

INTELIGENCIA EN EL NEGOCIO DEL CAFÉ

BUSINESS INTELLIGENCE IN COFFEE

Hugo Leonardo Acero Rodríguez
Universidad pedagógica y
tecnológica de Colombia
hugo.acero@uptc.edu.co

Tipo de artículo: revisión

Recibido: 2014-03-25

Aceptado: 2014-06-05

RESUMEN

Actualmente, es importante para las organizaciones y empresas de cualquier tipo, el contar con herramientas que les permitan conocer tanto su productividad actual, como la futura. Es allí, donde la inteligencia de negocios juega un papel fundamental en el análisis de datos, ya que permite evaluar mediante diversas técnicas como la Minería de Datos, un conjunto específico de información, para de tal forma extraer conocimiento de valor, y tomar en base a este, las decisiones pertinentes. Es por esto, que el presente proyecto, utiliza una técnica de minería de datos conocida como Regresión Lineal Simple, para hallar mediante una muestra de datos, la proyección de venta esperada para el segundo semestre de 2013 en un establecimiento dedicado a la venta de café.

Palabras clave: Inteligencia de negocios (IB), KDD, Minería de datos (MD), Regresión lineal.

INTRODUCCIÓN

Hoy en día, es imprescindible en cualquier tipo de empresa sea pequeña o grande, el contar con herramientas que le permitan clarificar a sus directivos, no solamente como se encuentra actualmente, sino más importante aún, cómo y en qué lugar estará ella a futuro. Conocer que tan rentable y viable será un negocio en un tiempo más allá de nuestro presente, resulta sumamente interesante, puesto que no solo refleja una repercusión importante en el hecho del conocimiento fidedigno de ciertas cifras que vislumbran la rentabilidad del negocio, sino también y más significativo aun, brinda la posibilidad de saber con certeza si políticas y acciones actuales están bien encaminadas o es necesario replantearlas en busca de garantizar que estos datos futuros demuestren además de un sostenimiento adecuado de la empresa, un crecimiento en las ventas y utilidades de la misma.

Es por esto, que el presente proyecto pretende aplicar **Inteligencia de Negocios** como principal herramienta para el tratamiento y

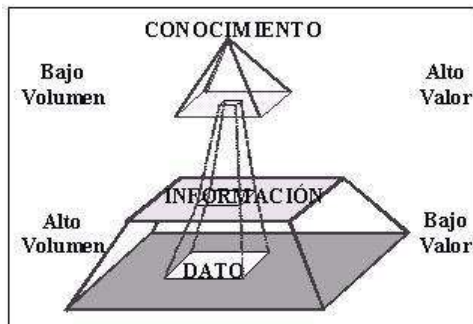
análisis de la información de un establecimiento dedicado a la venta de productos de cafetería. Para esto, se usará una **Regresión Lineal**, técnica de varias existentes de **Data Mining** (minería de datos), con la cual se podrán estudiar los datos que se tienen de la empresa, transformarlos en información fidedigna y debidamente ordenada y finalmente extraer el conocimiento debido de ella, resultado que arrojará las tendencias futuras de ventas que indiquen si el camino que sigue la empresa es el indicado o no y si las utilidades de esta no solamente se mantendrán en el tiempo, sino también si tendrán un incremento o decremento paulatino, que garantice la rentabilidad del negocio o alerte a tomar medidas al respecto.

MARCO TEORICO

En este apartado, se citaran conceptos de los temas más importantes a tratar, con el fin de introducir con facilidad al lector en la temática y desarrollo de la investigación.

- **Inteligencia de Negocios**

Inteligencia de negocios es una práctica que busca transformar los datos de las compañías en información y conocimiento, para que los funcionarios o personas que toman decisiones, cuenten con un apoyo de información actual e histórica del estado de la compañía y así determinar en donde están con respecto a sus objetivos y cuál es el camino más fácil para llegar a ellos. [1], [5]



Inteligencia de Negocios es muchas veces comparado con el GPS. Este muestra a qué velocidad promedio se está andando (información histórica), muestra a qué velocidad y en qué dirección se está a cada momento (información actual) y dice el tiempo de llegada y el camino más rápido a los objetivos establecidos.

La información cada día toma más fuerza como el activo más importante en las

compañías, a veces generando ventajas competitivas usando el conocimiento para descubrir oportunidades de negocio.

