

CALIDAD+IA, SOFTWARE BASADO EN INTELIGENCIA ARTIFICIAL PARA LA GESTIÓN DE LA CALIDAD EN LA PRODUCCIÓN DE HABANOS.

Ing. Godersi Alonso Rodríguez

Fábrica de Tabaco Torcido de Camajuaní, Cuba

Dra. Yanet Rodríguez Sarabia

Centro de Estudios de Informática, Universidad Central de Las Villas, Carretera Camajuaní, Villa Clara, Cuba

Resumen:

En todo proceso productivo tiene gran importancia evitar que los defectos que se pueden eliminar a tiempo carguen el siguiente eslabón del proceso. En una fábrica de tabaco, para cada lote de producción, se verifica el cumplimiento de los parámetros requeridos. Esta actividad la realizan los técnicos de calidad y como resultado se discrimina el tabaco que no cumpla con los estándares para su exportación. Al final del día un experto - persona con muchos años de experiencia en este ramo - diagnostica la causa que incidió con más fuerza en los defectuosos para su eliminación e incidencia favorable en la calidad de la producción del día siguiente. Cuando se cuenta con cientos de tabaqueros con producciones de 200 unidades como promedio, este diagnóstico se retarda y la corrección muchas veces se puede realizar cuando ya se ha hecho la mitad de la tarea del siguiente día.

Este trabajo presenta un software que aplica técnicas de la Inteligencia Artificial para simular el conocimiento de un experto, lo cual permite multiplicar el saber de los mismos a fin de garantizar diagnósticos confiables y rápidos. Además ha sido enriquecido con gráficos que ayudan a descubrir comportamientos, pudiendo hacerse valoraciones personalizadas por producto, intervalo de fechas, trabajador, etc. Ha sido desarrollado como código abierto en Java permitiendo que otros desarrolladores puedan ampliar y/o mejorar su desempeño, siendo esto una política de nuestro estado, en el que el conocimiento no se considera propiedad de nadie sino de la humanidad.

Palabras Claves:

IA, Inteligencia Artificial, Calidad, tabaco, experto, árbol de decisión, reglas, RNA, ConFuCiuS.

Introducción:

Para todo proceso productivo es de una importancia crucial ofrecer la máxima calidad en sus producciones, por ello es trascendental dejar de enarbolar como una consigna la gestión de la calidad en cada punto de producción sin llegarla a materializar y ni siquiera saber que significa evitar que los defectos que se pueden eliminar a tiempo carguen el siguiente eslabón del proceso. Por ello todo lo que se haga en la dirección de aumentar la calidad y disminuir el por ciento de las producciones defectuosas será muy bien recibido en todas las empresas, máxime si esta está en Perfeccionamiento Empresarial y está además sumida en un proceso de establecimiento de las ISO9000 que la hará más competitiva y adaptable en las arenas del mercado internacional. Este es el caso de la Fábrica de Tabaco Torcido de Camajuaní, donde se desarrolló el presente trabajo, la cual es una Unidad Empresarial de Base (UEB) de la Empresa Provincial Tabaco Torcido de Santa Clara que se dedica a la fabricación del famoso puro cubano.

En la UEB existen diferentes departamentos, unos productivos y otros no, entre los productivos se destaca por su importancia la galera, cuyo nombre proviene del hecho de que los trabajadores se disponen del mismo modo que los remeros en las antiguas embarcaciones romanas, en este taller es que se tuerce el tabaco a partir de sus materiales primarios, en el mismo trabajan verdaderos maestros artesanos, muy orgullosos de ser parte de la identidad cubana, estos son los tabaqueros.

La producción de cada tabaquero es revisada minuciosamente para comprobar que cumpla con los parámetros pactados con los compradores extranjeros; por tanto en el Departamento de Control Técnico, que es donde se realiza esta acción, trabajan especialistas que se ocupan de medir ciertas magnitudes unas con instrumentos y otras son mediciones sensoriales. A partir de los resultados de esta actividad es acep-

tada o rechazada la producción del tabaquero, los datos de la cantidad por tipo de defectos es almacenada en una base de datos Access a partir de la cual se calcula un índice de calidad que constituye una de las bases para el pago de los trabajadores, sin embargo estos datos no son usados para más ningún otro análisis.

Las producciones rechazadas no pueden seguir el proceso productivo, constituyendo un gasto, por ello es de una importancia crucial disminuir este índice de rechazo lo más posible y con la celeridad máxima, para ello se cuenta con la sabiduría de un experto, quien es una persona de vastísima experiencia empírica, este “gurú” en la materia diagnóstica la causa que provocó la no conformidad de la producción de cada trabajador para poder tomar medidas encaminadas a eliminar o disminuir en el máximo grado posible la causa que más influyó en la aparición de los defectos. Este diagnóstico, cuando se cuenta con cientos de producciones diarias y cada una a su vez formadas por cientos de unidades, se hace muy engorroso y el resultado está cuando ya se ha realizado la mitad de la producción del siguiente día sin aún poderse corregir las causas de los defectos del día anterior.

Estos argumentos fundamentan la formulación del problema científico siguiente:

El desarrollo de un producto de software para diagnosticar la causa que más incidencia tuvo en el rechazo de la producción, cuyo modelo computacional sea aprendido a partir de la información histórica almacenada en la Base de Datos del “Sistema de Control de la Calidad” que se utiliza en la UEB para el cálculo del índice de calidad para el pago a los trabajadores. Específicamente se pretende que a partir de la cantidad de los tipos de defectos encontrados en la producción se diagnostique la causa que los provocó lo más rápido y exacto posible, pudiendo reproducir el comportamiento de un experto de la materia, quienes generalmente son personas que han acumulado una gran experiencia empírica a causa de los incontables años dedicados a esa actividad, por tanto son obreros de un valor inestimable y muy escasos.

El cual delimita como objetivo general del presente trabajo:

Desarrollar un sistema computacional para diagnosticar lo más rápido y exactamente posible la causa que más influyó en la no aceptación de la producción de un lote de tabacos torcido a partir de las cantidades de los tipos de defectos mediante la aplicación de varias técnicas de la inteligencia artificial.

