



CIATEC

ANÁLISIS DE LA DISTRIBUCIÓN Y VENTA DE LOS PRODUCTOS DE
LA PURIFICADORA “LA NORIA” MEDIANTE LA CADENA DE VALOR,
DATA MINING Y BIN PACKING.

Tesis

QUE PARA OBTENER EL GRADO ACADEMICO DE

Maestro en Ciencia y Tecnología
en la Especialidad de Ingeniería
Industrial y de Manufactura

PRESENTA

Yazmani García Ortiz.



CIENCIA Y TECNOLOGIA

León, Guanajuato, México Octubre del 2010

RESUMEN.

La Minería de Datos así como los Algoritmos Evolutivos son herramientas importantes usadas para la resolución de diferentes problemas. Los principios de dichas herramientas parten de las ideas del modelo de evolución natural que fue propuesto por Charles Darwin y han sido implementados dentro del cómputo evolutivo para abordar problemas complejos de búsqueda y optimización que surgen en las ingenierías y los campos científicos como problemas de planificación de tareas, horarios, tráfico aéreo y ferroviario, búsqueda de caminos óptimos, acomodo de objetos entre otros.

Dentro del rubro del agua purificada un punto importante es determinar las rutas de distribución y venta así como la forma de empacar sus productos, estas se determinan generalmente en base a la una herramienta importante como lo es la Logística, la cual ha sido muy utilizada al grado que dentro de las empresas existen áreas específicamente para desarrollarla, sin embargo esto implica costos que sobre todo para las Pymes, es difícil solventar. Derivado de esto, encontramos que no hay otra forma de establecer el ruteo y mucho menos en una purificadora de agua.

Surge así la oportunidad de la creación de nuevas herramientas utilizando dichas heurísticas, que tenga la capacidad de realizar esta actividad basándose exclusivamente en la ideología de la Logística mas no así en su formulación.

En esta tesis se propone la combinación de la Minería de Datos con Algoritmos Culturales para determinar las rutas óptimas así como la carga apropiada para los vehículos de distribución de la purificadora “La Noria”.

Finalmente los resultados obtenidos dan prueba de que estas herramientas, pueden suplir en forma práctica y económica esta área de la Logística.

Índice

CAPÍTULO 1

1	Introducción.....	2
1.2	Planteamiento del Problema.....	4
1.3	Estado del Arte.....	5
1.3.1	Trabajo Previo de la Minería de Datos y Cómputo Evolutivo.....	6
1.3.2	Aplicación de la Minería de Datos.....	7
1.3.2.1	En el Gobierno.....	7
1.3.2.2	En la Empresa.....	8
1.3.2.3	En la Universidad.....	10
1.3.2.4	En Investigaciones Especiales.....	10
1.3.2.5	En Clubes Deportivos.....	11
1.3.3	Trabajo Previo de Algoritmos Evolutivos.....	12
1.3.3.1	Esquema General de Algoritmos Evolutivos.....	13
1.4	Marco Teórico.....	18
1.4.1	Herramientas de Calidad.....	18
1.4.2	Lluvia de Ideas.....	19
1.4.3	Diagrama Ishikawa.....	20
1.4.4	Interpretación de Diagrama Ishikawa.....	21
1.4.5	Cadena de Valor.....	22
1.4.6	Minería de Datos.....	24
1.4.7	K-means.....	27
1.4.8	Optimización de Espacio.....	28
1.4.9	Computación Evolutiva.....	31

1.4.10	Elitismo dentro de los Algoritmos Evolutivos.....	32
1.4.11	Algoritmos Culturales.....	32
1.4.12	Logística.....	36
1.4.13	Problema de Ruteo de Vehículos.....	39
1.4.14	Diseño de Experimentos (Análisis Experimental de los Algoritmos).....	40
1.4.14.1	Objetivo de Diseño de Experimentos.....	40
1.4.14.2	Interpretación y Reporte de los Resultados.....	41

CAPÍTULO 2

2.1	Justificación.....	43
2.2	Objetivos Generales.....	44
2.3	Objetivos Particulares.....	44
2.4	Hipótesis.....	45

CAPÍTULO 3

3.1	Desarrollo de la Investigación.....	47
3.2	Diagrama Causa Efecto.....	48
3.3	Aplicación de la Cadena de Valor.....	49
3.4	Aplicación de la Minería de Datos.....	49
3.4.1	Matriz de Datos.....	51
3.4.2	Matriz de Centróides.....	52
3.4.3	Matriz de Distancias.....	52
3.4.4	Matriz de Pertenencias.....	53

3.5	Aplicación del Algoritmo Cultural.....	54
3.6	Presentación de la Interfaz.....	56
3.7	Aplicación del Bin Parking.....	61
3.8	Resumen de Ambas Herramientas.....	63

CAPÍTULO 4.

4.1	Análisis de los Resultados.....	66
4.2	Resultados de la Minería de Datos.....	66
4.3	Resultados del Bin Parking.....	69
4.4	Diseño de Experimentos para las herramientas desarrolladas.....	73
4.5	Resumen de Resultados de DOE para Minería de Datos.....	88
4.5.1	Respecto al K-means.....	88
4.5.2	Respecto al Algoritmo Cultural.....	89
4.6	Comparativa del Algoritmo propuesto contra otro similar.....	89
4.6.1	Plan de Prueba.....	90

CAPÍTULO 5

5.1	Conclusiones y Futuros Trabajos.....	96
5.2	Bibliografía.....	99
5.3	Anexos.....	104

Índice de Figuras.

Fig. 1.- Pseudocódigo posible de un algoritmo evolutivo.....	14
Fig. 2.- Diagrama conceptual de mejora continua.....	19
Fig. 3.- Metodología Diagrama Ishikawa.....	21
Fig. 4.- Diagrama de la Cadena de Valor.....	23
Fig. 5.- Relación entre dato, información y conocimiento.....	24
Fig. 6.- Diagrama Conceptual de Minería de Datos (Data Mining).....	26
Fig. 7.- Descripción del problema de Bin Packing.....	30
Fig. 8.- Pseudocódigo de un algoritmo cultural.....	34
Fig. 9.- Diagrama Conceptual de los Algoritmos Culturales.....	35
Fig. 10.- Logística en el Servicio.....	37
Fig. 11.- Diagrama Causa Efecto para Mala Distribución y Venta nos muestra que la principal falla detectada por el personal se encuentra en los métodos.....	48
Fig. 12.- Determinación de la Cadena de valor en la que forma parte la purificadora y la parte concerniente a la distribución.....	49
Fig. 13.- Mapa de Fresnillo dividido en Cuatro Cuadrantes para el agrupamiento de clientes.....	50
Fig. 14.- Colonias por zona acorde a sus respectivas coordenadas.	50
Fig. 15.- Algoritmo Cultural Propuesto.....	55
Fig. 16.- Función del Algoritmo Cultural.....	56

Fig. 17.- Ícono de Vendedor.....	57
Fig. 18.- Ícono del Cliente.....	58
Fig. 19.- Ícono del Producto.....	59
Fig. 20.- Ícono de Visitas.....	59
Fig. 21.- Ícono de Ventas.....	60
Fig. 22.- Ícono Para Actualizar Porcentajes.....	61
Fig. 23.- Determinación de Ruta.....	66
Fig. 24.- Pantalla de resultados de la ejecución del sistema.....	67
Fig. 25.- Gráfico de convergencia de herramienta de Minería de Datos.....	68
Fig. 26.- Estado de Resultados 1er Cuatrimestre 2009.....	71
Fig. 27.- Estado de Resultados 1er Cuatrimestre 2010.....	72
Fig. 28.- Combinación de Agentes e Iteraciones para lograr la combinación optima de ejecución del software de Minería de Datos.....	74
Fig. 29.- Gráfico de Dispersión de distancias mayores y menores de prueba.....	76
Fig. 30.- Gráfico de secuencia cronológica de la distancia menor.....	76
Fig. 31.- Gráfico de secuencias cronológicas de distancias mayores.....	77
Fig. 32.- Gráfico de Dispersión de las distancias menores.....	78
Fig. 33.- Diagrama de Caja.....	78
Fig. 34.- Gráfico de Probabilidad Normal para Menor Distancia.....	79
Fig. 35.- Gráfico de Probabilidad Normal para Ruta A.....	83

Fig. 36.- Gráfico de Dispersión para Ruta A.....	83
Fig. 37.- Gráfico de Secuencias Cronológicas.....	84
Fig. 38.- Gráfico de Probabilidad Normal para Ruta B.....	86
Fig. 39.- Grafico de Dispersión para Ruta B.....	87
Fig. 40.- Grafico de secuencia cronológica para Ruta B.....	87
Fig. 41.-Análisis comparativo de un Algoritmo Genético y un Algoritmo Cultural, para el Algoritmo de Bin Packing para una instancia de 3 camionetas considerando su desempeño con base en la Tabla 12.....	91
Fig. 42.- Comparativo para 10 corridas de presentación de 20lts entre el algoritmo cultural y el genético.....	94

Índice de Tablas.

Tabla 1.- Descripción del Producto.....	62
Tabla 2. Metodología utilizada para el desarrollo de la investigación.....	64
Tabla 3. Comparativo de Acomodo de Producto.....	69
Tabla 4. Acomodo de producto en camionetas con panel.....	70
Tabla 5 Resultados totales de tablas de prueba.....	75
Tabla 6.- Resultados de corridas de software de Minería de Datos para 5 agentes y 2 iteraciones.....	80
Tabla 7.- Resultados de corridas de software de Minería de Datos para 850 agentes y 850 iteraciones.....	80
Tabla 8.- Frecuencias para Ruta A.....	81
Tabla 9.- Resumen Estadístico para Ruta A.....	82
Tabla 10.- de Frecuencias para Ruta B.....	85
Tabla 11.- Resumen Estadístico para Ruta B.....	85
Tabla 12.-Indicadores propuestos en el Trabajo de Cruz [3] y que sirvieron de comparativa para el análisis de resultados de ambos algoritmos desarrollados.....	90
Tabla 13.- Comparativo Genético vs Cultural.....	93

CAPÍTULO 1

1 Introducción.

La Purificadora “La Noria” es una empresa situada en Fresnillo Zacatecas, el cuál es el municipio más grande y con la mayor actividad económica, se ubica a 60 km al noreste de la capital, su población se dedica principalmente a la minería, comercio e industria.

El rubro de la purificación de agua, se encuentra muy competido, esto debido a varios factores, primeramente las grandes empresas que cuentan con un gran poder mercadológico y por tanto de ventas; así mismo, las llenadoras, lugares a los cuales la gente acude a llenar su garrafón por costos muy bajos sin constatar la calidad del agua y finalmente las empresas locales con mayor antigüedad y que cuentan también con un sistema de reparto y arraigo dentro del municipio. Por tanto, el buscar posicionar el producto, implica una mayor complejidad por lo ya expuesto.

Otro factor importante es la falta de concientización de las personas para utilizar agua purificada, ya que mucha gente prefieren tomarla directamente del grifo ya que no aceptan el cambio hacia el agua purificada ya sea por costumbre o por falta de recurso.

Por tanto en esta tesis se ataca el problema de Ruteo y Acomodo de producto dentro de una purificadora de agua, buscando así la solidificación de la misma, para lo cual, se utiliza una metodología basada Minería de Datos en conjunto con los algoritmos evolutivos con los cuales se generaron herramientas de trabajo, las cuales se van desglosando una a una para que al final la conjunción de las mismas nos den la solución del problema.

La Minería de Datos (*Data Mining*) es una técnica que reúne las ventajas de varias áreas como Estadística, Inteligencia Artificial, Regresión Lineal, entre otras, usando como materia prima las bases de datos; una definición tradicional es la siguiente: *“Conjunto de técnicas o herramientas aplicadas al proceso no trivial de extraer y presentar el proceso implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de*

grandes conjuntos de datos, con el objeto de predecir de forma automatizada tendencias y comportamientos previamente desconocidos” [1, 5, 8,14].

Esta tesis está dividida en 5 capítulos, los cuales se conforman de la siguiente manera:

En el capítulo 1 se describe el problema y con ello se habla acerca del estado del arte en base a trabajos realizados en base a Minería de Datos y Algoritmos Evolutivos que se relacionan con la investigación presentada en esta tesis. Se proporcionan los conceptos básicos utilizados en nuestra investigación y finalmente, se describen detalladamente las herramientas utilizadas para la resolución de nuestro problema.

En el capítulo 2 se mencionan los argumentos que dan sustento a la realización de este proyecto, los principales objetivos tanto generales como particulares y las hipótesis que se plantearon para solucionar el problema.

En el capítulo 3 se explican las técnicas propuestas, en primer plano, se hizo uso de algunas herramientas de calidad y se describe la cadena de valor de la cual forma parte la purificadora de agua así como las herramientas basadas en Minería de Datos en combinación con el Cómputo Evolutivo, en este caso los Algoritmos Culturales.

En el capítulo 4 se realiza un análisis detallado de los resultados de los programas para la elaboración de rutas y empaque basado en minería de datos, mostrándose un antes y un después. Finalmente se realiza diseño de experimentos sobre nuestro software de Minería de Datos y un comparativo del Bin Packing.

En el capítulo 5 se dan las conclusiones del trabajo realizado y se detalla el posible trabajo futuro a realizar.

1.2 Planteamiento del Problema.

La ciudad de Fresnillo cuenta con un clima semiseco con lluvias en verano e invierno, lo cual teóricamente propicia un lugar apropiado para la venta y consumo de agua purificada. Son bien sabidos los problemas de escases de agua potable, esto debido a la sobreexplotación de pozos los cuales no son suficientes para satisfacer a la ciudad en constante crecimiento, así como endeudamientos de la comisión de agua, aunado a la contaminación de mantos acuíferos que se tienen en la región por motivos de las compañías mineras establecidas.

Con lo anterior se podría intuir que el negocio de la distribución y venta de agua purificada es muy redituable, sin embargo al implementar vehículos de distribución dentro de la purificadora, se encontraron diferentes factores que no se tenían contemplados y que impidieron el rápido desarrollo de la misma, lo cual llevo a la formulación de 2 preguntas:

¿Por qué no se ha satisfecho el servicio de reparto del agua purificada, en base a las características de compra que tiene la población de Fresnillo y comunidades circunvecinas?

¿Cómo se debe optimizar la red de distribución, así como la optimización del espacio con que se cuenta, para hacer más redituable el producto?

La Purificadora “La Noria” a lo largo de su historia había tenido ventas aceptables, por lo menos así lo consideraban sus empresarios, ya que no se contaba con vehículos de distribución, sin embargo, con la implementación de los mismos lejos de incrementarse fue en decremento, esto por diferentes motivos:

- La falta de experiencia para elaborar rutas de distribución: Lo anterior sucedía en consecuencia de que no se tenían rutas establecidas, los pocos clientes que se tenían no se les daba el trato adecuado, es decir, eran descuidados ya que las visitas eran muy poco frecuentes o bien,

sobre visitados aunado a ello el consumo de gasolina se incrementaba en tanto que las ventas eran mínimas.

- La competencia agresiva y voraz por parte de las diferentes empresas ya establecidas y con un amplio mercado cubierto.
- Rotación de personal, ya que al trabajar bajo comisión muchos deciden dejar el empleo y otros asignándoles un sueldo se conforman con el mismo y el desempeño y venta no son los deseables.
- Falta de recurso para implementar un programa mercadológico adecuado para dar a conocer el producto.

En relación a la problemática antes mencionada, se buscaron las siguientes soluciones:

- La ayuda de personal especializado para el desarrollo de su empresa.
- Estructuración de rutas de distribución para el reparto oportuno del producto.
- La búsqueda de nuevos clientes y de esta manera, la ampliación de la cobertura de venta.
- Reestructuración en el sistema de ventas.

1.3 Estado del Arte.

La Ingeniería Industrial tiene por objetivo el diseño, el mejoramiento, la instalación y el perfeccionamiento de sistemas integrados por personas, materiales, equipo, recursos financieros y de información, que den una solución adecuada a necesidades reales que presenta la sociedad. Este objetivo se logra por el uso de destrezas y conocimientos especiales, relacionado a las ciencias matemáticas, físicas y sociales, junto a los principios de diseño y análisis propios de la ingeniería, lo que permite especificar, predecir y evaluar los resultados que se obtendrán de tales sistemas, implementándose así los resultados.

Las probabilidades de éxito en la toma de decisiones dependen, en gran parte, de la calidad de la información manejada. Desde la perspectiva estratégica, es imprescindible conocer la información sobre los competidores y la posición de la empresa para el desarrollo de ventajas competitivas sostenibles. De ahí que el análisis de la cadena de valor se presente como una herramienta valiosa y al alcance incluso de las pequeñas y medianas empresas para el proceso de diseño de estrategias, ya que aporta información cuantitativa y cualitativa sobre la empresa.

Ciertamente, dentro del sector industrial las PYMES presentan una serie de características que hacen imprescindible como alternativa para el desarrollo de un país, al contribuir con la tasa de crecimiento económico.

1.3.1 Trabajo Previo en Minería de Datos y Cómputo Evolutivos.

En la actualidad todas las empresas generan grandes cantidades de datos, los cuales son almacenados en grandes bases denominados repositorios o Data Warehouses; sin embargo, no siempre es útil o no se sabe explotar.

Para lograr el descubrimiento de información oculta que aporte para la solución de las problemáticas de las empresas, es posible gracias a la Minería de Datos, que entre otras sofisticadas técnicas aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad, pero es el descubrimiento del conocimiento (KDD, por sus siglas en inglés) que se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados [1].

Así el valor real de los datos reside en la información que se puede extraer de ellos, para ayudar a tomar decisiones o mejorar nuestra comprensión de los fenómenos que nos rodean. Hoy, más que nunca, los métodos analíticos

avanzados son el arma secreta de muchos negocios exitosos. Empleando métodos analíticos avanzados para la explotación de datos, los negocios incrementan sus ganancias, maximizan la eficiencia operativa [32], reducen costos e incrementan la satisfacción al cliente.

1.3.2 Aplicaciones de la Minería de Datos.

Las aplicaciones de la Minería de Datos año con año se va incrementando, surgen nuevas aplicaciones sobre este campo emergente que combina técnicas de aprendizaje-máquina, reconocimiento de patrones, estadística, bases de datos y visualización para extraer automáticamente conceptos interrelacionados, y patrones de interés desde grandes bases de datos. [8]:

1.3.2.1 En el Gobierno:

El FBI analiza las bases de datos comerciales para detectar terroristas. A principios del mes de julio de 2002, el director del Federal Bureau of Investigation (FBI), John Aschcroft, anunció que el Departamento de Justicia comenzará a introducirse en la vasta cantidad de datos comerciales referentes a los hábitos y preferencias de compra de los consumidores, con el fin de descubrir potenciales terroristas antes de que ejecuten una acción [5].

Se busca que con esta información, el FBI unirá todas las bases de datos probablemente mediante el número de la Seguridad Social y permitirá saber si una persona fuma, qué talla y tipo de ropa usa, su registro de arrestos, su salario, las revistas a las que está suscrito, su altura y peso, sus contribuciones a la Iglesia, grupos políticos u organizaciones no gubernamentales, sus enfermedades crónicas (como diabetes o asma), los libros que lee, los productos de supermercado que compra, si tomó clases de vuelo o si tiene cuentas de banco abiertas, entre otros. La inversión inicial ronda los setenta

millones de dólares estadounidenses para consolidar los almacenes de datos, desarrollar redes de seguridad para compartir información e implementar nuevo software analítico y de visualización.

1.3.2.2 En la Empresa:

- Detección de fraudes en las tarjetas de crédito.

En 2001, las instituciones financieras a escala mundial perdieron más de 2.000 millones de dólares estadounidenses en fraudes con tarjetas de crédito y débito. El Falcon Fraud Manager es un sistema inteligente que examina transacciones, propietarios de tarjetas y datos financieros para detectar y mitigar fraudes. En un principio estaba pensado, en instituciones financieras de Norteamérica, para detectar fraudes en tarjetas de crédito. Sin embargo, actualmente se le han incorporado funcionalidades de análisis en las tarjetas comerciales, de combustibles y de débito [14].

El sistema Falcón ha permitido ahorrar más de seiscientos millones de dólares estadounidenses cada año y protege aproximadamente más de cuatrocientos cincuenta millones de pagos con tarjeta en todo el mundo, aproximadamente el 65% de todas las transacciones con tarjeta de crédito.

- Descubriendo el porqué de la deserción de clientes de una compañía operadora de telefonía móvil.

Este estudio fue desarrollado en una operadora española que básicamente situó sus objetivos en dos puntos: el análisis del perfil de los clientes que se dan de baja y la predicción del comportamiento de sus nuevos clientes. Se analizaron los diferentes históricos de clientes que habían abandonado la operadora (12,6%) y de clientes que continuaban con su servicio (87,4%). También se analizaron las variables personales de cada cliente (estado civil, edad, sexo, nacionalidad, entre otros.).

De igual forma se estudiaron, para cada cliente, la morosidad, la frecuencia y el horario de uso del servicio, los descuentos y el porcentaje de llamadas locales, interprovinciales, internacionales y gratuitas. Al contrario de lo que se podría pensar, los clientes que abandonaban la operadora generaban ganancias para la empresa; sin embargo, una de las conclusiones más importantes radicó en el hecho de que los clientes que se daban de baja recibían pocas promociones y registraban un mayor número de incidencias respecto a la media. De esta forma se recomendó a la operadora hacer un estudio sobre sus ofertas y analizar profundamente las incidencias recibidas por esos clientes. Al descubrir el perfil que presentaban, la operadora tuvo que diseñar un trato más personalizado para sus clientes actuales con esas características. Para poder predecir el comportamiento de sus nuevos clientes se diseñó un sistema de predicción basado en la cantidad de datos que se podía obtener de los nuevos clientes comparados con el comportamiento de clientes anteriores [11].

- Hábitos de compra en supermercados.

Un estudio muy citado detectó que los viernes había una cantidad inusualmente elevada de clientes que adquirirían a la vez pañales y cerveza. Se detectó que se debía a que dicho día solían acudir al supermercado padres jóvenes cuya perspectiva para el fin de semana consistía en quedarse en casa cuidando de su hijo y viendo la televisión con una cerveza en la mano. El supermercado pudo incrementar sus ventas de cerveza colocándolas próximas a los pañales para fomentar las ventas compulsivas [42].

- Prediciendo el tamaño de las audiencias televisivas.

La British Broadcasting Corporation (BBC) del Reino Unido emplea un sistema para predecir el tamaño de las audiencias televisivas para un programa propuesto, así como el tiempo óptimo de exhibición. El sistema utiliza redes neuronales y árboles de decisión aplicados a datos históricos de la cadena para determinar los criterios que participan según el programa que hay que presentar. La versión final se desempeña tan bien como un experto humano con la ventaja de que se adapta más fácilmente a los cambios porque es constantemente reentrenada con datos actuales [22].

1.3.2.3 En la Universidad

- Conociendo si los recién egresados de una universidad llevan a cabo actividades profesionales relacionadas con sus estudios.

Se hizo un estudio sobre los recién titulados de la carrera de Ingeniería en Sistemas Computacionales del Instituto Tecnológico de Chihuahua II, en México. Se quería observar si sus recién titulados se insertaban en actividades profesionales relacionadas con sus estudios y, en caso negativo, se buscaba saber el perfil que caracterizó a los exalumnos durante su estancia en la Institución. El objetivo era saber si con los planes de estudio de la universidad y el aprovechamiento del alumno se hacía una buena inserción laboral o si existían otras variables que participaban en el proceso. Dentro de la información considerada estaba el sexo, la edad, la escuela de procedencia, el desempeño académico, la zona económica donde tenía su vivienda y la actividad profesional, entre otras variables [40].

Mediante la aplicación de conjuntos aproximados se descubrió que existían cuatro variables que determinaban la adecuada inserción laboral, que son citadas de acuerdo con su importancia: zona económica donde habitaba el estudiante, colegio de donde provenía, nota al ingresar y promedio final al salir de la carrera. A partir de estos resultados, la universidad tendrá que hacer un estudio socioeconómico sobre grupos de alumnos que pertenecían a las clases económicas bajas para dar posibles soluciones, debido a que tres de las cuatro variables no dependían de la universidad [38].

1.3.2.4 En Investigaciones Espaciales.

- Proyecto SKYCAT.