Por otro lado, la globalización y las ya casi inexistentes barreras comerciales entre los diferentes países obligan a compañías nacionales a fortalecerse y volverse más eficientes para poder sobrevivir ante las gigantes multinacionales. La inteligencia de negocios juega un papel vital en este proceso de volver a las compañías más eficientes porque les muestra que unidades de negocio son las más rentables, lo que ayuda a una asignación de los recursos más eficiente y potencializa el desempeño de la compañía.

Por último, esta práctica reemplaza la elaboración de reportes, cálculo y construcción de indicadores, y disminuye al mínimo la manipulación de información que genera diversas fuentes de la verdad en las compañías.

Beneficios:

- Información en el momento, lugar y formato correcto: al tener toda la información en tiempo actual y de la forma que se necesita, los gerentes tienen la ventaja de poder tomar medidas correctivas acertadas en momentos críticos.
- "Investigar" de lo general a lo particular: los usuarios tendrán la capacidad de llegar a las fuentes específicas de los problemas. Muchas veces los gerentes saben que hay un problema pero no conocen con certeza cuál es el origen de este por la cantidad de variables involucradas. Inteligencia de negocios soluciona este problema con solo un par de "clicks".
- Visualización de indicadores y alarmas: al implementar una solución de inteligencia de negocios, los usuarios pueden visualizar fácil y prácticamente en tableros "dashboard" el estado de su compañía, establecer parámetros en el comportamiento de indicadores y programar

alarmas para tomar acciones preventivas en caso que el comportamientos de estos indicadores se salgan de sus parámetros.

- Alinear las decisiones hacia los objetivos: esta solución usa información histórica y la compara con la actual, lo que permite que los gerentes o los tomadores de decisiones de las instituciones trazar un camino lógico para llevar a la compañía a lograr sus objetivos. [1].
- Una solución de inteligencia de negocios completa permite:
 - Observar ¿qué está ocurriendo?
 - Comprender ¿por qué ocurre?
 - Predecir ¿qué ocurriría?
 - Colaborar ¿qué debería hacer el equipo?
 - Decidir ¿qué camino se debe seguir?

Entre las herramientas de bases de datos de mayor uso en la BI están.[4]

- La minería de datos.
- Las herramientas OLAP
- Data Warehouses
- Data Marts

➤ Query & Report

Aunque todas estas herramientas son de gran importancia y cada una de ellas brinda ciertos beneficios mutuos y excluyentes con las otras, solamente se estudiara con más profundidad la Minería de Datos puesto que como se verá más adelante en su definición, se adapta de manera oportuna a nuestra necesidad.

• Minería de datos

Una definición tradicional es la siguiente: Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos (Fayyadyotros, 1996). Desde el punto de vista empresarial, se expresa como: 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 (Molina y otros, 2001). [2]

La revolución digital ha hecho posible que la información digitalizada sea fácil de capturar, procesar, almacenar, distribuir, y transmitir. Con el importante progreso en

informática, el internet y las tecnologías relacionadas y la expansión de su uso en diferentes aspectos de la vida, se continúa recogiendo y almacenando en bases de datos gran cantidad de información.

Descubrir conocimiento de este enorme volumen de datos es un reto en sí mismo. La minería de datos (MD) es un intento de buscarle sentido a la explosión de información que actualmente puede ser almacenada. [3]

El Data Mining (DM) o minería de datos es una técnica orientada a extraer información desde grandes volúmenes de datos y corresponde a una de las claves de la inteligencia de negocios. El DM es un proceso conocido como KDD (Knowledge Discovery in Databases) que consta de tres fases: preparación de los datos, extracción de información e interpretación de los resultados. [4]

Normalmente, el DM se realiza con las llamadas “técnicas de verificación” -en donde se prueba una hipótesis del

usuario- y los “métodos de descubrimiento” -búsqueda de patrones (incluyendo las técnicas de predicción)-. Lo anterior se traduce en análisis exploratorios, los cuales se apoyan en herramientas de visualización de datos y construcción de modelos, con el objetivo de realizar predicciones utilizando los patrones detectados en los datos conocidos con información del pasado. De allí que el DM no es una herramienta orientada sólo a las asociaciones estadísticas, sino una tecnología que es fundamentalmente prospectiva, esto es, que permite descubrir nuevos factores, tendencias y asociaciones desconocidas previamente, llegando incluso a ser predictiva.