Para lograr este objetivo general se plantean los objetivos específicos siguientes:

1. Desarrollar el proceso de descubrimiento de conocimiento en la Base de Datos “Control de la Calidad”
2. Validar la efectividad de los modelos MLP, ConFuCiuS, Reglas y Árboles de Decisión en el diagnóstico de la causa que más influyó en la no aceptación de una producción.
3. Utilizar Gráficos de Pareto y Gráficos de Control para visualizar y analizar de un modo más eficiente comportamientos que sin estas herramientas hoy se pasan por alto.
4. Desarrollar el producto final como *open source* utilizando el lenguaje Java y los objetos que publica Weka ¹ (Waikato Environment for Knowledge Analysis).

Hipótesis de Investigación:

- ✓ El conocimiento minado a partir de la Base de Datos del “Sistema Control de Calidad” aporta información útil para el diagnóstico de la causa que más influyó en la no aceptación de un lote de tabaco torcido a mano.
- ✓ La cantidad y tipos de defectos determinan la causa que provocó la no conformidad de una producción.
- ✓ Las técnicas de Inteligencia Artificial muestran ventajas sobre técnicas tradicionales en el desarrollo de un sistema computacional que dé un diagnóstico aceptable de la causa que más intervino en el rechazo de un lote de tabaco a partir de información recogida por el “Sistema Control de Calidad”.

Capítulo 1.

En este capítulo se aborda el descubrimiento de conocimiento en bases de datos, y su conexión con el aprendizaje automático. Se presta especial atención al aprendizaje de modelos basados en técnicas de Inteligencia Artificial, los cuales han sido ampliamente aplicados para desarrollar diferentes tipos de sistemas basados en el conocimiento, combinando en algunos las ventajas de diferentes enfoques en los llamados modelos híbridos.

¹ WEKA es una herramienta de código abierto escrita en Java. Está disponible en <http://www.cs.waikato.ac.nz/~ml/weka/> bajo licencia pública GNU.

1.1 El descubrimiento de conocimiento en bases de datos

El descubrimiento de conocimiento (KDD) generalmente se entiende como la búsqueda de modelos interesantes, nuevos y útiles, en datos que normalmente se encuentran en una gran base de datos. Otra definición dada por (FAYYAD, y otros, 1996), se refiere al proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y fundamentalmente comprensibles en los datos. Nótese que la primera definición deja claro que los datos se almacenan en bases de datos, mientras que la segunda hace énfasis en identificar patrones comprensibles. Luego, estos modelos (o patrones) pueden ser utilizados por las organizaciones como base para la toma de decisiones comerciales (HALGAMUGE, y otros, 2005).

1.1.1 Proceso de KDD

Según (KAUFMANN, 2000) el proceso de KDD consiste en una sucesión reiterativa de los pasos siguientes:

1. Limpieza de datos: eliminar los datos ruidosos o irrelevantes.
2. Integración de datos: combinar los datos desde múltiples fuentes
3. Selección de datos: recuperar desde la base de datos los datos relevantes para la tarea objeto de análisis.
4. Transformación de los datos: transformar los datos o llevarlos a formas apropiadas para extraer el conocimiento.
5. Minería de datos: aplicar métodos inteligentes para extraer los patrones de los datos.
6. Evaluación de patrones: identificar los patrones verdaderamente interesantes representando conocimiento basado en medidas de interés.
7. Representación del conocimiento: utilizar técnicas de representación y visualización para mostrar al usuario el conocimiento extraído.

Otra manera de ver el proceso KDD, aunque en esencia se refiere a los mismos pasos pero agrupados, se presenta en la Figura 1 (AGUILAR RUIZ, 2008):

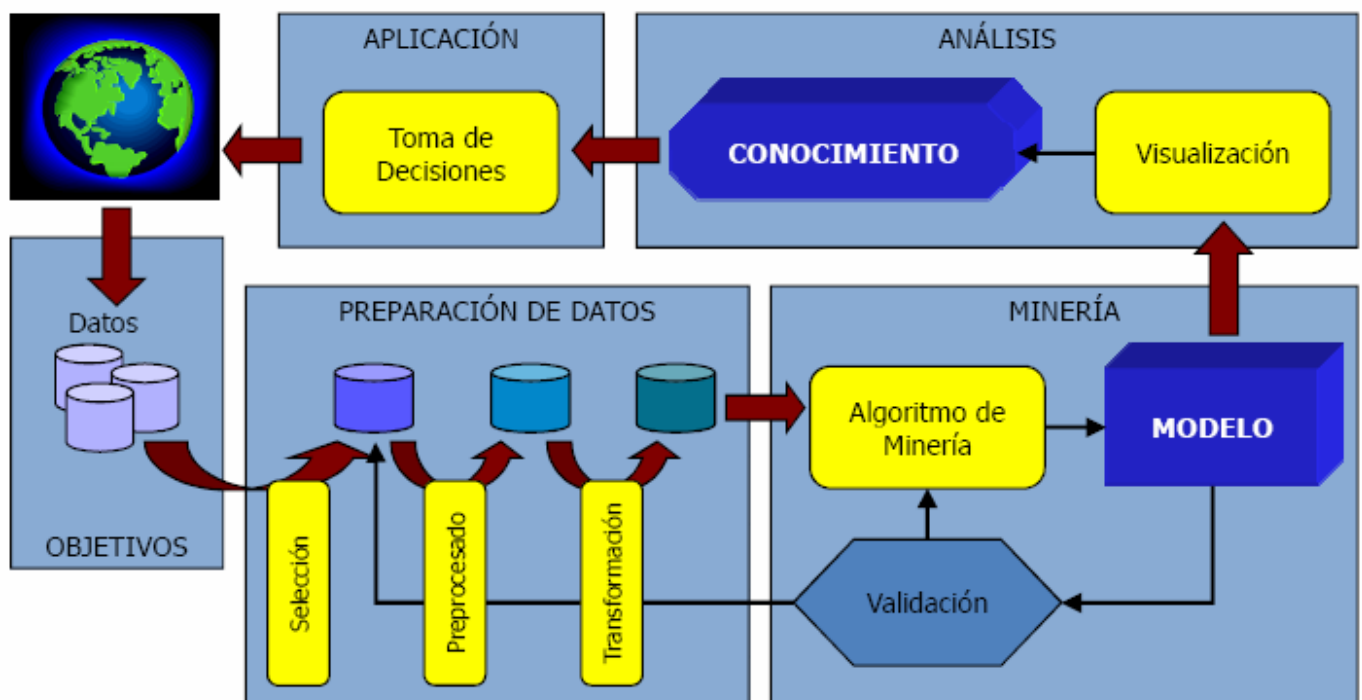


Figura 1.1. Proceso KDD según (AGUILAR RUIZ, 2008)