Durante seis años, el Second Palomar Observatory Sky Survey (POSS-II) coleccionó tres terabytes de imágenes que contenían aproximadamente dos

millones de objetos en el cielo. Tres mil fotografías fueron digitalizadas a una resolución de 16 bits por píxel con 23.040 x 23.040 píxeles por imagen. El objetivo era formar un catálogo de todos esos objetos. El sistema Sky Image Cataloguing and Analysis Tool (SKYCAT) se basa en técnicas de agrupación (clustering) y árboles de decisión para poder clasificar los objetos en estrellas, planetas, sistemas, galaxias, etc. Con una alta confiabilidad (Fayyad y otros, 1996). Los resultados han ayudado a los astrónomos a descubrir dieciséis nuevos descubrimientos con corrimiento hacia el rojo que los incluye entre los objetos más lejanos del universo y, por consiguiente, más antiguos. Estos descubrimientos son difíciles de encontrar y permiten saber más acerca de los orígenes del universo [22].

1.3.2.5 En los Clubes Deportivos.

- Los equipos de la NBA utilizan aplicaciones inteligentes para apoyar a su cuerpo de entrenadores.

El Advanced Scout es un software que emplea técnicas de Minería de Datos y que han desarrollado investigadores de IBM para detectar patrones estadísticos y eventos raros. Tiene una interfaz gráfica muy amigable orientada a un objetivo muy específico: analizar el juego de los equipos de la National Basketball Association (NBA).

El software utiliza todos los registros guardados de cada evento en cada juego: pases, encestes, rebotes y doble marcaje a un jugador por el equipo contrario, entre otros. El objetivo es ayudar a los entrenadores a aislar eventos que no detectan cuando observan el juego en vivo o en película.

Un resultado interesante fue uno hasta entonces no observado por los entrenadores de los Knicks de Nueva York. El doble marcaje a un jugador puede generalmente dar la oportunidad a otro jugador de encestar más fácilmente. Sin embargo, cuando los Bulls de Chicago jugaban contra los Knicks, se encontró que el porcentaje de encestes después de que al centro de

los Knicks, Patrick Ewing, le hicieran doble marcaje era extremadamente bajo, indicando que los Knicks no reaccionaban correctamente a los dobles marcajes.

Para saber el porqué, el cuerpo de entrenadores estudió cuidadosamente todas las películas de juegos contra Chicago. Observaron que los jugadores de Chicago rompían su doble marcaje muy rápido de tal forma que podían tapar al encestador libre de los Knicks antes de prepararse para efectuar su tiro. Con este conocimiento, los entrenadores crearon estrategias alternativas para tratar con el doble marcaje.

La temporada pasada, IBM ofreció el Advanced Scout a la NBA, que se convirtió así en un patrocinador corporativo. La NBA dio a sus veintinueve equipos la oportunidad de aplicarlo. Dieciocho equipos lo están haciendo hasta el momento obteniendo descubrimientos interesantes [38].

1.3.3 Trabajo Previo de Algoritmos Evolutivos.

La forma más común de tratar este tipo de problemas es mediante Los Algoritmos Evolutivos (AEs), estos son una técnica de resolución de problemas de búsqueda y optimización inspirada en la teoría de la evolución de las especies y la selección natural [4]. Estos algoritmos reúnen características de búsqueda aleatoria con características de búsqueda dirigida que provienen del mecanismo de selección de los individuos adaptados. La unión de ambas características les permite abordar los problemas de una forma muy particular, ya que tienen capacidad para acceder a cualquier región del espacio de búsqueda, capacidad de la que carecen otros métodos de búsqueda exhaustiva, a la vez que exploran el espacio de soluciones de una forma mucho más eficiente que los métodos puramente aleatorios.

Indudablemente, un algoritmo diseñado de forma específica para la resolución de un problema concreto será más eficiente que un algoritmo evolutivo, que es

una técnica general de resolución pero existen muchas situaciones en la que no es posible contar con tales algoritmos.

Los algoritmos evolutivos proporcionan un esquema general para la resolución de problemas es decir, tenemos el algoritmo diseñado para el problema que nos ocupa y sólo tenemos que especificar la forma de ciertos componentes.

Incluso en el diseño de estos componentes, hay patrones de diseño que se pueden aplicar a toda una clase de problemas y que facilitan la construcción del algoritmo evolutivo. En general, la parte más dependiente del problema específico considerando es la definición de la función de adaptación, que se utiliza en el proceso de selección del algoritmo evolutivo.

Estos algoritmos no siempre garantizan una solución exacta al problema abordado, sino una aproximación cuya calidad en el resultado dependerá de los recursos dedicados a la búsqueda, es decir, tiempo y memoria, aparte, claro está, del diseño adecuado de los componentes del algoritmo.

1.3.3.1 Esquema General de un Algoritmo Evolutivo.

Los distintos algoritmos evolutivos que se pueden formular responden a un esquema básico común, y comparten una serie de propiedades:

- Procesan simultáneamente, no una solución al problema, sino todo un conjunto de ellas. Estos algoritmos trabajan con alguna forma de representación de soluciones potenciales al problema, que se denominan individuos. El conjunto de todos ellos forma la población con la que trabaja el algoritmo.
- La composición de la población se va modificando a lo largo de las iteraciones del algoritmo que se denominan generaciones. De generación en generación, además de variar el número de copias de un mismo individuo en la población, también pueden aparecer nuevos

individuos generados mediante operaciones de transformaciones sobre individuos de la población anterior.

- Cada generación incluye un proceso de selección, que da mayor probabilidad de permanecer en la población y participar en las operaciones de reproducción a los mejores individuos. Los mejores individuos son aquellos que dan lugar a los mejores valores (ya sea máximos o mínimos) de la función de adaptación del algoritmo. Es fundamental para el funcionamiento de un algoritmo evolutivo que este proceso de selección tenga una componente aleatoria, de forma que individuos con baja adaptación también tengan oportunidades de sobrevivir, aunque su probabilidad sea menor. Es esta componente aleatoria la que dota a los algoritmos evolutivos de capacidad para escapar de óptimos locales y de explorar distintas zonas del espacio de búsqueda [2].

En el siguiente esquema se muestra un pseudocódigo general posible en un algoritmo evolutivo:

```
funcion Algoritmo_Genético()
{
    TPoblacion pob; //población
    TParametros parámetros// tamaño población

    obtener_parametros(parametros);
    pob = población_inicial();
    evaluación(pob, tam-pob, pos_mejor, sumadaptacion);

    // bucle de evolución
    mientras no se alcanza condición de terminación hacer{
        seleccion (pob, parámetros);
        reproduccion(pob, parámetros);
        evaluación (pob, parámetros, pos_mejor, sumadapatcion);
    }
    devolver pob [pos_mejor]
}
```

Fig. 1 Pseudocódigo de un algoritmo evolutivo [2].

El algoritmo procesa un conjunto de individuos que forman la población *pob*. Al comienzo del algoritmo se obtiene los datos de entrada al problema (*obtener_parámetros*) y se genera la población inicial, cuyos individuos se evalúan mediante la función de adaptación del algoritmo. El resto del algoritmo

consiste en un bucle, cada una de cuyas iteraciones es una generación en la que se produce un proceso de selección, que da mayores probabilidades de tener copias en la nueva población a los individuos más adaptados, seguido de un proceso de reproducción en el que se generan nuevos individuos a partir de los de la población mediante operaciones de mezcla y pequeñas alteraciones, y finalmente una evaluación de la nueva población. En muchas ocasiones se utilizan pequeñas variantes de este esquema. Así, por ejemplo, a veces se selecciona un subconjunto de la población que es el único que participa en las operaciones de reproducción.

Una de las principales propiedades que necesita cumplir un algoritmo evolutivo para producir resultados de calidad es contar con suficiente diversidad en la población. Esta diversidad se refiere tanto a los genotipos de los individuos como a los valores de adaptación que les corresponden. Cuando falta diversidad, los operadores genéticos no son capaces de crear nuevos individuos progenitores. Así mismo, si los valores que toma la función de adaptación son muy similares, la selección no funciona adecuadamente. Si el valor de adaptación de los mejores individuos se diferencia muy poco del resto, la naturaleza probabilística del mecanismo de selección hará que la composición de la población sea aleatoria, sin conseguir favorecer a los más adaptados.

Por otra parte, diferencias demasiado grandes de los valores de adaptación impiden que el mecanismo de selección funcione adecuadamente. El mecanismo de selección hace que los individuos con un valor de adaptación muy superior a la media consigan una cantidad de copias de sí mismos en la siguiente generación muy superior al resto. Esto hace que los restantes individuos tiendan a desaparecer, llegando nuevamente a una situación de falta de diversidad.

En situaciones de falta de diversidad en la población, un algoritmo evolutivo tiende a converger prematuramente. La convergencia que fue introducida por De Jong, se refiere a una evolución en una situación de uniformidad, bien en la composición de los individuos, bien en los valores de adaptación de la

población. Una vez que el algoritmo ha convergido, la exploración del espacio de búsqueda se minimiza o se detiene. Si el algoritmo evolutivo funciona correctamente, debe tender a converger hacia el óptimo de los valores de la función de adaptación.

Siguiendo este esquema general se han desarrollado distintas variantes de algoritmos evolutivos, cuya principal diferencia se encuentra en la representación de los individuos. Los operadores genéticos que se utilizan para la reproducción en cada caso dependen de la representación adoptada. Entre algunas variantes conocidas son las siguientes:

- Los Algoritmos Genéticos [15]: utilizan una representación binaria o entera.
- Los Programas de Evolución [21]: los individuos son cualquier estructura de datos de tamaño fijo.
- La Programación Evolutiva [16,17]: se hace evolucionar “una población de 2 programas” para resolver un problema en general. Los programas que evolucionan pueden adoptar distintas formas, aunque la más usual es un árbol. En cualquier caso, se trata de estructuras de datos de tamaño variable, es decir, no todos los individuos tienen el mismo tamaño. Este tipo de Algoritmo Evolutivo presenta una diferencia fundamental con el resto: no buscan la solución a una instancia concreta de un problema, sino una estrategia capaz de resolver instancia de ese problema.
- Los algoritmos Culturales [35]: son descritos como un conjunto de individuos cada uno de ellos posee su propio comportamiento, y están agrupados en Comunidades, las cuales van cambiando su paradigma que comparten permitiendo adaptarse mejor a su entorno; es decir, a mayor cultura, mejor se resuelve el problema al que tratan de contribuir.

- Las redes neuronales [9]: Las redes neuronales como su nombre lo indica pretenden imitar a pequeñísima escala la forma de funcionamiento de las neuronas que forman el cerebro humano. Todo el desarrollo de las redes neuronales tiene mucho que ver con la neurofisiología, no en vano se trata de imitar a una neurona humana con la mayor exactitud posible. Entre los pioneros en el modelado de neuronas se encuentra Warren McCulloch y Walter Pitts. Estos dos investigadores propusieron un modelo matemático de neurona. En este modelo cada neurona estaba dotada de un conjunto de entradas y salidas. Las redes neuronales como su nombre lo indica pretenden imitar a pequeñísima escala la forma de funcionamiento de las neuronas que forman el cerebro humano. Todo el desarrollo de las redes neuronales tiene mucho que ver con la neurofisiología, no en vano se trata de imitar a una neurona humana con la mayor exactitud posible. Entre los pioneros en el modelado de neuronas se encuentra Warren McCulloch y Walter Pitts. Estos dos investigadores propusieron un modelo matemático de neurona. En este modelo cada neurona estaba dotada de un conjunto de entradas y salidas. Cada entrada está afectada por un peso. La activación de la neurona se calcula mediante la suma de los productos de cada entrada y la salida es una función de esta activación. La principal clave de este sistema se encuentra en los pesos de las diferentes entradas. Como se ha visto, las entradas son modificadas por el peso y las salidas son función de estas modificaciones. Esto nos lleva a concluir que los pesos influyen de forma decisiva en la salida y por lo tanto pueden ser utilizados para controlar la salida que se desea.
- Algoritmo de Hormigas [28]: se inspiran en la estrategia utilizada por las colonias de hormigas para buscar alimentos. Cuando una hormiga se encuentra un camino hacia una fuente de alimento, deposita en el trayecto una sustancia llamada feromona. La cantidad de feromona depositada depende de la longitud del camino y de la calidad del alimento encontrado. Si una hormiga no detecta la presencia de feromona se mueve aleatoriamente; pero si la detecta, decidirá moverse

a los trayectos con mayor cantidad lo que a su vez provocará un aumento de la feromona depositada en esa zona.

1.4 Marco Teórico.

El siguiente capítulo muestra el conjunto de principios teóricos que guían la investigación estableciendo las unidades más relevantes para cada problema a investigado.

1.4.1 Herramientas de Calidad.

Es bien sabido que en el trabajo que se realiza día con día, se requiere resolver las variaciones que van surgiendo en los diferentes procesos de producción: reducir los defectos. El camino hacia la Calidad Total, en cualquiera de sus rubros, requiere el establecimiento de una filosofía de calidad, crear una nueva cultura, mantener un liderazgo, desarrollar al personal y trabajar en equipo, entre otras cosas.

Para resolver estos problemas o variaciones y mejorar la calidad, es necesario basarse en hechos y no dejarse guiar solamente por el sentido común, la experiencia o la audacia.

En la actividad de mejora continua sea cual sea la metodología elegida, siempre podemos identificar una serie de actividades comunes en todas ellas [39]:

Generación de Ideas: Recoger muchas ideas y/o organizarlas.

Análisis de Procesos: Conocer y entender cómo funciona un proceso o una parte del mismo.

Planificación: Planificar que hacer.

Evaluación: Evaluar una situación de partida o una situación final. Seleccionar una opción como la mejor.

Recogida y Análisis de Datos: Recoger datos y extraer información de ellos.



Fig.2.- Diagrama conceptual de mejora continua.

Cuando es necesario expandir el pensamiento para incluir todas las posibles dimensiones de un problema o de su solución y se desea la participación de todos los miembros del grupo, se recomienda la utilización de la Lluvia de Ideas, ya que esta herramienta proporciona gran cantidad de información en corto período de tiempo.

Los pasos para la aplicación de la misma es la siguiente:

1.4.2 Lluvia de Ideas (Metodología).

- 1.- Definir con claridad y precisión el tema.
- 2.- Nombrar moderador.

- 3.- Cada participante elabora lista por escrito sobre el tema.
- 4.- Acomodo en forma circular y debatir sobre cada idea de la lista.
- 5.- El moderador pregunta si hay puntos adicionales.
- 6.- Agrupar ideas similares y elaborar un diagrama Ishikawa.
- 7.- Una vez concluido el debate, se pregunta si se ha omitido alguna idea, de ser así se agrega.
- 8.- Se realiza una discusión abierta y respetuosa para ir eliminando aquellos factores de menor o nula relevancia.
- 9.- Se eligen las causas o ideas más importantes mediante: datos, consenso o votación.
- 10.- Darle énfasis a las acciones tomadas para evitar redundancia de reuniones.

Cabe señalar que la Lluvia de Ideas no proporciona directamente resultados, solamente es un apoyo para la resolución de la problemática tratada y así, tener una lista que sirva como punto de partida para el análisis.

1.4.3 Diagrama Ishikawa o Causa-Efecto.

El diagrama causa-efecto, es una forma de organizar y representar las diferentes teorías propuestas sobre las causas de un problema generado en el proceso de Lluvia de Ideas [7]. Se conoce también como diagrama Ishikawa o diagrama de espina de pescado y se utiliza en las fases de Diagnóstico y Solución de la causa.

Este diagrama tiene como propósito expresar en forma gráfica el conjunto de factores causales que intervienen en una determinada característica de calidad.

La forma de realizarlo es la siguiente:

Para empezar, decide cual característica de calidad, salida o efecto quieres examinar y continua con los siguientes pasos [13]:

- 1.- Decidir la característica de calidad, salida o efecto que se quiere examinar.
- 2.- Dibujar un diagrama en blanco.
- 3.- Escribir de forma breve el problema o efecto.
- 4.- Escribir las categorías que se consideran apropiadas al problema: maquina, mano de obra, materiales, métodos, son los más comunes y aplican en muchos procesos.
- 5.- Realizar una lluvia de ideas.

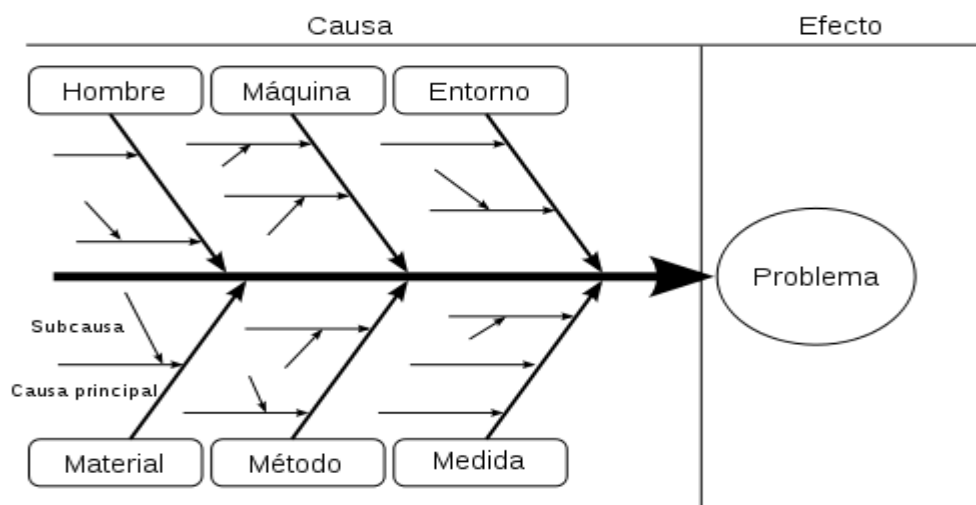


Fig. 3. Metodología Diagrama Ishikawa

- 6.- Preguntar ¿por qué? a cada causa, no más de dos o tres veces.
- 7.- Empezar por enfocar las variaciones en las causas seleccionadas como fácil de implementar y de alto impacto.

1.4.4 Interpretación de un Diagrama Causa-Efecto.

Es un vehículo para ordenar, de forma muy concentrada, todas las causas que supuestamente pueden contribuir a un determinado efecto. Nos permite, por tanto, lograr un conocimiento común de un problema complejo, sin ser nunca sustitutivo de los datos. Es importante ser conscientes de que los diagramas de

causa-efecto presentan y organizan teorías. Solo cuando estas teorías son contrastadas con datos, podemos probar las causas de los fenómenos observables. Errores comunes son construir el diagrama antes de analizar globalmente los síntomas, limitar las teorías propuestas enmascarando involuntariamente la causa raíz, o cometer errores tanto en la relación causal como en el orden de las teorías, suponiendo un gasto de tiempo importante.

1.4.5 Cadena de Valor.

La cadena de valor industrial se define como el conjunto interrelacionado de actividades creadoras de valor, la cual va desde la obtención de fuentes de materias primas, hasta que el producto terminado es entregado al consumidor final, incluyendo actividades post venta en las cuales participan varias empresas [10].

Con lo anterior podemos observar que la mayoría de las empresas son sólo una parte o componente de la cadena de valor; el análisis de la cadena de valor es un método utilizado para descomponer el conjunto en las actividades que lo conforman, con la finalidad de facilitar la toma de decisiones estratégicas, al ubicar a la empresa frente a sus clientes, proveedores y competidores.

Uno de los métodos de análisis estratégico más importante es el denominado Cadena de Valor. Dentro de este método se determina la existencia de vínculos externos o “eslabones verticales”, dicho método fue propuesto por Michel Porter [32] quien afirmó que la forma de operación de los proveedores y canales de distribución afectan el costo y la forma de realizar las operaciones de la empresa, ya que, el nivel de costos y calidad de suministros proporcionados por los proveedores, depende de los costos incurridos y el grado de diferenciación frente a sus clientes. Igualmente, su aceptación en el mercado, depende de los mecanismos de entrega y de los niveles de costos incurridos por los distribuidores [23].

Toda organización requiere clientes si desea sobrevivir y prosperar. Los clientes desean obtener algún tipo de valor de los bienes y servicios que compran o utilizan y estos usuarios finales determinando que tiene valor [35].

Los elementos básicos de una cadena de valor son los siguientes:

ACTIVIDADES PRIMARIAS: Aquella que tiene que ver con el desarrollo del producto, su producción, las logísticas, comercialización y los servicios de post-venta.

ACTIVIDADES SECUNDARIAS: Como lo son la administración de recursos humanos, las de compras de bienes y servicios, desarrollo tecnológico, e infraestructura empresarial.

EL MARGEN: Diferencia entre el valor total y los costos totales incurridos por la empresa para desempeñar las actividades generadoras de valor.

Lo anterior lo podemos ver en forma clara mediante el diagrama propuesto por Michel Porter Figura 4.



Figura 4.- Diagrama de la cadena de valor.

1.4.6 Minería de Datos (Data Mining)

La Minería de Datos surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos. De forma general, los datos son la materia prima bruta. En el momento que el usuario les atribuye algún significado especial pasan a convertirse en información [30]. Cuando los especialistas elaboran o encuentran un modelo, haciendo que la interpretación del confronto entre la información y ese modelo represente un valor agregado, entonces nos referimos al conocimiento.

En la Figura 4, se ilustra la jerarquía que existe en una base de datos entre dato, información y conocimiento. Se observa igualmente el volumen que presenta en cada nivel y el valor que los responsables de las decisiones le dan en esa jerarquía. El área interna dentro del triángulo representa los objetivos que se han propuesto. La separación del triángulo representa la estrecha unión entre dato e información, no así entre la información y el conocimiento. La Minería de Datos trabaja en el nivel superior buscando patrones, comportamientos, agrupaciones, secuencias, tendencias o asociaciones que puedan generar algún modelo que nos permita comprender mejor el dominio para ayudar en una posible toma de decisión [29].

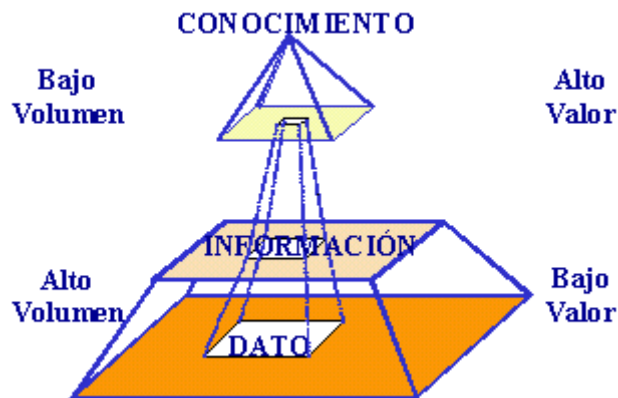


Fig. 5. Relación entre dato, información y conocimiento.

Aunque desde un punto de vista académico el término Minería de Datos es una etapa dentro de un proceso mayor llamado extracción de conocimiento en bases de datos en el entorno comercial, así como en este trabajo, ambos términos se usan de manera indistinta. Lo que en verdad hace la Minería de Datos es reunir las ventajas de varias áreas como la Estadística, la Inteligencia Artificial, Computación Gráfica y el Procesamiento Masivo, principalmente usando como materia prima las bases de datos.

La Minería de Datos es una tecnología compuesta por etapas que integra varias áreas y que no se debe confundir con un gran software. Durante el desarrollo de un proyecto de este tipo se usan diferentes aplicaciones software en cada etapa que pueden ser estadísticas, de visualización de datos o de inteligencia artificial, principalmente [1].

Algunas definiciones tradicionales sobre Minería de Datos son las siguientes: "Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos". Otra nos dice que la Minería de Datos es "la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión" [23].

En la actual sociedad de la información, donde día a día se multiplica la cantidad de datos almacenados casi de forma exponencial, la minería de datos es una herramienta fundamental para analizarlos y explotarlos de forma eficaz para los objetivos de cualquier organización [40]. Recordemos que en la minería, el oro se extrae de grandes cantidades de roca, obteniendo solo pequeñas partes que, finalmente, son las que no interesan por ejemplo (Oro); quizá, el nombre más apropiado para la minería de datos sería "El conocimiento de la Minería de Datos".

La Minería de Datos hace uso de todas las técnicas que puedan aportar información útil, desde un sencillo análisis gráfico, pasando por métodos estadísticos más o menos complejos, complementados con métodos y algoritmos del campo de la inteligencia artificial y el aprendizaje automático que

resuelven problemas típicos de agrupamiento automático, clasificación, predicción de valores, detección de patrones, asociación de atributos, entre otros [10]. Es, por tanto, un campo multidisciplinar que cubre numerosas áreas y se aborda desde múltiples puntos de vista, como la estadística, la informática (cálculo automático) o la ingeniería [37].