Los modelos de minería de datos, se usan con éxito en aplicaciones de control de procesos productivos, como herramienta de ayuda a la planificación y a la decisión en marketing, finanzas, entre otras. Asimismo, la minería de datos es fundamental en la investigación científica y técnica, como herramienta de análisis y descubrimiento de conocimiento a partir de datos de observación o de resultados de experimentos.

Ejemplos de uso de la minería de datos:

- Negocios.
- Hábitos de compra en supermercados.
- Patrones de fuga.
- Fraudes.
- Recursos humanos.
- Comportamiento en Internet.
- Terrorismo.
- Juegos.
- Ciencia e Ingeniería Genética.
- Análisis de gases

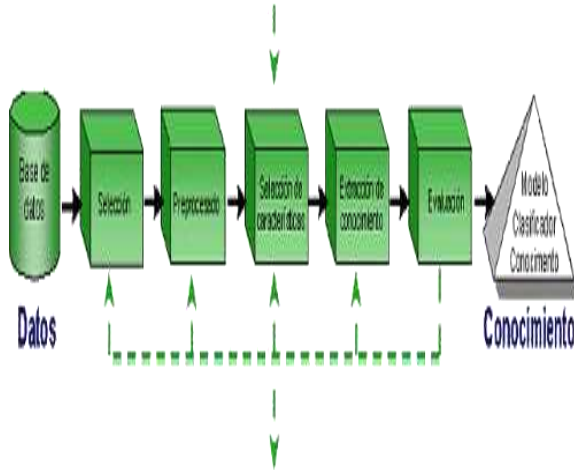
Un proyecto de minería de datos tiene varias fases necesarias que son, esencialmente:

- Comprensión del negocio y el problema que se quiere resolver.
- Determinación, obtención y limpieza de los datos necesarios.
- Creación de modelos matemáticos.
- Validación, comunicación, entre otras de los resultados obtenidos.

- Integración, si procede, de los resultados en un sistema transaccional o similar.

Proceso general del Data Mining:

- Selección del conjunto de datos
- Análisis de las propiedades de los datos.
- Transformación del conjunto de datos de entrada.
- Seleccionar y aplicar la técnica de minería de datos.
- Extracción de conocimiento.
- Interpretación y evaluación de datos



[6]

Aquí es importante destacar la técnica de minería de datos que se espera implementar. Estas provienen de la Inteligencia artificial y de la estadística y no son más que algoritmos, más o menos

sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados.

Técnicas más representativas:

- Redes neuronales.
- Regresión lineal.
- Árboles de decisión.
- Modelos estadísticos.
- Agrupamiento o Clustering.

Para continuar con el análisis se ha decidido implementar la técnica de regresión lineal y entre sus tipos la simple, pues, como se verá a continuación, es la más apropiada según el tipo de información con la que se cuenta para hacer el estudio.

• Regresión Lineal

Es la técnica más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.

El análisis de regresión lineal es una técnica estadística utilizada para estudiar la relación entre variables. Se adapta a una amplia variedad de

situaciones. En la investigación social, el análisis de regresión se utiliza para predecir un amplio rango de fenómenos, desde medidas económicas hasta diferentes aspectos del comportamiento humano. En el contexto de la investigación de mercados puede utilizarse para determinar en cuál de diferentes medios de comunicación puede resultar más eficaz invertir; o para predecir el número de ventas de un determinado producto.

En física se utiliza para caracterizar la relación entre variables o para calibrar medidas. Tanto en el caso de dos (2) variables (regresión simple) como en el de más de dos variables (regresión múltiple), el análisis de regresión lineal puede utilizarse para explorar y cuantificar la relación entre una variable llamada dependiente o criterio (Y) y una o más variables llamadas independientes o predictoras (X_1, X_2, \dots, X_k), así como para desarrollar una ecuación lineal con fines predictivos. Además, el análisis de regresión lleva asociados una serie de procedimientos de diagnóstico (análisis de los residuos, puntos de influencia) que informan sobre la estabilidad e idoneidad

del análisis y que proporcionan pistas sobre cómo perfeccionarlo. [7]

Debido a la sencillez de los datos y relaciones que serán analizados, teniendo claro que solo se cuenta con información global de las ventas diarias del negocio sin tener especificación alguna por producto, lo cual deja una única variable independiente con la cual contar para el estudio, el tipo de Regresión que se utilizara será por tanto la **Regresión Lineal simple**.