1. Determinación de Objetivos: Precisar qué objetivos quieren cumplirse desde el punto de vista del usuario, los datos y las técnicas a usar.
2. Preparación de los datos: se requiere cuando los datos tienen ambigüedades, ruidos o no están en el formato adecuado, y define como subfases las siguientes:

- ✓ Selección: distinguir los subconjuntos de datos significativos.
 - ✓ Preprocesado: tratar los valores ausentes y el ruido presente en los datos.
 - ✓ Transformación: Representación adecuada (ej. Normalizar, discretizar, etc).
3. Minería de Datos: elección del método de minería es fundamental en el proceso KDD. Incluye la elección del algoritmo de aprendizaje y la validación del modelo.
 4. Análisis: Estudia, interpreta y evalúa el modelo de conocimiento.
 5. Aplicación: Integración del conocimiento adquirido al campo real de aplicación mediante la toma de decisiones, así como la comparación con el conocimiento previo a la aplicación del proceso.

El KDD no es un campo aislado, sino la convergencia de otros campos tales como se muestra en la Figura 1.2 tomada de (AGUILAR RUIZ, 2008): el Aprendizaje Automático (Machine learning), la Estadística y las Bases de Datos y en particular el trabajo con los almacenes de datos (Data Warehouse).

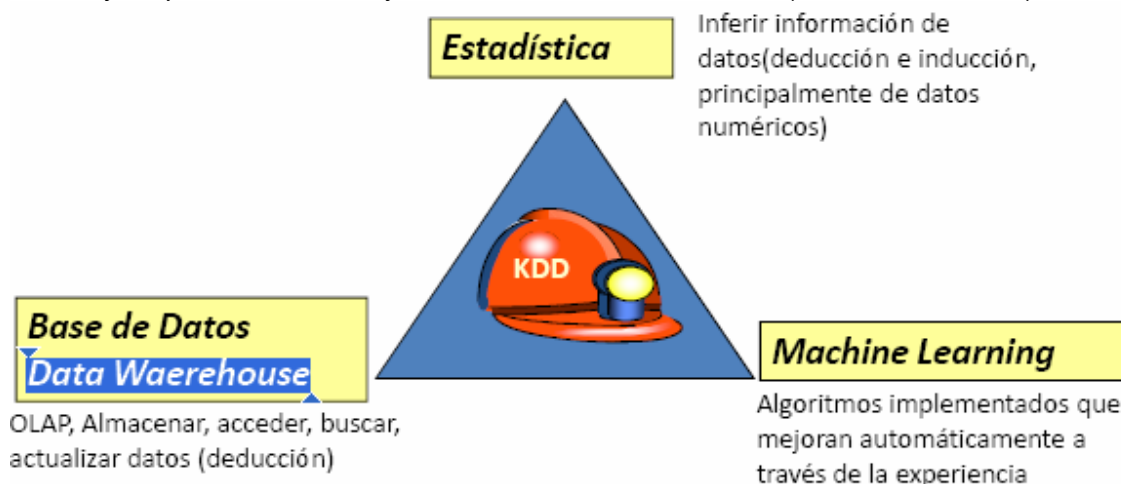


Figura 1.2. Relación del KDD con otros campos

Es cierto que hay que contar con los datos antes de poder aplicar la minería de los mismos; sin embargo, no siempre es suficiente con tener los datos, se requiere tener en cuenta si los datos son suficientes y si éstos son correctos. Es necesario tener el criterio de una persona que tiene un conocimiento del dominio para seleccionar los datos correctos. Este conocimiento puede usarse en varias maneras diferentes en la minería de datos, sobre todo en la fase de preparación de datos que es fundamental en el proceso de KDD (BIGUS, 1996).

Atendiendo a los objetivos del presente trabajo, a continuación se profundiza en los modelos basados en técnicas de la Inteligencia Artificial.

1.2 Los sistemas inteligentes

Desde que se acuñó el término Inteligencia Artificial (IA) en el año 1956, esta área de la Computación ha sido un tema importante de gran discusión y estudio en el desarrollo de nuevas tecnologías. Según (RUSSELL, y otros, 1995) la IA es la habilidad que tienen las computadoras de realizar funciones que normalmente requieren inteligencia humana. Es un campo de estudio que abarca las técnicas computacionales que permiten la ejecución de tareas que caracterizan el pensamiento humano (TANIMOTO, 1987).

Un Sistema Basado en el Conocimiento (SBC) en términos generales se refiere a “un sistema computarizado que usa conocimiento sobre un dominio para arribar a una solución de un problema de ese dominio. Esta solución es esencialmente la misma que la obtenida por una persona experimentada en el dominio del problema cuando se enfrenta al mismo problema” (GÁLVEZ, 1998). Los SBC modelan la pericia de los expertos del dominio y la representan explícitamente de una manera declarativa. Por este medio, el sistema no es simplemente capaz de generar las explicaciones entendibles, sino también le permite al diseñador mejorar al modelo experto (AGHA, y otros, 1999).

El conocimiento representado en los SBC generalmente se obtiene a partir de los expertos humanos. Un experto es una persona que posee habilidades que le permiten sacar conclusiones de experiencias pasadas y llegar rápidamente a la raíz de un problema dado. El experto, debido a su experiencia, ha adquirido un conjunto de relaciones de causa y efecto poderosas llevando a que aumente la posibilidad de éxito en la búsqueda de solución a un problema. Utiliza este conocimiento básico para reconocer rápidamente ras-

gos sobresalientes del problema, clasificar este de acuerdo a estas características y buscar una solución (GÁLVEZ, 1998).

