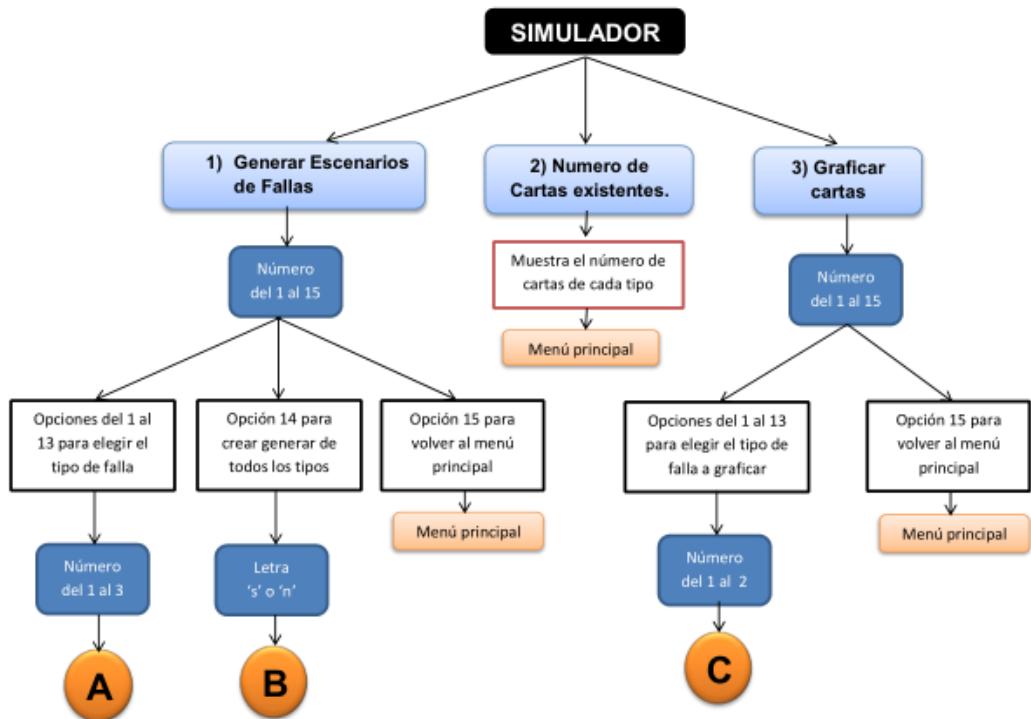


Figura 32. Diagrama del sistema de simulación. Elaboración Propia.

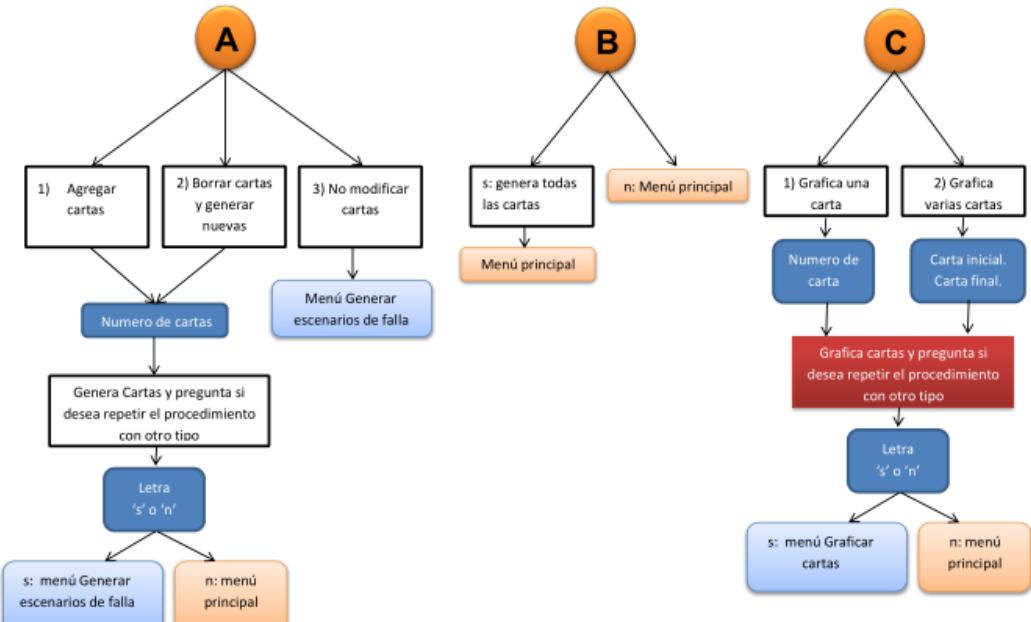


Figura 33. Diagrama del sistema de simulación. Elaboración Propia.

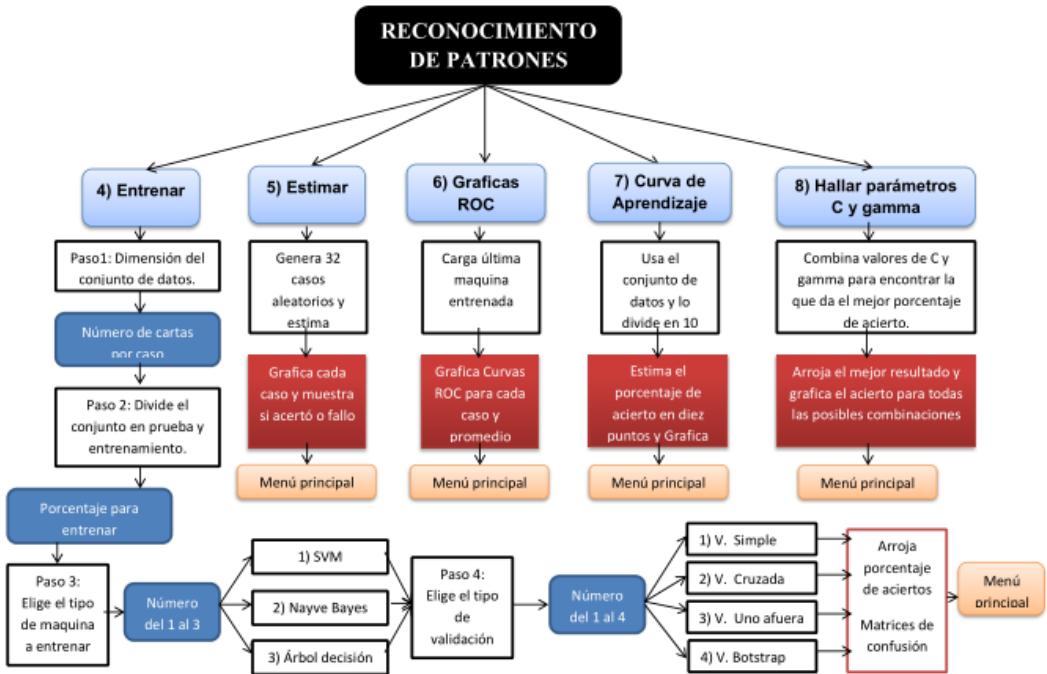


Figura 34. Diagrama del sistema de Reconocimiento de patrones. Elaboración Propia.

a) Simulador

El reconocimiento de patrones parte de una data generada de forma controlada pero aleatoria. Esto se logra a través de condiciones, estados, comparaciones, conocimientos de campo e interpretaciones, tomando como referencia cartas amperimétricas reales. Por ello es necesario en primer lugar, definir el número de fallas a emular por el sistema. En total se delimitó a un comportamiento normal y 12 escenarios de falla:

1. *Comportamiento normal:* Se modela a partir de un valor nominal de corriente en torno a la cual se generan valores aleatorios delimitados, se genera un pico de arranque aleatorio.

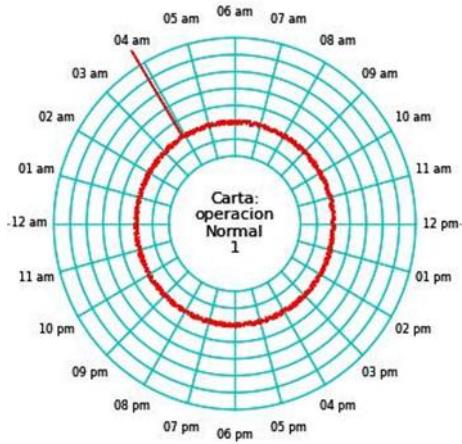


Figura 35. Carta operación Normal. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

2. *Falla picos de corriente:* se modela a partir de una carta de comportamiento normal a la cual se agregan un número de picos aleatorios (delimitados) en posiciones aleatorias y con amplitudes aleatorias (delimitadas)

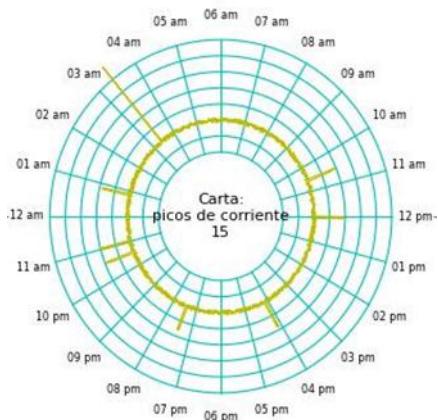


Figura 36. Carta con picos de corriente. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

3. *Falla apagado por gas en la bomba:* está formada por varios ciclos. En cada ciclo: se simula la corriente en comportamiento normal hasta cierto intervalo, en un segundo intervalo se aumenta el rango en que oscilarán los valores de corriente en torno al valor nominal y en un tercer intervalo se lleva la corriente a su condición de apagado. El número y la posición de los ciclos, así como la duración

de cada intervalo se definen de manera aleatoria.

