“Sistema de Predicción de la Demanda de Papel Fotográfico para Ecuacolor”

AUTORES:

Teodora Sofía Argoti Doylet[1], Soraya Freire Oliveros[2], Carmen Orozco Roggiero[3], Juan Alvarado O.[4]

[1] Egresada de Licenciada en Sistemas de Información 2005.
[2] Egresada de Licenciada en Sistemas de Información 2005.
[3] Egresada de Licenciada en Sistemas de Información 2005.
[4] Director de Tópico. Master en Administración Pública 1996 ESPAE, Ingeniero en Electricidad y Computación, especialidad en computación 13 Diciembre 1991 ESPOL, Profesor de LSI-FIEC-ICM ESPOL.

RESUMEN

Versión Español:

Este proyecto fue desarrollado como requisito del tópico de graduación previo a la obtención del título de Licenciado en Sistemas de Información, su objetivo es utilizar los conocimientos adquiridos en Minería de Datos, aplicándolos en la predicción de la demanda futura de papel fotográfico para Ecuacolor.

Como resultado del proyecto se elaboró un módulo de software que podrá ser utilizado, cada mes por el encargado de compras de Ecuacolor, para estimar la demanda de papel en el periodo siguiente. El software permite aplicar el algoritmo de minería de datos a un conjunto de datos históricos y cambiar ciertos parámetros a discreción del usuario.

English Version:

This project was developed as the final requirement of the graduation process to obtain the degree of “Licenciado en Sistemas de Información”, its objective is to use the knowledge acquired in Data Mining, to be applied in the prediction of the future demand of photographic paper for Ecuacolor.

As result of the project, a software module was developed. This module could be used every month by the person in charge of purchases in Ecuacolor to estimate the demand of paper in the next period. The software allow apply the Data Minig algorithm to a historical data set and change some parameters as the user require.

INTRODUCCIÓN

El objetivo de este proyecto es predecir la necesidad de papel fotográfico para Ecuacolor, tomando como base el comportamiento histórico de la producción fotográfica, para la aplicación de un algoritmo de minería de datos que permita determinar la cantidad de papel a importar para un determinado período. La implementación de este proyecto permitirá a Ecuacolor contar con una herramienta que le ayude a determinar la cantidad apropiada a importar, y de esta forma optimizar los niveles de inventarios de su principal materia prima que es el papel fotográfico.

El material que constituye la documentación del proyecto está compuesto de cinco capítulos y dos anexos, organizados de la siguiente manera:

***El Capítulo 1:* Descripción del Negocio.-** da una descripción general de la empresa Ecuacolor, para la cual se aplica el proyecto, así como, las necesidades de información actuales y la solución propuesta en lo relacionado al tema del proyecto.

***El Capítulo 2:* Descripción de los Datos.-** Describe las bases de datos existentes y la información que contienen, las mismas que serán utilizadas para obtener los datos tomados como base para la predicción.

***El Capítulo 3:* Preprocesamiento de los Datos.-** Describe cada uno de los procesos realizados sobre los datos, para la preparación del archivo final que será sometido a la minería de datos.

***El Capítulo 4:* Selección y Evaluación del modelo.-** Analiza los datos disponibles y los resultados deseados, y explica la selección e implementación del modelo de minería de datos seleccionado

***El Capítulo 5:* Implementación del sistema.–** Explica las consideraciones hechas para la programación e implementación del modelo seleccionado y muestra el uso del sistema en el inicio y periódicamente.

***El Anexo 1.* Selección de atributos relevantes.-** Muestra los resultados obtenidos de aplicar la técnica de selección de atributos previo al uso mismo del algoritmo de predicción.

***El Anexo 2.* Manual de Usuario.-** Contiene información respecto a como funciona y se opera el "Sistema de predicción de la demanda de papel fotográfico". Ayuda a identificar los Usuarios que pueden utilizar el Sistema y que conocimientos mínimos debe tener para lograr una compresión exitosa de cada una de las funciones del mismo.