La función de regresión lineal simple es un método estadístico para detectar relaciones entre dos variables, X e Y. Mientras que a la variable X se le denomina independiente, caracterizada por tener valores fijos y ordenados, medidos sin error. La variable dependiente es Y, sus valores son aleatorios y distribuidos a través del eje de variables dependientes, por encima de la recta X. Y es la variable que se desea predecir. Se puede aplicar la función de regresión lineal cuando se sabe que existe una correlación lineal entre las variable. [8]

La representación de la función es la siguiente:

$$Y = a + b X$$

Dónde:

a es el valor de la ordenada donde la línea de regresión se intercepta con el eje Y.

b es el coeficiente de regresión poblacional (pendiente de la línea recta.

Aplicando el método de mínimos cuadrados para resolver la a y b de la función: [8]

$$a = \bar{Y} - b \bar{X}$$

$$b = \frac{\sum(x_i - \bar{X})(Y_i - \bar{Y})}{\sum(x_i - \bar{X})^2} = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2}$$

De esta forma, se cuenta ya a esta altura del proyecto, con los fundamentos necesarios para desarrollar mediante la aplicación de las técnicas escogidas, los procedimientos y procesos necesarios

para obtener la respuesta al objetivo planteado.

DESARROLLO

La metodología que se usara para desarrollar la investigación está enmarcada por supuesto en las fases y procedimientos establecidos en la Minería de Datos.

➤ **Comprensión del negocio y el problema que se quiere resolver.**

En qué lugar se efectuara el estudio y cuáles son sus características?

El establecimiento en el cual se va a realizar el estudio es un Café internet llamado “TU NORTE *café internet*” ubicado en la ciudad de Tunja/Boyacá/Colombia en el costado norte de la esta, y cuyos principales usuarios en un 70% (según datos ofrecidos por el propietario del establecimiento) son los estudiantes de la Universidad de Boyacá ubicada a aproximadamente 200 metros del establecimiento. El porcentaje restante de clientes está constituido por el personal y usuarios de la clínica Medilaser ubicada a tan solo 60

metros del establecimiento y los habitantes del sector.

El establecimiento cuenta con zona de fumadores y no fumadores con un total de 12 mesas en las que se pueden albergar 50 personas además de una zona para la prestación del servicio de internet, la cual cuenta con 7 computadores. Entre los productos y servicios ofrecidos por el establecimiento están los de cafetería (cafés, empanadas, gaseosas, golosinas, dulcería, limonadas, jugos naturales, pasabocas, entre otros) y los servicios derivados de internet (minutos, recargas, impresiones, escáner, fotocopias, papelería, venta de películas, alquiler de módems, entre otros), los cuales buscan suplir principalmente las necesidad del porcentaje más grande de clientes como lo son los estudiantes universitarios.

➤ **Determinación, obtención y limpieza de los datos necesarios.**

Selección del conjunto de datos

Los datos con los cuales se cuenta para el estudio son proporcionados por el propietario del establecimiento en físico y corresponden al total diario de ventas y gastos y la diferencia entre el uno y el otro

que se determina como la ganancia o utilidad diaria del establecimiento. La cantidad de datos proporcionados está determinada por los movimientos del último año en el establecimiento lo cual arroja más o menos 360 registros de ventas y el mismo número en gastos descartando cinco (5) registros que corresponden a días festivos de importancia en lo que no se laboró en el establecimiento. En la siguiente grafica se muestra una pequeña porción de los datos suministrados por el usuario.