Con respecto a otros métodos de solución convencionales, los SBC ofrecen algunas ventajas importantes. Entre ellas se encuentra la gran accesibilidad de la base y la fácil modificación ya que el conocimiento es explícito y accesible, así como la distribución y preservación de la pericia. Permiten evaluar el efecto de nuevas estrategias añadiendo o modificando conocimiento. Dan soluciones incluso a problemas que incluyen datos incompletos dando explicación a dicha solución justificando sus conclusiones. Se logra consistencia en la solución obtenida. Un SBC constituye un entrenador en el dominio de aplicación (GÁLVEZ, 1998). Estos también tienen algunas desventajas que hay que considerar: no siempre dan respuestas correctas y puede presentar la ausencia de sentido común, además no reconocen el límite de su conocimiento el cual es limitado al dominio de aplicación.

En general un SBC se considera compuesto por una base de conocimiento, uno o más mecanismos de inferencia y una interfaz de usuario. La más importante de estas componentes es la base de conocimiento, la cual contiene el conocimiento específico del dominio requerido para resolver un problema. Esta se conforma generalmente como resultado de un proceso de ingeniería del conocimiento, donde no es necesario que el ingeniero de conocimiento sea hábil en el dominio en que el sistema está desarrollándose, pero sí debe tener un conocimiento general y familiaridad con el dominio de aplicación (KRISHNAMOORTHY, y otros, 1996).

A estos sistemas que hacen uso de las técnicas de la IA, se le denominan sistemas inteligentes o expertos. Una parte esencial de un sistema experto es el conocimiento que este utiliza en la solución de un determinado problema.

Los mecanismos de la inferencia son estrategias de control o técnicas de búsqueda que exploran a través de la base de conocimiento para llegar a decisiones. La base de conocimiento es el espacio de estado y el mecanismo de inferencia es un proceso de búsqueda. Los dos métodos populares de inferencia son encadenamiento hacia atrás (backward chaining) y el encadenamiento hacia delante (forward chaining) (BELLO, 1993).

Los SBC se diferencian esencialmente en la Forma de Representación del Conocimiento (FRC), en el Método de Solución del Problema (MSP), y en la forma en que se explica una solución obtenida. En algunos casos los SBC incluyen también algunos módulos opcionales: módulo explicativo, módulo de cálculo de certidumbre y módulo de autoaprendizaje (GARCÍA, 2003).

1.2.2 Tipos de Sistemas Basados en el Conocimiento

Los paradigmas a emplear en el desarrollo de SBC deben considerar los cambios en la representación del conocimiento (FRC) y en el mecanismo de inferencia (o MSP) a los efectos de su tratamiento. Este epígrafe describe algunos de ellos.

Sistemas Basados en Reglas

Los Sistemas Basados en Reglas (SBR) son un tipo de SBC que utiliza como FRC las reglas de producción. Las reglas utilizan un formato IF - THEN para representar el conocimiento, la parte IF de una regla es una condición o premisa y la parte THEN de la regla es la acción o conclusión la cual permite inferir un conjunto de hechos nuevos si se verifican las condiciones de la parte IF. El MSP utilizado por los SBR sigue el método conocido como Módus Ponens, creando una cadena de inferencias que constituye un camino entre la definición del problema y la solución. Este camino normalmente es creado por una búsqueda primero en profundidad dirigido por objetivo o datos en dependencia de la cantidad de datos de entrada y de conclusiones posibles entre otras cosas (BELLO, 2002).

Sistemas Basados en Casos

Los sistemas basados en casos (SBCA) son un tipo de SBC que utiliza los casos como FRC y emplean el Razonamiento Basados en Casos como mecanismo de inferencia. El RBC es un proceso de considerar casos anteriores y entonces llegar a una decisión en dependencia de la comparación entre la situación actual y los casos viejos. Dado un nuevo caso para clasificar, un razonador basado en casos verifica primero si existe un caso de entrenamiento idéntico. Si encuentra uno, retorna la solución acompañando a ese caso. Si no encuentra ningún caso idéntico, el razonador buscará casos de entrenamiento que tienen componentes similares a aquellas del nuevo caso. El sistema es capaz de aprender a través de la agregación de nuevos casos (KAUFMANN, 2000).

Sistemas conexionistas

Un sistema conexionista es aquel que emplea una Red Neuronal Artificial (RNA) como MSP. Una RNA es un modelo computacional que pretende simular el funcionamiento del cerebro a partir del desarrollo de una arquitectura que toma rasgos del funcionamiento de este órgano sin llegar a desarrollar una réplica del mismo (BELLO, 1993). El modelo computacional de una red neuronal está compuesto de varios nodos, los cuales constituyen las neuronas artificiales, que están conectadas por arcos dirigidos. Cada arco tiene asociado un peso numérico w_{ij} que representa la intensidad de la interacción entre las unidades u_i y u_j . Los pesos son los medios primarios de almacenamiento a largo plazo en las redes neuronales, y generalmente el aprendizaje ocurre al actualizarlos.

La idea es que cada nodo hace un cómputo local basado en las entradas de sus vecinos, pero sin la necesidad de cualquier control global sobre el conjunto entero de unidades.

El aprendizaje no es una habilidad individual de una sola neurona, es un proceso colectivo de la red neuronal entera y resultado de un proceso de entrenamiento (KASABOV, 1998).

Sistemas híbridos basados en la lógica borrosa

La lógica borrosa y los conjuntos borrosos se han desarrollado como una semántica para representar, manipular y utilizar la imprecisión presente en aplicaciones del mundo real (ZADEH, 1975). A esta idea responde la definición de Soft Computing dada por Zadeh (ZADEH, 2001) al especificar que este concepto se refiere a “una mezcla de distintos métodos que de una forma u otra cooperan desde sus fundamentos con el principal objetivo de aprovechar la tolerancia que conllevan la imprecisión y la incertidumbre, para conseguir manejabilidad, robustez y soluciones de bajo costo”. Por ejemplo, en la combinación de la Lógica Borrosa, Neurocomputación y Razonamiento Probabilístico, la primera se ocupa principalmente de la imprecisión y el razonamiento aproximado; la Neurocomputación del aprendizaje, y el Razonamiento Probabilístico de la incertidumbre y la propagación de las creencias”.