CONTENIDO

1. DESCRIPCION DEL NEGOCIO

Laboratorios Fotográficos Ecuacolor es una empresa que nació en la ciudad de Quito hace 35 años. gracias a la iniciativa y visión de los señores Luis Orrantia G. y Enrique Martínez Q., Presidente y Gerente General de Comandato de aquel entonces, quienes decidieron crear una empresa dedicada a la venta y procesamiento de las películas en blanco y negro, bajo la razón social de Ecuacolor Laboratorios Fotográficos.

Ecuacolor se dedica en la actualidad a la captura, reproducción, conservación y comunicación de imágenes que son los más preciados recuerdos y sentimientos del ser humano. Gracias a la utilización de tecnología de punta a la experiencia de su recurso humano y a la capacitación continua, ha logrado colocarse como líder en la comercialización y distribución de productos y servicios fotográficos.

El avance tecnológico de la producción fotográfica y la incursión de los medios digitales ha ocasionado que Ecuacolor incluya nuevos servicios, acompañados de múltiples promociones con el fin de mantener su posición en el mercado. La gran demanda de producción en ciertas temporadas del año ha ocasionado que se pierdan ventas por la falta de materiales, sobre todo de papel fotográfico, pues el principal proveedor es Kodak, a quien se debe emitir órdenes de compra con por lo menos 1 mes de anticipación.

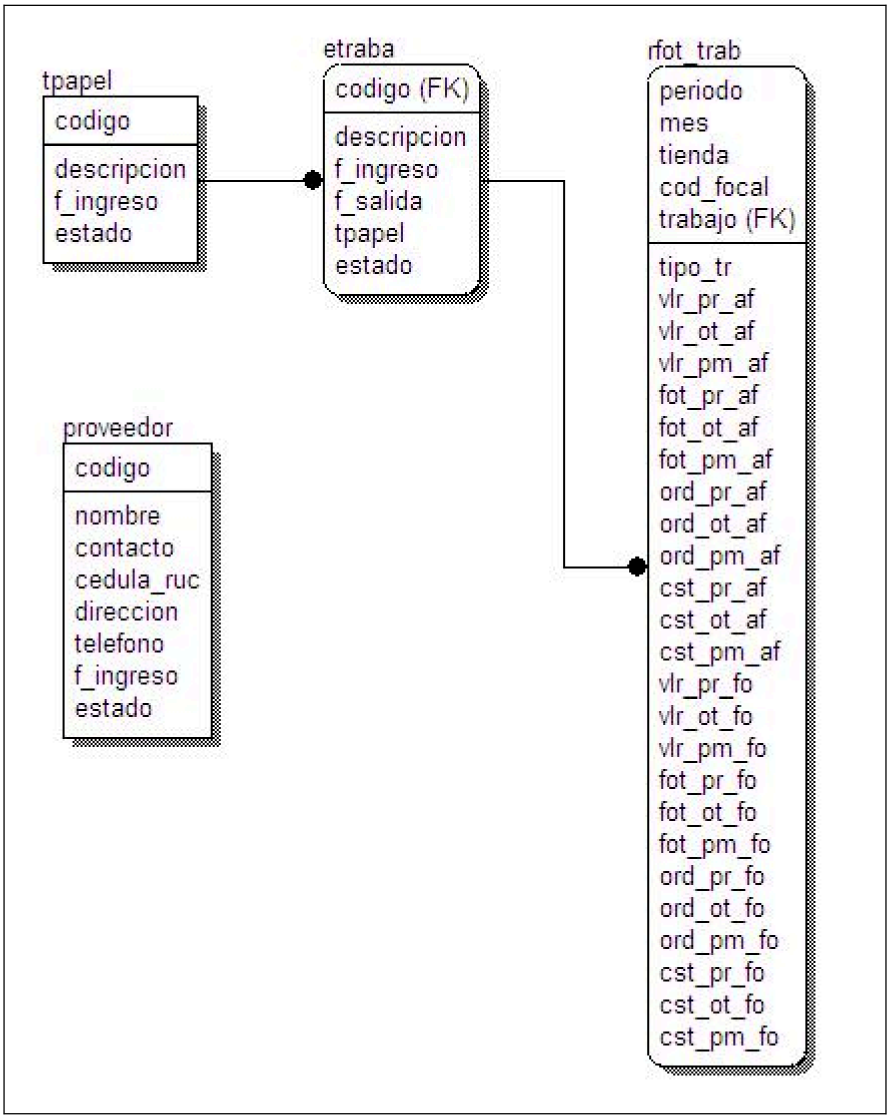
Una vez analizada la situación actual de la empresa, se ha determinado que uno de los problemas existentes es la adquisición de papel fotográfico, básico para su labor de producción, por lo que se considera necesaria la implementación de un sistema que permita determinar la demanda futura en base a los datos históricos de producción, aplicando técnicas de Minería de Datos.