Año	Mes	semana/Dia	Dia L	Neto del Cargo	rec. Saldo in rec.	Saldo Fin	venta recabado in caja	Saldo Fin	Ver totales	Ver - net	Herty		
2013	Enero	1	3 Lunes	52,500	52,500		52,500	0		0	-52,500		
2013	Enero	1	4 Viernes	157,500	157,500		157,500	0		0	-157,500		
2013	Enero	1	5 Sábado	52,500	52,500		52,500	0		0	-52,500		
2013	Enero	1	6 Domingo			0	0	0		0	0		
2013	Enero	1	7 Lunes	52,500	52,500		52,500	0		0	-52,500		
2013	Enero	2	8 Martes	157,500	157,500	380,000	46,500	540,000	346,700	334,170	305,950	346,550	58,000
2013	Enero	2	9 Miércoles	180,000	180,000	48,000	40,000	96,700	342,000	38,300	334,650	274,050	348,000
2013	Enero	2	10 Jueves	180,000	153,000	95,000	58,000	238,000	374,400	275,000	294,200	238,100	40,000
2013	Enero	2	11 Viernes	81,000	81,000	38,000	67,000	94,400	402,000	48,000	375,400	308,600	348,000
2013	Enero	2	12 Sábado	233,000	233,000	110,000	138,000	162,000			162,000	348,000	
2013	Enero	2	13 Domingo			133,000	94,000	30,000	276,750	98,000	275,550	237,550	338,000
2013	Enero	3	14 Lunes	94,000	94,000	0	94,000	148,750	498,000	298,300	434,550	348,550	348,000
2013	Enero	3	15 Martes	52,500	52,500	0	52,500	238,000	81,000	527,000	464,000	412,400	0
2013	Enero	3	16 Miércoles			0	0	81,000	438,400	35,300	432,000	418,000	508,000
2013	Enero	3	17 Jueves			0	0	138,400	308,000	338,400	467,000	467,000	35,000
2013	Enero	3	18 Viernes			0	0	5,800	412,041	238,000	637,241	637,241	358,000
2013	Enero	3	19 Sábado			0	0	53,041	438,000	68,000	438,959	426,959	348,000
2013	Enero	3	20 Domingo			0	0	78,000	238,381	313,300	274,000	274,000	98,000
2013	Enero	4	21 Lunes			0	0	348,200	638,000	387,750	877,450	877,450	448,000
2013	Enero	4	22 Martes			0	0	438,000	578,000	778,000	894,450	894,450	408,000

Muestra del proceso carga a una base de datos de Excel, de los datos proporcionados por el propietario del establecimiento

Teniendo en cuenta que el objetivo del estudio es generar una proyección

de ventas para el segundo semestre del 2013, la muestra que se tomara para el respectivo análisis corresponderá únicamente a los datos recolectados en el primer semestre del año en mención (1 de enero a 2 de junio de 2013), pues por ser la información más cercana al periodo de tiempo que se desea proyectar, contiene los datos más precisos acerca del promedio de venta actual del establecimiento.

Análisis de las propiedades de los datos.

Algo importante para destacar en los datos obtenidos para el análisis, es que no permiten conocer las ventas por productos específicos pues como se mencionó anteriormente, son totales globales de venta con los que cuenta el proyecto, por cuanto es imposible obtener datos por producto de manera individual y hacer algún tipo de análisis al respecto, de tal forma que la finalidad será la obtención de totales proyectados de venta para el segundo semestre de 2013.

Otro aspecto de interés, es que de los datos proporcionados por el propietario, faltan alrededor de 25 registros de

diferentes días de los cuales este afirma no tener manera de recuperarlos.

También es importante resaltar que las ventas del establecimiento en los meses en que los estudiantes universitarios se encuentran en vacaciones, se reducen notoriamente, lo cual afecta de manera significativa el promedio mensual de ventas en cada semestre.

Transformación del conjunto de datos de entrada.

Para solucionar los inconvenientes presentes en la información, expuestos en el párrafo anterior, se realizaron los siguientes ajustes:

1. Para enmendar la falta de algunos de los datos necesarios para el análisis, se determinó promediar la información de días iguales al faltante pero de semanas diferentes, y así trasladar el resultado al campo faltante. Para entender con facilidad el proceso que efectuado se

mostrara a continuación un ejemplo:

Añc.	Mes	semar.	Di.	Día L	festivc.	saldo ini caj.	saldo f caj.	Gasto.	Ven totale.	Ven - rer.	Efecty.
2013	Marzo	9	1	Viernes		251,000	979,800	280,490	1,009,290	907,790	775,000
2013	Marzo	9	2	Sábado		204,800	368,100	386,900	550,200	550,200	270,000
2013	Marzo	9	3	Domingo		98,100	387,900	83,300	373,100	373,100	194,000
2013	Marzo	10	4	Lunes		193,900	691,200	552,200	1,049,500	965,500	440,000
2013	Marzo	10	5	Martes		251,200			-251,200	-324,700	
2013	Marzo	10	6	Miércoles		0	587,400	619,050	1,206,450	1,122,450	330,000
2013	Marzo	10	7	Jueves		257,400	921,200	381,950	1,045,750	749,750	660,000