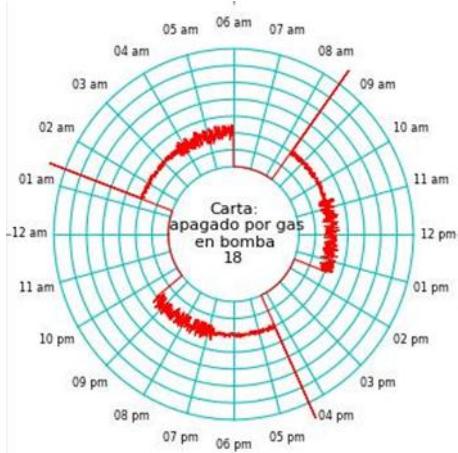


Figura 37. Carta apagado por gas en bomba. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES

4. *Falla por gas libre en la bomba:* se simula la corriente en comportamiento normal por un corto intervalo, en un segundo intervalo de mayor duración se aumenta el rango en que oscilará dicha corriente. Finalmente se determina si la corriente cae en condición de apagado o no. Los intervalos y el inicio de los mismos se generan de manera aleatoria

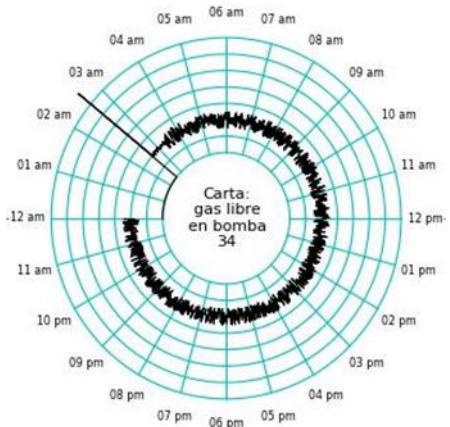


Figura 38. Carta gas libre en la bomba. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

5. *Carta corriente en baja carga*: inicia con corriente en comportamiento normal, le sigue un intervalo en que desciende hasta el valor límite donde se mantiene y luego se apaga, las características son aleatorias

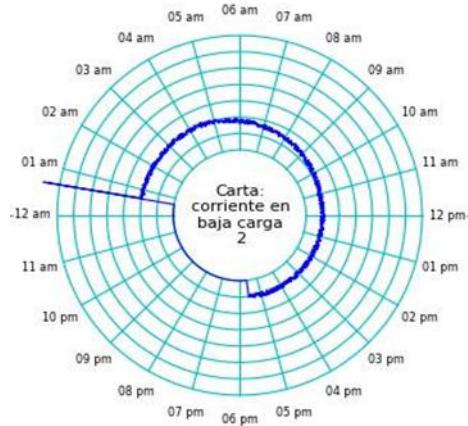


Figura 39. Carta corriente en baja carga. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

6. *Carta corriente en sobrecarga*: inicia con corriente en comportamiento normal, le sigue un intervalo en que asciende hasta el valor límite donde se mantiene y luego se apaga, las características son aleatorias

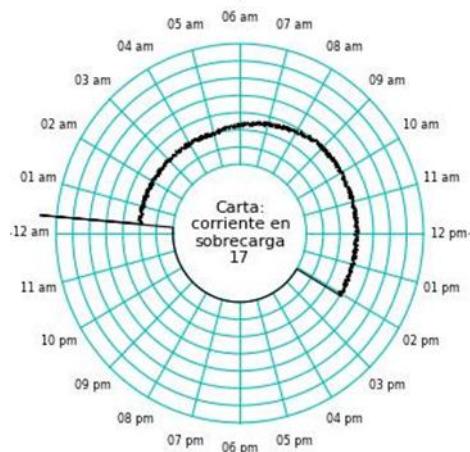


Figura 40. Carta corriente en sobre carga. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES

7. *Carta descarga de fluido*: la corriente inicia a un valor superior al nominal, en el siguiente intervalo desciende hasta llegar a su valor nominal y entrar en comportamiento normal, características aleatorias

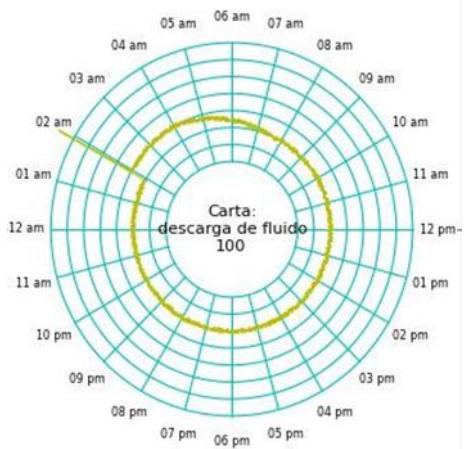


Figura 41. Carta descarga de fluido. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES

8. *Carta bajo nivel de fluido con gas*: condiciones similares a la carta de apagado por gas con la salvedad que la corriente desciende por debajo de su valor nominal antes de la presencia del gas

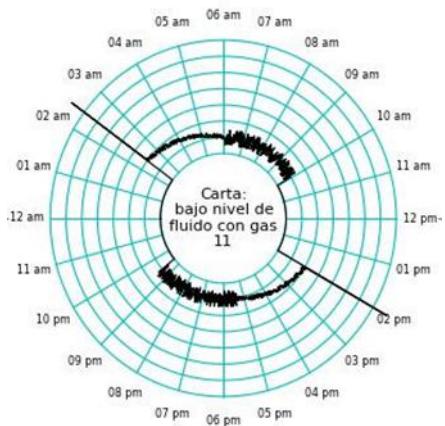


Figura 42. Carta bajo nivel de fluido con gas. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

9. *Carta bajo nivel de fluido sin gas:* mismas condiciones que el caso anterior con la excepción de que el motor se apaga sin fluctuaciones de corriente debido a la presencia de gas

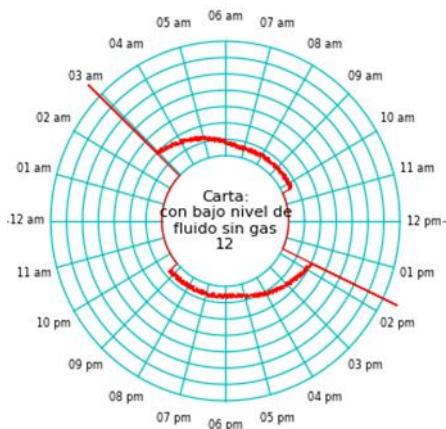


Figura 43. Carta bajo nivel de fluido sin gas. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

10. *Carta número excesivo de arranques:* se generan intentos de arranques aleatorios desde la condición de apagado, al ser intentos fallidos el motor electrosumergible no logra arrancar y se mantiene en condición de apagado

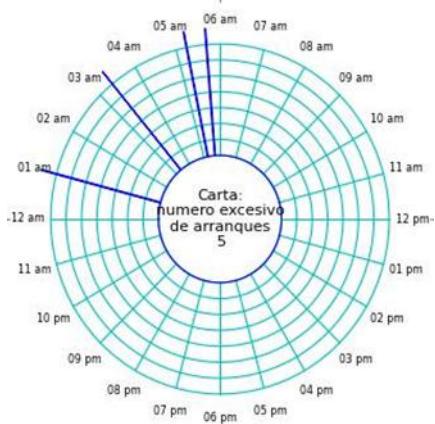


Figura 44. Carta número excesivo de arranques. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

11. Carta excesivos ciclos de operación: se presentan varios grupos de intervalos, cada grupo tiene: un valor de arranque seguido de un intervalo en donde disminuye el valor de corriente, en el intervalo final la corriente se encuentra en condición de apagado. El número de grupos, la posición de los mismos y los intervalos son generados aleatoriamente

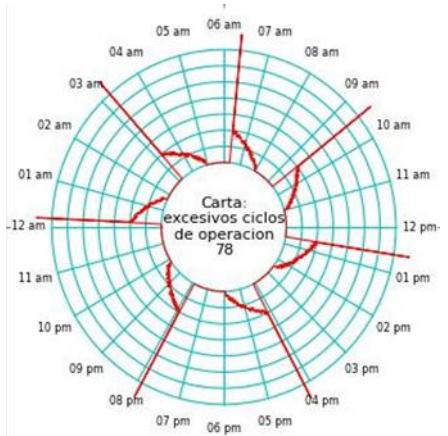


Figura 45. Carta excesivos ciclos de operación. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

12. Carta cargas o emulsiones en superficie: la corriente parte de su valor nominal y se incrementa hasta cierto nivel para caer nuevamente a su valor nominal, estos ciclos se repiten un número de veces aleatorio

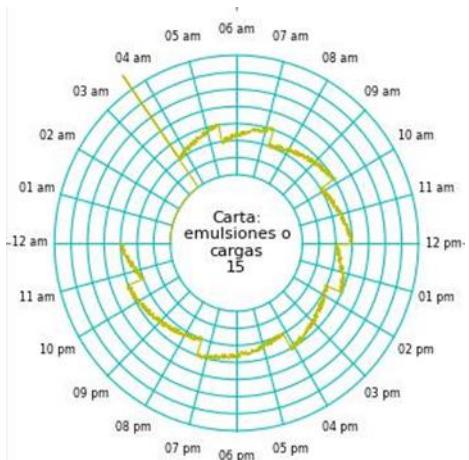


Figura 46. Carta cargas o emulsiones en superficie. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

13. Carta presencia de sólidos en la bomba: en una carta de comportamiento normal se introducen perturbaciones; el número de perturbaciones, su duración y ubicación son generados aleatoriamente.

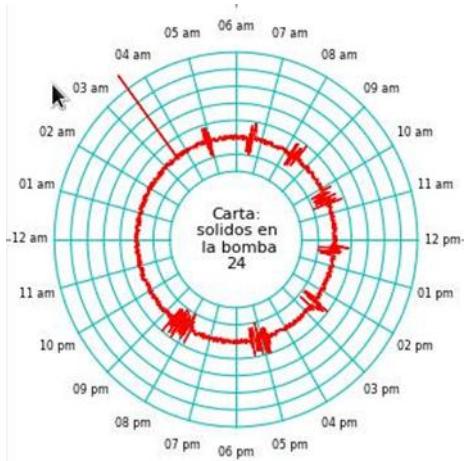


Figura 47. Carta presencia de sólidos en la bomba Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

Una vez se desarrolló el código para cada escenario, mediante el uso de listas, anidaciones y selecciones se definen los métodos que han de generar las cartas, su cantidad y tipo. Los menús del programa son:

1. Generar escenarios de fallas (cartas amperimétricas) en el sistema BES:

Acá se da la opción de crear un numero de cartas definidos por el usuario para cada caso (opciones del 1 al 13) y una opción (14) para generar la misma cantidad de cartas para todos los casos. Al ingresar en alguna de las primeras trece opciones se muestran a su vez 3 opciones: la primera agrega cartas al conjunto existente, la segunda borra todas las cartas existentes y generar un nuevo conjunto y la tercera dejar el conjunto como está.

- La primera opción es útil a la hora del entrenamiento, si se desea entrenar con mayor número de casos es preciso generar más cartas.

- La opción 2 sirve para redefinir características particulares de cada escenario, se usa cuando se tienen nuevos procedimientos y consideraciones que permiten describir de mejor manera el comportamiento real de algún escenario. Cuando se presentan estas situaciones es preciso modificar parcialmente el código. Al hacer esto las condiciones cambian, por lo tanto, el conjunto de cartas actual no es representativo, siendo necesario descartarlo y crear uno nuevo. Ya que el proceso de simulación no se rige por métodos o modelos matemáticos estrictos sino por reglas, experiencias, análisis e interpretaciones es preciso ajustar con regularidad. Todo esto tiene como finalidad mejorar y adaptar a condiciones reales de operación.

2. Consultar número de cartas existentes actualmente:

Muestra el número de cartas existente para cada caso y su ubicación. Una manera rápida y sencilla de consulta.