2. DESCRIPCION DE LOS DATOS

Ecuacolor cuenta actualmente con un sistema integrado para el área administrativa, el cual le permite tener información general de las operaciones de la empresa.

Con el crecimiento de la producción fotográfica, se ha desarrollado un Sistema de producción anexo al sistema principal, que permite controlar el volumen producido así como la rentabilidad que se genera en cada una de las fototiendas a nivel nacional.

Del Sistema de Producción actual se han tomado los datos de la utilización de papel en períodos anteriores con el fín de usarlos como base para la predicción de la demanda futura. El sistema cuenta con información histórica desde Enero de 2.000. La estructura general de los datos seleccionados es la siguiente:



3. PREPROCESAMIENTO DE LOS DATOS

De la base de datos del Sistema de Producción, se puede obtener información referente a los consumos de papel, sin embargo esta debe ser preprocesada con el fin de seleccionar los datos que servirán para la minería.

Se dispone del número de fotos y el correspondiente valor de venta totalizados a nivel de periodo – tienda – tipo de trabajo fotográfico, mediante el uso del diagrama entidad – relación, se establece las relaciones pertinentes y se realizan las conversiones necesarias para permitir generar el nivel de resumen adecuado para realizar la minería de datos que es tipo de papel–periodo.

Con el fin de someter a prueba el algoritmo de minería de datos a utilizar se realizó una serie de pasos manuales, que posteriormente serían incorporados al sistema desarrollado. Fue necesario crear estructuras provisionales y completar algunos datos que no estaban disponibles en el Sistema de producción existente.

4. SELECCIÓN Y EVALUACION DEL MODELO

La tabla ***Produccion_papel*** contiene el uso de cada tipo de papel fotográfico en el periodo a analizar y en cada uno de los doce meses inmediatamente anteriores a dicho período, por lo cual decimos que tiene una estructura de Serie de Tiempo.

El primer paso en el análisis de series de tiempo, consiste en graficar la serie. La determinación de los tipos de variaciones encontradas en la serie de tiempo usada en el proyecto permite hacer nuevas consideraciones para lograr el resultado deseado, que es la predicción de la cantidad de papel a usar en un determinado periodo. Sin embargo, este análisis no es tan fácil de obtener con la simple observación de los gráficos de los datos, se requiere para ello la utilización de algún algoritmo de minería de datos que nos permita automatizar dicho análisis.

Para este proyecto contamos con la opción de hasta 14 atributos numéricos y se desea predecir una clase también numérica. Varios modelos de minería de datos, pueden procesar variables numéricas directamente o discretizando atributos numéricos a nominales por diferentes métodos. Sin embargo, hay modelos de minería de datos que trabajan más naturalmente con atributos numéricos, como la regresión lineal. La regresión lineal es un método simple y excelente para la predicción de clases numéricas y ha sido ampliamente usada en estadística por décadas.