Muestra de datos faltantes en el día Martes 5 de Marzo

Año	Mes	semana	Dia	Día L	festivos	saldo ini caja	saldo f caja	Gastos	Ven totales	Ven - rec	Efecty
2013	Enero	4	22	Martes		458,900	576,600	778,950	896,650	896,650	400,000
2013	Enero	5	29	Martes		397,900	605,100	614,350	821,550	741,550	400,000
2013	Febrero	6	5	Martes		256,400	278,500	908,900	931,000	830,000	40,000
2013	Febrero	7	12	Martes		244,400	657,800	589,650	1,003,050	920,550	400,000
2013	Febrero	8	19	Martes		451,500	418,400	961,000	927,900	810,900	150,000
2013	Febrero	9	26	Martes		257,500	155,400	900,750	798,650	724,650	0
2013	Marzo	11	12	Martes		675,400	779,600	901,850	1,006,050	930,750	500,000
2013	Marzo	12	19	Martes		508,200	782,100	723,850	997,750	894,750	520,000
2013	Abril	14	2	Martes		215,700	497,642	708,531	990,472	922,472	290,000
2013	Abril	15	9	Martes		224,100	447,600	681,100	904,600	856,600	230,000
2013	Abril	16	16	Martes		245,000	416,350	694,900	866,250	790,250	210,000
2013	Abril	17	23	Martes		203,000	379,500	673,200	849,700	821,700	190,000
2013	Abril	18	30	Martes		265,100	580,800	795,350	1,111,050	1,059,050	430,000
2013	Mayo	19	7	Martes		504,200	658,900	819,950	974,650	878,650	420,000
2013	Mayo	20	14	Martes		223,400	725,400	373,700	875,700	849,700	510,000
2013	Mayo	21	21	Martes		187,000	326,100	793,600	932,700	860,700	140,000
2013	Mayo	22	28	Martes		181,200	391,800	649,700	860,300	820,300	250,000
Promedio día Martes							510,447	739,372	926,354	859,366	298,824

Método para extraer el promedio de los martes restantes contenidos en la muestra

tanto, la muestra resultante y

Añc.	Mes	seman	Di	Día L	festivc	saldo ini caj	saldo f caj	Gastor	Ven totale	Ven - rec	Efecty
2013	Marzo	9	1	Viernes		251,000	979,800	280,490	1,009,290	907,790	775,000
2013	Marzo	9	2	Sábado		204,800	368,100	386,900	550,200	550,200	270,000
2013	Marzo	9	3	Domingo		98,100	387,900	83,300	373,100	373,100	194,000
2013	Marzo	10	4	Lunes		193,900	691,200	552,200	1,049,500	965,500	440,000
2013	Marzo	10	5	Martes		251,200	510,447	739,372	998,619	925,119	298,824
2013	Marzo	10	6	Miércoles		211,623	587,400	619,050	994,827	910,827	330,000
2013	Marzo	10	7	Jueves		257,400	921,200	381,950	1,045,750	749,750	660,000
2013	Marzo	10	8	Viernes		261,200	686,200	918,100	1,343,100	1,242,100	500,000

Inclusión de los promedios obtenidos en el proceso anterior

2. En relación a los días de vacaciones y su posible afectación en los promedios de venta y de igual forma en las proyecciones, la solución planteada permitió descartar los datos de las semanas en las cuales la Universidad de Boyacá se encontraba en receso, producto de los periodos de vacaciones establecidos normalmente. Así, se suprimieron de la muestra a analizar, los datos correspondientes a los meses de enero y junio 2013 respectivamente, pues son periodos en que los estudiantes de la Universidad en mención se encuentran en vacaciones. Por

definitiva toma 120 días de datos, traducidos en 18 semanas o 4 meses lo que constituye el primer

3. semestre académico de los estudiantes adscritos a esta institución, datos homogéneos para realizar un adecuado estudio y proyección final de ventas para el segundo semestre académico de 2013.

4. Se efectuaron pruebas de aplicación del método elegido analizando los datos contenidos en la muestra de manera individual (diaria) y en totales semanales y mensuales. Se determinó consecuentemente, que no es posible analizar los datos individualmente o agrupándolos de forma semanal, debido a

que la Regresión Lineal Simple busca tendencias de incremento secuencial en los datos, es decir que requiere según nuestro caso de aplicación, que los datos tengan un incremento paulatino en sus cifras, no necesariamente de manera significativa, pero si progresiva, pues de lo contrario, las constantes calculadas para hallar los datos proyectados de venta, carecen de exactitud determinando errores en el resultado.