De esta manera, el desarrollo de sistemas comprensibles y confiables sigue una modelación de caja gris (CASILLAS, y otros, 2003) por la combinación de la modelación de caja blanca (garantizan comprensibilidad) y de caja negra (confiabilidad).

Las operaciones y notaciones de la lógica borrosa son basadas en la lógica clásica y el cálculo proposicional. Dos principios importantes de lógica clásica declaran que una proposición debe ser o verdadera o falsa, y que una proposición no puede ser verdadera y falsa al mismo tiempo (SILER, y otros, 2005).

El núcleo de la teoría de los conjuntos borrosos es la idea de una variable lingüística. Una variable lingüística es aquella que puede tomar por valor términos del lenguaje natural, como mucho, poco, bajo, alto, etc., que son las etiquetas en un conjunto borroso (MARTÍN, y otros).

Modelo con aprendizaje basado en instancias utilizando la distancia entre asociaciones borrosas (ConFuCiuS, Connectionist Fuzzy Case-based System)

El modelo ConFuCiuS (RODRIGUEZ, y otros, 2006) es un modelo conexionista borroso para desarrollar SBCA. Este modelo se define como una instancia del clasificador basado en los vecinos más cercanos, pero usando una función de distancia que utiliza los pesos del modelo FSIAC (RODRIGUEZ, 2007).

Para comparar dos instancias (en el espacio m -dimensional) se utiliza la distancia Euclídeana que emplea el k -NN estándar; y considerando como criterio de comparación a nivel de rasgo una de las funciones de distancia propuestas en (RODRIGUEZ, 2007). Cuando se utiliza la métrica basada en la diferencia de valores borrosos (FVDM, Fuzzy Value Difference Metric) el modelo se referenciará como ConFuCiuS. Cuando se utiliza la Métrica basada en la diferencia de correlaciones borrosas (FCDM, Fuzzy Correlation Difference Metric), se especifica que se trabaja con ConFuCiuScc.

Capítulo 2.

2.1. Descubrimiento de conocimiento a partir de la base de datos del control de la calidad.

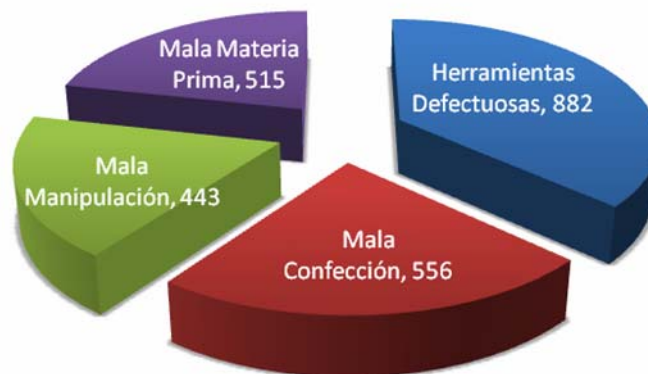
La solución de este problema se desarrolla a partir de la base de datos del “Sistema de Control de la Calidad” donde se guarda la cantidad de tabacos que incumple con los parámetros pactados según el tipo de defecto (NC ISO 2859-1, 2003). Se identificaron diez posibles rasgos predictores los cuales son Peso, Diámetro, Dimensión, Roto, Capa, Perilla, Duro, Fofo, Baches, Empalme. Algunos de ellos son medidos con instrumentos muy sencillos, en cambio otros son medidos basados en la gran experiencia de los técnicos de calidad.

El rasgo objetivo es la causa del defecto que tiene 4 clases, a saber:

- ✓ Mala Confección: El tabaquero elabora mal los tabacos, sobre esta causa es importante incidir lo antes posible.
- ✓ Herramientas Defectuosas: La guillotina está mal ajustada, o el molde no tiene las dimensiones específicas.
- ✓ Mala Materia Prima: La materia prima no es de la calidad adecuada para el tipo de vitola que se pretende hacer con ella.
- ✓ Mala Manipulación: Los tabacos son tratados con descuido, no respetándose las normas dictadas que regulan su trasiego, almacenamiento y manipulación.

Por tanto atendiendo a lo anteriormente expresado estamos hablando de un problema de clasificación.

La Base de Casos cuenta inicialmente con 2396 casos distribuidos de la siguiente forma:



A partir de estos rasgos se creó una nueva tabla con un campo adicional que se denominó categoría y el cual recibió valores según el criterio de un experto, esta información fue recopilada manualmente e introducida a la BD. Los valores de los rasgos se llevaron a base de mil para que el tamaño del lote no influyera en nuevas clasificaciones.

No todos los rasgos son igualmente importantes con respecto a la confiabilidad de la predicción, en este sentido Weka tiene implementado un conjunto de métodos de preprocesamiento de datos.

Utilizando el evaluador CfsSubsetEval como método de evaluación de los rasgos y empleando el algoritmo BestFirst (primero el mejor) como método de búsqueda se obtuvo que los rasgos Peso y Empalme no inflúan en la predicción, esto concordó totalmente con el criterio de los expertos, ya que estos defectos generalmente son corregidos con la inspección primaria en la galera, llegando muy rara vez hasta el Control Técnico.

Ya definidos los rasgos se pueden realizar una serie de pruebas para determinar los modelos de mejor desempeño en el problema. Para realizar estas pruebas se utiliza la validación cruzada estratificada con 10 particiones (10-fold cross validation) y se emplea el modo experimentador de Weka con 10 repeticiones. Los modelos aplicados son los siguientes: ConFuCiuS ($k=9$), MLP(RNA), RJ48(Árbol de Decisión), JRip(Reglas). Para poder aplicar el modelo ConFuCiuS, es necesario convertir los atributos numéricos en variables lingüísticas. Este proceso se lleva a cabo utilizando el filtro MDLFuzzyfier. El desempeño de clasificación correcta estuvo en el siguiente orden: MLP(98.87 %), RJ48(98.16 %), ConFuCius(98.11 %), JRip(98.08 %). Se puede apreciar que el comportamiento de los cuatro modelos es excelente, siendo el mejor de todos el MLP, sin embargo se decide implementar los cuatro modelos para tener más de un modo de evaluar un caso dudoso contando con un tribunal de cuatro jueces.