3. Graficar cartas:

Permite ver la gráfica según el tipo de carta, dando la opción de mostrar una carta, o un conjunto de cartas:

- En la primera opción se pide el número de carta.
- En la segunda opción se pide el número de la primera y última carta del conjunto.

b) Reconocimiento de patrones

Para el reconocimiento de patrones se toma como referencia el

procedimiento KDD (Knowledge Discovery from Databases) en las fases que sean pertinentes.

1. *Integración y recopilación de datos:* como la data es generada y se encuentra en el mismo sitio agrupada en carpetas según sus características no aplica este paso (la fase aplica cuando la data es tomada de distintos lugares, en distintos espacios de tiempo, u ambos)
2. *Selección, limpieza y transformación:* se aplica implícitamente en el proceso simulación. En algunos casos se generan cartas erradas o que presentan inconsistencias, es decir, toman valores fuera de rango o son creadas como cartas en blanco, estas se capturan con excepciones donde se descartan y se generan nuevas. La opción graficar combinada con la de agregar nuevas cartas permite identificar de manera visual aquellas que sean consideradas por el usuario como situaciones no representativas.

Fue necesario generar y visualizar los distintos tipos de carta, comparar con cartas reales y ajustar en los casos que ameriten. También se hizo una revisión de la aleatoriedad entre cartas del mismo tipo. Como se usan Maquinas de soporte vectorial con kernels gaussianos no se realiza una transformación a las características de entrada, dejando la opción abierta en caso de que el desempeño de la maquina no resultase satisfactorio.

3. *Minería de datos:* en esta parte se ejecutan las acciones necesarias para implementar cualquier máquina de aprendizaje y las específicas para las máquinas de soporte vectorial:
 - *Paso 1 generar el conjunto de datos:* se pide el número de cartas de cada tipo a usar (si se usan 20 cartas por ejemplo, el set de datos tendrá $13*20=260$ cartas, se toma la misma cantidad para que las

clases estén balanceadas). Una vez se cargan las cartas se desordena la data aleatoriamente 100 veces gracias a un ciclo anidado. Posteriormente se separan las características de entrada en una matriz y las salidas discretas en un vector.

- *Paso 2 se divide el conjunto de datos en dos:* un conjunto de entrenamiento, con el cual se hallan los parámetros que necesita el modelo para realizar clasificaciones y estimaciones. Un conjunto de prueba en el cual se evalúa el porcentaje de acierto de la maquina al comparar salidas estimadas con salidas reales conocidas. La proporción de cada set es definida por el usuario.
- *Paso 3 Seleccionar la máquina de entrenamiento:* el trabajo está enfocado en la Maquina de soporte vectorial con kernels gaussianos, sin embargo, se presentan 2 opciones de máquinas: arboles de decisión para clasificación y nayve bayes gaussiano. Estas solo son opciones de referencia y no se acondicionan.

En la Maquina de soporte vectorial se realiza el entrenamiento considerando 3 parámetros propios de ella:

- El valor de C, que incide en el sobreajuste, subajuste y acierto.
- El kernel o núcleo, que en este caso es un kernel gaussiano y define la función de similitud necesaria para transformar las características de entrada no lineales (no separables por una función de decisión lineal). Con ello se obtiene una nueva dimensión de características que permiten ser separadas por hiperplanos.

- el valor de gamma: que incide directamente en la función de similitud gaussiana al establecer cuál es la diferencia mínima a la cual se considera un patrón similar o diferente de otro patrón.
4. *Evaluación e interpretación:* una vez entrenada la máquina de aprendizaje se tienen:
- 4 alternativas de validación: simple, cruzada, uno afuera y bootstrap. Ellas arrojan indicadores que estiman el acierto del modelo y reentrenan la maquina según sus propios requerimientos. Cada uno de los métodos muestra una matriz de confusión y porcentaje de patrones bien clasificados.
 - Gráficas ROC que muestran el desempeño para cada clase según la rata de aciertos positivos vs rata de desaciertos positivos y una gráfica ROC promedio que indica la rata general del clasificador múltiple.
 - Curva de aprendizaje en donde se muestra el porcentaje de acierto de la maquina en función de la dimensión del conjunto usado.
 - Comparación del desempeño de la máquina de soporte vectorial ante distintas combinaciones de C y gamma que arroja, de las posibles combinaciones propuestas los valores óptimos que garanticen el mayor porcentaje de aciertos
5. *Difusión y uso:* esta etapa requiere tiempo y el uso del sistema por varios usuarios para obtener valoraciones, medir resultados y expectativas.

Inicialmente se distribuye el código al tutor industrial, al tutor académico y al equipo de ingenieros del distrito expertos en el área.

Etapa IV: Realizar pruebas del sistema computacional para reconocimiento de fallas

Se realizaron pruebas del sistema computacional para encontrar errores de ejecución, bucles no deseados, opciones de menús y submenús, ingreso de valores u opciones no previstas. Se observó la manera en que se presenta la información, la funcionalidad de las opciones, los mensajes explicativos al usuario según sea el caso. En esta etapa se buscaron dos objetivos: que el programa no entre en una condición de bug y que se comporte de acuerdo a lo estipulado. La prueba desde el punto de vista de resultados, rendimiento, comportamientos, predicción, valores reales vs valores esperados se deja para una etapa posterior.

Etapa V: Desarrollar escenarios de prueba para la evaluación y análisis estadístico del sistema

En esta etapa se analiza que tan bien se comporta el sistema desde el punto de vista de acierto/error, también se buscan las características que den el mejor desempeño. En la primera prueba se utiliza la curva de aprendizaje, la cual indica el acierto del clasificador en función del número de ejemplos tomados. Se trabaja con una data de 1300 ejemplos, la misma se divide en 10 partes, se toma el acierto en cada caso.

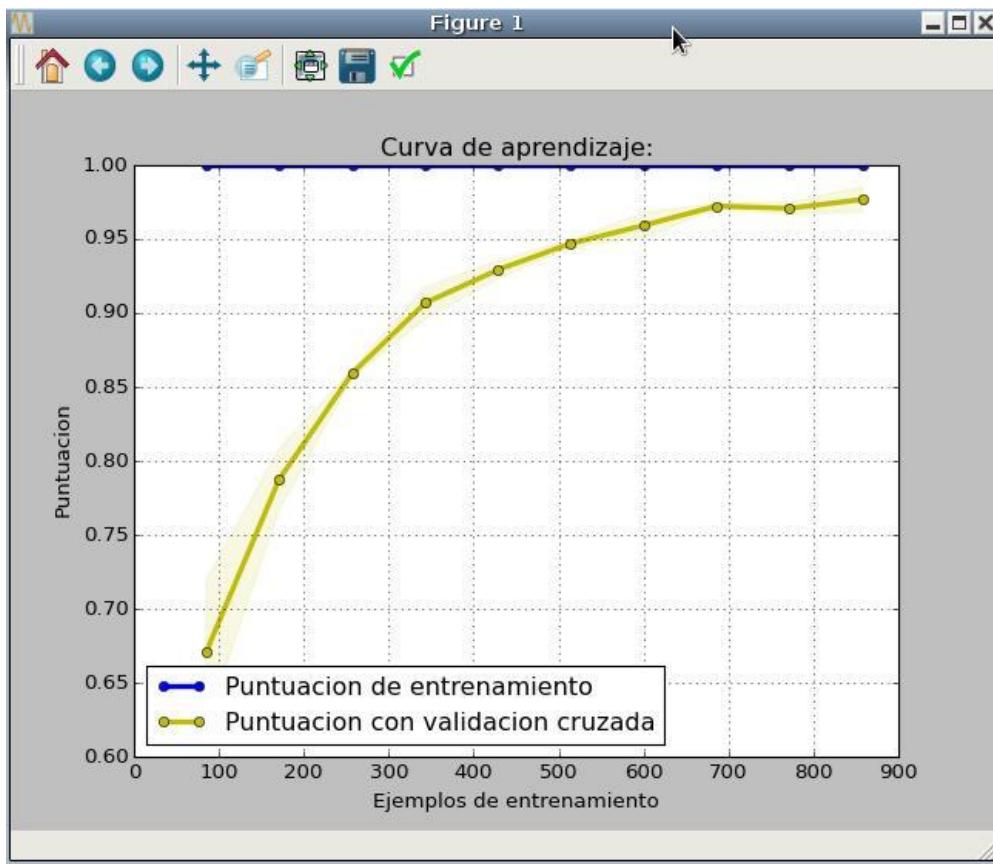


Figura 48. Curva de aprendizaje para MSV. Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia.

En la gráfica se observa que, para un conjunto de datos mayor a 600 se espera un acierto superior al 95%, es decir, con un promedio de 40 a 50 cartas por caso el sistema presenta un buen porcentaje de clasificación

En la segunda prueba se hallan C y gamma, se definen valores y se realizan todas las combinaciones posibles, esto con la intención de elegir aquella que dé mejor desempeño.

Para $C = 1, 2, 5, 10, 15, 20$

Para $\gamma = 0.0001, 0.0005, 0.001, 0.005, 0.01$

Es importante aclarar que este grupo de valores se elige luego de múltiples pruebas, en las cuales se usaron valores de C desde 1 hasta 10000 y de gammas

desde 1E-7 hasta 1. Gracias a ello se logra delimitar de manera experimental al espacio expuesto.

Los valores óptimos encontrados para un set de datos de 1300 son:

C= 5

Gamma= 0.001

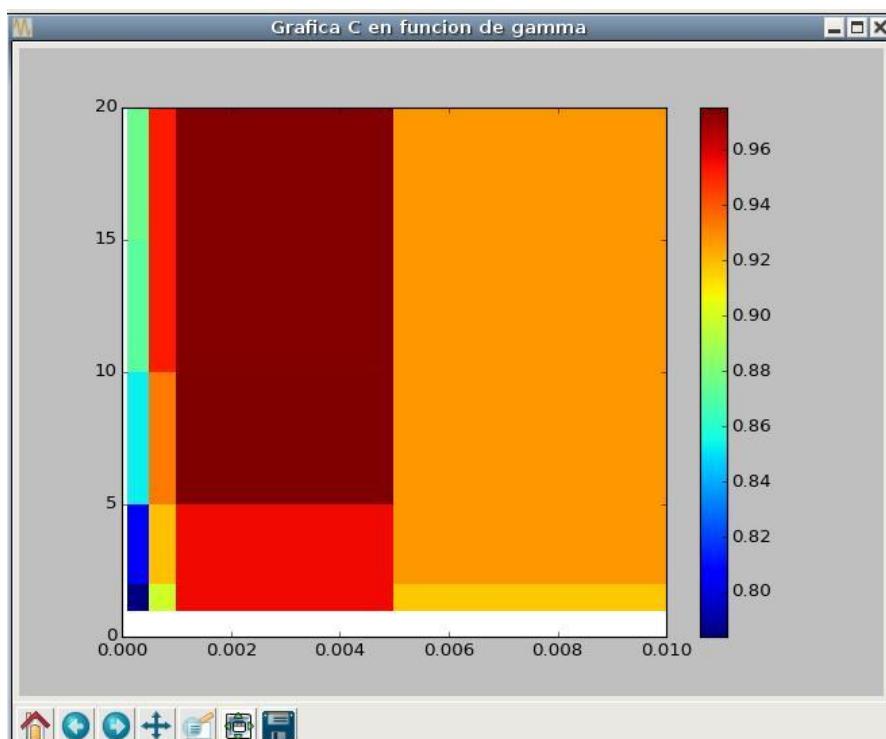


Figura 49. Grafica C vs Gamma Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia.