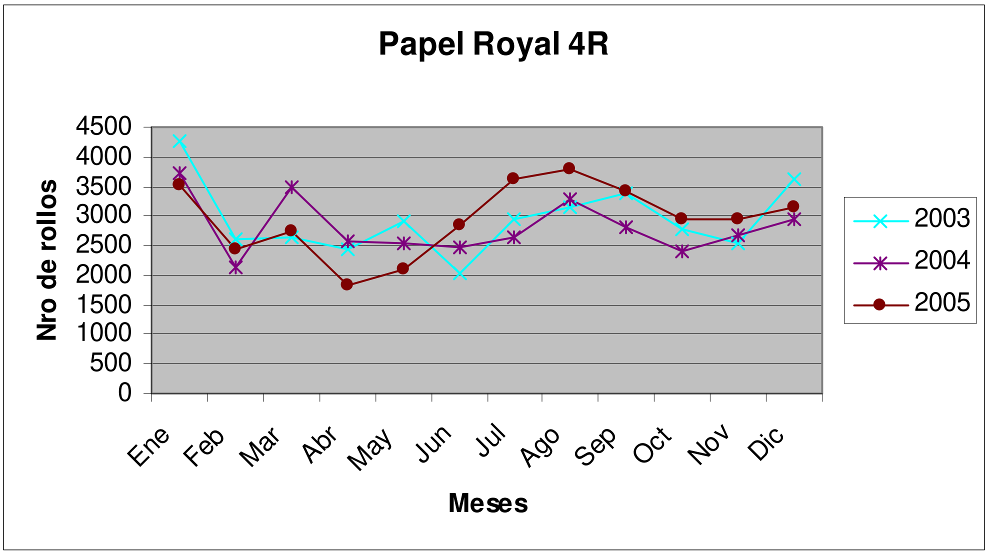
Previo a la elaboración del sistema se probó exclusivamente el algoritmo de regresión lineal, para lo cual se usó datos de producción de Ene.2003 a Dic.2005 del tipo de papel 1 (Papel Royal 4R). Originalmente se utilizó las catorce variables disponibles, y posteriormente se realizó la selección de las variables relevantes usando Weka, lo que generó mejores resultados, por lo cual se incluyó entre las característicos del programa la selección de las variables relevantes de forma individual para cada tipo de papel.

Se probó usando la desviación estándar de la serie total de los datos históricos, esta desviación estándar generó un valor muy alto, producto de las altas variaciones generadas principalmente por la parte estacional, por lo cual se decidió mejorar la estimación, descomponiendo la serie en el componente estacional y el componente aleatorio, con el fin de aplicar únicamente la desviación estándar del componente aleatorio y no la del componente estacional, según se muestra en los cuadros siguientes:

<u>Papel Royal 4R</u>

Serie completa

Anio	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Año	Desviac. Estándar
2003	4247	2591	2637	2441	2899	2029	2944	3130	3380	2776	2540	3607	**35221**	
2004	3723	2142	3488	2563	2549	2462	2629	3295	2799	2407	2689	2955	**33701**	
2005	3524	2453	2745	1830	2085	2849	3628	3785	3431	2941	2934	3159	**35364**	**548**



Se puede apreciar una variación estacional en los tres años analizados.

Componente estacional

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Año
3881	2427	2997	2307	2543	2480	3106	3448	3244	2744	2758	3286	**35221**
3714	2322	2868	2207	2433	2373	2972	3299	3104	2625	2639	3145	**33701**
3897	2437	3009	2316	2553	2490	3119	3462	3257	2755	2769	3300	**35364**