Estos resultados hacen que el software desarrollado CALIDAD+IA finalmente conste de:

- ✓ Una única base de conocimiento contenida en la BD Control de la Calidad que se denomina SetEntrenamientoFinal y que almacena todos los ejemplos de entrenamiento.
- ✓ Como métodos de soluciones de problemas se emplean los modelos MLP, RJ48, ConFuCius, JRip.
- ✓ Una interfaz para hacer selecciones de intervalos de fechas, tipo de vitola, trabajador, etc. de las instancias de la tabla Primario contenida en la BD Control de la Calidad y evaluarla según el método deseado, esta tabla es la que contiene las mediciones de los rasgos predictores.

- ✓ Interfaz para mostrar con la utilización de Gráficos de Pareto las cantidades de defectos y las causas de los mismos por categorías.
- ✓ Interfaz para mostrar con la utilización de Gráficos de Control como se ha comportado el índice de defectuosos para la selección dada.
- ✓ Interfaz para definir los valores de los rasgos predictores manualmente y diagnosticar según el modelo escogido.

Al ejecutar la aplicación, se muestra la siguiente ventana inicial antes de mostrar la ventana principal. En esta ventana aparece el nombre de la aplicación, los nombres del autor y de la tutora, también se puede apreciar una línea negra que va llenándose de rojo a modo de barra de progreso indicando las acciones que van ocurriendo en segundo plano. El nombre de la acción que se está ejecutando aparece debajo de esta línea.



La ventana principal muestra dos pestañas con las etiquetas Varios Lotes y Un Lote.

La pestaña Un Lote permite definir las características de una producción y presionando el botón Diagnosticar obtener los resultados según el modelo escogido en el cuadro desplegable Paradigma a Utilizar. Estos resultados constan de los valores iniciales, la certidumbre de la respuesta y la causa. Finalmente presionando Dibujar Pareto despliega un diagrama de este tipo muy importante para la toma de decisiones. Esta pestaña activada se muestra a continuación.

Calidad

Paradigma a Utilizar:

MLP

Varios Lotes

Un Lote

Configurar:

Produccion:

100

Rasgos:

Diametro:

0

Perilla:

0

Dimension:

1

Duro:

2

Rotos:

0

Fofo:

0

Capa:

0

Baches:

0

Diagnosticar

Dibujar Pareto

Resultados:

Prod	Diam	Dim	Roto	Capa	Per	Duro	Fofo	Bach	MMan	MMP	Conf	Herr	-> Causa
100	0	1	0	0	0	2	0	0	0.0	0.12	98.19	1.69	-> Mala Confeccion

Cerrar

La pestaña Varios Lotes permite definir las características de una selección según intervalo de fechas, tipo de tabaco, trabajador y presionando el botón Diagnosticar obtener los resultados según el modelo escogido en el cuadro desplegable Paradigma a Utilizar. Estos resultados constan de los mismos valores descritos anteriormente. La selección que se va realizando es descrita en el área de texto Acción. Esta pestaña activada se muestra a continuación.

Calidad

Paradigma a Utilizar:
MLP

Varios Lotes Un Lote

Configurar Selección:

Días
☒ Desglosar Desde: 02/03/09 Hasta: 17/04/09

Vitolas
☒ Desglosar
 Vitola:

Trabajadores
☒ Desglosar
 Trabajador:

Accion:
 Deseo diagnosticar las producciones realizadas entre los días 2/3/2009 y 17/4/2009 desglosadas por días, desglosadas por productos, desglosadas por trabajadores. Utilizando MLP.

(485/1325)= 37 %

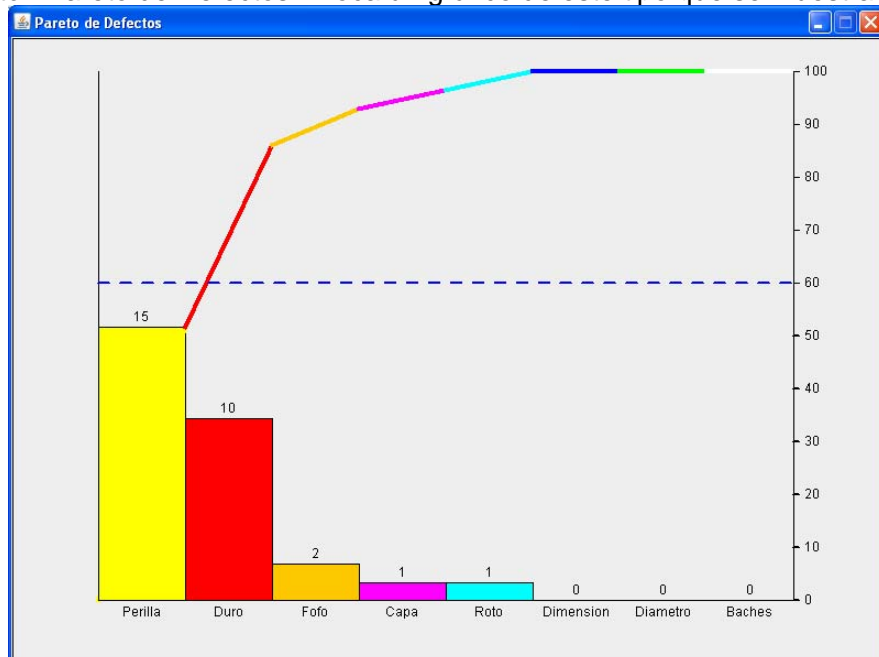
Grafico de Control 100p Pareto de Causas Pareto de Defectos Abortar

Resultados:

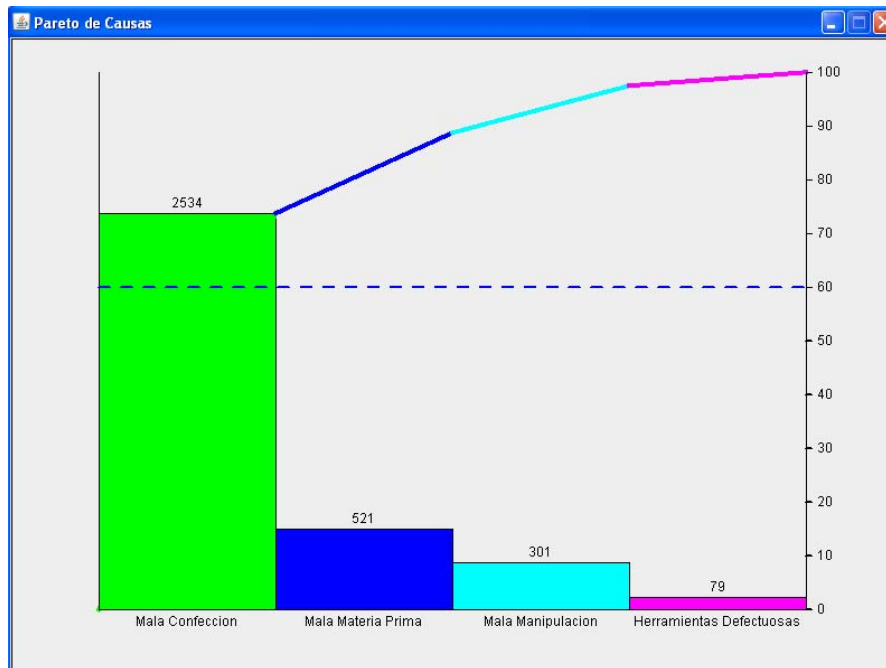
Nombre	Vitolas	Fecha	Prod	Diam	Dim	Roto	Capa	Per	Duro	Fofo	Bach	MMan	MMP	Conf	Herr	-> Causa
Adonis Laffita	Brevas	Con 2009-03-02	140	0	0	0	0	0	0	0	1	0.04	0.55	99.42	0.0	-> Mala Confecc
Alexei Ferrer M	Brevas	Con 2009-03-02	230	0	0	0	2	0	0	0	0	0.0	99.94	0.06	0.0	-> Mala Materi
Basilio Bárbaro	Brevas	Con 2009-03-02	200	0	0	0	2	0	0	0	0	0.0	99.94	0.06	0.0	-> Mala Materi
Daineyis Blanco	Brevas	Con 2009-03-02	150	0	0	0	1	0	0	0	0	0.0	99.92	0.08	0.0	-> Mala Materi
Digna Iris Lópe	Brevas	Con 2009-03-02	200	0	0	0	0	0	0	0	2	0.04	0.48	99.47	0.0	-> Mala Confecc
Dignora Manso N	Brevas	Con 2009-03-02	150	0	0	0	0	0	0	0	1	0.03	0.56	99.41	0.0	-> Mala Confecc
Eloisa Hernández	Brevas	Con 2009-03-02	240	0	0	0	0	0	0	0	2	0.04	0.51	99.45	0.0	-> Mala Confecc
Manuel Dimas Mo	Brevas	Con 2009-03-02	150	0	0	0	0	0	0	0	1	0.03	0.56	99.41	0.0	-> Mala Confecc
Mildrey Portal	Brevas	Con 2009-03-02	150	0	0	0	0	0	0	0	1	0.03	0.56	99.41	0.0	-> Mala Confecc
Mirian Rodríguez	Brevas	Con 2009-03-02	240	0	0	0	2	0	0	0	0	0.0	99.94	0.06	0.0	-> Mala Materi
Redel Salazar C	Brevas	Con 2009-03-02	250	0	0	0	0	0	0	0	2	0.04	0.52	99.44	0.0	-> Mala Confecc
Rosalba Gonzále	Brevas	Con 2009-03-02	230	0	0	0	1	0	0	0	1	0.0	99.7	0.3	0.0	-> Mala Materi
Wendy Falcón Pé	Brevas	Con 2009-03-02	190	0	0	0	0	0	0	0	2	0.05	0.48	99.48	0.0	-> Mala Confecc
Yisel Leon Jaco	Brevas	Con 2009-03-02	150	0	0	0	1	0	0	0	0	0.0	99.92	0.08	0.0	-> Mala Materi

Cerrar

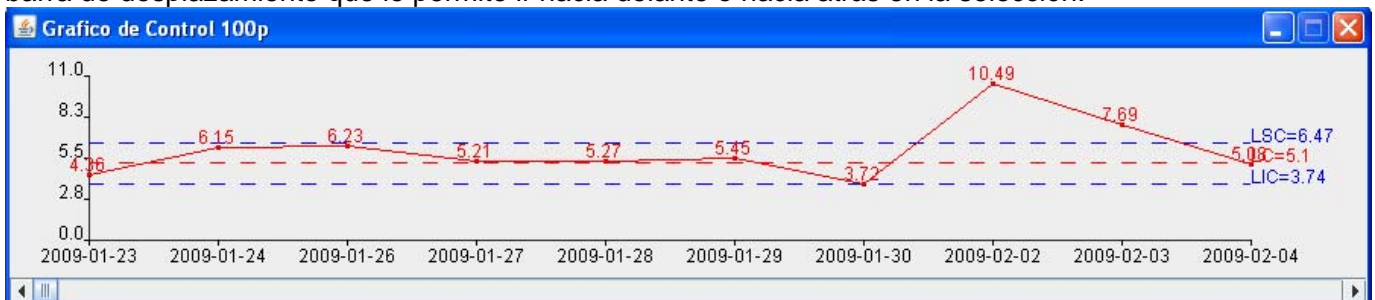
Presionando el botón Pareto de Defectos invoca un gráfico de este tipo que se muestra a continuación:



Presionando el botón Pareto de Causas invoca un gráfico de este tipo que se muestra a continuación:



Presionando el botón Gráfico de Control 100p invoca un gráfico que muestra cronológicamente el comportamiento del por ciento de defectuosos por muestra como se muestra a continuación, el gráfico tiene una barra de desplazamiento que le permite ir hacia delante o hacia atrás en la selección:



Todos estos gráficos son de una importancia extraordinaria para la toma de decisiones y permite reconocer patrones de comportamiento más fácilmente.