Validación Simple:

Tomando un 70% para entrenar y un 30% de la data para estimar se obtuvo un acierto de 98.71794871794873 %

Matriz de confusión:

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>F6</i>	<i>F7</i>	<i>F8</i>	<i>F9</i>	<i>F10</i>	<i>F11</i>	<i>F12</i>	<i>F13</i>
<i>F1</i>	30	0	0	0	0	0	0	0	0	0	0	0	0
<i>F2</i>	0	29	0	0	0	0	0	0	0	0	0	0	2
<i>F3</i>	0	0	31	0	0	0	0	0	0	0	0	0	0
<i>F4</i>	0	0	0	29	0	0	0	0	0	0	0	0	1
<i>F5</i>	0	0	0	0	27	0	0	0	0	0	0	0	0
<i>F6</i>	0	0	0	0	0	32	0	0	0	0	0	0	0
<i>F7</i>	0	0	0	0	0	0	29	0	0	0	0	0	0
<i>F8</i>	0	0	0	0	0	0	0	25	0	0	0	0	0
<i>F9</i>	0	0	0	0	0	0	0	0	30	0	0	0	0
<i>F10</i>	0	0	0	0	0	0	0	0	0	30	0	0	0
<i>F11</i>	0	0	0	0	0	0	0	0	0	0	37	0	0
<i>F12</i>	0	0	0	0	0	0	0	0	0	0	0	29	0
<i>F13</i>	0	0	0	0	0	0	0	0	0	0	0	0	27

Tabla 1. Matriz de confusión para la validación simple

Validación cruzada:

Al dividir el set de datos en 10 partes y realizar el mismo número de validaciones:

<i>Numero de validación</i>	<i>Porcentaje de aciertos</i>
1	98.46153846153847
2	100.0
3	98.46153846153847
4	99.23076923076923
5	100.0
6	98.46153846153847
8	98.46153846153847
9	96.15384615384616
10	98.46153846153847
<i>Promedio total:</i>	98.61538461538461

Tabla 2. Porcentaje de aciertos en validación cruzada

Matriz de confusión validación 9

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>F6</i>	<i>F7</i>	<i>F8</i>	<i>F9</i>	<i>F10</i>	<i>F11</i>	<i>F12</i>	<i>F13</i>
<i>F1</i>	12	0	0	0	0	0	0	0	0	0	0	0	0
<i>F2</i>	0	16	0	0	0	0	0	0	0	0	0	0	4
<i>F3</i>	0	0	8	0	0	0	0	0	0	0	0	0	0
<i>F4</i>	0	0	0	11	0	0	0	0	0	0	0	0	0
<i>F5</i>	0	0	0	0	9	0	0	0	0	0	0	0	0
<i>F6</i>	0	0	0	0	0	8	0	0	0	0	0	0	0
<i>F7</i>	0	0	0	0	0	0	7	0	0	0	0	0	0
<i>F8</i>	0	0	0	0	0	0	0	10	0	0	0	0	0
<i>F9</i>	0	0	0	0	0	0	0	0	8	0	0	0	0
<i>F10</i>	0	0	0	0	0	0	0	0	0	10	0	0	0
<i>F11</i>	0	0	0	0	0	0	0	0	0	0	8	0	0
<i>F12</i>	0	0	0	0	0	0	0	0	0	0	0	6	0
<i>F13</i>	0	0	0	0	0	0	0	0	0	0	0	0	12

Tabla 3. Matriz de confusión en validación cruzada

Validación Bootstrap:

Al realizar validación bootstrap 10 veces y promediar se obtuvo lo siguiente

<i>Numero de validación</i>	<i>Porcentaje</i>		
	<i>Acierto pesimista</i>	<i>Acierto optimista</i>	<i>Bootstrap 0.636</i>
1	95.61586	98.38461	96.62369
2	96.70103	98.76923	97.45385
3	96.56652	98.76923	97.36830
4	97.13114	98.92307	97.78340
5	97.64957	99.15384	98.19712
6	97.89473	99.23076	98.38105
7	97.20430	99.00000	97.85793
8	97.72727	99.15384	98.24654

9	97.40259	99.07692	98.01205
10	96.04166	98.53846	96.95050
Promedio	96.99347	98.90000	97.68744

Tabla 4. Porcentaje de aciertos en validación bootstrap

Matriz de confusión pesimista

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>F6</i>	<i>F7</i>	<i>F8</i>	<i>F9</i>	<i>F10</i>	<i>F11</i>	<i>F12</i>	<i>F13</i>
<i>F1</i>	29	0	0	0	0	0	0	0	0	0	0	0	0
<i>F2</i>	0	32	0	0	0	0	0	0	0	0	0	0	6
<i>F3</i>	0	0	30	0	0	0	0	0	0	0	0	0	0
<i>F4</i>	0	0	0	38	0	0	0	0	0	0	0	0	2
<i>F5</i>	0	0	0	0	34	0	0	0	0	0	0	0	0
<i>F6</i>	0	0	0	0	0	36	0	0	0	0	0	0	0
<i>F7</i>	0	0	0	0	0	0	46	0	0	0	0	0	0
<i>F8</i>	0	0	0	0	0	0	0	41	0	0	0	0	0
<i>F9</i>	0	0	0	0	0	0	0	0	34	0	0	0	0
<i>F10</i>	0	0	0	0	0	0	0	0	0	38	0	0	0
<i>F11</i>	0	0	0	0	0	0	0	0	0	0	38	0	0
<i>F12</i>	0	0	0	0	0	0	0	0	0	0	0	36	0
<i>F13</i>	7	4	0	0	0	0	0	0	0	0	0	0	29

Tabla 5. Matriz de Confusión pesimista en validación bootstrap

Matriz de confusión optimista

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>F6</i>	<i>F7</i>	<i>F8</i>	<i>F9</i>	<i>F10</i>	<i>F11</i>	<i>F12</i>	<i>F13</i>
<i>F1</i>	100	0	0	0	0	0	0	0	0	0	0	0	0
<i>F2</i>	0	94	0	0	0	0	0	0	0	0	0	0	6
<i>F3</i>	0	0	100	0	0	0	0	0	0	0	0	0	0
<i>F4</i>	0	0	0	98	0	0	0	0	0	0	0	0	2
<i>F5</i>	0	0	0	0	100	0	0	0	0	0	0	0	0
<i>F6</i>	0	0	0	0	0	100	0	0	0	0	0	0	0

<i>F7</i>	0	0	0	0	0	0	100	0	0	0	0	0	0
<i>F8</i>	0	0	0	0	0	0	0	100	0	0	0	0	0
<i>F9</i>	0	0	0	0	0	0	0	0	100	0	0	0	0
<i>F10</i>	0	0	0	0	0	0	0	0	0	100	0	0	0
<i>F11</i>	0	0	0	0	0	0	0	0	0	0	100	0	0
<i>F12</i>	0	0	0	0	0	0	0	0	0	0	0	100	0
<i>F13</i>	7	4	0	0	0	0	0	0	0	0	0	0	89

Tabla 6. Matriz de Confusión optimista en validación bootstrap

Los resultados arrojados en la validación simple, cruzada y bootstrap son similares, tal como se esperaba.

Análisis ROC:

Se tiene el comportamiento de cada clase en función de rata de aciertos positivos vs rata de desaciertos negativos, por debajo de la línea punteada el desempeño es pobre, las curvas indican un buen desempeño tal como se esperaba.

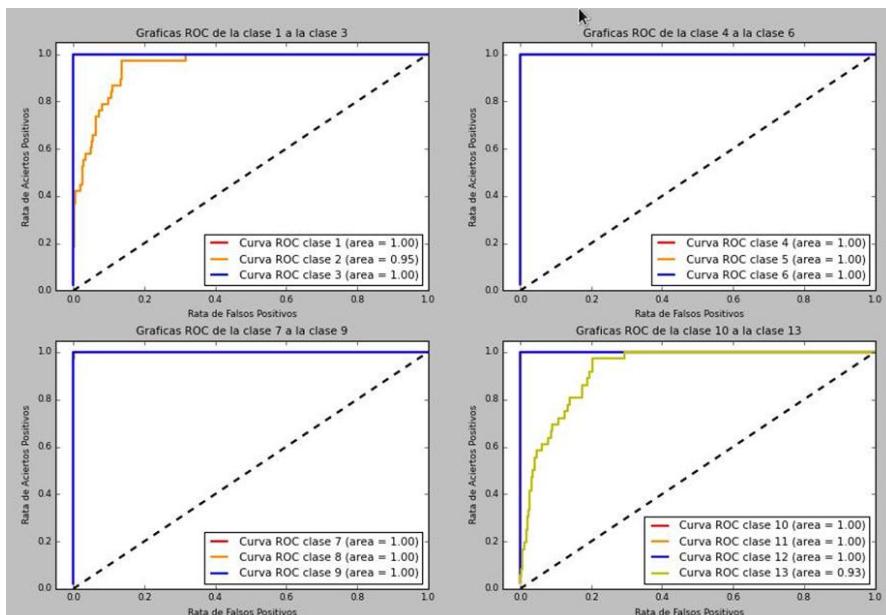


Figura 50. Graficas ROC por clase Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

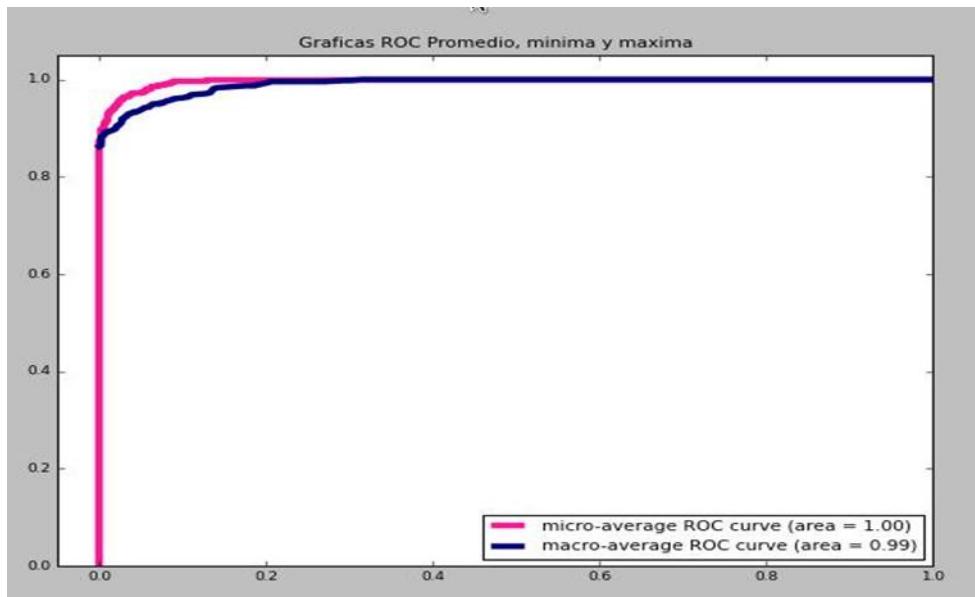


Figura 51. Grafica ROC promedio Fuente: Sistema de simulación y reconocimiento de patrones de fallas en el sistema BES. Elaboración propia

Etapa VI: Elaborar la documentación relacionada con la instalación y uso del sistema

Se creó un manual con las pautas para instalar el programa, se describe cada una de las opciones y se incluyen ejemplos sobre lo que se espera ver en el mismo.