Un proceso típico de minería de datos consta de los siguientes pasos generales:

1. **Selección del conjunto de datos**, tanto en lo que se refiere a las variables dependientes, como a las variables objetivo, como posiblemente al muestreo de los registros disponibles.
2. **Análisis de las propiedades de los datos**, en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).
3. **Transformación del conjunto de datos de entrada**, se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema.
4. **Seleccionar y aplicar la técnica de minería de datos**, se construye el modelo predictivo, de clasificación o segmentación.
5. **Evaluar los resultados** contrastándolos con un conjunto de datos previamente reservado para validar la generalidad del molde.

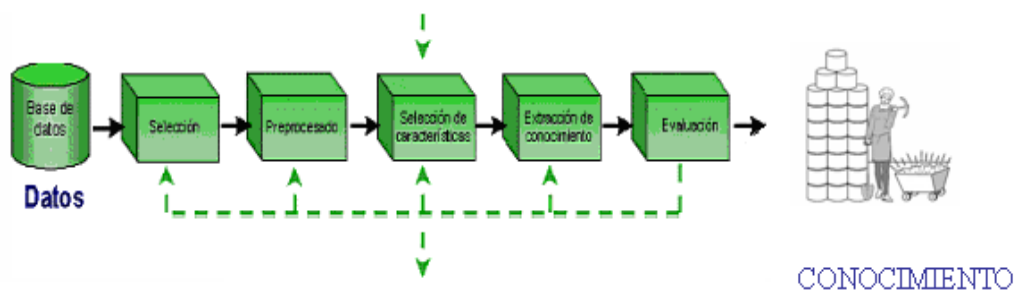


Fig. 6. Diagrama Conceptual de Minería de Datos (Data Mining).

Si el modelo final no superara esta evaluación el proceso se podría repetir desde el principio o, si se cuenta con algún experto y lo considera oportuno, a partir de cualquiera de los pasos anteriores. Esta retroalimentación se podrá repetir cuantas veces se considere necesario hasta obtener un modelo válido.

Una vez validado el modelo, si resulta ser aceptable (proporciona salidas adecuadas y/o con márgenes de error admisibles) éste ya está listo para su explotación. Los modelos obtenidos por técnicas de minería de datos se aplican incorporándolos en los sistemas de análisis de información de las organizaciones, e incluso, en los sistemas transaccionales [14].

Tradicionalmente, las técnicas de minería de datos se aplicaban sobre información contenida en almacenes de datos. De hecho, muchas grandes empresas e instituciones han creado y alimentan bases de datos especialmente diseñadas para proyectos de minería de datos en las que centralizan información potencialmente útil de todas sus áreas de negocio. No obstante, actualmente está cobrando una importancia cada vez mayor la minería de datos desestructurados como información contenida en ficheros de texto, en Internet, etc.

Actualmente existen aplicaciones o herramientas comerciales de la Minería de Datos muy poderosas que contienen un sinfín de utilerías que facilitan el desarrollo de un proyecto. Sin embargo, casi siempre acaban complementándose con otra herramienta que de mayor soporte y confiabilidad [5].

1.4.7 K-means

El K-means es uno de los algoritmos de aprendizaje supervisado más simple que resuelve problemas de agrupamiento. El procedimiento sigue una fácil y sencilla manera de clasificar un determinado conjunto de datos a través de un cierto número de datos fijados a priori. La idea principal es definir n centroides,

uno para cada grupo, los cuales, deben colocarse de tal manera que la ubicación sean diferentes entre si. Por lo tanto, la mejor opción es colocarlos lo más lejos posible entre ellos. El siguiente paso es tomar cada punto que pertenece a un determinado conjunto de datos y asociarlo al centroide más cercano [28].

1.4.8 Optimización de Espacio.

El problema de explicar formalmente porqué un algoritmo sigue cierto comportamiento al resolver un conjunto instancias de un problema se puede describir de la siguiente manera.

Dados:

1. un algoritmo metaheurístico A para resolver un problema P ,
2. un conjunto de instancias $I = \{i_1, i_2, \dots, i_n\}$ del problema P ,
3. un conjunto de métricas que caracterizan factores del problema y del algoritmo que impactan en el desempeño $F = \{f_1, f_2, \dots, f_m\}$.

Se busca:

1. un conjunto $R = \{r_1, r_2, \dots, r_p\}$, para $1 \leq p \leq m$, tal que cada r_i establece una relación causal entre los elementos de G_i , donde $G_i \subseteq F$,
2. aplicar el conocimiento adquirido, a través de R , en el rediseño del algoritmo A .

De tal manera que lo que se busca, es identificar relaciones entre los factores críticos para el desempeño algorítmico que permitan explicar porqué el algoritmo objeto de estudio sigue cierto comportamiento al resolver un conjunto de casos de prueba y aplicar el conocimiento adquirido en la mejora de desempeño del algoritmo [33].

El Problema del (Bin Packing), que en lo siguiente lo llamaremos BPP, es una variante del problema de la mochila; consiste en dado un conjunto de objetos los cuales tienen un peso y un beneficio ó utilidad, se desea encontrar un subconjunto de objetos que maximice el beneficio ó utilidad total de los objetos seleccionados sin sobrepasar la capacidad de los contenedores ó depósitos, ver figura 7, el objetivo de problema puede representarse formalmente a través de la fórmula 1 que se muestra a continuación:

$$\max \sum_{j=1}^n p_j x_j \text{ sujeto a } \sum_{j=1}^n r_j x_j \leq c \quad (1)$$

Donde:

p_j es el beneficio de seleccionar el objeto j ,
 x_j vale 1 ó si elemento j existe en la solución,
 r_j es el espacio que ocupa el objeto,

La sumatoria del peso de los objetos seleccionados debe ser \leq a la capacidad del contenedor C .

El BPP puede ser tratado también como un problema multiojetivo debido a que por un lado se debe maximizar la utilidad de los objetos y por el otro minimizar el peso de los objetos para no sobrepasar la capacidad de los contenedores de tamaño fijo [12].



Fig. 7. Descripción del problema de Bin Packing.

La complejidad del problema de distribución de objetos en contenedores hace prácticamente imposible el uso de algoritmos exactos para su solución. Estos procedimientos incluyen heurísticas sencillas que se diferencian por la manera en que los objetos son tratados antes de ser acomodados y por la forma en que se elige el contenedor que almacenará cada objeto. Las principales estrategias deterministas para BPP son descritas a continuación [31]:

- *Primer Ajuste* (First Fit, FF): Cada objeto considerado es colocado en el primer contenedor que tenga suficiente capacidad disponible. Si ningún contenedor parcialmente lleno puede almacenarlo, el objeto es colocado en un contenedor nuevo (vacío). Una variación a este método se establece cuando los objetos son tomados según el orden decreciente de sus pesos (los objetos son ordenados de mayor a menor peso antes de ser acomodados), dicha variante es conocida como *Primer Ajuste Decreciente* (First Fit Decreasing, FFD).
- *Mejor Ajuste* (Best Fit, BF): Cada objeto es acomodado en el contenedor más lleno que lo pueda almacenar, agregando contenedores cuando sea necesario. De igual manera que con FF, existe una variación, llamada

Mejor Ajuste Decreciente (Best Fit Decreasing, BFD), que considera los objetos en orden decreciente de sus pesos.

- *Peor Ajuste* (Worst Fit, WF): Contrario a la estrategia anterior (Best Fit, BF), cada objeto en consideración es almacenado en el contenedor menos lleno con capacidad residual suficiente para contenerlo. La variante que toma los objetos según el orden decreciente de sus pesos es denominada *Peor Ajuste Decreciente* (Worst Fit Decreasing, WFD).
- *Best 3-Fit Decreasing* (B3FD) [34]: Inicialmente se abre un número límite de contenedores. Si existe un contenedor vacío, se selecciona y se coloca el objeto actual, de otro modo, se intenta llenar cada contenedor exactamente con objetos que no han sido seleccionados y por pares suman la capacidad residual del contenedor. Para el resto de los objetos, el elemento actual es insertado en el contenedor más lleno en el que ajuste (como en BF). Si no existe contenedor con capacidad suficiente un nuevo contenedor es agregado a la solución.

1.4.9 Computación Evolutiva

La computación evolutiva está basada en las ideas de la selección natural, y su aplicación en un ambiente artificial [4]. La selección natural es vista como un proceso de optimización, en el que paulatinamente los individuos de la población se van mejorando para adaptarse a su medio.

Para la computación evolutiva, un individuo es una solución potencial a un problema, codificada de acuerdo con el funcionamiento del algoritmo; y el medio donde se desenvuelve lo componen la función objetivo y las restricciones, las que nos dirán qué tan apto es el individuo para sobrevivir.

La computación evolutiva involucra métodos que son poblacionales, lo que significa que trabajan con varias soluciones a la vez, y no con una, como lo

hacen la mayoría de las otras heurísticas (por ejemplo, el recocido simulado), con lo que evitan quedar atrapados en óptimos locales.

Durante la ejecución de un algoritmo evolutivo, a la población se le aplican algunos operadores probabilísticos, con lo que se obtienen nuevas soluciones, las cuales se conservan o se descartan mediante un mecanismo de selección. Este proceso se repite por un número determinado de iteraciones, que en el contexto de computación evolutiva se denominan generaciones. El número de generaciones puede ser dado por el usuario o definido por el propio algoritmo [2].

1.4.10 Elitismo Dentro de los Algoritmos Evolutivos.

El elitismo consiste en asegurar la supervivencia de los mejores individuos de la población. El elitismo acelera la convergencia de funciones unimodales, es decir, con un único valor óptimo, y por tanto relativamente sencillas. Por lo tanto, debemos utilizar el elitismo cuidadosamente, teniendo en cuenta las características del problema [6].

1.4.11 Algoritmos Culturales.

Los algoritmos culturales fueron desarrollados por Robert G. Reynolds [25,26], como un complemento a la metáfora que usan los algoritmos de computación evolutiva, que se habían concentrado en conceptos genéticos, y de selección natural [33].

Los algoritmos culturales están basados en las teorías de algunos sociólogos y arqueólogos, que han tratado de modelar la evolución cultural. Tales investigadores indican que la evolución cultural puede ser vista como un proceso de herencia en dos niveles: el nivel micro-evolutivo, que consiste en el

material genético heredado por los padres a sus descendientes, y el nivel macro-evolutivo, que es el conocimiento adquirido por los individuos a través de las generaciones, y que una vez codificado y almacenado, sirve para guiar el comportamiento de los individuos que pertenecen a una población [36].

La cultura puede verse como un conjunto de fenómenos ideológicos compartidos por una población, pero por medio de los cuales, un individuo puede interpretar sus experiencias y decidir su comportamiento. En estos modelos se aprecia muy claramente la parte del sistema que es compartida por la población: el conocimiento, recabado por miembros de la sociedad, pero codificado de tal forma que sea potencialmente accesible a todos. De igual manera se distingue la parte del sistema que es individual: la interpretación de ese conocimiento codificado en forma de un conjunto de símbolos, y los comportamientos que trae como consecuencia su asimilación; también la parte individual incluye las experiencias vividas, y la forma en que éstas pueden aportar algo al conocimiento compartido.

Reynolds intenta captar ese fenómeno de herencia doble en los algoritmos culturales. El objetivo es incrementar las tasas de aprendizaje o convergencia, y de esta manera, que el sistema responda mejor a un gran número de problemas [19].

Reynolds utiliza la población de un algoritmo genético (junto con sus operadores) como espacio de la población, y los Espacios de Versiones como espacio de creencias. A este algoritmo cultural, Reynolds lo llamó Algoritmo Genético guiado por un Espacio de Versiones (*Version Space guided Genetic Algorithm, VGA*).

En la teoría de los algoritmos genéticos existe una expresión, llamada teorema de los esquemas, que representa una cota para la velocidad a la que se propagan los mejores esquemas entre la población. Reynolds utilizó algoritmos genéticos, para aprovechar esa expresión y, mediante una breve discusión, intuir cómo un espacio de creencias puede afectar el teorema. De ahí se concluye que, agregando un espacio de creencias a alguno de los métodos de

computación evolutiva, es posible mejorar su desempeño, al aumentar las velocidades de convergencia [29].

Los algoritmos culturales operan en dos espacios. Primero, el espacio de la población, como en todos los métodos de computación evolutiva, en el que se tiene un conjunto de individuos. Cada individuo tiene un conjunto de características independientes de los otros, con las que es posible determinar su aptitud. A través del tiempo, tales individuos podrán ser reemplazados por algunos de sus descendientes, obtenidos a partir de un conjunto de operadores aplicados a la población.

El segundo espacio es el de creencias, donde se almacenarán los conocimientos que han adquirido los individuos en generaciones anteriores. La información contenida en este espacio debe ser accesible a cualquier individuo, quien puede utilizarla para modificar su comportamiento.

Para unir ambos espacios se establece un protocolo de comunicación que dicta las reglas del tipo de información que se debe intercambiar entre los espacios. Dicho todo lo anterior, podemos ver la figura 8, pseudo-código de un algoritmo cultural.

```
PSEUDOCODIGO.  
Inicio  
T = 0;  
Inicializar Población POP (t);  
Inicializar Espacio de Creencias (t);  
Repetir  
Evaluar la población POP (t);  
    Actualizar el espacio de creencias (con los individuos aceptados).  
    Aplicar operadores de variación (bajo la influencia del  
espacio de creencias)  
    Evaluar cada hijo.  
    Realizar la selección.  
Mientras no se cumpla la condición de finalización  
Fin.
```

Figura 8.- Pseudocódigo de un algoritmo cultural.

La mayoría de los pasos de un algoritmo cultural corresponden con los de los algoritmos tradicionales de computación evolutiva, y se puede apreciar que las diferencias están en los pasos que incluyen al espacio de creencias. Por ejemplo, en los primeros pasos se encuentra la iniciación del espacio de creencias figura 9.



Figura 9.- Diagrama Conceptual de los Algoritmos Culturales.

En el ciclo principal, está la actualización del espacio de creencias. Es en ese momento donde el espacio de creencias incorpora las experiencias individuales de un grupo selecto de individuos. Tal grupo se obtiene entre toda la población con la función de aceptación.

Por otro lado, los operadores de variación de los individuos (como la recombinación o la mutación) son modificados por la función de influencia. La función de influencia ejerce cierta presión, para que los hijos resultantes de la variación se acerquen a los comportamientos deseables, y se alejen de los indeseables, según la información almacenada en el espacio de creencias.

Estas dos funciones, la de aceptación y la de influencia, son mediante las cuales se establece la comunicación entre los espacios de la población y de creencias. Estas interacciones entre los espacios pueden apreciarse en la figura 9 [24].

1.4.12 Logística.

Prácticamente desde el principio de los tiempos, los productos que la gente deseaba, se producían acorde a sus necesidades y características. Por aquel entonces, la comida y otros productos existían en abundancia sólo en determinadas épocas del año. Al principio, la humanidad tuvo que optar por consumir los productos en el lugar donde se encontraban o transportarlos a un lugar determinado y almacenarlos allí para uso posterior. Como no existía un sistema desarrollado de transporte y almacenamiento, el movimiento de los productos se limitaba a lo que una persona podía acarrear, y el almacenamiento de los productos perecederos era posible solamente un período corto. Este sistema de transporte y almacenamiento obligaba a las personas a vivir cerca de los lugares de producción y a consumir una gama bastante pequeña de productos o servicios.

Cuando los sistemas logísticos empezaron a mejorar, el consumo y la producción fueron separándose geográficamente. Las distintas zonas se especializaron en lo que podían producir más eficientemente. Así, el exceso de producción se pudo enviar de forma rentable a otras regiones y los productos que no se fabricaban en la zona pudieron importarse.

Muchas veces hemos escuchado hablar de la logística, sin embargo tenemos una idea vaga de lo que es. Logística se define como: “el arte y la ciencia de manejar y de controlar los flujos de mercancías, energía e información”, se utilizó para resolver el conflicto entre la falta de coordinación entre las ventas y la producción, estableciendo el nexo entre la demanda del mercado y las actividades de producción de la empresa.

Este término, surge en el área militar, para determinar cómo y cuándo movilizar recursos a los lugares donde sean necesarios, con el objetivo de mantener las líneas de suministro propias e interrumpir las del enemigo [29].

La explosión de productos y su comercialización desordenada propiciaron que se buscaran nuevas alternativas, así se evidencia la evolución del término

logística desde la necesidad de los militares, hasta actividades propias del mundo empresarial, tales como la compra, producción, transporte, almacenaje, manutención, organización y la planificación de estas actividades. El mérito le corresponde a Peter Drucker, en la década de 1960. El llamado gurú de la administración, identificó el problema y enfocó su atención en los retos y oportunidades que ofrecía el campo de la logística y la distribución, convirtiéndose en testigo de avances y progresos significativos en el campo de estos conceptos [20].

La distribución consiste en la unión de los aspectos relacionados con el transporte, almacenaje, embalaje, carga / descarga y distribución y un sistema de apoyo e información; esta se relaciona estrechamente con otras actividades de la logística que se identifican con el área de obtención, producción y ventas, es decir, con las entradas, procesos de transformación de dichas entradas y las salidas del sistema.



Fig. 10.- Logística en el servicio.

Hoy en día el tema de la logística es un asunto tan importante que las empresas crean áreas específicas para su tratamiento, se ha desarrollado a

través del tiempo y es en la actualidad un aspecto básico en la constante lucha por ser una empresa del primer mundo.

Anteriormente la logística era solamente, tener el producto justo, en el sitio justo, en el tiempo oportuno, al menor costo posible, actualmente éstas actividades aparentemente sencillas ha sido redefinido y ahora son todo un proceso.

La logística tiene muchos significados, uno de ellos, es la encargada de la distribución eficiente de los productos de una determinada empresa con un menor costo y un excelente servicio al cliente, sin embargo, también determina y coordina en forma óptima el producto correcto, el cliente correcto, el lugar correcto y el tiempo correcto. Si asumimos que el rol del mercadeo es estimular la demanda, el rol de la logística será precisamente satisfacerla.

Solamente a través de un detallado análisis de la demanda en términos de nivel, locación y tiempo, es posible determinar el punto de partida para el logro del resultado final de la actividad logística, atender dicha demanda en términos de costos y efectividad.

La función logística se encarga de la gestión de los flujos físicos (materias primas, productos acabados...) y se interesa a su entorno. El entorno corresponde en este caso a:

- recursos (humanos, consumibles, electricidad)
- bienes necesarios a la realización de la prestación (almacenes propios, herramientas, camiones propios, sistemas informáticos...)
- servicios (transportes o almacén subcontratados, ...)

La función logística gestiona directamente los flujos físicos e indirectamente los flujos financieros y de información asociados. Los flujos físicos son generalmente divididos entre los “de compra” (entre un proveedor y su cliente), “de distribución” (entre un proveedor y el cliente final), “de devolución” (logística inversa).

1.4.13 Problema de Ruteo de Vehículos.

El problema de ruteo de vehículos consiste en, dado un conjunto de clientes y depósitos dispersos geográficamente y una flota de vehículos determinar un conjunto de rutas de costo mínimo que comiencen y terminen en planta visitando a todos los clientes agendados.

Los Clientes.

Cada cliente tiene cierta demanda que deberá ser satisfecha por algún vehículo. En muchos casos la demanda es un bien que ocupa lugar en los vehículos y es usual que el mismo vehículo no pueda satisfacer la demanda de todos los clientes con una sola ruta.

En otros casos la demanda no es un bien sino un servicio donde el cliente simplemente debe ser visitado por un vehículo. Un vehículo podría potencialmente visitar a todos los clientes.

El Depósito.

Tanto los vehículos como las mercancías a distribuir suelen estar en un depósito o planta matriz, usualmente se pide que cada ruta inicie y termine en el mismo punto de partida.

Los Vehículos.

La capacidad de cada vehículo podría tener varias dimensiones, como por ejemplo peso y volumen. Cuando en un mismo problema existen diferentes productos a ser distribuidos, los vehículos podrían tener compartimientos de modo que la capacidad del vehículo dependa de la mercancía que traslada.

Algunas técnicas clásicas del ruteo de vehículos se pueden mencionar [31]:

- Problema del agente viajero.
- Problema con ventanas de tiempo.
- Problema con flota heterogénea.
- Algoritmos de ahorros.
- Etc.

1.4. 14 Diseño de experimento (Análisis experimental de los algoritmos).

Un Diseño de Experimentos es un esfuerzo organizado de un científico para adquirir conocimientos sobre un proceso natural o artificial. El diseño puede necesitar de muchos estudios individuales, cada uno con objetivos específicos. El Diseño de Experimentos debe llevar una secuencia completa de pasos tomados de antemano para asegurar que los datos apropiados se obtendrán de modo que permitan un análisis objetivo que conduzca a deducciones válidas con respecto al problema establecido.

Un experimento debe limitarse a investigaciones que establecen un conjunto particular de circunstancias, bajo un protocolo específico para observar y evaluar las implicaciones de las observaciones resultantes. El investigador determina y controla los protocolos de un experimento para evaluar y probar algo que en su mayor parte no se conoce hasta ese momento [18].

1.4.14. 1 Objetivo del diseño de experimentos.

- Reducir los datos registrados a forma numérica.
- Aplicar técnicas de la Estadística Matemática adecuadas.
- Proporcionar la máxima cantidad de información correspondiente al problema investigado.
- El diseño debe ser tan simple como sea posible.

- Se debe efectuar la mayor eficiencia (ahorro de tiempo, dinero, personal y material experimental).

1.4.14. 2 Interpretación y reporte de resultados.

Al momento de interpretar los resultados, es importante considerar todos los datos observados; a demás, se debe limitar las conclusiones a deducciones estrictas a partir de la evidencia obtenida.

Así mismo, se debe probar mediante experimentos independientes, las controversias que susciten los datos para así poder llegar a conclusiones respecto al resultado técnico de los resultados como a la significancia estadística [39].

Finalmente en la elaboración del reporte, se debe describir claramente el trabajo dando antecedentes, aclaraciones pertinentes del problema y del significado de los resultados; así mismo, se debe hacer uso de métodos gráficos y tabulares de tal forma que la presentación de los datos sea clara y eficiente de tal forma que la información sea suficiente para que el lector pueda verificar los resultados y sacar sus propias conclusiones.

CAPÍTULO 2

2.1 Justificación.

Generalmente cuando hablamos de incrementar el volumen y venta de algún producto o servicio, rápido nos viene a la mente conceptos o herramientas como:

- Logística, ya que es una herramienta que maneja y controla flujos de mercancía o información.
- Mercadotecnia, que busca satisfacer las necesidades de los clientes, y a la vez obtener un beneficio.

Ambos generalmente trabajan de la mano, sin embargo, en busca de lograr satisfacer al cliente; en este trabajo se desarrollo una herramienta basada en Minería de Datos, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos de venta reportados en la purificadora de agua en la cual se realiza el estudio.

De esta manera se implementa la Minería de Datos dentro de un rubro que no se había aplicado como lo es Logística de Servicio al Cliente, y así contribuir con el desarrollo de la empresa.

Además, con el uso de Algoritmos Culturales, los cuales al ser una herramienta de la Inteligencia Artificial relativamente nueva, se implementan por primera vez en problemas de optimización de espacio, aportando un nuevo uso a los mismos, y conllevando a su valía metodológica.

Lo anterior debe realizarse previamente un estudio de la cadena de valor que nos marca la pauta para ver las áreas generadoras de valor para de esta manera tener un sustento sólido a la implementación de la minería de datos y por ende un trabajo holístico.

2.2 Objetivos generales.

- Establecer un sistema de distribución del producto que permita optimizar los costos de la empresa.
- Determinar los factores que intervienen en la logística de distribución y venta del agua purificada en Fresnillo y las comunidades que lo circundan, para contribuir al desarrollo de la empresa.
- Proponer un sistema de optimización del espacio utilizado por los garrafones en los vehículos para cubrir la demanda del producto en las rutas de distribución.

2.2.1 Objetivos Particulares.

- Mostrar un sistema sólido de distribución del producto que permita brindar el servicio que el mercado merece y que a la vez esto respalde la confiabilidad del mismo.
- Identificar aquellas zonas en las que existe mayores posibilidades de venta considerando que la incursión y la permanencia sea factible dentro del área de cobertura de la empresa.
- Lograr un crecimiento sostenido en las ventas y para que así, se pueda obtener más ingresos económicos con relación a un periodo de tiempo previo.
- Lograr que el servicio se encuentre en las cantidades y condiciones adecuadas, para que de esta forma, el servicio se preste en los lugares y momentos precisos en el que los clientes lo necesitan o desean.

2.3 Hipótesis.

Al no realizarse un estudio apropiado de distribución y logística, la aceptación del producto no se dio con la fuerza que se requería.

Al establecerse un sistema sólido de distribución y servicio, se logrará el incremento y posicionamiento de la marca en la región.