Conclusiones:

En este trabajo se ha mostrado la factibilidad de aplicar KDD en la Base de Datos de Control de la Calidad, cuyo resultado aporta información útil a la toma de decisiones en la gestión de la calidad proceso de torcido de una fábrica de tabaco y permitió desarrollar un sistema computacional para diagnosticar la causa que más influyó en los defectos. Estos resultados se centran alrededor de los cuatro objetivos específicos planteados para los cuales se concluye que:

- El proceso de descubrimiento de conocimiento en la Base de Datos de Control de la Calidad utilizada permitió identificar patrones válidos y útiles y en particular:
 - ✓ Se conformó el conjunto de instancias que constituyen los ejemplos de entrenamiento, cuyo análisis corrobora que tiene sentido hablar de la tarea de diagnosticar la causa de los defectos a partir de la aparición de los mismos y su distribución como lo haría un experto.
 - ✓ Se identificaron desde la Base de Datos 10 rasgos en correspondencia con campos existentes en ésta, escogiéndose 8 como los más determinantes para el diagnóstico.
- Se obtiene una primera versión del producto de software CALIDAD+IA validando la efectividad de los modelos MLP, ConFuCiuS, Reglas y Árboles de Decisión en el diagnóstico de la causa que más influyó en la no aceptación de una producción a partir de que:

- ✓ Los cuatro algoritmos de aprendizaje automático utilizados en la etapa de “minería de datos” tuvieron precisiones mayores que 98.00 % en clasificar las instancias correctamente, lo que permite utilizarlos como jueces para decidir en un caso difícil de clasificar.
- ✓ Se obtiene un buen desempeño como resultado del proceso de validación utilizando una muestra de control.
- ✓ Además emplea un modelo comprensible para el usuario, y propicia un lenguaje más natural para los expertos humanos en este dominio de aplicación.
- Se utilizan Gráficos de Pareto y Gráficos de Control para visualizar y analizar de un modo más eficiente comportamientos que sin estas herramientas hoy se pasan por alto.
- La aplicación se implementa en Java como *open source* permitiendo a otros desarrolladores ampliar sus funcionalidades si resultara necesario.

Recomendaciones:

A partir de los resultados obtenidos se recomienda:

1. Convertir el software “CALIDAD+IA” en una herramienta imprescindible para la toma de decisiones encaminadas a disminuir el índice de defectuosos de tabaco torcido en la galera.
2. Hacer un estudio para extender el uso del software a otros departamentos productivos de una fábrica de tabaco y validar su desempeño en la nueva tarea.
3. Enriquecer la base de conocimiento con nuevos casos.
4. Extender el uso de esta herramienta a todas las fábricas de tabaco torcido cuyas condiciones lo permitan.
5. Estudiar la factibilidad de incluir nuevos rasgos predictores que permitan especificar más las causas de los defectos de las producciones.

Bibliografía:

- AGHA, G. A., M.CARLEY, K. y OTHERS. 1999. *Multiagent Systems. A Modern Approach to Distributed Modern Approach to Artificial Intelligence*. 1999.
- AGUILAR RUIZ, J. S. 2008. *Curso Preprocesado de datos y aprendizaje basado en reglas*. Santa Clara : s.n., 2008.
- BELLO, R. 2002. *Aplicaciones de la Inteligencia Artificial*. Guadalajara : Universidad de Guadalajara, 2002.
- . 1993. *Curso Introductorio a las Redes Neuronales Artificiales*. s.l. : UCLV, 1993.
- BIGUS, J. P. 1996. *Data Mining with Neural Networks*. 1996.
- CASILLAS, J. C., HERRERA, F. y MAGDALENA, L. 2003. *Interpretability Issues in Fuzzy Modeling*, Springer. 2003.
- FAYYAD, U. M., PIATETSKY-SHAPIO, G. y SMYTH, P. 1996. *Advances in Knowledge Discovery and Data Mining*. s.l. : The MIT Press., 1996.
- GÁLVEZ, D. 1998. *Sistemas Basados en el Conocimiento*. 1998.
- GARCÍA, A. 2003. *Sistema híbrido para el pronóstico del nivel de gravedad en pacientes pediátricos al ingresar en una Unidad de Cuidados Intensivos*. Santa Clara : UCLV, 2003.
- HALGAMUGE, S. K. y WANG, L. 2005. *Classification and Clustering for Knowledge Discovery*. 2005.
- KASABOV, N. K. 1998. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. 1998.
- KAUFMANN, M. 2000. *Data Mining. Concepts and Techniques*. 2000.
- . 2000. *Data Mining. Concepts and Techniques*. 2000.
- KRISHNAMOORTHY, C. S. y RAJEEV, S. 1996. *Artificial Intelligence and Expert Systems for Engineers*. 1996.
- MARTÍN, B. y SANZ, A. *Redes neuronales y sistemas difusos. Segunda Edición*.
- NC ISO 2859-1, 3. 2003. *PROCEDIMIENTOS DE MUESTREO PARA INSPECCION POR ATRIBUTOS*. 2003.
- RODRIGUEZ, Y. 2007. *Generalización de la métrica basada en la diferencia de valores (VDM) para variables lingüísticas y su aplicación en sistemas basados en el conocimiento*. s.l. : UCLV, 2007.
- RODRIGUEZ, Y., y otros. 2006. *Connectionist Fuzzy Case-Based Reasoning Model*. 2006.
- RUSSELL, S. J. y NORVIG, P. 1995. *Artificial Intelligence A Modern Approach*. 1995.

SILER, W. y BUCKLEY, J. J. 2005. *Fuzzy Expert Systems and Fuzzy Reasoning*. 2005.
TANIMOTO, S. L. 1987. *The Elements of Artificial Intelligence. An Introduction Using LISP*. 1987.
ZADEH, L. A. 2001. *Applied Soft Computing*. s.l. : ASOC, 2001.
—. 1975. *The concept of a linguistic variable and Its Application to Approximate Reasoning*. 1975.

Camajuaní, 26 de abril de 2009.