Capítulo V

Discusión y análisis de resultados

El sistema de simulación y reconocimiento de patrones de fallas en el método de levantamiento artificial por bombeo electrosumergible (BES) se desarrolló en lenguaje Python. Esta elección se debe a:

- Es un lenguaje de programación muy usado en el campo de la computación científica.
- Al ser software libre cuenta con una extensa gama de librerías y una comunidad de desarrolladores que regularmente realizan aportes en todas las áreas (esta es la filosofía con la que se creó Python).
- Es un lenguaje versátil que permite trabajar con varios enfoques de programación (orientada a objetos, estructurada, secuencial) y combinaciones de los mismos.
- Busca ser un lenguaje interpretativo, evitando el uso de pautas tipográficas que considera innecesarias como paréntesis, llaves, punto y coma.
- Es un lenguaje que se presta a ser implementado en sitios web, lo cual es una de las metas que persigue PDVSA AIT Corporativo con cada uno de sus proyectos.

En la simulación las cartas generadas presentan un comportamiento similar a las cartas reales (con las salvedades pertinentes). La misma se sustenta en un análisis descriptivo en lugar de modelos matemáticos o físicos, debido a que se trabaja con fallas. Un análisis de esta naturaleza para representar el comportamiento de la corriente eléctrica está basado en reglas, conocimiento experto, situaciones esperadas e interpretación de escenarios. Es importante dejar el simulador abierto a correcciones, consideraciones y estudios que le permita ajustar el modelado de fallas.

El programa se pensó con la idea de facilitar el análisis, ajuste y corrección

continúa, de allí las opciones de borrar y crear rápidamente nuevas cartas o escenarios simulados.

Inicialmente se organiza y selecciona la data representativamente. Luego del proceso de entrenamiento y validación se nota un pobre desempeño en la predicción, indistintamente del número de cartas usadas. Al analizar este resultado se detecta una situación particular: los casos en donde existe mala clasificación en su mayoría corresponden a las clases 12 y 13. La razón de ello es el estado del conjunto de datos, se encuentra ordenado. Es decir, si el set contiene 130 datos las filas del 1 al 10 están ocupadas por ejemplos del caso 1, las filas del 11 al 20 con ejemplos del caso 2 y así sucesivamente. Al hacer la partición de la data para prueba y validación se excluyen casos en el entrenamiento y se evalúa en estos mismos casos. Al aplicar una función de aleatoriedad que desordene las filas con la intención de estratificar el resultado es el esperado. El acierto mejoró a medida que se aumentó la data.

Las pruebas realizadas para hallar los valores de C y gamma empiezan con rangos de $C = 0.1$ hasta 100000 y de gamma desde 0.00001 hasta 1 . En la prueba macro, con un vector $C= [1, 10, 100, 1000, 10000, 1000000]$ y $\text{gamma}= [0, 0.00001, 0.0001, 0.001, 0.01, 0.1]$ se tienen 30 posibles combinaciones. En esta situación la gráfica no refleja una variación y presenta un color uniforme. En una segunda prueba se toman vectores $C = [1, 5, 10, 50, 100]$ y $\text{gama}= [0, 0.001, 0.005, 0.01, 0.05, 0.1]$ Acá se empieza a notar la variación que sufre el porcentaje de aciertos en función a los valores de compensación (C) y varianza (Gamma).

Finalmente se realiza la prueba con $C= [1, 5, 10, 15, 20]$ y $\text{gamma}= [0.0005, 0.001, 0.005, 0.01]$ la cual arroja la gráfica presentada en la figura 44. Para estas combinaciones se llega a un valor óptimo de $C=5$ y $\text{gamma}=0.001$, con un acierto superior al 96%. Es oportuno recordar que se trabajó en base a una data de 1300 ejemplos y que la elección de valores óptimos no es única y definitiva, se obtiene a través de pruebas de ensayo y error

Al trazar la curva de aprendizaje se observa que a partir de los 600 ejemplos el desempeño de la maquina se ubica arriba del 95 %, con este resultado se afirma que, ante una data con extensión igual o superior a la mencionada se puede entrenar una muy buena herramienta.

Con los valores encontrados de C, gamma y usando 1300 ejemplos se entrena la máquina de soporte vectorial. Al realizar el proceso de validación simple con un 70% para entrenamiento y un 30% para prueba se obtiene un acierto de 98.7879%, al realizar la validación cruzada en 10 y bootstrap en 10 se obtienen los resultados parciales y totales reflejados en las tablas 2 y 4. La validación uno afuera no se implementa ya que el costo computacional y de tiempo (requiere entrenar y probar la maquina 1300 veces) no compensa una estimación que ha de encontrarse cercana a los otros tres métodos.

Las matrices de confusión generadas en los procesos de validación presentan una tendencia: las cartas mal clasificadas suelen confundir las predicciones de la clase 13 con las clases 1 y 2 más que en cualquier otro caso. Aunque el índice de acierto/error sigue siendo favorable es un síntoma que podría corregirse en un futuro dándose un tratamiento a la data de entrada, o redefiniendo el kernel usado.

En las gráficas ROC de cada clase se refleja un buen desempeño del clasificador, la rata de aciertos positivos predomina significativamente sobre la racha de desaciertos positivos, igual ocurre en las gráficas ROC Promedio

Una vez entrenada la máquina y guardada en un archivo el proceso de carga-predicción es rápido, esto permite estudiar gran cantidad de cartas y encontrar tendencias que permitan realizar análisis del comportamiento de un pozo o grupo de ellos.

Conclusiones

El método de levantamiento artificial por bombeo electrosumergible es uno de los más utilizados en el sector petrolero, especialmente en pozos que presentan un buen nivel de producción. Unas de sus principales ventajas son: puede levantar grandes volúmenes de fluido, trabajar a baja presión de fondo y con el mantenimiento adecuado tienen una expectativa de vida larga. En contraparte su principal desventaja es que requiere una inversión inicial elevada en comparación a otros métodos, su diseño es complejo y necesita un tiempo mayor para corregir fallas, pues la infraestructura del bombeo (bomba-motor) se encuentran en el subsuelo.

El proceso requiere un estudio que analice costos-beneficios, una manera de garantizar la vida útil y buen funcionamiento del sistema BES es a través de métodos que permitan identificar anomalías. Uno de esos métodos es el estudio del comportamiento de la corriente a través de cartas amperimétricas.

Una carta amperimétrica muestra los valores de la corriente del motor 24 horas del día, a través de ella se analizan comportamientos, anomalías y se toman planes de acción. En base a la experiencia y observaciones se han logrado identificar situaciones y patrones que indican lo que está pasando en el sistema BES. El objeto del trabajo presentado es dar un primer paso en la identificación a nivel de datos, mediante el uso de máquinas de aprendizaje, específicamente máquinas de soporte vectorial de estas fallas.

Se observa la no linealidad del problema, por ello se recurre a las máquinas de soporte vectorial con kernels gaussianos; a través de múltiples pruebas, ajustes y restricciones se concluye que un método de esta naturaleza aplicado a la detección de fallas, brinda un buen desempeño, además permite la

generalización de condiciones a múltiples pozos cosa que no siempre es posible. ¿Por qué permite una generalización de este tipo? El estudio se fundamenta en variables eléctricas del motor y aunque los motores empleados en distintos pozos presentan condiciones de potencia, voltaje y dimensión distintas el comportamiento no cambia (los valores de corriente pueden diferir entre sí, pero esto se solventa con una normalización previa).

El estudio es una fase experimental y se debe tomar como un primer intento de aplicar las máquinas de aprendizaje no solo en el sistema BES sino en cualquier otro problema del que se cuente con datos y situaciones que no siempre puedan ser modeladas a través de procedimientos matemáticos y/o físicos en el DST PDVSA AIT Corporativo

Como todo proceso requiere ajustes, correcciones y reestructuraciones para ser concluyente, por tal razón el programa deja múltiples alternativas para todas estas opciones. Se dejan comentarios y consideraciones tomadas a la hora de simular escenarios en los códigos o scripts, con la intención de que futuros desarrolladores se sientan con la libertad de aportar ideas o corregir situaciones pasadas por alto.

Recomendaciones

Se recomienda consultar la opinión de expertos de campo, de personal con experiencia en la lectura y análisis de cartas amperimétricas para obtener una respuesta objetiva y real que permita al estudio pasar de situaciones simuladas y pruebas controladas al plano de aplicación propiamente dicho

Se sugiere un contacto continuo con personal de campo, para conocer los requerimientos puntuales, situaciones particulares y contar con una base de datos y de información que permitan un estudio y análisis más profesional.

Sería bueno aumentar el número de reuniones técnicas a la hora de ir mostrando avances, definir planes a seguir y compartir opiniones.