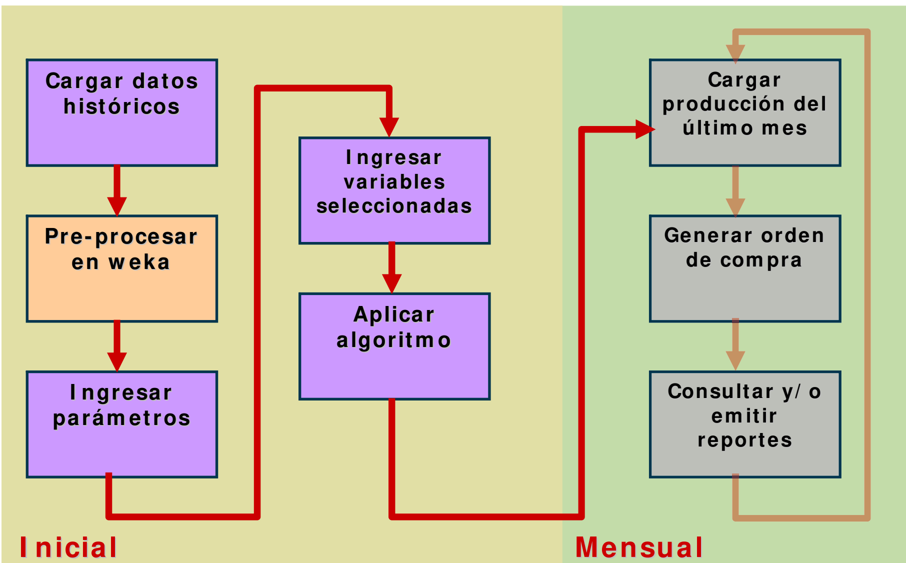
Componente aleatorio

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	Año	Desviac. Estándar
366	164	-360	134	356	-451	-162	-318	136	32	-218	321	**0**	
9	-180	620	356	116	89	-343	-4	-305	-218	50	-190	**0**	
-373	16	-264	-486	-468	359	509	323	174	186	165	-141	**0**	**296**

Finalmente se evaluó la oportunidad de fallar la predicción estimando en más o en menos la demanda, decidiéndose incluir un factor de ajuste calculado con la relación entre el margen de ganancia y el costo financiero, para reflejar que probablemente es más costoso dejar de vender por una subestimación de la demanda (pérdida del margen), que pagar un costo financiero por una sobreestimación de la demanda y su consecuente sobre stock, para lo cual se usará la siguiente fórmula en el programa.

$$\text{Factor (P)} = \frac{1}{\dfrac{\text{Costo Fin}}{\text{Margen}} + 1}$$

5. IMPLEMENTACION DEL SISTEMA



El sistema considera dos puntos de proceso en el tiempo:

Inicial

- Para iniciar la utilización del programa se beberá cargar la información histórica de producción, que es la base para la aplicación del algoritmo, este proceso genera la tabla ***Produccion_papel*** y se hace procesando mes a mes toda la información existente.
- Antes de la ejecución del algoritmo en el sistema, se extrae la tabla ***Produccion_papel*** y se carga en Weka, donde se emplea el método de evaluación *CfsSubsetEval – GeneticSearch* para obtener las variables seleccionadas como relevantes para cada tipo de papel.
- Se marcan las variables seleccionadas como relevantes para cada tipo de papel, seleccionando entre las 14 variables disponibles.
- Se ingresa los parámetros (margen y costo financiero) requeridos por el sistema para ajustar la predicción resultante de aplicar los pesos generados por el algoritmo.
- Se ejecuta el algoritmo, el cual almacena los pesos para cada tipo de papel, que luego serán usados para calcular las órdenes de compra.

Todos estos pasos con excepción del primero pueden ser ejecutados posteriormente a la primera vez, cuando se desee cambiar los pesos obtenidos como resultado del algoritmo.