Las ventas son mejoradas al encontrar el patrón oculto entre las variables de distribución y venta.

CAPÍTULO 3

3.1 Desarrollo de la Investigación.

Mediante las herramientas descritas anteriormente, se realizó una Lluvia de Ideas con la finalidad de encontrar las posibles causas que el personal atribuía a la problemática de la distribución y de las ventas de la planta y se logró obtener el siguiente listado:

Posibles Causas de la purificadora “La Noria” sobre la mala distribución.

- Falta de Organización.
- Malos Sueldos.
- Vehículos en Mal Estado.
- Pereza en los obreros.
- Falta de control sobre los obreros.
- No hay capacitación.
- Clima.
- Competencia.
- Mala Mercadotecnia.
- Mala distribución de rutas.
- Distancias entre clientes.
- Mala comunicación (Patrón Empleado).
- Rotación constante de personal.
- Desconocimiento de la implementación de rutas.
- Desconocimiento de clientela por parte de los patrones.
- Garrafrones en mal estado.
- Etiquetas rotas.
- Pérdida de clientes.
- No existen bitácoras.
- Falta de pruebas de admisión para repartidores (Test).
- Desconocimiento de tiempo de recorrido por ruta.
- Mal sellado de garrafrones.

En base a este listado es necesario darle otro formato de tal manera que sea de fácil comprensión para el personal y que se pueda trabajar en base a ello, para lo cual se elaboró el diagrama Causa Efecto (Ishikawa).

3.2 Desarrollo de Diagrama Causa Efecto (Ishikawa) [41].

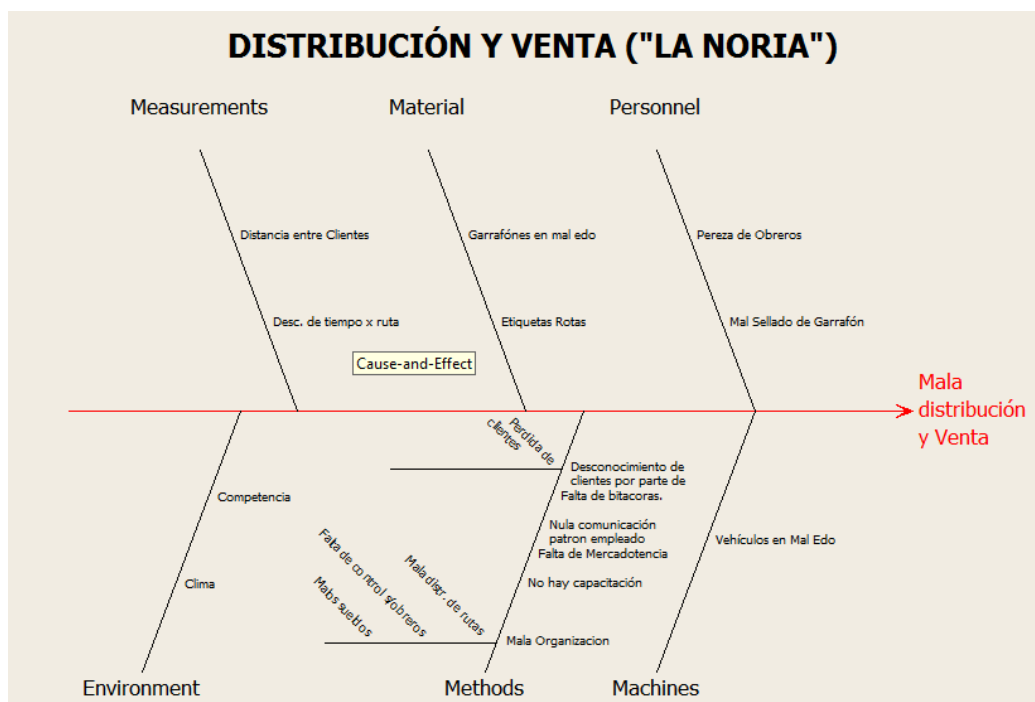


Fig.11.- Diagrama Causa Efecto para Mala Distribución y Venta nos muestra que la principal falla detectada por el personal se encuentra en los métodos.

El Diagrama nos muestra que en base a la percepción de los empleados externada en la lluvia de ideas, la problemática se centra principalmente en los métodos empleados para la distribución y venta que se transforma en el servicio al cliente que es finalmente el objetivo de cualquier cadena de valor.

3.3 Aplicación de la Cadena de Valor.

Por lo anterior, se procedió a la determinación de la Cadena de Valor de la Purificadora, y por tanto las actividades que general valor y en consecuencia ganancia y así definir si la aplicación de la Minería y Algoritmos Evolutivos son viables para el desarrollo de la tesis.

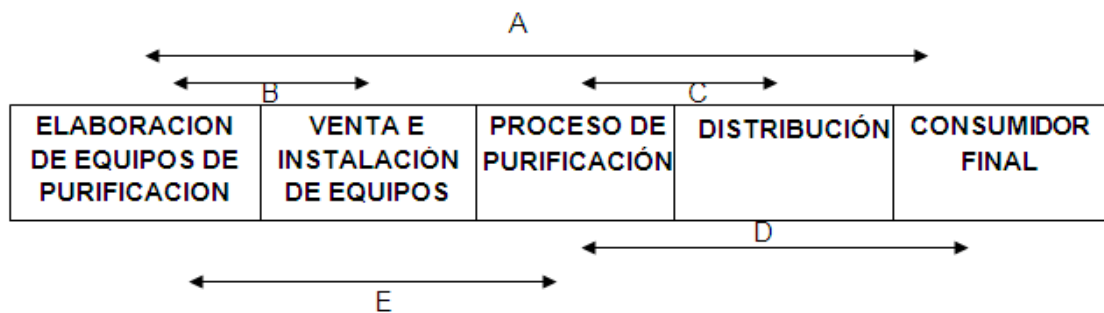


Fig.12.- Determinación de la Cadena de valor en la que forma parte la purificadora y la parte concerniente a la distribución.

La bibliografía clásica nos menciona que existen Actividades Primarias y Actividades Secundarias, como se puede observar en la figura 12, la parte correspondiente a la distribución en base a esto, corresponde a una actividad primaria.

3.4 Aplicación de la Minería de Datos.

Como primer paso se obtuvo un mapa de Fresnillo con la finalidad de ubicar a cada cliente que se tenía registrado en la base de datos para determinar las áreas en la cual se tenía mayor número de venta.

Una vez realizado esto, se dividió el mapa de la ciudad en 4 cuatro cuadrantes y por ende 4 zonas, Figura 13, para a partir de esto, asignarle coordenadas a cada cliente con que se contaba así como los nuevos.



Fig. 13.- Mapa de Fresnillo dividido en Cuatro Cuadrantes para el agrupamiento de clientes.

Una vez que se agruparon en zonas por coordenadas, se procedió a la clusterización (agrupar) para poder aplicar la Minería de Datos con el método de k-means.

COORDENADAS DE COLONIAS

UBICACIÓN DE LA PLANTA (-5,-2)		ZONA 1		ZONA 2		ZONA 3		ZONA 4			
1	PERIODISTAS	(-1,0)	21	MAGISTERIAL	(-2,0)	38	JOSE MARIA VAZQUES	(1,1)	56	BARRIO ALTO	(4,-2)
2	INDUSTRIAL	(-4,-1)	22	FRANCISCO GOYTIA	(0,1)	39	FRACC. 18 DE JULIO	(2,0)	57	EL VERGEL	(6,-1)
3	SECTOR LAGUNILLA	(-2,-3)	23	TECNOLOGICA	(-2,1)	40	BENITO JUAREZ	(1,2)	58	COLONIA FELIPE ANGELES	(9,-1)
4	DEL PARQUE	(-4,-4)	24	DEL FRESNO	(-1,2)	41	LAS FLORES	(5,1)	59	FRACC. LAS HACIENDAS	(9,-2)
5	LAS AMERICAS	(-6,-4)	25	LOMAS DE PLATEROS	(-2,3)	42	JESUS ORTEGA	(7,0)	60	EL OLIVAR	(4,-3)
6	MIGUEL HIDALGO	(-11,-3)	26	PLUTARCO ELIAS CALLES	(-2,3.5)	43	COLONIA DEL SOL	(10,1)	61	CENTRO	(1,-4)
7	LAS JOYAS	(-11,-2)	27	FRANCISCO YMADERO	(-1,6)	44	FELIPE MONREAL	(9,2)	62	LAS MARAVILLAS	(4,-5)
8	PROGRESO	(-10,-1)	28	PATRIA Y LIBERTAD	(-3,5)	45	OBRAERA	(7,2)	63	ELECTRICISTAS	(2,-6)
9	DEL VALLE	(-12,-5)	29	COL AZTECA	(-3,6)	46	LUIS DONALDO COLOSIO	(5,3)	64	FRACC. MARAVILLAS	(4,-7)
10	LA FORTUNA	(-8,-6)	30	MEXICO	(0,7)	47	EMILIANO ZAPATA	(3,3)	65	FRACC. QUINTA DEL COBRE	(4.5,-6.5)
11	FRACC LA FORTUNA	(-7,-7)	31	AMPLEACION AZTECA	(-1,7)	48	ARBOLEDAS	(7,4)	66	VENUSTIANO CARRANZA	(6,-6)
12	LA PAZ	(-5,-7)	32	LAZARO CARDENAS	(-3,8)	49	SOLIDARIDAD	(3,6)	67	FRANCISCO VILLA	(8,-4)
13	DR. MONREAL	(-5,-6.5)	33	PLAN DE AYALA	(-5,8)	50	GOMEZ	(4,5.5)	68	FRACC. JESUS ORTEGA	(9,-5)
14	BUENOS AIRES	(-5,-6)	34	NUEVA ESPERANZA	(-4,6)	51	EMILIANO ZAPATA SECTOR 1	(1,5)	69	FRACC. VENUSTIANO CARRANZA	(10,-8.8)
15	REAL DE MINAS	(-5,-5)	35	LINDAVISTA	(-6,3)	52	FRACC. REAL DEL FRESNO	(8,6)	70	FRACC. SN FRANCISCO	(10.5,-6)
16	FRACC EL VERGEL	(-6,-5)	36	LIENZO CHARRO	(-10,1)	53	FRACC. LOS BALCONES	(9,6)	71	COLONIA PETROLEOS	(10,-7)
17	ESPARZA	(-2,-5)	37	UNIVERSIDAD	(-9,5)	54	FRACC. VILLA DE PLATEROS	(8,7)	72	FRACC. VILLA JARDIN	(9.5,-8.5)
18	MINERA	(-1,-6)				55	COLONIA EL PUENTE	(7,12)	73	ECOLOGISTA	(12,-4)
19	LOS SOTOLES	(-12,-7)							74	MESOAMERICA	(14,-5)
20	FOVISSSTE	(-8,-5)							75	DEL BOSQUE	(13,-6)
									76	POLVADERAS	(15,-6)
									77	MAGISTERIAL	(14,-6.2)
									78	FRACC. IMPRESIONISTAS	(16,-6)
									79	FRACC. VICTORIA 2	(13,-7)
									80	LOS OLIVOS	(14,-7)
									81	FRACC. PROVIDENCIA	(12,-7)
									82	FRACC. MURALISTAS	(13,-8)
									83	BELEÑA	(1,-11)

Fig. 14.- Colonias por zona acorde a sus respectivas coordenadas.

Es bien sabido que existen diferentes técnicas que permiten aplicar minería de datos, sin embargo, el algoritmo de “k-means” es el referente principal entre los diversos métodos para seleccionar grupos representativos entre los datos a diferencia de otros que se utilizan para el proceso de clusterización.

El algoritmo genera agrupamiento de datos, sin tener clases predefinidas, basándose en una función de similitud de los valores que poseen sus diferentes atributos, realizándose de forma no supervisada (es decir, descubren patrones o tendencias en los datos).

K-means es un método particional de clustering (comienza del todo a lo particular), en donde se realiza el particionamiento de una base de datos de n objetos en un conjunto de k grupos, buscando optimizar el criterio elegido de particionamiento. En K-means cada agrupación de datos está representada por un centroide, a demás, intenta formar k grupos, con k predeterminado antes del inicio del proceso. El objetivo es minimizar la varianza total intra-grupo.

Existen una serie matrices que constituyen el fundamento para la implementación de este tipo de algoritmo, sus diferentes variantes se basan fundamentalmente en la forma de medir distancias entre los datos y los grupos, el criterio para definir la pertenencia de los datos a cada grupo y la forma de actualizar dichos grupos.

A continuación mencionaremos las matrices utilizadas para la implementación del algoritmo:

3.4.1 Matriz de Datos:

Esta matriz de orden $R \times N$,

$$\begin{matrix} & \mathbf{C1} & \mathbf{C2} & \dots & \mathbf{Cn} \\ \mathbf{x} & \begin{bmatrix} 1 & 2 & \dots & n \end{bmatrix} & & & \\ \mathbf{y} & \begin{bmatrix} 2 & 3 & & n \end{bmatrix} & & & \end{matrix} \quad (2)$$

Donde:

N= No de Clientes

R= Rasgos para identificar y agrupar a cada cliente.

Para cada cliente los rasgos consisten en su ubicación en el eje de coordenadas propuesto basados en su posición x e y.

3.4.2 Matriz de Centroides.

Es de orden NxN, los centroides que se generen durante la evolución del algoritmo de clustering van siendo almacenados en la matriz.

$$\begin{array}{c} \begin{array}{c} x \\ y \end{array} \begin{array}{|c|c|c|c|} \hline c1 & c2 & \dots & cn \\ \hline 1 & 2 & \dots & n \\ \hline 2 & 3 & & n \\ \hline \end{array} \end{array} \quad (3)$$

Donde:

N= número de rasgos que caracterizan a cada cluster

C= Número de Centroides que se desea tener.

Para este caso en particular se utilizaron 3 centroides, debido a que se deseaba determinar 3 rutas distintas de distribución.

3.4.3 Matriz de Distancias.

Es de orden cxN, aquí se define la pertenencia a uno u otro grupo. Esta matriz define la eficiencia del algoritmo de clustering, pues será la base para actualizar los valores de los centroides.

	C1	C2	...	Cn
c1	1.5	1.8	...	n
c2	2.3	2.4	...	n
cn	n	n		n

(4)

Donde:

N= número de rasgos que caracterizan a cada cluster

C= Número de Centroides que se desea tener.

En esta matriz se almacenarán las distancias de cada uno de los clientes a cada uno de los centroides, lo cual se hará mediante la fórmula de distancias euclídeana, cabe mencionar que hay otros métodos, pero para este caso se utilizó este tipo ya que brinda buenos resultados y facilidad de uso.

3.4.4 Matriz de Pertenencias.

Es de orden CxN, Se determina en base a la matriz de distancias, es la matriz base para la iteración del algoritmo ya que a partir de sus valores se determinara la actualización de la matriz de centroides hasta llegar a converger.

	C1	C2	...	Cn
c1	0	0	...	1
c2	1	0	...	0
cn	0	1		0

(5)

Donde:

C: Es igual al número de centroides.

N: Es igual al número de clientes.

3.5 Aplicación del Algoritmo Cultural.

Con lo anterior se cumplen las condiciones para la aplicación de Minería de Datos mediante el k-means, sin embargo, falta el complemento que permita a la herramienta ser capaz de tomar decisiones para determinar las rutas de distribución.

La siguiente parte se realizó mediante un Algoritmo Cultural Figura 15, el cual maneja distintas entradas, el único que controla directamente el usuario es el tamaño de la población o número de agentes, al respecto se espera que a que se incremente, mayor será el conocimiento e influencia cultural que se tenga obteniendo así una mejor y más rápida optimización.

El Algoritmo Cultural genera una población de n agentes (un agente, es la simulación computacional de una persona), que conforman una sociedad basada en algoritmos culturales, los cuales se encargarán de determinar, con el paso de las épocas, el recorrido óptimo.

En la época cero (es cuando se inicializa el programa, y los agentes poseen un espacio de creencias vacío), todos los agentes obtienen la información generada de cada colonia, cada uno de ellos propondrá un recorrido, ello dará lugar a la negociación entre los agentes para seleccionar cuál es el mejor recorrido propuesto en determinada época.

El espacio de creencias será actualizado sólo cuando el recorrido propuesto sea mejor que el previamente almacenado anteriormente, iniciando un ciclo de mejoras que será interrumpido cuando ocurran m cantidad de épocas (iteraciones en el comportamiento de los agentes) sin mejoras a los recorridos o cuando una condición de paro sea realizada.

PSEUDOCODIGO Cultural
Inicio

Inicializar población

// Recibir como parámetro el numero de agentes

// Generar recorridos propuestos por cada agente

Inicializar los datos del espacio creencia inicial

Evaluar (Iteración)

MIENTRAS ESPCREENCIA SIN CAMBIOS < 7 (RECORRIDO MAS PEQUEÑO QUE EL ACTUAL)

Si la distancia propuesta para el agente < distanciaEspCreencias

Pasar los datos del agente al espCreencias

// IdAgente

// Recorrido propuesto

// Distancia propuesta

// Arreglo de clientes

Generar nuevo recorrido para el agente a partir del arreglo que encontró la mejor ruta hasta ahora

Incrementar época

Fin

Fig. 15 Algoritmo Cultural Propuesto.

Una vez que se determinó la ubicación, se emprendió la construcción del Data Warehouse, organizado en los siguientes campos:

- Algoritmo Cultural.
- Vendedor.
- Cliente.
- Producto.
- Visita.
- Ventas.
- Actualización del Porcentaje.

Cada uno de estos campos conforman la parte del menú del programa, su funcionamiento se explicara detalladamente en la siguiente sección.

3.6 Presentación de la Interfaz.



Fig. 16.- Función del Algoritmo Cultural.

La pantalla principal muestra el menú con el que cuenta figura 16, la opción **Algoritmo Cultural**, permite la ejecución del programa y así obtener las rutas sugeridas en base a las ventas reportadas.

Dado que actualmente se cuenta con 3 vehículos de distribución, al momento de ejecutarse el software arroja 3 recorridos propuestos (Uno por cada vehículo), las rutas propuestas son las que cuentan con el mejor recorrido y el menor tiempo. Cabe señalar que las rutas se modificaran en base a las ventas reportadas y a los clientes que se van dando de alta.

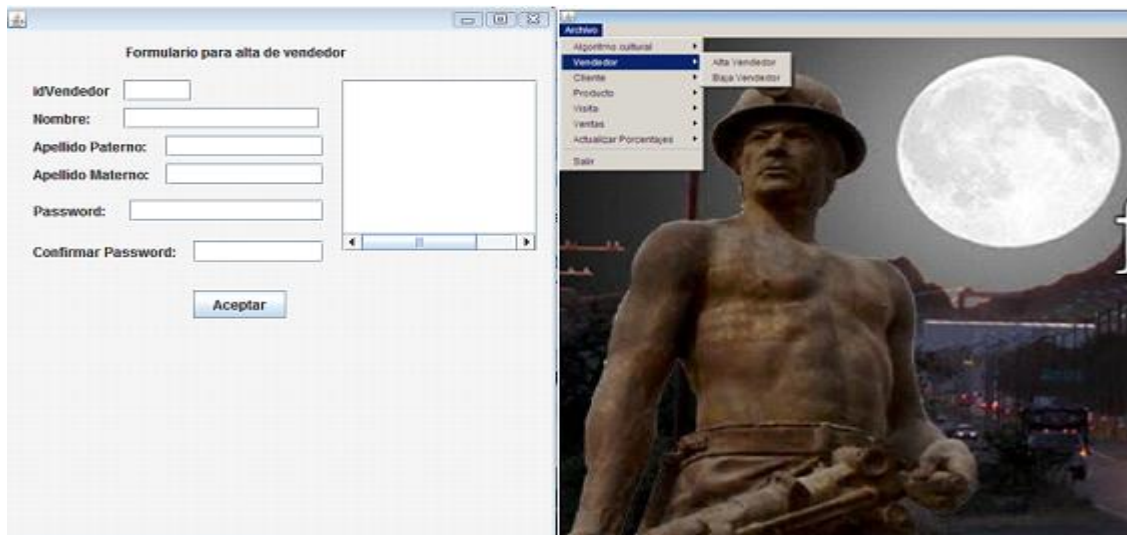


Fig. 17.- Icono de Vendedor

La opción **Alta de Vendedor**, permite como su nombre lo indica, ingresar a la base de datos a un vendedor, para lo cual es necesario asignar un idvendedor (clave) Figura 17.

El software registra a demás, el nombre, incluyendo apellidos y un password el cual elije el vendedor, y así, al momento de acceso al sistema es registrada su venta.

De la misma manera en esta opción se manejan otra opción **Baja Vendedor**, dicha opción permite eliminar a un vendedor del software mas no así sus históricos de venta ya que de ser así perjudicaría en la generación de las rutas.

La siguiente opción que forma parte del menú es referente a los clientes, la Figura 18, muestra la forma en que se despliega esta opción:

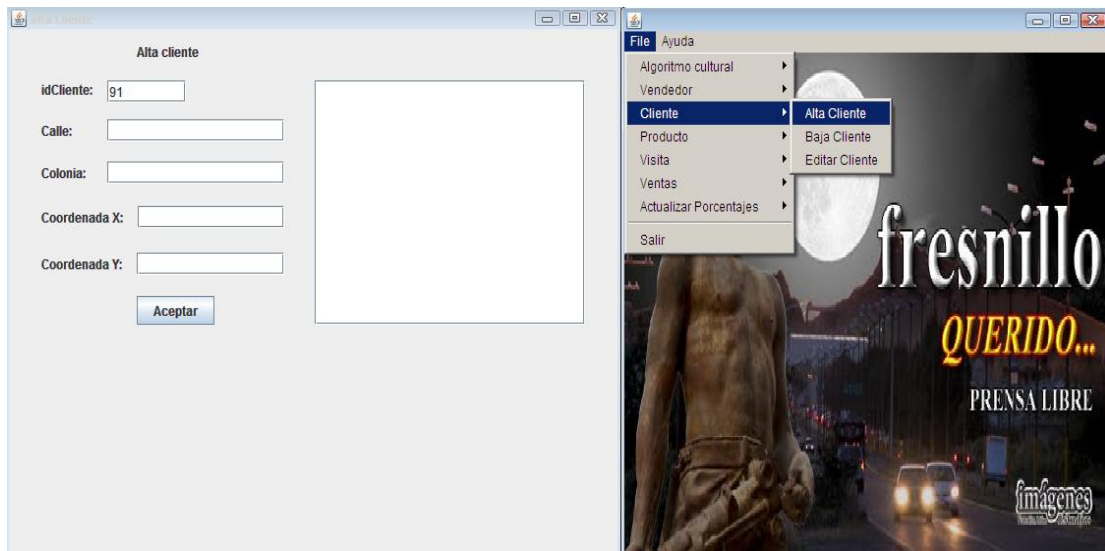


Fig.18.- Icono del Cliente.

La primera que se menciona es **Alta Cliente**, esta opción permite dar de alta a un cliente nuevo cuando así sea necesario, los requisitos que pide el software son: Id Cliente el cual se lo asigna el mismo software en base al consecutivo que maneja, Calle, Colonia y Coordenadas en las cuales se ubica en base al mapa que se utilizó para la asignación de coordenadas, una vez que le damos aceptar, el cliente está dado de alta y listo para cuando realice una compra, el vendedor pueda dar de alta su venta.

De la misma manera se maneja **Baja de Cliente**, el cual se utiliza cuando un cliente permanece determinado tiempo sin compra o bien ya no vive en su domicilio, dicho cliente puede ser eliminado de la base de datos mediante esta opción.

Por último se tiene **Editar Cliente**, este se utiliza cuando algún cliente cambio de domicilio pero sigue adquiriendo nuestro producto, es necesario actualizar su nueva dirección para que el software determine si se continúa con su ruta de visita o bien si es cambiado de ruta.