Referencias Bibliográficas

- Aguilar, J. (2010) *Introducción a Minería de Datos, Metodologías y Técnicas de Minería de datos*. Diapositivas para CEMISID, Escuela de Sistemas Facultad de Ingeniería Universidad de Los Andes Mérida (Venezuela)
- Amao M (2014) *Electrical Submersible Pumping (ESP) Systems Artificial Lift Methods and Surface Operation.*
- Arcentales, D. & Caizapanta, M. (2011). *Aplicación del análisis nodal para la evaluación del sistema de levantamiento artificial por bombeo electrosumergible del campo ESPOL III* Tesis de pregrado. Escuela Superior Politécnica del Litoral, Guayaquil (Ecuador)
- Arias, F. (1999). El proyecto de Investigación]. Caracas: Editorial Episteme.
- Camargo, E., Aguilar, J., Ríos, A. & Rivas, F. (2009). Un modelo de producción de pozos por levantamiento artificial utilizando análisis nodal. *Ciencia e Ingeniería*, 30(1), paginas 23-28
- Camargo E. (2009). *Supervisión Inteligente en la Producción Industrial*. Tesis doctoral. Universidad De Los Andes (ULA) Mérida
- Cardozo N. y Pabón L. (2009) *análisis de falla de pozos de bombeo electro sumergible en el campo cantagallo* Tesis de pregrado, Universidad Industrial De Santander. Bucaramanga (Colombia)
- Carrillo J. (2010) *Diseño de una metodología en la determinación de una familia de criterios para la selección de sistemas de levantamiento artificial*. Trabajo especial de postgrado. Universidad Católica Andrés Bello (UCAB) Caracas

González, L (2010) *Modelos de Clasificación basados en Máquinas de Vectores Soporte*. Artículo. Universidad de Sevilla. Sevilla (España)

Guillén M. Paredes J. y Camacho O. (2010 abril-Julio) Un enfoque para la detección y el diagnóstico de fallas en la instrumentación de un proceso usando reconocimiento de patrones en el dominio wavelet. *Revista Ciencia e Ingeniería*. Vol. 31(2), paginas. 83-90.

Leal R. Camargo E. Aguilar J. Ríos A. Travé-MassuyèsL (2015 septiembre) Análisis de Diagnosticabilidad y Localización de Sensores en un Pozo de Extracción de Petróleo por Inyección de Gas. *Revista Politécnica* Vol. 36 (1)

Machete, E. (2007). *Análisis del desempeño de dos sistemas de levantamiento artificial bombeo electrosumergible y bombeo mecánico VSH2 utilizados en el campo Cano Limón*. Tesis de pregrado. Universidad Industrial de Santander, Bucaramanga (Colombia)

Mitchell, T. (1997). *Machine Learning*. Washington: McGraw-Hill

Witten, I. Frank, E. y Hall, M. (2011) Data Mining: Practical Machine Learning Tools and Techniques USA: Morgan Kaufmann

Padrón, J. (1998). La Estructura de los Procesos de Investigación. [Documento en línea]

Ramírez. M, ESP Oil International Training Group (2004) *Bombeo Electrosumergible: Análisis, Diseño optimización y trouble shooting*. Curso de adiestramiento dictado en el workshop internacional. Maturín (Venezuela)

Ruiz, S. (2013) “*Aprendizaje de Maquina y reconocimiento de patrones*” Curso dictado en Centro de investigaciones en matemáticas. Verano de la computación. México.

Sabino Carlos (2002) *El proceso de investigación*, Caracas: Editorial Panapo.

Vásquez, J. (2011) *Evaluación del método de levantamiento artificial por bombeo electrosumergible para su aplicación en el área mayor de oficina del distrito gas Anaco*. Tesis de pregrado. Universidad de Oriente (UDO), Barcelona (Venezuela)

Zumoffen D. (2008). *Desarrollo de sistemas de diagnóstico de fallas integrado al diseño de control tolerante a fallas en procesos químicos*. Tesis doctoral, Universidad Nacional del Rosario (UNR) Ciudad de Rosario (Argentina)

Anexo: Manual del programa.



DIRECCIÓN EJECUTIVA DE AUTOMATIZACIÓN, INFORMÁTICA Y TELECOMUNICACIONES
GERENCIA DE DESARROLLO SOCIALISTA PRODUCTIVO
DISTRITO SOCIALISTA TECNOLÓGICO

MANUAL.

SISTEMA DE SIMULACIÓN Y RECONOCIMIENTO DE PATRONES DE FALLAS EN EL MÉTODO DE EXTRACCIÓN DE CRUDO POR BOMBEO ELECTROSUMERGIBLE (BES) MEDIANTE CARTAS AMPERIMÉTRICAS

Angel Leonardo Duarte Montes

10
AÑOS

Estado Bolivariano de Miranda | Ministerio del Poder Popular para la Defensa | 2018 | 10 AÑOS DE SERVICIOS AL PUEBLO

ANTES DE INICIAR

- Debe tener instalado python 3 en su equipo, si no lo tiene y es usuario Linux escriba el siguiente comando en la terminal

```
apt-get install python3
```

- una vez instalado haga lo siguiente

```
apt-get install python3-setuptools  
apt-get install python3-pip
```

- Si ha seguido los pasos anteriores con éxito puede instalar la librería scikit-learn necesaria para ejecutar el programa; la librería requiere :

Python (>= 2.7 or >= 3.3),
NumPy (>= 1.8.2),
SciPy (>= 0.13.3).

por lo general NumPy y SciPy vienen incluidos, sin embargo puede comprobarlo:

```
$python3  
>>> import numpy  
>>>import scipy
```

si libera el prompt (>>>) las librerías están instaladas, en caso de no tenerlas se mostrará una traza de error similar a la siguiente



```
Traceback (most recent call last):
File "<stdin>", line 1, in <module>
ImportError: No module named 'numpy'
ImportError: No module named 'scipy'
```

si tiene éxito puede proseguir, en caso contrario instale las librerías faltantes, el pip y el easyinstall de python pueden resultarle de ayuda

- Instale la librería scikit-learn

```
pip3 install -U scikit-learn
```

al terminar de descargar e instalar se prueba que la librería este instalada correctamente

```
python3
>>>import sklearn
```

al igual que en el caso anterior, si la librería no se encuentra instalada correctamente mostrara una traza de error:

```
Traceback (most recent call last):
File "<stdin>", line 1, in <module>
ImportError: No module named 'sklearn'
```

Si ha tenido éxito con los pasos anteriores: Felicidades!, ya puede ejecutar el programa.



EJECUTAR EL PROGRAMA:

El programa debe estar contenido en la carpeta SimuladorDetector y tiene los siguientes scripts:

- *Principal.py*
- *Training_Set.py*
- *Menus.py*
- *Manage_Files.py*
- *Load_And_Form_Data.py*
- *Generate_Ammeter_Charts.py*

Una vez compruebe el contenido de la carpeta siga los siguientes pasos:

- 1) copie y pegue la carpeta en su equipo
- 2) busque la carpeta, una vez sea la carpeta SimuladorDetector su directorio actual ejecute :

python3 Principal.py

```
angelf@Angel-Leonardo:~$ cd SimuladorDetector/
angelf@Angel-Leonardo:~/SimuladorDetector$ ls
Generate_Ammeter_Charts.py  Manage_Files.py  Principal.py
Load_And_Form_Data.py       Menus.py        Training_Set.py
angelf@Angel-Leonardo:~/SimuladorDetector$ python3 Principal.py
```



10
AÑOS

Debería ver lo siguiente:

SIMULACION Y RECONOCIMIENTO DE PATRONES DE FALLA EN EL SISTEMA BES

Angel Leonardo Duarte Montes
V-19134002
Universidad Nacional Experimental del Táchira UNET
PDVSA AIT Corporativo, DST Mérida-Venezuela

¡BIENVENIDO!!, que desea hacer?

- 1) Generar escenarios de fallas (cartas amperimétricas) en el sistema BES
- 2) Consultar numero de cartas existentes actualmente
- 3) Graficar cartas
- 4) Entrenar el modelo de Reconocimiento
- 5) Estimar tipo de falla con el modelo d Reconocimiento
- 6) Graficas ROC para Maquina Soporte Vectorial
- 7) Curva de Aprendizaje
- 8) Salir

IMPORTANTE!!: si se ejecuta por primera vez en su equipo es necesario ingresar a la opción 1, se mostrará un sub-menú :

OPCIÓN 1: Que tipo de carta desea generar

- 1) Operación normal
- 2) Falla picos de corriente
- 3) Falla apagado por gas en la bomba
- 4) Falla por gas libre en la bomba
- 5) Falla por corriente en bajacarga
- 6) Falla por corriente en sobrecarga
- 7) Falla por descarga de fluido
- 8) Falla por bajo nivel de fluido con gas en la bomba
- 9) Falla por bajo nivel de fluido sin gas en la bomba
- 10) Falla por numero excesivo de arranques
- 11) Falla por excesivos ciclos de operación
- 12) Falla por emulsiones o cargas en superficie
- 13) Falla por sólidos en la bomba
- 14) Generar cartas de todos los tipos
- 15) Regresar al menu anterior



Una vez en ingrese a este submenu elegir la opción 14:

Se vaciaran todas las carpetas y se crearan nuevas cartas, esta de acuerdo? s/n

Elegir la opción : s ; se mostrara lo siguiente:

cartas borradas
Cuantas cartas desea que contenga cada carpeta?

En esta parte pedirá cuantas cartas o escenarios de cada tipo generará el programa, el numero queda a decisión del usuario (entre 50 y 150 suelen ser suficientes).

Ingrese el número, presione ENTER y espere (es un proceso rápido, la consola indicara las cartas o escenarios de fallas que se están generando), al terminar arrojará el siguiente mensaje:

Cartas generadas con exito!, presione enter para continuar

al presionar ENTER devolverá al menú principal, puede comprobar las cartas generadas con la opción 2:

```
OPCIÓN 2: NUMERO DE CARTAS EXISTENTES

1 ) existen 100 cartas de Operacion normal
almacenadas en /home/angel/SimuladorDetector/1_operacion_normal

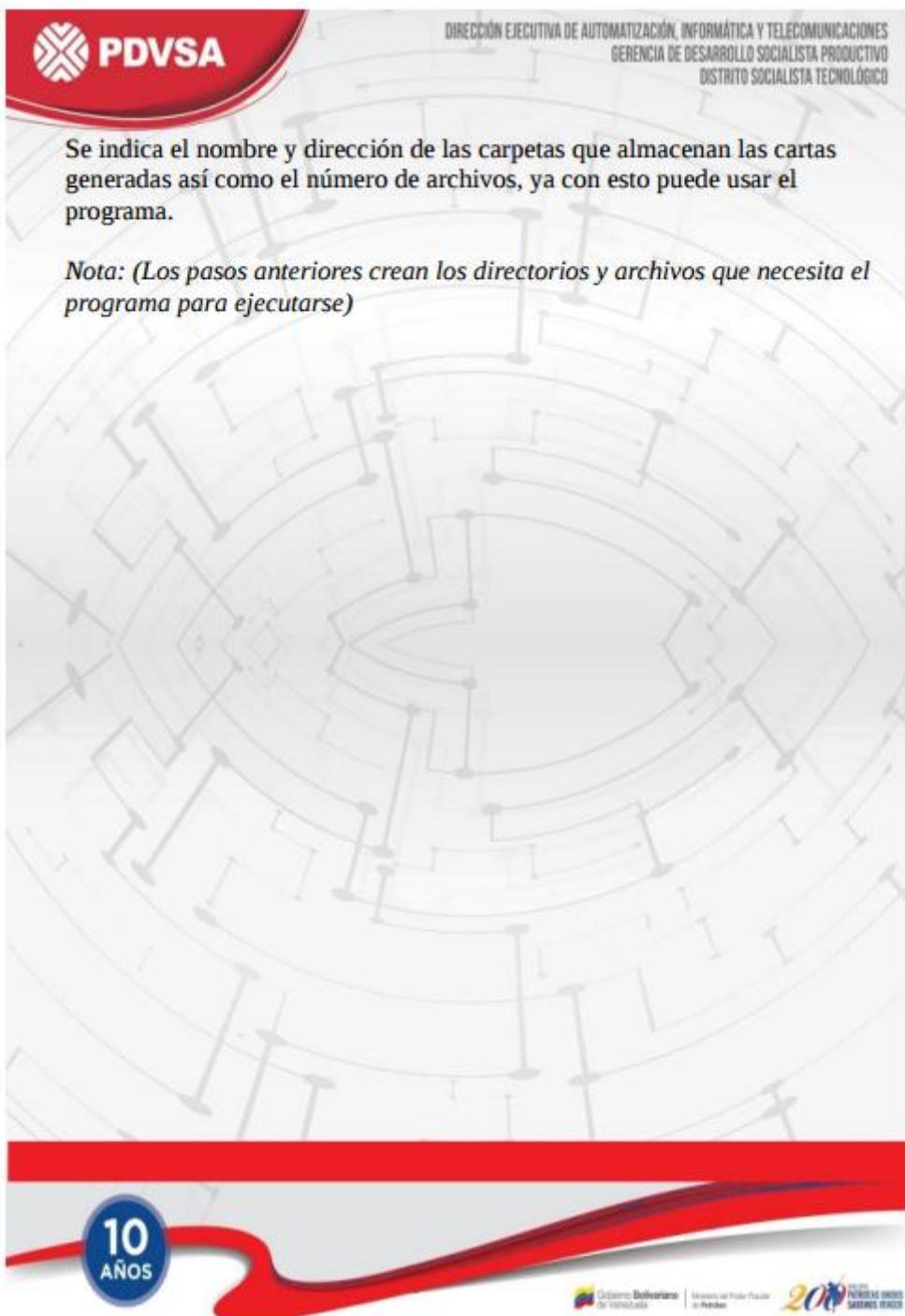
2 ) existen 100 cartas de Falla picos de corriente
almacenadas en /home/angel/SimuladorDetector/2_con_picos

3 ) existen 100 cartas de Falla apagado por gas en la bomba
almacenadas en /home/angel/SimuladorDetector/3_apagado_por_gas
```



Se indica el nombre y dirección de las carpetas que almacenan las cartas generadas así como el número de archivos, ya con esto puede usar el programa.

Nota: (Los pasos anteriores crean los directorios y archivos que necesita el programa para ejecutarse)



OPCIONES DEL PROGRAMA

OPCIÓN 1: Generar escenarios de fallas (cartas amperimétricas) en el sistema BES

Permite crear de manera aleatoria un conjunto de escenarios enmarcados en trece tipos: un comportamiento normal y doce comportamientos de falla, se incluye una opción(14) para crear el mismo número de escenarios para cada tipo. Cada caso presenta las siguientes opciones:

```
el fichero /12_cargas_en_superficie contiene 100 cartas
a) si desea agregar mas cartas al fichero inserte 1
b) si desea borrar las cartas existentes y crear un nuevo set presione 2
c) si desea dejar el set de cartas actual y salir presione 3
```

Opción a) al seleccionar pedirá el número de cartas nuevas que se desean crear y agregar a las actuales

Opción b) al seleccionar borrara todas las cartas existentes y pedirá el número de cartas nuevas a crear.

Opción c) al seleccionar deja el set actual y retorna al menú

OPCIÓN 2: Consultar número de cartas existentes actualmente

Muestra el número de cartas de cada tipo existentes



10
AÑOS

OPCIÓN 3: Graficar cartas

Permite mostrar las cartas existentes

OPCION 3: GRAFICAR CARTAS
que tipo de carta desea graficar?
1) Operacion normal
2) Falla picos de corriente
3) Falla apagado por gas en la bomba
4) Falla por gas libre en la bomba
5) Falla por corriente en bajacarga
6) Falla por corriente en sobrecarga
7) Falla por descarga de fluido
8) Falla por bajo nivel de fluido con gas en la bomba
9) Falla por bajo nivel de fluido sin gas en la bomba
10) Falla por numero excesivo de arranques
11) Falla por excesivos ciclos de operacion
12) Falla por emulsiones o cargas en superficie
13) Falla por solidos en la bomba
14) Fallas Mixtas (no implementado)
15) Volver al menu anterior

Al seleccionar una opción (de la 1 a la 13) se muestra el siguiente submenu:

¿Cuantas cartas de /4_gas_en_bomba desea mostrar?
1)una carta
2)varias cartas

1) una carta:

Que numero de carta (/4_gas_en_bomba) desea graficar?
entre 1 y 100

Pide el número de grafica (de las existentes) a graficar



2) varias cartas:

Se ingresa la primera carta a mostrar

Cual es el valor inicial (/3_apagado_por_gas) que desea graficar
(recuerde que la carpeta tiene 100 cartas)

Se ingresa la última carta a mostrar

Cual es el valor final de carta /3_apagado_por_gas que desea graficar

Una vez se ingresan los valores el programa pedirá el formato para mostrar la información (como se dividirá la gráfica):

Número de filas

ingrese el numero de filas

Número de columnas

ingrese el numero de columnas

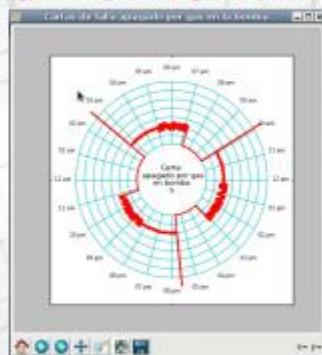
Una vez seleccionados los valores requeridos el programa preguntara si desea mostrar otro conjunto de graficas:

desea cargar y ver otro tipo de carta? s/n
(Todas las cartas se mostraran al final)

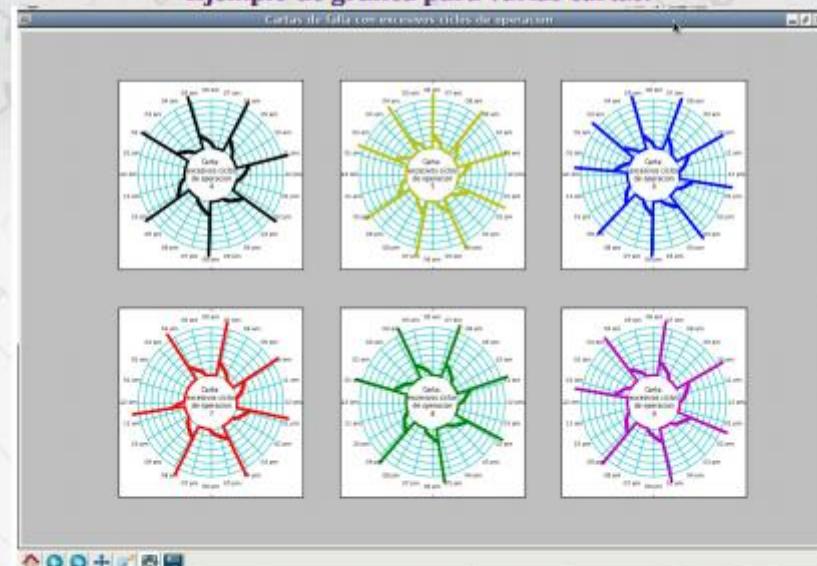


10
AÑOS

Ejemplo de gráfica para una carta:



Ejemplo de gráfica para varias cartas:



**10
AÑOS**

OPCIÓN 4: Entrenar el modelo de Reconocimiento

En esta opción es, donde a través de una serie de pasos y las cartas generadas se entrena la máquina de aprendizaje para que pueda analizar patrones y clasificarlos de manera correcta

Paso 1: define la dimensión del conjunto de datos

PASO 1: Definir el set de datos.

cuantas cartas de cada tipo de falla ha de tener el set de datos?

Carga las cartas para armar una matriz de datos que será usada para entrenamiento y prueba

Paso 2: define el porcentaje en que se divide el conjunto de datos: una parte para entrenamiento y la otra para prueba

PASO 2: Definir el set de entrenamiento

Cuantos ejemplos para el set de entrenamiento? (en porcentaje)

Paso 3: Se dan las opciones de maquina a entrenar (El sistema está enfocado a la máquina de soporte vectorial, sin embargo se presentan 2 opciones mas, clasificación por árbol de decisión y Naive Bayes Gaussiano)

PASO 3: Elegir Maquina de Aprendizaje y entrenar

Que tipo de maquina de aprendizaje quiere entrenar?

- 1)Maquina Soporte Vectorial
- 2)Clasificador por arbol de decision
- 3) Naive Bayes Gaussiano



Paso 4: Acá se proponen 4 maneras de probar el acierto de la maquina escogida, se arroja un porcentaje de éxito y la matriz de confusión correspondiente

```
PASO 4: Validar Resultados

que tipo de validacion?
1) Simple
2) Cruzada
3) Uno afuera
4) Bootstrap
```

Ejemplo de salida:

```
Bootstrap numero 1 de 5
a) Acierto pesimista
Matriz de Confusión:
[[42  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2 26  0  0  0  0  0  0  0  0  0  0  7]
 [ 0  0 37  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0 37  0  0  0  0  0  0  0  0  4]
 [ 0  0  0  0 34  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 48  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 39  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0 41  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  2 35  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 34  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 36  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 23  0  0]
 [ 0  4  0  0  0  0  0  0  0  0  0 29]]
```

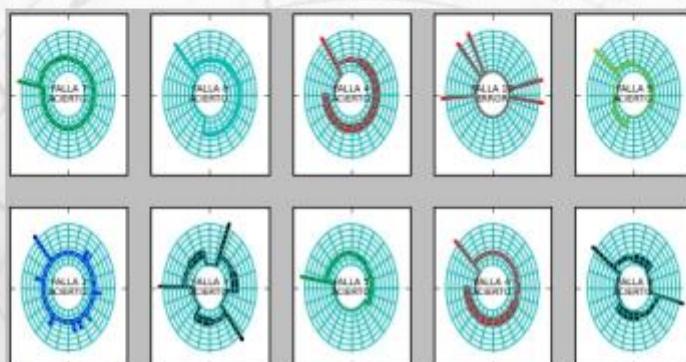
Acierto: de un 96.04166666666667 %

10
AÑOS

OPCIÓN 5: Estimar tipo de falla con el modelo de Reconocimiento

Una vez entrenada la máquina y calculado su acierto se estiman 32 valores obtenidos de manera aleatoria (se puede estimar cualquier carta y cantidad)