Mensual

- Se carga la información del mes que se está cerrando. Este proceso adiciona registros a la tabla ***Produccion_papel***, que serán considerados en la emisión de las nuevas órdenes de compra
- Se genera la(s) orden(es) de compra para los diferentes proveedores. Para este proceso se toma la información histórica actualizada (solo los atributos relevantes seleccionados para el tipo de papel), los pesos generados por el algoritmo de regresión lineal, la desviación estándar de los datos históricos (considerando únicamente la variación aleatoria y no la estacional) y los factores de ajuste ingresados (margen de utilidad y costo financiero). El margen y el costo financiero modifican la predicción generada por el algoritmo para reflejar que probablemente es más costoso dejar de vender por una subestimación de la demanda (pérdida del margen), que pagar un costo financiero por una sobreestimación de la demanda y su consecuente sobre stock.
- Se puede hacer consultas y /o emitir reportes en cualquier momento.

CONCLUSIONES

El sistema desarrollado para Ecuacolor ayudará a estimar la demanda de papel fotográfico, permitiéndole determinar la cantidad apropiada a importar, y de esta forma optimizar los niveles de inventario.

Se seleccionó la regresión lineal como el algoritmo de minería de datos para la predicción de la demanda futura de papel fotográfico.

Se hizo algunas consideraciones adicionales, producto del análisis gráfico de los datos disponibles, según el cual se pudo determinar lo siguiente:

- Los datos históricos muestran cero en la producción de algunos periodos, probablemente por falta de stock, por si estos casos se llegan a presentar, el programa considera calcular el valor promedio de los últimos 12 meses y colocar este valor como producción del periodo.
- Existe un comportamiento estacional mensual en el consumo de papel, el cual está relacionado con los feriados que se dan en los diferentes meses del año.
- La presencia de variaciones aleatorias
- No se considera el manejo de datos aberrantes y tendencias, porque en los datos históricos analizados (6 años) no se pudieron detectar.

El programa permite seleccionar los atributos relevantes para cada tipo de papel de manera individual.

Los pesos generados por el algoritmo de regresión lineal se usan para calcular la demanda de papel para el periodo deseado, valor que se modifica con:

- La desviación estándar de los datos históricos, considerando únicamente la variación aleatoria y no la estacional.
- Un factor de ajuste calculado con la relación entre el margen de ganancia y el costo financiero, para reflejar que probablemente es más costoso dejar de vender por una subestimación de la demanda (pérdida del margen), que pagar un costo financiero por una sobreestimación de la demanda y su consecuente sobre stock-Sin embargo, la poca cantidad de información histórica (30 registros) disponible al momento de la implementación del sistema puede degradar en algo la estimación, por lo cual se recomienda ejecutar nuevamente el algoritmo al finalizar el año 2006, con el fin de contar con un mayor número de datos de entrenamiento y obtener mejores resultados.

Adicionalmente, el sistema proporciona opciones de consultas gráficas, con lo cual se puede monitorear la precisión de la predicción e ir haciendo ajustes, tales como: la selección de nuevas variables relevantes, ejecuciones periódicas del algoritmo para generar nuevos pesos y cambios en los factores de ajuste (margen y costo financiero).

Se recomienda ejecutar nuevamente el algoritmo al finalizar cada año, con el fin de contar con un mayor número de datos de entrenamiento y obtener mejores resultados.

REFERENCIAS

Data Mining
Practical Machine Learning Tools and Techniques
Second Edition
Ian H. Witten - Eibe Frank
Elsevier Inc.- Morgan Kaufmann Publishers
2005 - San Francisco, CA, USA

Data Mining Techniques
For Marketing, Sales, and Customer Relationship Management
Second Edition
Michael J.A. Berry - Gordon S. Linoff
Wiley Publishing, Inc.
2004 - Indianapolis, Indiana, USA

Preparación y evaluación de Proyectos
Segunda Edición
Nassir Sapag Chain - Reinaldo Sapag Chain
Editorial McGraw Hill Latinoamericana S. A.
1989 – Bogotá Colombia

Métodos y modelos de investigación de operaciones
Vol.2 Modelos Estocásticos
Segunda Reimpresión
Juan Prawda
Editorial LIMUSA S. A. de CV
1984 – Méjico

Ing. Juan Alvarado O.
Director de Tesis