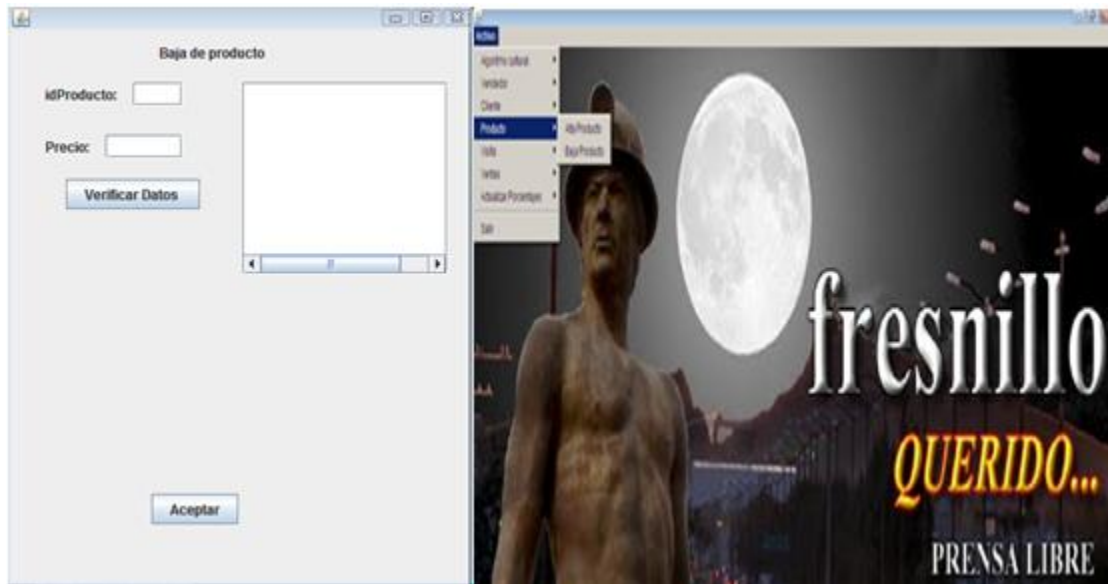


Fig. 19.- Icono del Producto.

En lo que se refiere al producto se manejan 4 presentaciones, a las cuales se le a asignado un idProducto, para poder dar de alta al producto nos ubicamos en la opción Producto, **Alta de Producto**, ahí únicamente asignamos el id así como su costo, con ello al momento de dar de alta la venta nos pedirá el idproducto para registrar la venta.

Al igual que en las anteriores opciones podemos dar de baja algún producto mediante la opción **Baja Producto**, esto lo muestra la Figura 19.

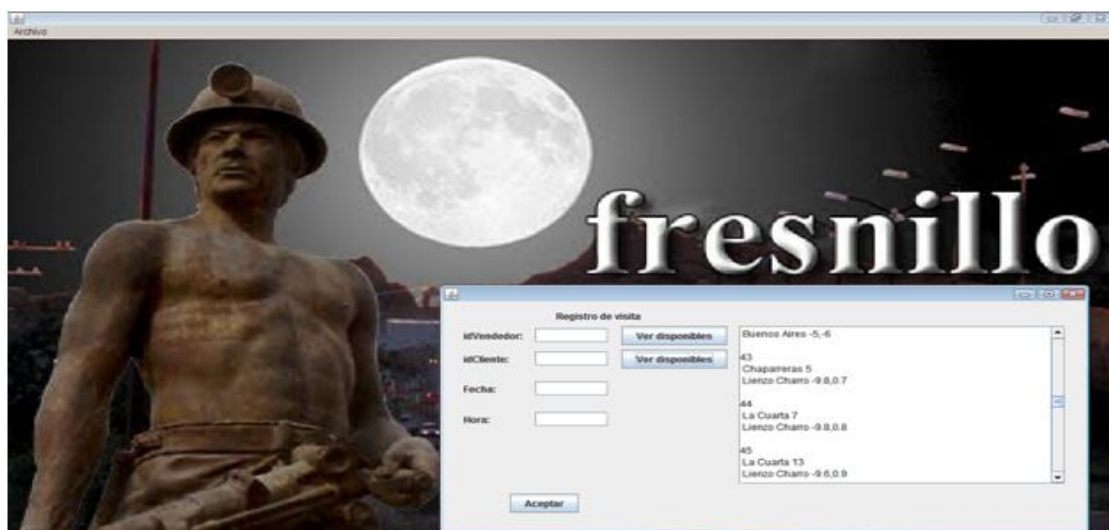


Fig. 20.- Icono de Visitas.

Cuando se generan las rutas, es importante señalar que el software tal como se explico en el proceso de algoritmo cultural, funciona en base a una *influencia cultural*, en este caso deduce quiénes son los potenciales clientes, con ello no quiere decir que todos los clientes generados en la ruta vayan a comprar, sino que hay una mayor posibilidad de venta, por tanto cuando se visita y no se obtuvo alguna venta, el repartidor reporta la visita mas no la venta, esto lo hace en el menú mediante la opción de **Alta de Visita**, en la que da el idcliente y el idvendedor para reportar que se visito pero no se obtuvo venta, lo anterior lo muestra la Figura 20.

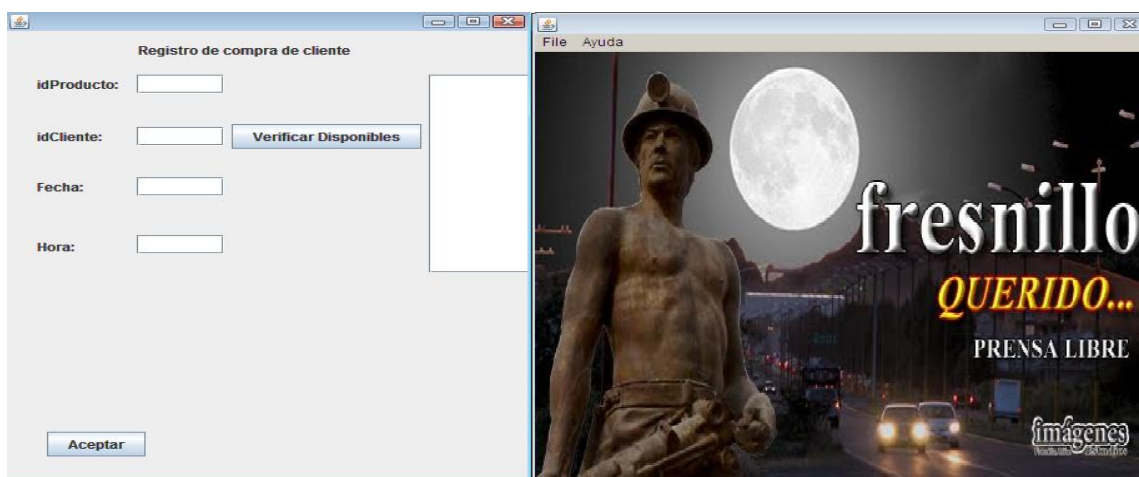


Fig. 21.- Icono de Ventas.

La Figura 21, muestra la opción de **Alta de Ventas**, dicha opción es la más utilizada, ya que aquí se registran las ventas diarias de cada vendedor, lo único que se necesita es dar de alta el producto, el idcliente, fecha y la hora a la cual se registra la venta, con ello se va actualizando el histórico de ventas, dicho de otra manera, es la boca de la herramienta, ya que si no se dan de alta las ventas, las rutas generadas no serian reales, por tanto las ventas es el alimento del sistema.



Fig. 22.- Icono Para Actualizar Porcentajes.

Finalmente, el software está programado con cierto elitismo, según lo establece la literatura de algoritmos culturales, lo cual nosotros delimitamos con un porcentaje de compra, para que el cliente salga sorteado en las rutas, es necesario que tenga un porcentaje de compra, para nuestro caso, hemos elegido un porcentaje del 50%, por lo cual, al momento de generar nuevas rutas se **Actualizan Porcentajes** y con ello se eligen a los mejores clientes, esto se realiza mensualmente para determinar las rutas en base a la actualización de ventas y por tanto de la base de datos.

3.7 Aplicación del BPP.

En lo que respecta a la optimización de espacio, se construyó un software con capacidad de elegir el mejor acomodo, esto basado en el concepto de Bin Packing pero utilizando algoritmos culturales, en complejidad más sencilla, pero no por ello menos capaz.

La principal diferencia consiste en que **no** se tomó en cuenta el peso de los objetos para nuestro caso, las presentaciones del producto (Agua Purificada), sino el volumen de los mismos para en base a ellos determinar su mejor acomodo dentro de los vehículos de distribución.

Como primer paso se realizó la toma de medidas del vehículo de distribución sobre el cual se realizó el estudio para determinar el espacio con el cual se dispone en m³, así como la medición de las diferentes presentaciones y su respectivo volumen, para de esta manera poder plantear el problema y sus respectivas restricciones.

También se tomo en cuenta la demanda de las diferentes presentaciones, para de esta manera, determinar con mayor precisión la utilidad del producto.

TABLA 1.- Descripción del Producto.

PRODUCTO	CAPACIDAD	% DEMANDA	VOLUMEN	UTILIDAD
1	20 lt	45	36500cm ³	\$ 10.00
2	1.5 lt	15	26731cm ³	\$ 26.50
3	1 lt	15	18435cm ³	\$ 21.60
4	.500lt	25	18177cm ³	\$ 38.50

Una vez que se realizaron las mediciones, fue necesaria la creación de un algoritmo capaz de encontrar la combinación adecuada en cuanto a la carga, de manera tal que se optimice la ganancia de la camioneta y por ende de la empresa, dicho algoritmo utiliza una población como base, e inicializa otra como espacio de creencia que en ese momento es desconocido su valor.

Se realiza el ajuste de la demanda para de esta manera ajustar con mayor precisión el valor obtenido, esto en base al porcentaje de venta por presentación del producto, el cual se muestra en la tabla 1.

Se evalúa la población inicial en base al problema y restricción del mismo como se muestra en la tabla.

$$\text{Max } z = r_1m_1+r_2m_2+\dots+r_nm_n$$

Sujeto a $v_1m_1+v_2m_2+\dots+v_nm_n \leq V$ (6)

$$m_1, m_2, \dots, m_n \geq 0 \text{ y entero}$$

$$V = 1138425 \text{ cm}^3$$

Donde:

r: Utilidad por unidad.

v: Volumen de cada unidad.

m: Unidades de cada tipo de producto.

V: Máxima capacidad de volumen.

Toda condición que viole las restricciones, deberá ser penalizada de tal manera que solo se obtengan las mejores combinaciones y que así, podamos sacar un promedio por medio del cual podremos tener los mejores individuos e influenciar las siguientes generaciones en base a las medias de los individuos.

El resultado obtenido (Época) será la solución propuesta, la condición de paro será a partir de que se repitan 8 épocas sin cambio, es decir, sean = ó > a las anteriores.

Para esta Herramienta no se tiene una interfaz, únicamente se alimenta el algoritmo con los datos requeridos y nos proporciona la combinación adecuada los resultados se muestran en el siguiente capítulo.

3.8 Resumen de ambas herramientas.

Una vez que se tienen desarrolladas las dos herramientas, se aplicaron dentro de la purificadora “La Noria”, con esto se determinan rutas en base a la herramienta de Minería de Datos en combinación con los Algoritmos Culturales

y por otro lado se determina la carga óptima de los vehículos de distribución, los resultados se muestran en el siguiente capítulo.

Los puntos más importantes de este capítulo se muestran en la Tabla 2.

Tabla 2. Metodología utilizada para el desarrollo de la investigación.

ETAPA	HERRAMIENTA	RESULTADO
Determinación de la posible problemática.	<ul style="list-style-type: none"> • Lluvia de Ideas. • Diagrama Ishikawa. 	Identificación de los posibles factores que influyen en la distribución y venta de la purificadora.
Ubicación del problema.	Cadena de Valor	Ubicación de la Cadena de Valor de la purificadora.
Desarrollo de Software Inteligente para Toma de Decisiones.	Minería de Datos. Algoritmos Culturales.	Determinación de Rutas.
Desarrollo de Software Inteligente para optimización de Espacio.	Algoritmos Culturales. BPP.	Determinación de Carga Óptima.
Aplicación	Sistemas Desarrollados.	Implementación de Rutas con cargas óptimas en la purificadora “La Noria”.

CAPÍTULO 4

4.1 Análisis de los resultados.

En este capítulo se analizarán los resultados obtenidos con la aplicación de las herramientas desarrolladas en la purificadora “La Noria”, y un diseño de experimentos realizado para verificar la eficacia de los programas.

4.2 Resultados de Minería de Datos.

Con la implementación de la herramienta desarrollada con Minería de Datos y Algoritmos Culturales, se ha logrado determinar las rutas operativas de la purificadora “La Noria”.

Dado que en el capítulo anterior se explicó la forma en cómo se desarrolló el sistema, se mostrará la forma de ejecución y obtención de resultados del mismo. Se utilizó la base de clientes de la purificadora para realizar el ejemplo. La Figura 23, muestra la interfaz del programa.



Fig. 23.- Determinación de Ruta.

En el ejemplo se muestra la ejecución para una combinación de 800 agentes y 800 iteraciones, con ello al momento de ejecutarlo, se ponen en marcha las funciones del sistema, el cual realiza una búsqueda interna entre la base de datos de ventas, buscando aquellos clientes que tuvieron un porcentaje de compra por encima del 50 % y por tanto mayor probabilidad de compra; estos clientes son tomados por agentes inteligentes que generan recorridos aleatorios proponiendo diferentes rutas y tiempos de recorrido, para ello toman en cuenta sus distancias, las cuales fueron calculadas en base a sus coordenadas mediante el algoritmo de K-means.

El mejor recorrido es almacenado y no realiza cambio mientras no se decreta que existe otro recorrido mejor, si después de 8 épocas no se encuentra otra mejor opción, el sistema converge y muestra la ruta propuesta.

La figura 24, muestra la pantalla de resultados generada por el sistema:

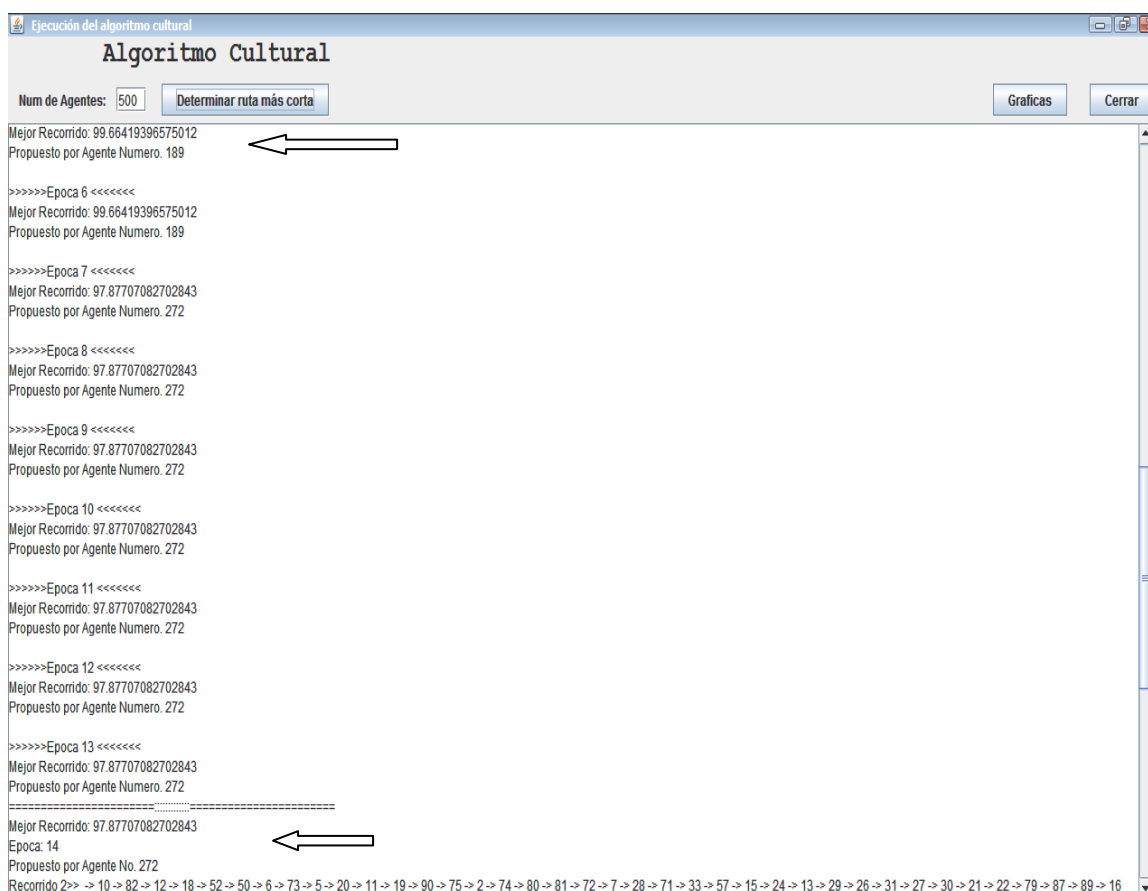


Fig. 24.- Pantalla de resultados de la ejecución del sistema.

Como se puede observar, Figura 24, muestra la pantalla que arroja los resultados, los cuales nos indican las rutas propuestas, en el ejemplo podemos ver que la flecha mayor nos indica la distancia mayor, en tanto que la flecha pequeña nos muestran la distancia propuesta, la época de convergencia y el agente que propuso dicha ruta.

Es importante señalar que son unidades predefinidas por el software ya que la alimentación de las coordenadas no son reales, son a escala y por tanto no se tienen las distancias reales.

El sistema menciona a demás, el agente que propuso el recorrido ideal, la numeración corresponde al idcliente, y el orden en que debe ser realizado el recorrido, las cuales son interpretadas por el departamento de ventas y se entrega el itinerario de visita a los repartidores, el listado incluye el idCliente, Nombre y Colonia.

El sistema también proporciona la opción de un gráfico, para el ejemplo la gráfica se muestra en la Figura 25:

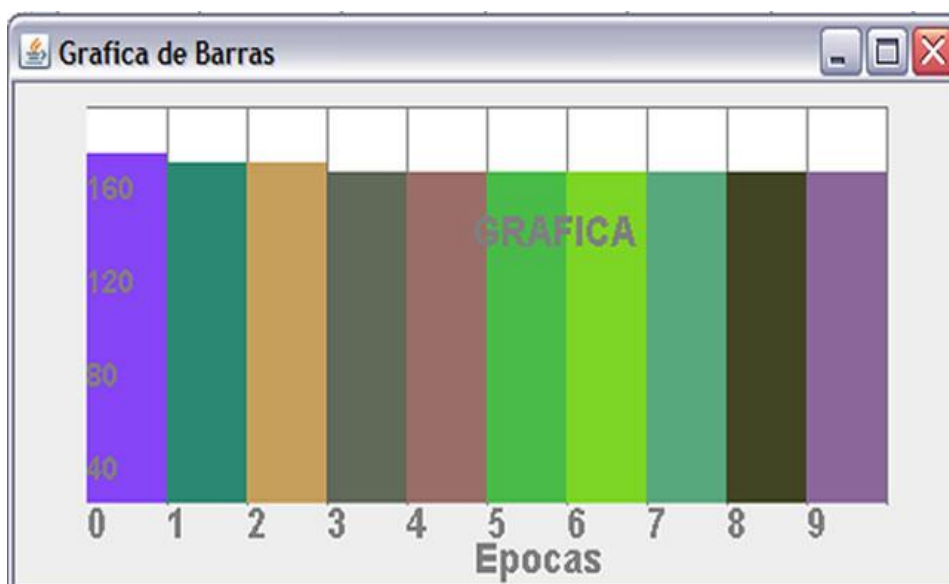


Fig. 25.- Gráfico de convergencia de herramienta de Minería de Datos.

La gráfica nos indica que después de la época 3, no hubo ninguna propuesta de un mejor recorrido por tanto convergió en la época 9.

El sistema actualmente está programado para determinar tres rutas, ya que son los vehículos de distribución con que actualmente se cuenta, sin embargo, si se requiriera añadir más rutas, sería necesario modificar el código.

4.3 Resultados de BPP.

La bibliografía clásica nos menciona que BPP busca almacenar de objetos de diferentes tamaños o pesos, para este caso, se tomo en cuenta el primer término. Lo que se busco fue determinar la carga óptima de las camionetas en base a la demanda de cada presentación.

A continuación se muestran 2 tablas mostrando los resultados obtenidos.

Es importante mencionar que inicialmente solo se contaba con 1 camioneta, motivo por el cual el comparativo de la Tabla 3 solo esta contemplando este vehículo.

Tabla 3. Comparativo de Acomodo de Producto.

PRESENTACION	19lt	1.5lts	1lt	.500ml	TOTAL
CANTIDAD	33	0	0	1	34
UTILIDAD	10	26.5	21.6	38.5	96.6
TOTAL	330	0	0	38.5	368.5

Anterior al la implementación de BPP

PRESENTACION	19lt	1.5lts	1lt	.500ml	TOTAL
CANTIDAD	18	6	6	10	40
UTILIDAD	10	26.5	21.6	38.5	96.6
TOTAL	180	159	129.6	385	853.6

Ahora con la implementación de BPP

Las tablas nos muestran la diferencia de utilidad respecto a la misma camioneta y en base a la demanda de los productos que es como trabaja el sistema; como se puede ver la diferencia es de 485.1, lo cual quiere decir que

se incremento la utilidad a más del doble; es importante resaltar que el acomodo anterior no tenía ningún patrón, es decir, solo se llenaba la camioneta con garrafones de 19lts y si se solicitaba alguna otra presentación se recargaba.

Ahora tenemos el acomodo que se realiza en las otras 2 camionetas, aquí es importante señalar que ambas camionetas cuentan con paneles especiales, lo que condiciona a una carga mayor de productor de 19 lts, que finalmente es el producto con mayor demanda.

Tabla 4. Acomodo de producto en camionetas con panel.

PRESENTACION	19lt	1.5lts	1lt	.500ml	TOTAL
CANTIDAD	32	4	4	2	42
UTILIDAD	10	26.5	21.6	38.5	96.6
TOTAL	320	106	86.4	77	589.4

CAMIONETA 1

PRESENTACION	19lt	1.5lts	1lt	.500ml	TOTAL
CANTIDAD	34	6	6	2	48
UTILIDAD	10	26.5	21.6	38.5	96.6
TOTAL	340	159	129.6	77	705.6

CAMIONETA 2

Como se puede ver la implementación del algoritmo de BPP a colaborado a incrementar la utilidad, quizás hubiese sido muy bueno el haber tenido las 3 camionetas inicialmente para en base a ello determinar la mejora total, sin embargo la camioneta 1, nos muestra los resultados de la aplicación del sistema a demás, de que la demanda también forma un factor fundamental en el acomodo de las mismas así como el número de paneles con que cuentan las otras 2 camionetas.

Se tiene también un comparativo económico de las utilidades del primer cuatrimestre del año 2009 con respecto al año 2010.

Purificadora "La Noria"		
Estado de Resultados		
Del <u>2</u> de <u>Enero</u> al <u>30</u> de <u>Abril</u> del <u>2009</u>		
Ventas	Ventas	46,318.90
Menos:	Dev. Desc y Bonif. s/Ventas	<u>4524</u>
Igual:	Ventas Netas	41,794.90
Menos:	Costo de Ventas	<u>6000</u>
Igual:	Utilidad Bruta	35,794.90
Menos:	Gastos de Operación:	26350
	Gastos de Venta	26200
	Gastos de Admón	<u>150</u>
Igual:	Utilidad de Operación	9,444.90
Igual:	Utilidad Neta	9,445.00

Fig. 26 Estado de Resultados 1er Cuatrimestre 2009.

Purificadora "La Noria"			
Estado de Resultados			
Del <u>2</u> de <u>Enero</u> al <u>16</u> de <u>Abril</u> del <u>2010</u>			
Ventas	Ventas		143,958.00
Menos:	Dev. Desc y Bonif. s/Ventas		<u>18100</u>
Igual:	Ventas Netas		125,858.00
Menos:	Costo de Ventas		<u>18000</u>
Igual:	Utilidad Bruta		107,858.00
Menos:	Gastos de Operación:		72350
	Gastos de Venta	72200	
	Gastos de Admón	<u>150</u>	
Igual:	Utilidad de Operación		35,508.00
Igual:	Utilidad Neta		35,508.00

Fig. 27 Estado de Resultados 1er Cuatrimestre 2010.

La diferencia de utilidades entre un bimestre y otro es de \$26,000, esto pues nos muestra que la aplicación de las diferentes herramientas esta dando resultados importantes.

Para dar un mayor soporte se realizó un Diseño de Experimentos para darle mayor soporte a los sistemas desarrollados e implementados.

4.4 Diseño de experimentos para las herramientas desarrolladas.

A pesar de que los ACs datan de los años 90s, su utilización ha mostrado un comportamiento errático, y sólo se cuenta con un poco más de 67 publicaciones de este tópico.

Al no existir un Benchmarking de pruebas, los investigadores en el tema han desarrollado sus propios Diseños de Experimentos, que en lo posterior se nombraran como DOEs, acordes a lo que se pretende realizar.