Extracto de la salida:



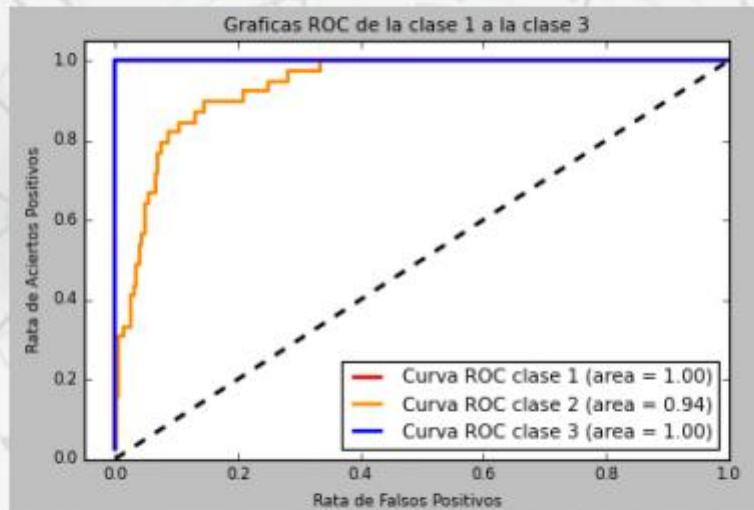
```
[ '1) Falla 11.0' '2) Falla 4.0' '3) Falla 11.0' '4) Falla 7.0'
 '5) Falla 6.0' '6) Falla 4.0' '7) Falla 11.0' '8) Falla 9.0'
 '9) Falla 2.0' '10) Falla 7.0' '11) Falla 6.0' '12) Falla 2.0'
 '13) Falla 3.0' '14) Falla 5.0' '15) Falla 4.0' '16) Falla 8.0'
 '17) Falla 3.0' '18) Falla 11.0' '19) Falla 13.0' '20) Falla 1.0'
 '21) Falla 7.0' '22) Falla 12.0' '23) Falla 13.0' '24) Falla 2.0'
 '25) Falla 9.0' '26) Falla 2.0' '27) Falla 9.0' '28) Falla 6.0'
 '29) Falla 11.0' '30) Falla 9.0' '31) Falla 4.0' '32) Falla 2.0']
```

10
AÑOS

OPCIÓN 6: Gráficas ROC para Maquina Soporte Vectorial

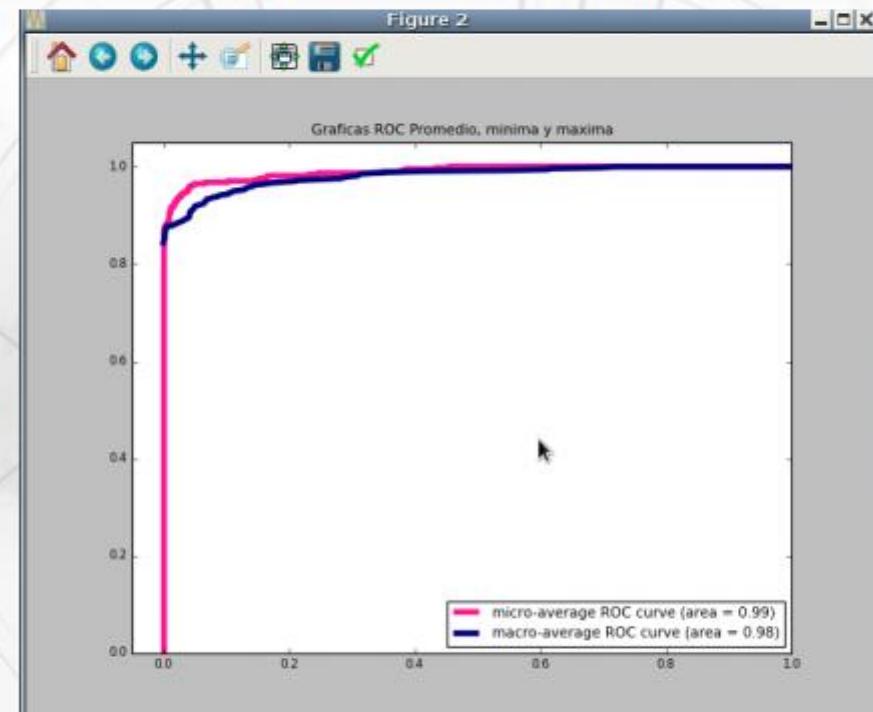
Con la última maquina entrenada (siempre que sea de soporte vectorial) muestra las gráficas ROC de cada clase y el promedio de las mismas, estas graficas sirven para evaluar el desempeño del clasificador

Ejemplo de gráfica ROC por clase



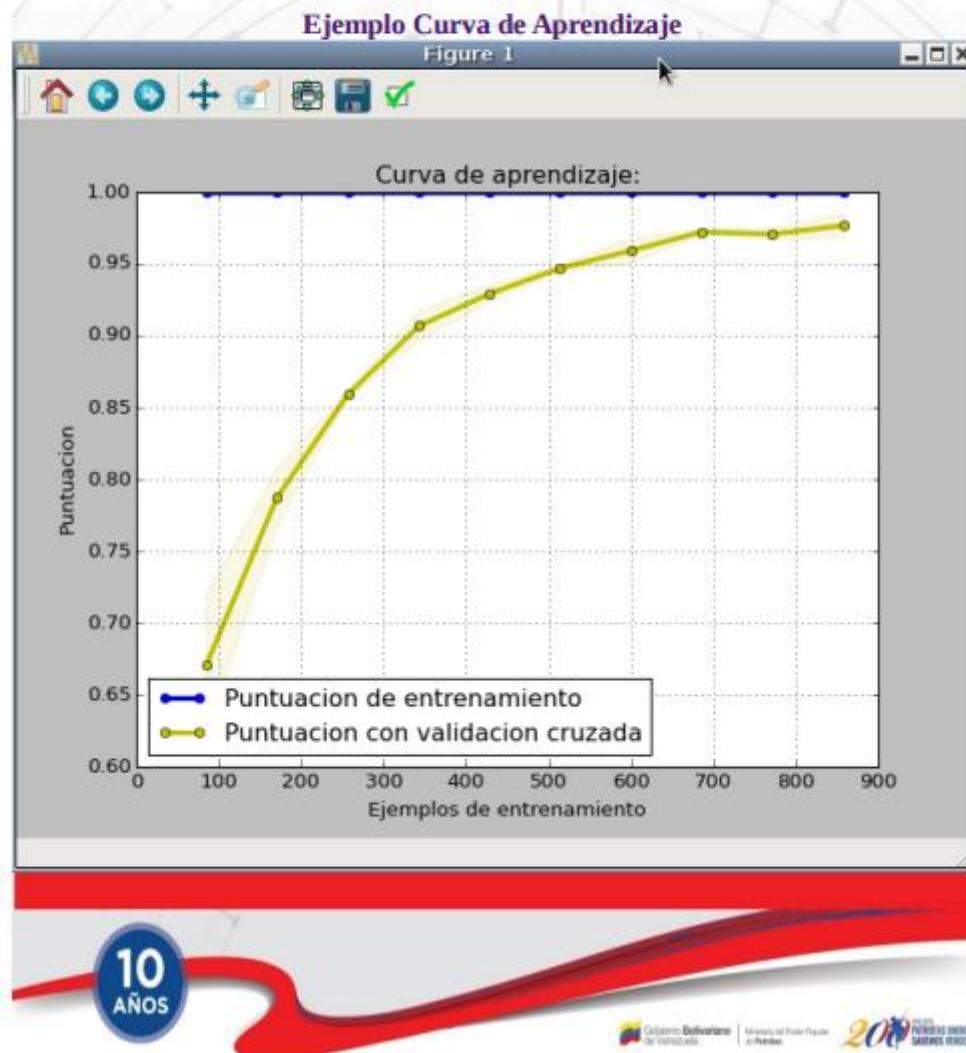
10
AÑOS

Ejemplo de gráfica ROC Promedio

10
AÑOS

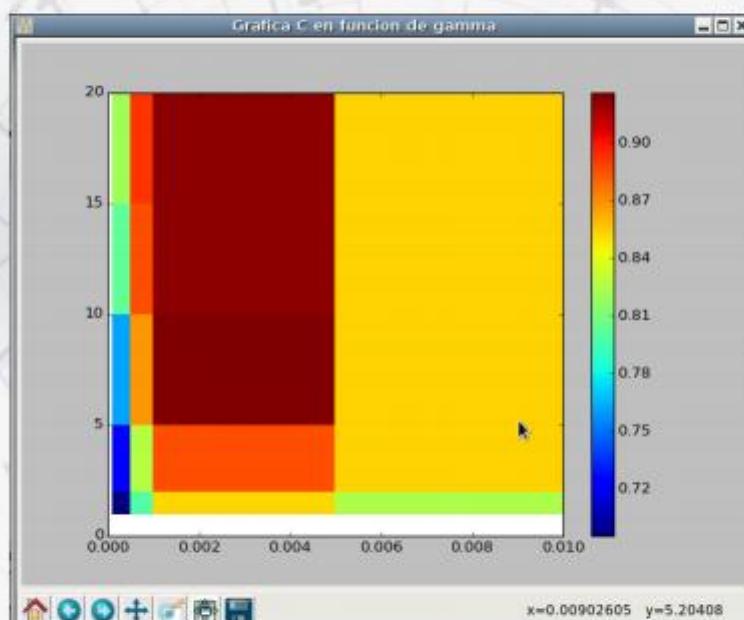
OPCIÓN 7: Curva de aprendizaje

Indica cómo evoluciona la máquina en función al número de ejemplos que se toman para el entrenamiento, usa la última máquina entrenada para la curva



OPCIÓN 8: Calcular parámetros C y gamma para MSV

Calcula los valores óptimos de C y gamma para entrenar la máquina de aprendizaje en base a un conjunto de valores definidos por el programador, también arroja una gráfica que me relaciona los aciertos con cada posible combinación de C vs gamma

Ejemplo gráfico acierto C vs Gamma**OPCIÓN 9: Salir**



DIRECCIÓN EJECUTIVA DE AUTOMATIZACIÓN, INFORMÁTICA Y TELECOMUNICACIONES
GERENCIA DE DESARROLLO SOCIALISTA PRODUCTIVO
DISTRITO SOCIALISTA TECNOLÓGICO

CUALQUIER DUDA O SUGERENCIA ESCRIBIR AL CORREO:

angel.duarte@unet.edu.ve

**10
AÑOS**

Gobierno Bolivariano de Venezuela | Ministerio del Poder Popular para la Defensa | 20 AÑOS FABRICAS UNIDAS SABERES VENEC