El principal objetivo del diseño de experimentos es calibrar los parámetros del algoritmo para que encuentre las mejores soluciones, por tanto, para el funcionamiento de los presentes algoritmos culturales que se aplican en conjunto con un algoritmo de minería de datos, se calibraron los parámetros de ambos algoritmos de manera que en conjunto produjeran los mejores resultados, para el caso en cuestión la optimización de distancias recorridas por las rutas de distribución de la purificadora “La Noria”.

El primer paso fue elaborar diferentes corridas para determinar la combinación que optimizara al máximo el sistema, en total fueron 55 corridas con 91 clientes; Los parámetros que se han tomado en cuenta para este análisis son Número de Iteraciones del kmeans, Número de Agentes, Ruta, Distancia Total Recorrida y Épocas en que converge; estas pruebas se muestran en las siguientes tablas:

Iteraciones		Tabla 1				
Agentes		2				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	53	47	6	9		
2	141	141	0	8		
3	34	34	0	8		
1	57	47	10	13		
2	149	149	0	8		
3	27	27	0	8		
	461	445	16	54		

Iteraciones		Tabla 2				
Agentes		10				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	44	44	0	8		
2	143	128	15	15		
3	23	23	0	8		
1	45	41	4	15		
2	132	130	2	14		
3	24	24	0	8		
	411	390	21	68		

Iteraciones		Tabla 3				
Agentes		50				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	42	40	2	11		
2	121	121	0	8		
3	22	22	0	8		
1	44	40	4	15		
2	127	118	9	9		
3	23	23	0	8		
	379	364	15	59		

Iteraciones		Tabla 4				
Agentes		100				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	41	41	0	8		
2	137	115	22	13		
3	22	22	0	8		
1	43	38	5	9		
2	130	113	17	15		
3	21	21	0	8		
	394	350	44	61		

Iteraciones		Tabla 5				
Agentes		100				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	37	37	0	8		
2	81	75	6	14		
3	44	44	0	0		
1	38	34	4	15		
2	88	83	5	16		
3	40	40	0	8		
	328	313	15	61		

Iteraciones		Tabla 6				
Agentes		200				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	36	32	4	21		
2	88	77	11	14		
3	39	39	0	8		
1	38	33	5	13		
2	83	80	3	14		
3	45	45	0	8		
	329	306	23	78		

Iteraciones		Tabla 7				
Agentes		400				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	34	33	1	14		
2	83	76	7	11		
3	38	38	0	8		
1	34	31	3	17		
2	81	77	4	12		
3	36	36	0	8		
	306	291	15	70		

Iteraciones		Tabla 8				
Agentes		600				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	36	31	5	11		
2	85	70	15	10		
3	37	37	0	8		
1	33	32	1	11		
2	86	73	13	9		
3	40	40	0	8		
	317	283	34	57		

Iteraciones		Tabla 9				
Agentes		800				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	33	31	2	9		
2	84	73	11	21		
3	37	37	0	8		
1	35	31	4	11		
2	87	72	15	29		
3	37	37	0	8		
	313	281	32	86		

Iteraciones		Tabla 10				
Agentes		850				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	34	28	6	9		
2	82	68	14	12		
3	35	35	0	8		
1	33	30	3	10		
2	80	63	17	9		
3	37	37	0	8		
	301	261	40	56		

Iteraciones		Tabla 11				
Agentes		855				
Rutas	Mayor	Menor	Optimizacion	Epocas		
1	33	31	2	9		
2	82	78	4	9		
3	38	38	0	8		
1	34	29	5	11		
2	81	79	2	9		
3	39	39	0	8		
	307	294	13	54		

Fig. 28.- Combinación de Agentes e Iteraciones para lograr la combinación óptima de ejecución del software de Minería de Datos.

La tabla 5 nos muestra un concentrado con los totales de la Figura 28, la primer columna nos muestra la combinación de agentes e iteraciones, la segunda las distancias mayores en tanto que la tercera las distancias menores, finalmente la cuarta columna nos muestra la diferencia entre las dos anteriores.

Tabla 5 Resultados totales de tablas de prueba.

Combinación	D Mayor	D Menor	Diferencia
2A, 1i	461	445	16
10A, 1i	411	390	21
50A, 1i	379	364	15
100A, 1i	394	350	44
100A, 100i	328	313	15
200A, 200i	329	306	23
400A, 400i	306	291	15
600A, 600i	317	283	34
800A, 800i	313	281	32
850A, 850i	301	261	40
855A, 855i	307	294	13

Estos datos fueron analizados estadísticamente, primeramente se determino el diagrama de dispersión de las distancias mayores y menores.

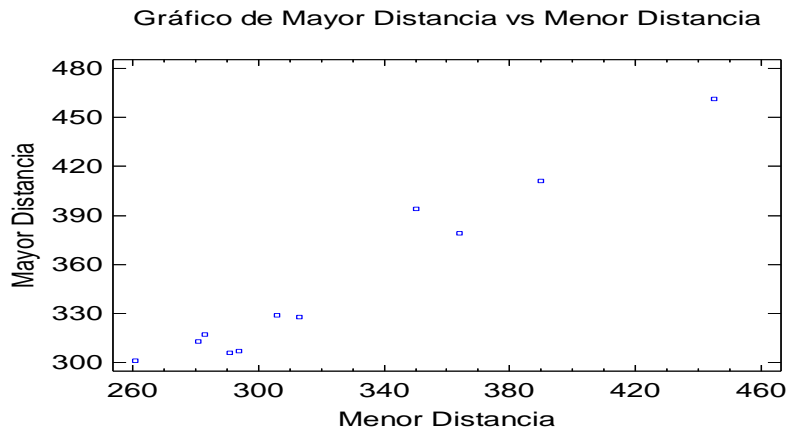


Fig. 29.- Gráfico de Dispersión de distancias mayores y menores de prueba.

La figura 29 nos muestra una correlación de los datos positiva, en la parte inferior izquierda podemos observar que se encuentra el mayor acumulado de puntos, esto es, a medida que se incrementaba el número de agentes e iteraciones, se mejoraban las distancias de recorrido, dando soporte con esto a las conclusiones ya explicadas.

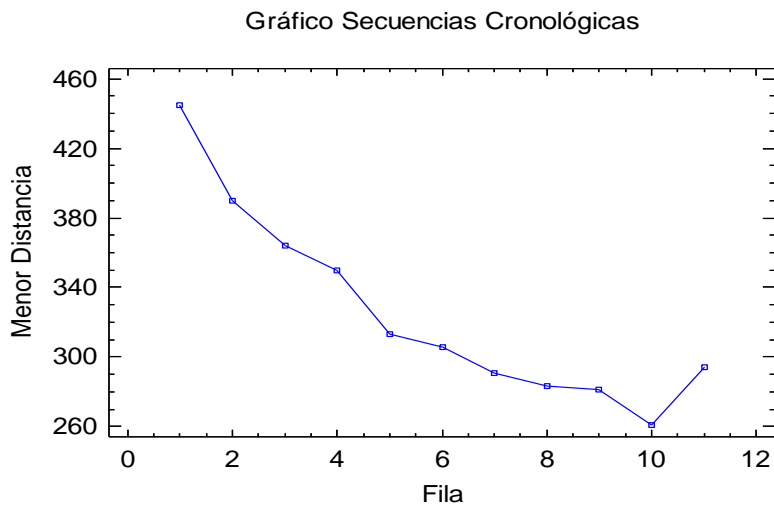


Fig. 30.- Gráfico de secuencia cronológica de la distancia menor.

La figura 30 muestra el comportamiento final de las distancias menores, el punto 10, es la combinación de 850 agentes y 850 iteraciones, y es allí donde se obtuvo el óptimo de las corridas, el punto 11 es la combinación de 855 agentes y 855 iteraciones rompiendo la normalidad del decrecimiento gráfico.

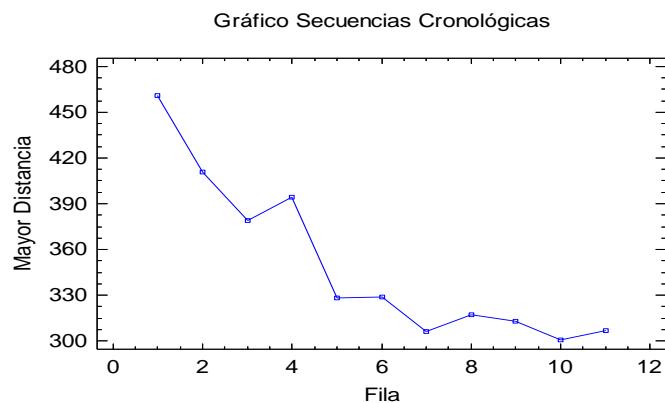


Fig. 31.- Gráfico de secuencias cronológicas de distancias mayores.

En la Figura 31 muestra el gráfico de secuencia cronológica de datos de mayor distancia, el cual tiende a descender excepto en el punto cuatro, ya que el número de agentes provocó ese incremento repentino, sin embargo, se puede observar algunos pequeños altibajos, lo cual se explica en base a las condiciones de paro del kmeans y por tanto al número iteraciones.

Gráfico de Dispersión

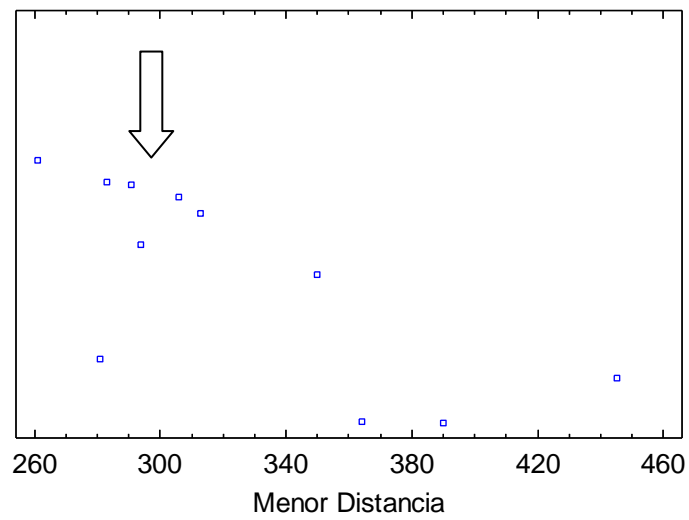


Fig. 32.- Gráfico de Dispersión de las distancias menores.

El gráfico de dispersión para distancias menores, muestra la tendencia de los puntos a la izquierda, es decir menores distancias, caso contrario nos mostraría una ineficacia del sistema.

Gráfico de Caja y Bigotes

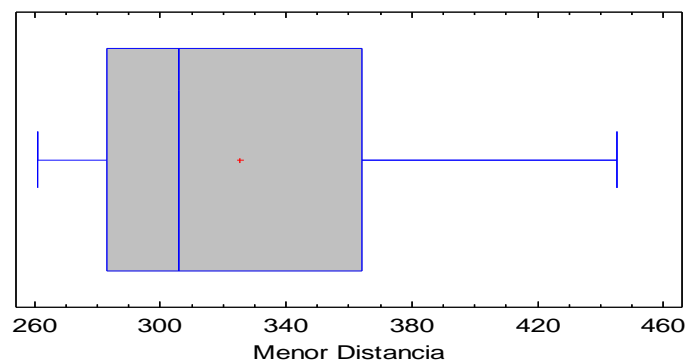


Fig. 33.- Diagrama de Caja.

El diagrama de caja y bigotes nos muestran que la mediana tiende a la izquierda, por tanto no es una distribución simétrica, lo cual sigue dando mayor soporte a los resultados.

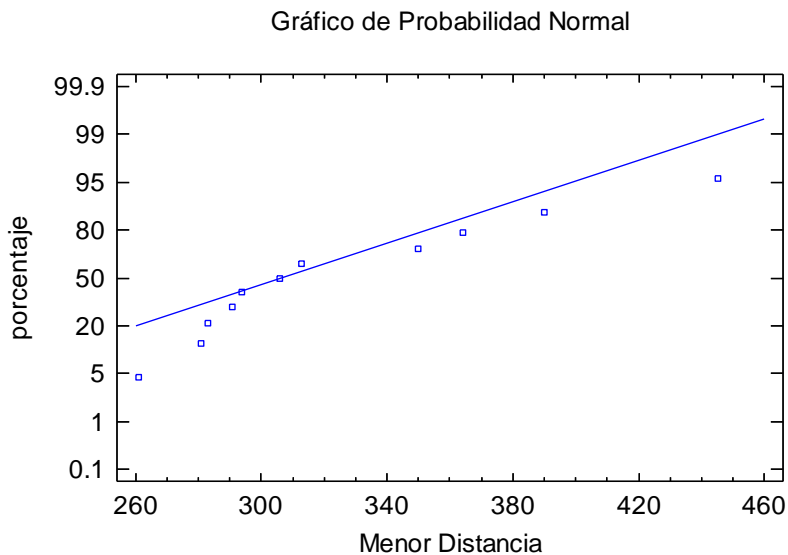


Fig. 34.- Gráfico de Probabilidad Normal para Menor Distancia.

La Figura 34 muestra que existe normalidad en los datos generados, a demás, se redujo el porcentaje de menores distancias por arriba del 50%.

Con estos resultados se puede validar a la funcionalidad del sistema a parte de los resultados económicos ya comentados.

Una vez encontrada la combinación óptima se realizo otro experimento, en el cual se tomaron 30 observaciones con unas combinaciones de agentes e iteraciones pequeños y otro con la combinación optima encontrada de 850 agentes y 850 iteraciones.

Tabla 6.- Resultados de corridas de software de Minería de Datos para 5 agentes y 2 iteraciones.

No.	Rutas	Mayor	Menor	Optimización	Épocas	Rutas	Mayor	Menor	Optimización	Épocas	Rutas	Mayor	Menor	Optimización	Épocas
1	1	41	41	0	8	2	102	89	13	9	3	52	52	0	8
2	1	46	36	10	29	2	90	90	0	8	3	55	55	0	8
3	1	44	36	8	11	2	95	95	0	8	3	57	57	0	8
4	1	45	41	4	10	2	99	90	9	17	3	45	45	0	8
5	1	41	41	0	8	2	104	101	3	9	3	53	53	0	8
6	1	48	37	11	9	2	101	97	4	9	3	42	42	0	8
7	1	43	38	5	13	2	105	86	19	22	3	52	52	0	8
8	1	35	35	0	8	2	103	92	11	10	3	61	61	0	8
9	1	48	40	8	15	2	112	91	21	91	3	41	41	0	8
10	1	46	40	6	13	2	95	95	0	8	3	52	52	0	8
11	1	38	37	1	10	2	110	92	18	13	3	70	70	0	8
12	1	45	41	4	10	2	113	89	24	20	3	55	55	0	8
13	1	45	40	5	12	2	106	97	9	12	3	53	53	0	8
14	1	43	36	7	11	3	53	53	0	8	3	59	59	0	8
15	1	45	38	7	13	2	94	90	4	11	3	58	58	0	8
16	1	47	37	10	18	2	104	90	14	12	3	49	49	0	8
17	1	50	36	14	16	2	104	90	14	11	3	62	62	0	8
18	1	46	40	6	14	2	100	100	0	8	3	55	55	0	8
19	1	38	37	1	10	2	99	99	0	8	3	52	52	0	8
20	1	46	38	8	16	2	107	87	20	17	3	56	56	0	8
21	1	45	39	6	12	2	109	87	22	11	3	54	54	0	8
22	1	51	39	12	9	2	107	90	17	14	3	61	61	0	8
23	1	44	40	4	9	2	100	89	11	11	3	62	62	0	8
24	1	47	41	6	10	2	94	94	0	10	3	56	56	0	8
25	1	49	41	8	14	2	109	89	20	12	3	39	39	0	8
26	1	47	38	9	15	2	111	92	19	10	3	49	49	0	8
27	1	46	39	7	12	2	113	89	24	21	3	55	55	0	8
28	1	48	37	11	11	2	90	85	5	9	3	55	55	0	8
29	1	46	36	10	10	2	110	84	26	17	3	61	61	0	8
30	1	44	37	7	22	2	102	86	16	15	3	57	57	0	8

La tabla 6 muestra los resultados de las 30 corridas del sistema de Minería de Datos con una combinación baja de agentes e iteraciones.

Tabla 7.- Resultados de corridas de software de Minería de Datos para 850 agentes y 850 iteraciones.

No. 1	Rutas	Mayor	Menor	Optimizacion	Epocas	Rutas	Mayor	Menor	Optimizacion	Epocas	Rutas	Mayor	Menor	Optimizacion	Epocas
1	1	36	30	5	14	2	81	76	5	9	3	31	31	0	8
2	1	36	31	5	10	2	80	78	2	11	3	36	36	0	8
3	1	34	30	4	17	2	81	75	6	11	3	41	41	0	8
4	1	33	29	5	18	2	79	76	3	8	3	37	37	0	8
5	1	34	29	4	13	2	84	74	10	10	3	32	32	0	8
6	1	33	33	0	13	2	75	72	3	12	3	36	36	0	8
7	1	34	32	2	12	2	81	74	7	16	3	37	37	0	8
8	1	35	30	5	9	2	86	78	8	12	3	39	39	0	8
9	1	34	31	3	10	2	83	73	10	18	3	37	37	0	8
10	1	32	32	0	8	2	79	79	0	14	3	37	37	0	8
11	1	35	29	6	12	2	85	75	10	12	3	32	32	0	8
12	1	34	32	2	14	2	79	72	7	9	3	32	32	0	8
13	1	35	33	2	10	2	81	71	10	20	3	33	33	0	8
14	1	36	31	5	11	2	80	72	8	17	3	37	37	0	8
15	1	32	30	2	18	2	78	76	2	9	3	31	31	0	8
16	1	33	32	1	9	2	81	73	8	12	3	34	34	0	8
17	1	32	32	0	8	2	83	72	11	11	3	36	36	0	8
18	1	33	32	1	13	2	87	73	14	9	3	35	35	0	8
19	1	30	30	0	8	2	81	75	6	16	3	40	40	0	8
20	1	36	28	8	18	2	83	76	7	16	3	31	31	0	8
21	1	35	34	1	9	2	78	78	0	8	3	35	35	0	8
22	1	34	28	6	14	2	78	72	6	14	3	33	33	0	8
23	1	32	32	0	8	2	81	72	9	13	3	41	41	0	8
24	1	34	32	2	9	2	82	74	8	22	3	34	34	0	8
25	1	33	31	2	13	2	81	75	6	15	3	33	33	0	8
26	1	36	30	6	12	2	81	69	12	13	3	35	35	0	8
27	1	34	29	5	10	2	83	72	11	14	3	36	36	0	8
28	1	31	28	3	15	2	75	72	3	10	3	36	36	0	8
29	1	37	31	6	10	2	84	71	13	12	3	35	35	0	8
30	1	36	31	5	15	2	75	74	1	14	3	36	36	0	8

Para el análisis de estas, se tiene que la sumatoria del recorrido menor en la Tabla 6 es de 1152 unidades, en tanto que en la Tabla 7 el recorrido mayor es 1018 que da un total de 134 unidades de recorrido menos, en tanto que el recorrido menor da un total de 922 unidades, que restado a el recorrido menor de la Tabla 6 da un total de 230 unidades menos.

La diferencia es bastante obvia, solo resta determinar algunos parámetros y gráficos que ratifiquen esto.

Tabla 8 de Frecuencias para Ruta A

<i>Clase</i>	<i>Límite Inferior</i>	<i>Límite Superior</i>	<i>Punto Medio</i>	<i>Frecuencia</i>	<i>Frecuencia Relativa</i>	<i>Frecuencia Acumulada</i>	<i>Frecuencia Rel. Acum.</i>
	menor o igual	34		0	0	0	0
1	34	35.3333	34.6667	1	0.0333	1	0.0333
2	35.3333	36.6667	36	5	0.1667	6	0.2
3	36.6667	38	37.3333	10	0.3333	16	0.5333
4	38	39.3333	38.6667	3	0.1	19	0.6333
5	39.3333	40.6667	40	5	0.1667	24	0.8
6	40.6667	42	41.3333	6	0.2	30	1
	mayor de	42		0	0	30	1

Media = 38.4 Desviación Estándar = 1.92264

La tabulación muestra una tabla de frecuencias que se va dividiendo el rango de Ruta A en intervalos del mismo ancho, y contando el número de datos en cada intervalo. Las frecuencias muestran el número de datos en cada intervalo, mientras que las frecuencias relativas muestran las proporciones en cada intervalo.

La Tabla 8 muestra que los datos tienen una media de 38.4, lo que indica que los recorridos propuestos para esta ruta oscilarán en este promedio.

Tabla 9.- Resumen Estadístico para Ruta A

Recuento	30
Promedio	38.4
Mediana	38
Moda	
Varianza	3.69655
Desviación Estándar	1.92264
Coefficiente de Variación	5.01%
Mínimo	35
Máximo	41
Rango	6
Sesgo	0.0074857
Sesgo Estandarizado	0.0167385
Curtosis Estandarizada	-1.54458

Esta tabla muestra los estadísticos de resumen para Ruta A. Incluye medidas de tendencia central, medidas de variabilidad y medidas de forma. De particular interés aquí son el sesgo estandarizado y la curtosis estandarizada, las cuales pueden utilizarse para determinar si la muestra proviene de una distribución normal. Valores de estos estadísticos fuera del rango de -2 a +2 indican desviaciones significativas de la normalidad, lo que tendería a invalidar cualquier prueba estadística con referencia a la desviación estándar. En este caso, el valor del sesgo estandarizado se encuentra dentro del rango esperado para datos provenientes una distribución normal. El valor de curtosis estandarizada se encuentra dentro del rango esperado para datos provenientes de una distribución normal.

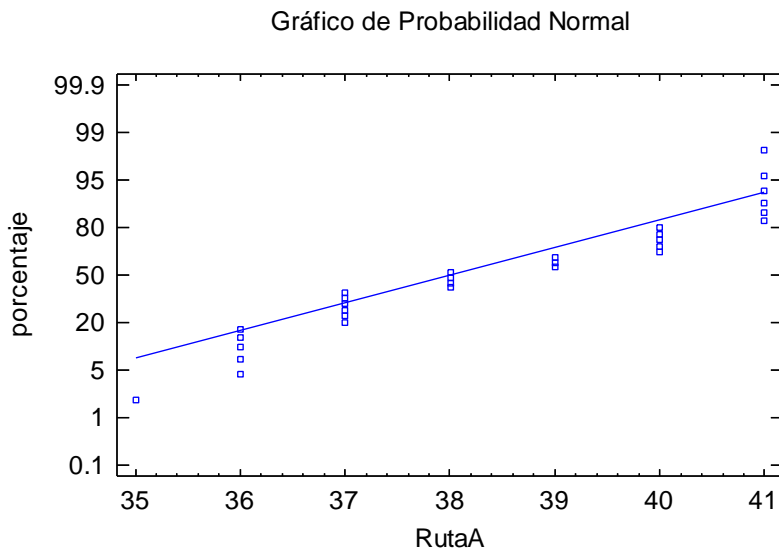


Fig.- 35 Gráfico de Probabilidad Normal para Ruta A.

Este gráfico nos muestra que existe normalidad en los datos analizados para la Ruta A y que aun cuando sus valores no son muy buenos, de cualquier forma tiende a minimizar los recorridos.

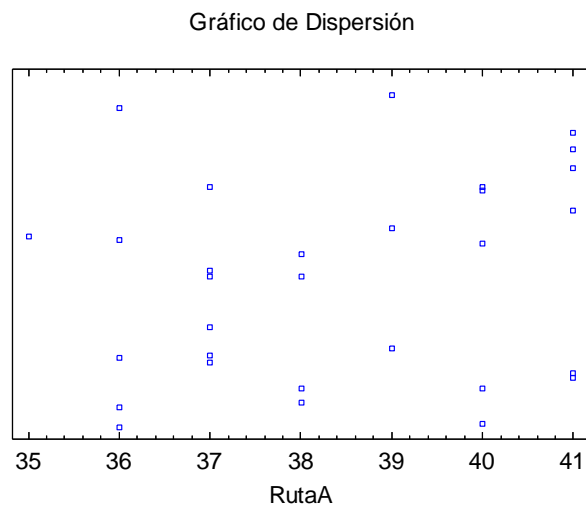


Fig. 36.- Gráfico de Dispersión para Ruta A.

La Figura 36 da la impresión de variabilidad, sin embargo, podemos observar que los valores están en un rango de 35 a 41 unidades.

Finalmente tenemos el análisis de las secuencias cronológicas para esta ruta Figura 37.

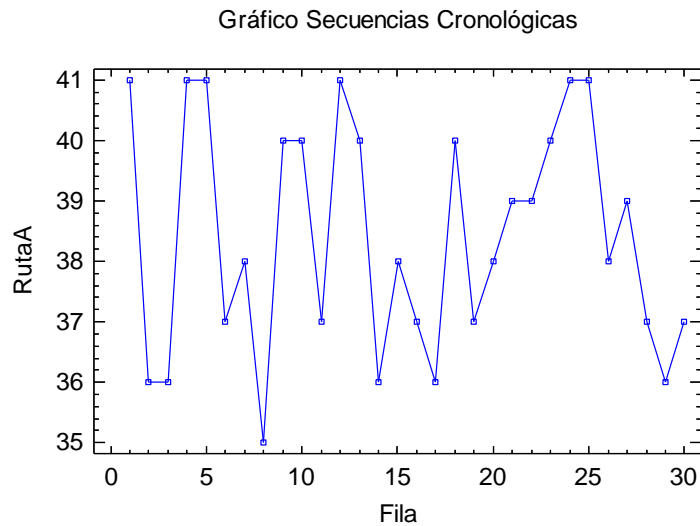


Fig. 37.- Gráfico de Secuencias Cronológicas.

El gráfico de la Figura 37 nos muestra el comportamiento cronológico de la ruta durante las 30 corridas, teniendo un límite superior de 41 unidades y un límite inferior de 35.

Con esto se puede mostrar que sin importar la combinación de agentes e iteraciones, el sistema de Minería de Datos muestra normalidad y coherencia en sus resultados, avalando así su aplicación.

En base a esto se realizó un análisis similar para la mejor combinación y los resultados se muestran a continuación:

Tabla 10.- de Frecuencias para Ruta B

	<i>Límite</i>	<i>Límite</i>			<i>Frecuencia</i>	<i>Frecuencia</i>	<i>Frecuencia</i>
<i>Clase</i>	<i>Inferior</i>	<i>Superior</i>	<i>Punto Medio</i>	<i>Frecuencia</i>	<i>Relativa</i>	<i>Acumulada</i>	<i>Rel. Acum.</i>
	menor o igual	27		0	0	0	0
1	27	28.3333	27.6667	3	0.1	3	0.1
2	28.3333	29.6667	29	4	0.1333	7	0.2333
3	29.6667	31	30.3333	12	0.4	19	0.6333
4	31	32.3333	31.6667	8	0.2667	27	0.9
5	32.3333	33.6667	33	2	0.0667	29	0.9667
6	33.6667	35	34.3333	1	0.0333	30	1
	mayor de	35		0	0	30	1

Media = 30.7333 Desviación Estándar = 1.57422

Esta tabulación de frecuencias al igual que para el análisis pasado se da dividiendo el rango de Ruta B en intervalos del mismo ancho, y contando el número de datos en cada intervalo. Las frecuencias muestran el número de datos en cada intervalo, mientras que las frecuencias relativas muestran las proporciones en cada intervalo. En contraste tenemos que la media de la prueba B, baja hasta 30 unidades, mostrando claramente las ventajas de utilizar combinaciones altas de agentes e iteraciones.

Con lo que respecta a las demás medidas de tendencia central, se tiene la siguiente tabla:

Tabla 11.- Resumen Estadístico para Ruta B

Recuento	30
Promedio	30.7333
Mediana	31
Moda	32
Varianza	2.47816
Desviación Estándar	1.57422
Coefficiente de Variación	5.12%
Mínimo	28
Máximo	34
Rango	6
Sesgo Estandarizado	-0.210448
Curtosis	-0.625626
Curtosis Estandarizada	-0.699471

Esta tabla muestra los estadísticos de resumen para Ruta B. Incluye medidas de tendencia central, medidas de variabilidad y medidas de forma. En este caso, el valor del sesgo estandarizado se encuentra dentro del rango esperado para datos provenientes una distribución normal. El valor de curtosis estandarizada se encuentra dentro del rango esperado para datos provenientes de una distribución normal. El promedio, mediana y moda se encuentran muy cercanos, en tanto que el rango entre el mayor y el menor dato es igual que en la muestra pasada por 6 unidades.

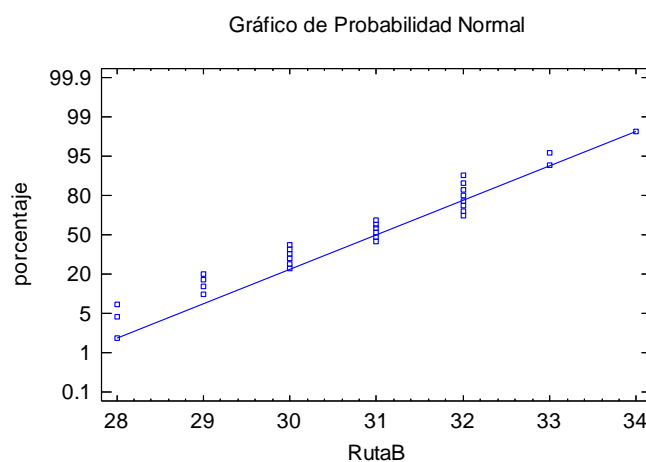


Fig. 38.- Gráfico de Probabilidad Normal para Ruta B.

El gráfico de la Figura 38, muestra la normalidad de los datos, aunado a esto, se detecta que se tiene una mayor acumulación de datos entre 30 y 32 unidades, lo cual reafirma la mediana que fue de 30 unidades.

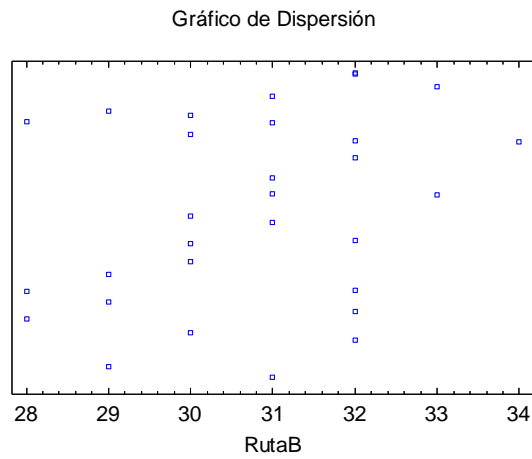


Fig. 39.- Grafico de Dispersión para Ruta B.

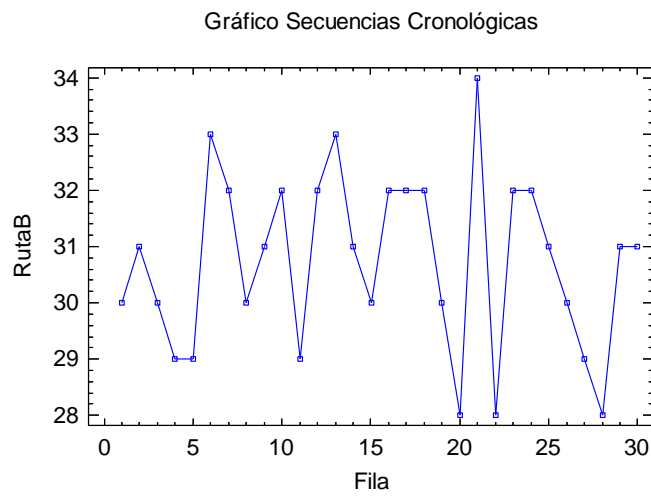


Fig. 40.- Grafico de secuencia cronológica para Ruta B.

El gráfico de dispersión así como el de secuencias cronológicas, reafirma la normalidad en los datos, los cuales se desplazan en un rango de 28 a 34 unidades.

Con lo anterior, al no tener un parámetro de comparación del sistema, estadísticamente se puede decir que el sistema está normalizado y que cumple las expectativas para lo cual ha sido creado.

4.5 Resumen del Sistema de Minería de Datos.

De acuerdo con los resultados obtenidos dentro de las corridas que se efectuaron se encuentra que la peor combinación es la tabla 1 con 2 agentes y una iteración, en tanto que la mejor combinación se dio con 850 agentes y 850 iteraciones; además se pueden comprobar dos tendencias:

4.5.1 Respecto al Kmeans:

El algoritmo del Kmeans se basa en la clusterización basada en los centroides que maneja el sistema los cuales se determinan inicialmente y se itera hasta poder alcanzar grupos homogéneos de acuerdo con ciertas características, para este caso en particular se puede decir que se tomaron en cuenta 3 de los clientes estrella para que fuesen los centroides iniciales y a partir de ahí establecer las rutas adecuadas.

El algoritmo tiene una condición de parada que consiste en cierto número de iteraciones los cuales se fueron asignando como un parámetro de funcionamiento del mismo.

Como se puede ver, entre más pequeño es el número de iteraciones, menos es el grado de equilibrio y menos efectivos los clusters resultantes ya que no se agrupan en forma adecuada provocando así, mayores recorridos; Por lo tanto a medida que se aumenta este número, es mejorado el proceso de clusterización, logrando con ello grupos más consistentes y distancias más cortas.

Por otro lado también se observa que los agentes tienden a optimizar el resultado a medida que se mejora el proceso de clusterización ya que facilita el trabajo de los mismos en términos de optimización.

4.5.2 Respecto al Cultural:

El parámetro a calibrar era el número de agentes, se esperaba que a medida que se incrementara éste, se tendría una mejor optimización, ya que la población sería más grande, tendrían más conocimiento a compartir, y por lo tanto tendería a converger de manera más rápida. Caso contrario cuando se tienen pocos agentes, debido al limitado conocimiento no optimiza lo suficiente en la búsqueda de un óptimo global.

Lo anterior se puede visualizar en la figura 28, tabla 1, que muestra una combinación de 1 iteración y 2 agentes, las distancias recorridas son mayores, el acumulado de la distancia total es igual a 461 unidades, y su recorrido final optimizado da un total de 445, solo una mejora de 16 unidades en 54 épocas.

Caso contrario encontramos con la combinación de 850 agentes y 850 iteraciones, la cual nos dio el óptimo de ejecución del sistema, con la cual se obtuvo un recorrido máximo total de 301 unidades, su recorrido optimizado fue de 261 unidades, es decir, una diferencia de 40 unidades en 56.

4.6 Comparativa del Algoritmo propuesto en contra de otro similar de Bin Packing.

Considerando la importancia que debe de existir entre una comparativa bajo el mismo dominio de aplicación, realizamos el análisis del software basado en AGs desarrollado en [33] y en donde se ubica la tabla siguiente [3], la cual detalla las ecuaciones concernientes a la implementación del Algoritmo de Bin packing, para ello mostramos los siguientes resultados considerando que la

parte más importante es el número de veces que es alcanzado el valor óptimo por los dos algoritmos a ser evaluados.

Tabla 12.-Indicadores propuestos en el Trabajo de Cruz [3] y que sirvieron de comparativa para el análisis de resultados de ambos algoritmos desarrollados.

Ecuación	Descripción
$p = \frac{n}{nmax}$	<p>p es el índice del tamaño del caso, donde n = número de objetos, $nmax$ = tamaño máximo solucionado.</p>
$t = \frac{\sum_{i=1}^n w_i / n}{c}$	<p>t es el índice de capacidad ocupada por un objeto promedio, donde w_i = peso del objeto i, c = capacidad del contenedor.</p>
$d = \sqrt{\frac{\sum_{i=1}^n \left[t - \left(\frac{w_i}{c} \right) \right]^2}{n}}$	<p>El índice de dispersión d expresa el grado de la dispersión del cociente del peso de los objetos entre el tamaño del contenedor.</p>
$f = \frac{\sum_{i=1}^n factor(c, w_i)}{n}$	<p>El índice de factores f expresa la proporción de objetos cuyo peso w_i es factor de la capacidad del contenedor, donde</p> $factor(c, w_i) = \begin{cases} 1 & \text{si } (c \bmod w_i) = 0 \\ 0 & \text{en caso contrario} \end{cases}$
$b = \begin{cases} 1 & \text{si } c \geq \sum_{i=1}^n w_i \\ \frac{c}{\sum_{i=1}^n w_i} & \text{en caso contrario} \end{cases}$	<p>El uso de contenedor b expresa la proporción del peso total que se puede asignar en un contenedor de capacidad c.</p>

4.6.1 Plan de Prueba.

Se utilizó una Computadora de Escritorio DELL procesador Intel Corel Duo, unidad C con capacidad de 232 GB.

Teniendo una instancia de 3 camionetas, con las siguientes características:

- Ford Ranger caja larga modelo 1998.
- Estaquita Nissan con panel modelo 1995.
- Chevrolet S10 modelo 1995 con panel.

Se generaron 500 experimentos, para los cuáles se iban acumulando las veces que realizaba un acomodo exitoso [33], como los AGs trabajan mediante generaciones y los ACs mediante épocas se procedió a incluir el número de veces que se tardaba en alcanzar dicho óptimo, generando la figura 41 sobre desempeño.

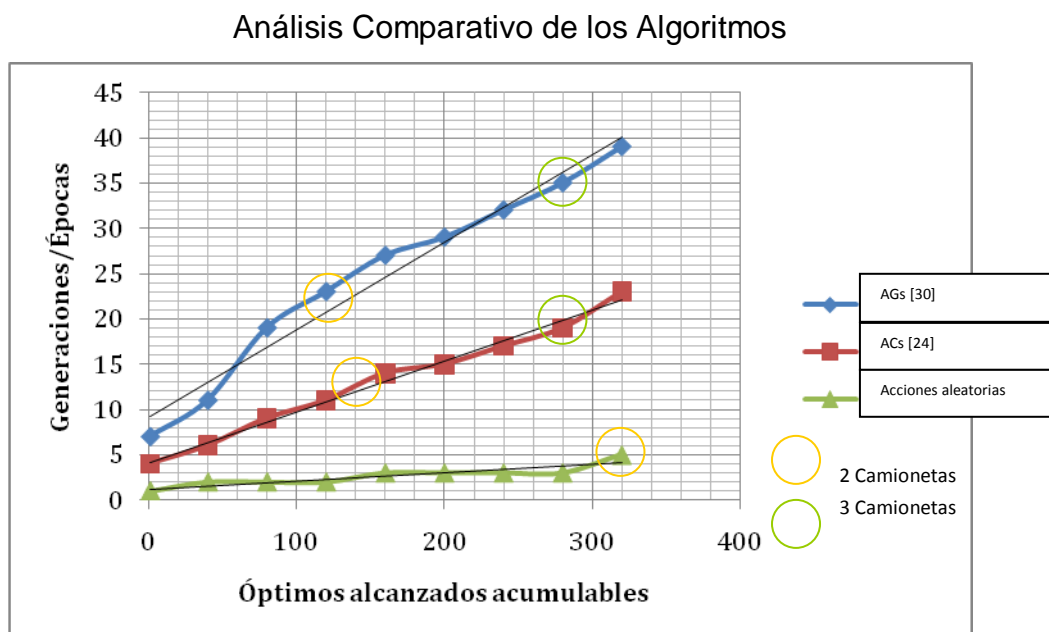


Figura 41.-Análisis comparativo de un Algoritmo Genético y un Algoritmo Cultural, para el Algoritmo de Bin Packing para una instancia de 3 camionetas considerando su desempeño con base en la Tabla 12.

El presente experimento ha intentado mostrar el desempeño de dos algoritmos diferentes bajo las condiciones de un mismo Dominio de Aplicación (Logística de Servicio) al incidir en compañías de objetos tridimensionales conteniendo líquidos y con acomodos basados en la demanda basados ambos en Cómputo Evolutivo. Los resultados obtenidos podrán ayudar a discriminar entre las diferentes posibles estrategias y estratagemas en la distribución del producto, al considerar que nuestro Algoritmo Cultural se basa en las consideraciones

iniciales generadas por el módulo de Data Mining. El conocimiento adicional y las heurísticas gobiernan la selección apropiada de patrones a seguir para un problema dado con especificaciones multirestrictivas permitiendo guiar en un balance adecuado entre itinerarios a realizar y el acomodo que deben de llevar cada una de las unidades de reparto permitiendo eliminar pobres resultados generados por acomodos aleatorios realizados por los empleados en la empresa donde se implementó este proyecto de investigación.

Varias consideraciones interesantes se han obtenido mediante el análisis de los resultados presentados aquí. Una mejor heurística se debe desarrollar para la selección del patrón apropiado para un horizonte de planeación de varias semanas en la empresa. Los patrones encontrados en la Logística de servicio deben distinguir entre temporadas de inclusión de clientes según si están comenzando, medio o un cliente establecido en el mercado del producto. Los procedimientos de evaluación todavía requieren de mucha mejora, dado que el AGs con el que nos comparamos no considera dichos patrones propuestos por la Minería de Datos. En una cuestión donde la opinión subjetiva del comprador genera la adquisición del producto, se debe hacer un esfuerzo especial para idear un procedimiento de la evaluación que proporcione un grado riguroso sin la interferencia con la actitud natural del mercado a cubrir y el tamaño de la compañía como proponente de esta clase de posicionamiento de un producto en una región específica

Finalmente se realizó un comparativo final con un algoritmo genético desarrollado en Matlab, se adecuó para el modelo de optimización de espacio en las camionetas de reparto y se realizaron 10 corridas con las mismas características, la única diferencia fue la metodología y el concepto de demanda. Los resultados se muestran en la Tabla 13:

Tabla 13.- Comparativo Genético vs Cultural.

Genético.						Cultural					
Corrida	20lt	1,5lt	1lt	,500lt	Total	Corrida	20lt	1,5lt	1lt	,500lt	Total
1	1	5	30	4	952	1	18	6	6	10	853,6
2	12	15	9	4	860	2	20	6	3	7	693,3
3	3	6	8	32	1578	3	17	7	5	6	694,5
4	3	3	4	23	1070	4	22	5	5	8	768,5
5	5	7	3	39	1780	5	18	7	5	11	897
6	2	5	8	38	1770	6	19	7	7	7	796,2
7	2	6	2	33	1474	7	21	6	5	8	785
8	6	10	3	26	1374	8	20	7	5	7	763
9	10	9	10	10	934	9	21	6	6	8	806,6
10	0	15	3	22	1292	10	23	5	8	9	881,8
13084						7940					

Los resultados parecieran ser contundentes con una diferencia de utilidad de \$5144, sin embargo la diferencia entre uno y otro radica en que el algoritmo genético no toma en cuenta la demanda de los productos de tal forma que al buscar la combinación óptima de carga optimiza en base a el producto que tiene mayor utilidad pero menor demanda, que en el caso es la presentación de .500lt.

La presentación con mayor demanda es la de 20lt, sin embargo irónicamente es el de menor utilidad, es por eso que el sistema de BPP ha dejado buenos resultados.

Con base en esto el Algoritmo Cultural resulto tener un mejor resultado frente al genético esto se observa en la figura 41:

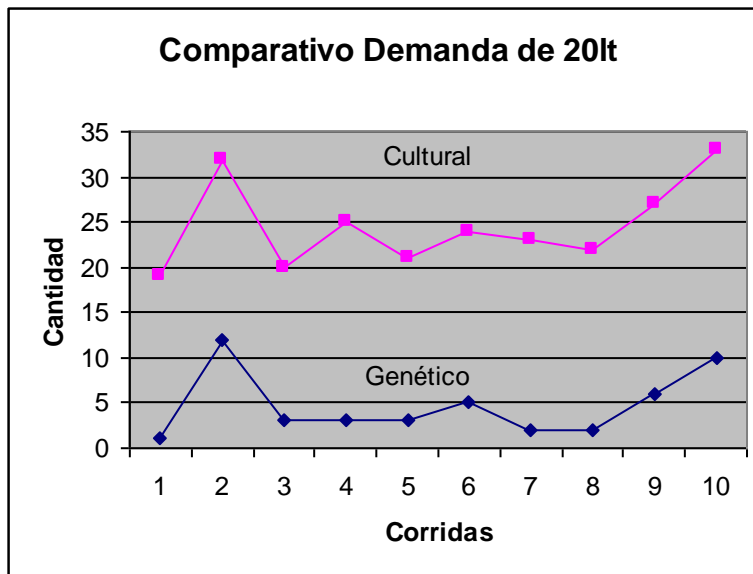


Fig. 41.- Comparativo para 10 corridas de presentación de 20lts entre el algoritmo cultural y el genético.

CAPÍTULO 5

5.1 Conclusiones y Trabajo Futuro.

En resumen, para realizar el trabajo dentro de la purificadora “La Noria” primeramente se tubo que identificar la problemática que estaba afectando a la mala distribución del producto, lo cual se realizo mediante una lluvia de ideas, posteriormente, se desarrollo un diagrama Ishikawa para visualizar los problemas encontrados y a partir de ahí empezar a trabajar en ellos.

Como primer etapa se tubo la elaboración de una herramienta inteligente basada en Minería de Datos y Algoritmos Culturales, la cual permite la generación de rutas en base a los históricos de venta; una vez que se tubo dicha herramienta y que se implemento, se procedió a la elaboración de otra pero esta vez el objetivo fue el acomodo dentro de los vehículos de distribución.

De esta manera se completo el sistema de distribución y venta de la purificadora “La Noria” en base a las características de clientes y demanda de productos.

Las técnicas de computación evolutiva han mostrado tener buen desempeño en muchos problemas de amplia complejidad, sin embargo se siguen explorando nuevas aplicaciones para estas heurísticas.

Con la implementación del proyecto así como sus resultados, se puede ver que ha sido benéfico para la empresa el desarrollo de herramientas inteligentes capaces de resolver un problema específico, aunado a ello se ha podido comprobar que la minería de datos en combinación con los algoritmos culturales conforman una buena herramienta híbrida para la elaboración de rutas.

Como se puede ver se desarrollo un sistema para la distribución de productos que permite optimizar costos en la empresa “La Noria” al crear rutas de distribución y decir que productos debe llevar consigo cada camión.

También hemos comprobado que el algoritmo cultural utilizando la ideología de BPP, es una correcta estrategia para optimizar el espacio de las unidades de reparto, en este caso dentro de nuestra flotilla de reparto de agua.

Entre las aportaciones más importantes de la investigación podemos resaltar:

- ❖ Un Modelo de Minería de Datos para analizar el patrón de consumo en la zona de distribución de la empresa (García et al. 2009) [25].
- ❖ Un Sistema Híbrido para mejorar la Logística de Reparto y la mejor distribución en el Parque Vehicular de la Empresa (García et al. 2010) [28].
- ❖ La implementación de un Algoritmo Cultural y su convalidación para el Algoritmo de acomodo BPP (García et al. 2010) [27].
- ❖ El análisis multivariable de las Rutas generadas por la Herramienta desarrollada y su puesta a punto por la Empresa (García et al. 2010) [26].
- ❖ La combinación de diferentes técnicas evolutivas aplicadas a Logística de Servicio.(Alberto Ochoa et al.2010) [30]
- ❖ Se ubicó el eslabón al cual pertenece la problemática tratada dentro de la cadena de valor.
- ❖ Se generaron rutas específicas de distribución del producto.
- ❖ La solidificación de la marca en Fresnillo y su región; ahora la purificadora “La Noria” es conocida y su cobertura va creciendo día con día.

Se logro a demás, identificar los principales factores que intervenían en la distribución y venta los cuales se mencionan:

- Nula Logística de Servicio.
- Mal Servicio al Cliente.
- Mala estructura de venta.
- Escaso conocimiento de comportamiento de la compra (Cliente).
- Falta de aprovechamiento de los recurso.

Con la implementación de las herramientas desarrolladas se determinaron aquellas zonas en las que existen mayores posibilidades de venta considerando que la incursión y la permanencia sean factibles dentro del área de cobertura de la empresa.

Como conclusión final, se tiene que se desarrollo un sistema sólido de distribución del producto que permite brindar el servicio que el mercado merece y que a la vez esto lo respalda la confiabilidad del mismo mediante la demanda del producto determinando así que la hipótesis planteada se acepta en base a los resultados obtenidos.

Finalmente es importante resaltar que aun cuando el proyecto está siendo ejecutado obteniendo resultados positivos, se tiene una gran área de oportunidad y por ende se puede realizar muchas aportaciones como trabajo futuro entre las cuales se mencionan las siguientes:

Adecuación en base a restricciones, es decir, ajustarse también a un horario que pueda manejar el cliente debido a su trabajo u ocupación particular.

Adecuación en base a la demanda del cliente, cuando en una ruta se tiene varios clientes que requieren el producto a un cierto horario, ajustar las rutas de tal manera que se satisfaga el servicio.

Asignación de categoría al cliente en base a la compra, si bien al momento de elaborar las rutas el sistema determina en mediante el porcentaje de compra del lo cliente el orden de visita, el sistema también podría también podría asignar una categoría (a, b, c...etc.).

Lo anterior solo por mencionar algunos ejemplos, ya que sin duda el área de oportunidad puede ser más amplia dado la flexibilidad de las heurísticas manejadas y la posibilidad de combinación con otras tecnicas evolutivas que pudieran ser desarrolladas e implementadas para continuar con el trabajo ya desarrollado.

5.2.- Bibliografía.

1. - Ajith A. et al. (Eds.): Swarm Intelligence in Data Mining (Book). Berlin, Germany; Series: Studies in Computational Intelligence, Vol. 34, Springer-Verlag, ISBN: 3-540-34955-3, 270 p. (September 2006).
- 2.- Araujo Lourdes, Cervigón Carlos.: Algoritmos Evolutivos., Un enfoque práctico. Ed. Alfa omega 2009.
- 3.- Cruz L. Caracterización de Algoritmos Heurísticos Aplicados al Diseño de Bases de Datos Distribuidas. Tesis de Doctorado, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Morelos, México, 2005.
4. - Darwin, C.: On the Origin of Species. John Murray, (1859).
5. - Dasarathy/Belur V: Data Mining and Knowledge Discovery: theory, tools, and Technology, Aprill (2001), Orlando.
6. - De Jong, K.A. An Analysis of the behavior of a class of genetic adaptative systems. PhD thesis, Ann Arbor, MI, USA, (1975).
- 7.- Escalante Eduardo J. Seis Sigma; Metodología y Técnicas, Lumusa, México 2008.
8. - Fernandez George.: Data Mining Using SAS Applications. CHAPMAN&HALL/CRC 2002.
- 9.- Flores López Raquel/Fernandez Fernandez José Miguel “Las Redes Neuronales Artificiales” Fundamentos Teóricos y aplicaciones. España 2008.
- 10.- Francés Antonio, Estrategia y Planes para la Empresa.; Con el cuadro de Mando Integral. Prentice Hall. Año (2006).

11. - Gill S. et al.: Data Warehousing. La Integración de la Información para la mejor toma de decisiones. Mexico: Prentice Hall (1996).
12. - Gonzalez Teofilo F.; Handbook of Approximation Algorithms and Metaheuristics. CHAPMAN & HALL/CRC 2007.
13. - Gutiérrez Pulido Humberto/ De la Vara Salazar Román;. Control Estadístico de Calidad y Seis Sigma. CIMAT, Ed. McGrawHill, Guanajuato México.
14. - Hand David. Mannila Heikki.; Smyth Padhraic; PRINCIPLES OF DATA MINING 2001.
15. - Holland, J.H. adaptation in Natural and Artificial Systems. Ann Harbor, MI: Univ. of Michigan Press, (1975).
16. - Koza, J.R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Mit Press, (1992).
17. - Koza, J. R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Mit Press, (1994).
18. - Kuehl Robert. O. Diseño de Experimentos Principios estadísticos. 2da. Edición. Thomson (2001).
- 19.- Landa-Becerra R.: Uso de Información del Dominio para Mejorar el Desempeño de un Algoritmo Evolutivo. CINVESTV PhD Thesis (2007).
- 20.- Long Douglas.; Logística Internacional; Administración de la Cadena de Abastecimiento Global. Editorial Limusa (2008).
- 21.- Michalewicz, Z. Genetic algorithms + Data Structures = Evolution Programs. Springer – Verlag, 2dn edition, (1994).

- 22.- Molina Félix Luis Carlos;. "Torturando los Datos hasta que confiesen"
México D.F. Noviembre 2002.
- 23.- Morillo, Marysela. "Análisis de la Cadena de Valor y de la Cadena de Valor Agregado para las Pequeñas y Medianas Empresas".; Actualidad Contable FACES, 2005 Mérida Venezuela.
24. - Muñoz A, Hernandez A, and Villa E: Constrained optimization via particle evolutionary swarm optimization algorithm (PESO). In: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2005, pp.209-216. Association for Computing Machinery, (2005).
25. - Ochoa Alberto, García Yazmani.: "Logistics Optimization Service using Artificial intelligence". Hybrid Artificial Intelligence Systems. 5th International Conference HAIS 2010 San Sebastian, Spain, June 2010 Proceedings, Part II.
- 26.- Ochoa Alberto, García Yazmani.: "*Logística de Servicios Mediante Minería de Datos e Inteligencia Artificial*" Congreso Nacional de Sistemas Computacionales "CONACICO" Universidad Autónoma de Zacatecas. Marzo-Abril 2009.
27. - Ochoa Alberto, García Yazmani.: "Logistics Optimization Service using Artificial intelligence". IEEE R8 International Conference on Computational Technologies in Electrical and Electronics Engineering. SIBIRCON 2010.
28. - Ochoa Alberto, García Yazmani, Yañez Javier, Teymanoglu Yaddik.: "Using Cultural Algorithms to Improve Intelligent Logistics", Hybrid Intelligence Systems, Tijuana México 2010.
- 29.- Ochoa A. & González S.: "Simulación Social de una Sociedad Artificial basada en Algoritmos Culturales", (IJSA 2011-0626) International Journal of South American Archeology. (2009).

30. - Ochoa Alberto "Crowdfunding to support a Social Networking in a Diorama Project to a Museum" Suiza septiembre 2010.
31. - Olivera Alfredo "Heurísticas para Problemas de Ruteo de Vehículos" Montevideo, Uruguay Agosto 2004.
- 32.- Pau Cos Jordi. De Navascues Ricardo.; Manual de Logística Integral. Ed. Díaz Santos S.A. Madrid España. Año (1998).
- 33.- Piatetski-Shapiro G.; Frawley W.J; Knowledge Discovery in Databases. Ed. AAI/MIT Press. (1991).
- 34.- Quiroz Castellanos Marcela.: Caracterización de Factores de Desempeño de Algoritmos de Solución de BPP, CD Madero, Tamps., México, Noviembre 2009.
- 35.- Restrepo Puerta Luis Fernando.; Interpretando a Porter. Centro Editorial. Año (1978).
36. - Reynolds G. R., Sverdlik W.: "Problem Solving Using Cultural Algorithms", International Conference on Evolutionary Computation Orlando, Florida, U.S.A. (1994).
37. - Reynolds, G. R., "An Introduction to Cultural Algorithms." In Proceedings of the 3rd Annual Conference on Evolutionary Programming, World Scientific Publishing. (1994).
- 38.- Rojas López Jesús Carlos: Uso de la minería de datos con fines predictorios de la infraestructura de seguridad de redes Monterrey, N. L.: (2004).
- 39.- Spiegel Murray R. Estadística Segunda Edición Shawn Ed. McGrawHill (1991).

- 40.- Vallejo Sofia "Minería de Datos" Corrientes Argentina Enero 2006.
- 41.- Vilar Barrio José Francisco.; Como Implantar y Gestionar la Calidad Total. 2da Edición. FUNDACION CONFEMENTAL. Año (1997).
42. - Witten H Ian & Eibe Frank: Data Mining: Practical Machine Learning Tools and Techniques. Second Edition ELSEVIER. Año (2005).

ANEXO A

LNAI 6077

Hybrid Artificial Intelligence Systems

5th International Conference, HAIS 2010
San Sebastián, Spain, June 2010
Proceedings, Part II

2
Part II

 Springer

Logistics Optimization Service improved with Artificial Intelligence

Alberto Ochoa^{1,2}, Yazmani Garcia², and
Javier Yañez²

¹Juarez City University, Mexico
¹CIATEC, Mexico
alberto.ochoa@uacj.mx

Abstract. Today the issue of logistics is a very important within companies to the extent that some have departments devoted exclusively to it. This has evolved over time and today is a fundamental aspect in the fight business seeking to consolidate or remain leaders in their field. With the above we know that logistics can be divided into different classes, however, in this regard, our study is based on the timely distribution to the customer with a lower cost, higher sales and better utilization of space resulting in excellent service. Finally, prepare a comparative analysis of the results with respect to another method of optimization solution space.

Keywords: Logistics, Data Mining, Cultural Algorithms, Population Space, Space of Beliefs, Protocol, Bin Packing, Simplex Method.

1 Introduction

Within the area of distribution of purified water, there isn't methodology to what the Service Logistics and space optimization in delivery vehicles, so that the service within the "La Noria" become at the logic or the need. But not have defined a pattern of optimal service logistics.

The optimization problems have been attacked widely in the area of evolutionary computation; this has been due largely to the kindness they have shown to solve such problems [2] to name a few.

This paper addresses the solution of Logistics Service Based on Data Mining in combination with other techniques such as evolutionary algorithms, as is cultural algorithm, once the tool is implemented together with the Bin Packing is to optimize for the space within distribution vehicles purified water.

At present there are many optimization techniques, such is the case of the Simplex Method, and Simplex Method is an iterative process that progressively allows an optimal solution to linear programming problems [15] to name a few, the main feature of this method is that it attacks the problem by restricting its maximum capacity through the vertices of the same [10] to name a few.

2 Methodology

Data mining is a process that uses several data analysis tools to discover patterns and relationships in data that can be used to make valid predictions [16] to name a few. The foundations of data mining is in the artificial intelligence and the statistical and using the models extracted using data mining techniques which addresses the solution to problems of prediction, classification and segmentation [14] to name a few.

The process for conducting the data mining are:

- Selecting the dataset.
- Analysis of the properties of the data.
- Transforming the input data set.
- Select and apply the technique of data mining.
- Evaluate the results. (See Figure 1).

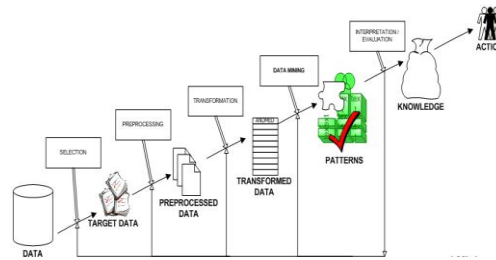


Fig. 1. Diagram of Data Mining

Data mining aims to generate information similar to that which could generate a human analyst: patterns, associations, changes, anomalies and significant structures.

The Cultural algorithms (CA's) are a class of computational models derived from observing the process of cultural evolution in nature as proposed by Reynolds in [11] to name a few. They consist of a population and a belief space as shown in Figure (2). The selected individuals from the population space contribute to cultural knowledge through the role of acceptance [8] to name a few. Cultural knowledge resides in the space of beliefs where it is stored and updated based on individual experiences of either success or failure. The knowledge in the belief space can also be used to influence their individual memories [9] to name a few. The (AC's) in addition to space and space population of beliefs, have a third component of importance: communication protocol, describes how knowledge is exchanged between the first two components [13] to name a few. The population space can support any population based on a computational model, such as Genetic Algorithms and Evolutionary Programming (See Figure 2).

Cultural Algorithms are a dual system of inheritance that characterizes the evolution of human culture in the macro-evolutionary level, which occurs within the space of beliefs, and micro-evolutionary level, which occurs in the area of population [4] to name a few. The knowledge produced in the population that the space in the micro-evolutionary level is accepted or to be passed to the belief space and used selectively to adjust the knowledge structures there [12] to name a few. This knowledge can then be used to influence the changes made by the population in the next age.

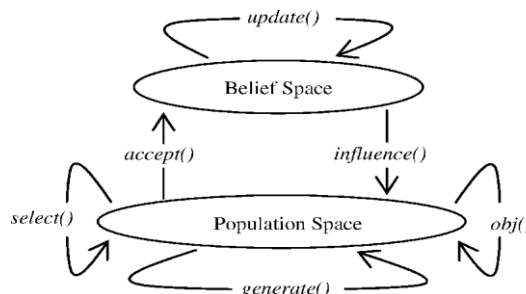


Fig. 2. Diagram Concept of Cultural Algorithms.

Cultural algorithms using five kinds of basic knowledge to generate an adequate solution in the search space where the problem is solved. The sources of expertise include regulatory knowledge (ranges of acceptable behavior), situational awareness (or copies of reports of successful or unsuccessful solutions, etc.), domain knowledge (knowledge of objects in the domain of relations they and their interactions), historical knowledge (temporal patterns of behavior), and topographical knowledge (spatial patterns of behavior) [5] to name a few. To make your programming is relatively simple, as shown in the pseudocode in (See Fig 3).

```

Begin
   $t = 0$ ;
  initialize  $B^t, P^t$ 
  repeat
    evaluate  $P^t \{obj()\}$ 
    update ( $B^t, accept(P^t)$ )
    generate ( $P^t, influence(B^t)$ )
     $t = t + 1$ ;
    select  $P^t$  from  $P^{t-1}$ 
  until (termination condition achieved)
End

```

Fig 3. Cultural Algorithm Pseudocode.

The main difference between Cultural Algorithms and Evolutionary Computation other techniques lies in the belief space utilization as well as the cultural influence of the same, as are those that guide us in obtaining optimum best [3] to name a few.

3 Tools developed

The prototype (see Figure 4) is an intelligent hybrid system developed with the Java programming language that combines data mining techniques to cultural algorithms. He got a map of Fresno and divided into 4 zones. To get the coordinates of each colony, then we started building the data warehouse, organized in the following fields: area, district, year, month, day, time and sales.

It was necessary the creation and implementation of algorithms capable of finding information in n dimensions, as well as a data clustering algorithm called K-Means [1] to name a few, which generates data pooling, without predefined classes, based on a function of similarity of the values that have different attributes, done in unsupervised [7] to name a few (i.e., discover patterns or trends in the data). K-Means is a partitional clustering method (i.e. we start altogether the particular), where partitioning is performed a database of n objects in a set of k groups, seeking to optimize the chosen partitioning criterion. In K-Means each cluster of data is represented by a centroid. K-Means is trying to form k groups with k predetermined before the start of the process. The goal is to minimize the within-group variance [6] to name a few.

With the use of these tools Data Mining, the prototype is able to determine (given a zone and a particular date) in which colonies are more likely to occur at a given time sale. Once these colonies is generated a population of n agents (an agent is the computer simulation of a person), which form a society based on cultural algorithms, which are responsible for determining, over the ages, the course optimal performance.

At the time zero (when initializing the program, and the agents beliefs have an empty space), all agents obtain the information generated from each colony, each propose a route, it will lead to negotiations between the agents to select the best route proposed at a certain time, the belief space will be updated only when the proposed route is better than the previously stored in the belief space, beginning a cycle of improvements that will be interrupted when they occur many times m (iterations in the behavior of agents) without improvements to the paths or when a stop condition is performed.

Once the data mining software, proceeded to the development of software for Bin Packing, the prototype is a hybrid intelligent system developed in Java, using the technique of Cultural Algorithms.

It began with the taking of measures of distribution vehicle on which the study was conducted to determine the space with which states in m^3 and measuring the various presentations and their respective volume, for in doing so raise the problem and their respective restrictions.

Also taken into account the demand of different presentations, to thus more accurately determine a product's usefulness.

Table 1. Product Description.

PRODUCT	CAPACITY	% DEMAND	VOLUME	UTILITY
1	20 lt	45	36500cm ³	\$ 10.00
2	1.5 lt	15	26731cm ³	\$ 26.50
3	1 lt	15	18435cm ³	\$ 21.60
4	500 lt	25	18177cm ³	\$ 38.50

Once the measurements were made, it was necessary to create an algorithm capable of finding the right mix in terms of cargo is concerned, so as to optimize the gain of the pickup and hence of the company, the algorithm uses a population basis, and initializes the other as an area of belief at that time its value is unknown.

It makes adjusting the demands for this way fine-tune the value, which based on the percentage of sales of products. Initial population is evaluated based on the problem and the same restriction as shown in the table 1.

$$\text{Max } z = r_1 m_1 + r_2 m_2 + \dots + r_n m_n$$

$$\text{Subject to } v_1 m_1 + v_2 m_2 + \dots + v_n m_n \leq V \quad (1)$$

$$m_1, m_2 \dots m_n = 0 \text{ Integer}$$

$$V = 1138425 \text{cm}^3$$

Where:

- r: Profit per unit.
- v: Volume of each unit.
- m: Are the units of each product type.
- V: Maximum volume capacity.

Any condition which violates the restrictions will be penalized so that only the best combinations are obtained and thus we get an average by which we can have the best individuals and thus influence the next generation based on the mean individuals.

The result (Epochs) is the proposed solution will stop condition from which is repeated 7 times without change, i.e., be = > to previous.

4 Results

The prototype used a database of sales generated at random, with it launched the system functions: information classification, clusters, the generation of routes (see Figure 4).

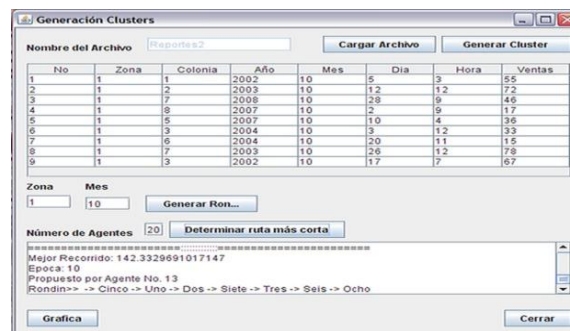


Fig 4. Generating optimal routes.

The system will determine the colonies in which sales have been registered within the specified date range. Based on the number of colonies and vehicle using the K-Means algorithm clusters to be generated immediately after being delivered to a society formed by artificial intelligent

agents (representatives of a group of individuals), which will determine the most optimal route for each cluster (see Fig 5).

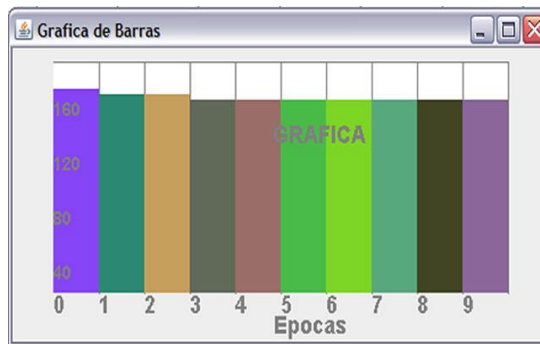


Fig 5. Graphic Convergence Times represented in épocas.

Moreover, the software for the Bin Packing use an initial population of 100 individuals and an area of beliefs with the same number of individuals, and that they could launch the system where the initial population is initialized, the Space of Beliefs, evaluated the results thus able to apply variation operators under the influence of belief space and get different times to achieve the status of unemployment or the convergence of results.

The program determines the best combination of load so as to maximize his profit, that based on the volumes handled for each presentation and the capacity of trucks in m^3 and the demand for each presentation (see figure 6).

```

// Código para saber si se mejora a través de las épocas y llegar a la
// convergencia para la condición de salida
if(temporal[4]==espacioCR){
    for(int fr=0; fr<temporal.length;fr++){
        System.out.println("temporal "+ temporal[fr]);
    }
    espacioCR=temporal[4];
    epoca=epoca+1;
}
    
```

```

ptres valeInfinity
pouatro valeInfinity
epoca 30
epocaSC 5
espacioCR 853.6
puno valeInfinity
pdos valeInfinity
ptres valeInfinity
pouatro valeInfinity
epoca 31
epocaSC 5
espacioCR 853.6
puno valeInfinity
pdos valeInfinity
ptres valeInfinity
pouatro valeInfinity
epoca 32
epocaSC 7
espacioCR 853.6
puno valeInfinity
pdos valeInfinity
ptres valeInfinity
pouatro valeInfinity
epoca 33
epocaSC 8
espacioCR 853.6
Process completed.
    
```

Fig 6. Generation of Bin Packing.

At the time 33 was found the best result after 8 times without change, that the following responses:

Table 2. Results.

PRESENTATION	20lt	1.5lt	1lt	500 lt
QUANTITY	16	6	6	9
NET PROFIT \$ 853.60				

However, making a comparison with what is the Simplex Method, we can see that this method gives us less effective results, as we see in Table 3.

Table 3. Result of the Simplex Method

14:37:33		Friday	April	17	2009			
Decision Variable	Solution Value	Unit Cost or Profit c(j)	Total Contribution	Reduced Cost	Basis Status	Allowable Min. c(j)	Allowable Max. c(j)	
1	X1	0.0225	10.0000	0.2250	0	basic	0	M
2	X2	0.1000	26.5000	2.6500	0	basic	0	M
3	X3	0.1500	21.6000	3.2400	0	basic	0	M
4	X4	0.5000	38.5000	19.2500	0	basic	0	M
Objective Function		(Max.) =	25.3650					
Constraint	Left Hand Side	Direction	Right Hand Side	Slack or Surplus	Shadow Price	Allowable Min. RHS	Allowable Max. RHS	
1	C1	0.4500	<=	0.4500	0	0.5000	0	615.8346
2	C2	0.1500	<=	0.1500	0	17.6667	0	63.1710
3	C3	0.1500	<=	0.1500	0	21.6000	0	61.0709
4	C4	0.2500	<=	0.2500	0	77.0000	0	31.1428
5	C5	15,348.1000	<=	1,138,425.0000	1,123,077.0000	0	15,348.1300	M

As we can see from the table, the maximum utility proposed by the Simplex Method is \$ 615.8346, i.e. less than the \$ 237.76 proposed by our program.

5 Conclusions

This research is being used within the purified "La Noria" trying to demonstrate that data mining can be used to increase sales and have a better logistics services, this software has a high value added for the generation and analysis logistics coupled with cultural algorithms responsible for the creation of routes, founded as a tool for decision making, based on the data generated daily by mobile sales. This is intended to provide the product to the larger population that requires service.

Similarly, using Bin Packing algorithm optimizes the space within the distribution units and thus provide the company a way to optimize new and creative through the use of this heuristic.

It concludes with the work that has been done for this kind of logistics in terms of logistics service and space optimization of delivery vehicles is satisfactory and allows the tool to see implemented are good choices. These tools have a high added value because it had not previously been used for this purpose, this further if we consider that until recently there was no practical implementation of these algorithms.

References

1. Ajith A. et al. (Eds.): Swarm Intelligence in Data Mining (Book). Berlin, Germany; Series: Studies in Computational Intelligence, Vol. 34, Springer-Verlag, ISBN: 3-540-34955-3, 270 p. (September 2006).
2. Araujo Lourdes, Cervigón Carlos: Algoritmos Evolutivos: un enfoque practico. Primera Edicion, Alfaomega Grupo Editor.México.
3. Calvillo Moreno Edgar Alan, Muñoz Arteaga Jaime, Alvarez Rodriguez Juan Manuel, Gonzalez Calleros, Rojano Rafael, De Casto Antonio: Acceso a Repositorios de Objetos de Aprendizaje a Través de un Sistema de Gestión de Contenidos, Conferencia Conjunta Iberoamericana sobre Tecnologías de Aprendizaje (CCITA09), July 6-10, 2009, Mèrida Yucatán, México.
4. Cowan George S. Reynolds Robert G. Acquisition of Software engineering knowledge. Vol. 14 Worl Cientific.
5. Dasarathy/Belur V: Data Mining and Knowledge Discovery: theory, tools, and Technology, April 2001, Orlando.

6. Gill S. et al.: Data Wherehousing. La integracion de la informacion para la mejor toma de decisions. Mexico: Prentice Hall (1996).
7. Landa-Becerra R.: Uso de Informacion del Dominio para Mejorar el Desempeño de un Algoritmo Evolutivo. CINVESTV PhD Thesis (2007).
8. Muñoz A, Hernandez A, and Villa E: Constrained optimization via particle evolutionary swarm optimization algorithm (PESO). In: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2005, pp.209-216. Association for Computing Machinery, (2005).
9. Muñoz Arteaga Jaime, Santaolaya Salgado Rene, Alvarez Rodri-guez Josh Squires, Acosta Ricardo, Mendoza Gonzalez Ricardo: Aprendizaje Multiculturales, Book Chapter titled: Topicos selectos de Tecnologia Educativa, Universidad de Colima, Compilador, Acosta Ricardo, April 2010.
10. Ochoa A. & Gonzalez S.: Simulación Social de una Sociedad Artificial basada en Algoritmos Culturales, (IJSA 2011-0626) Internationalt Journal of South American Archeology. (2009).
11. Prawda Juan: Méodos y Modelos de Investigación de Operaciones: Modelos Determinísticos, Vol 1 Ed. Limusa.
12. Reynolds G. R., Sverdlink W.: Problem Solving Using Cultural Algorithms, International Conference on Evolutionary Computation (1994).
13. Reynolds G. R.; Peng B. and Whallon R.: Emergent Social Structures in Cultural Algorithms (2008).
14. Rojas López Jesús Carlos: Uso de la minería de datos con fines predictorios de la infraestructura de seguridad de redes Monterrey, N. L.: 2004.
15. Taha, H.: Investigación de Operaciones. Séptima edición, México D.F., Prentice Hall. Pp. 71-90. (2004).
16. Witten H Ian & Eibe Frank: Data Mining: Practical Macbine Learning Tools and Techniques. Second Edition ELSEVIER.