Es así, pues analizando un poco los datos que producen errores en el momento de su aplicación como los totales diarios de venta, es claro que hay muchas situaciones en los que estos valores de ventas de fechas anteriores son mayores que algunos más recientes, hecho que es evidente pues puede haber más ventas en los primeros días del mes laborar (por causa del pago salarial), que en los últimos días del mes. De la misma forma sucede con los totales por semana.

Por tanto, los únicos grupos de datos susceptibles de ser analizados y que pueden brindar proyecciones y tendencias reales de ingresos para el caso de análisis, son los totales de venta por mes, pues cada uno de ellos envuelve en un único valor, todas las situaciones que pueden afectar individualmente los días contenidos en ellos.

La muestra por tanto quedara reducida a la siguiente tabla:

Meses 2013	Ventas por mes
Febrero	9,207,600
Marzo	9,711,289
Abril	10,390,850
Mayo	10,567,600
Total general	40,387,339

Tabla de ventas en tiempo de actividad académica, primer semestre 201()3

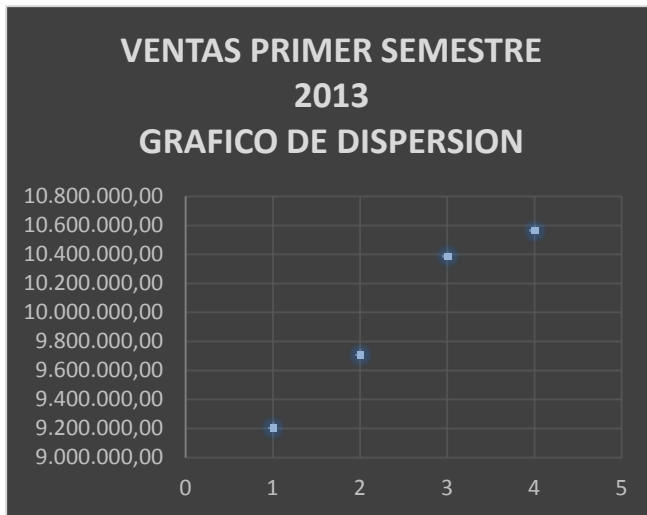


Gráfico de dispersión ventas (febrero, marzo, abril y mayo de 2013)

➤ **Creación de modelos matemáticos.**

Seleccionar y aplicar la técnica de minería de datos.

Para poder hacer un análisis certero de la información proporcionada, se usara como se ha mencionado anteriormente, el método de **Regresión Lineal Simple**, pues como se explicó en el capítulo anterior y a diferencia de otras técnicas (redes neuronales, regresión lineal múltiple, entre otras) que requieren contar con más de un tipo de dato, permite a partir de una única variable independiente X (que para el caso serán las ventas),

obtener una variable dependiente Y, que será finalmente la proyección de ventas.

Como ya vimos la formula general por medio de la cual se determina la proyección de venta esperada es la siguiente.

$$Y = a + b X$$

Para resolverla y conocer los valores resultantes, primero debemos hallar el valor de las variables **a** y **b**.

Dónde: “**a**” es el valor de la ordenada donde la línea de regresión se intercepta con el eje Y y “**b**” es el coeficiente de regresión poblacional (pendiente de la línea recta)

$$a = \bar{Y} - b \bar{X}$$

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum X_i Y_i - n \bar{X} \bar{Y}}{\sum X_i^2 - n \bar{X}^2}$$

Lo primero que haremos será hallar el valor de **b** pues como se aprecia en las formulas

es necesario tener este dato para conocer el resultado de la variable **a**. Para esto, hallamos los valores de las operaciones contenidas en la ecuación:

i	x (meses)	y (ventas)	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$
1	2	9,207,600.00	1,142,602.30	2.25
2	3	9,711,289.47	129,022.70	0.25
3	4	10,390,850.00	210,757.57	0.25
4	5	10,567,600.00	897,397.70	2.25
Suma			2,379,780.26	5

Por tanto,

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{2.379.780}{5} = 475.956$$

Ahora, reemplazamos **b** en la ecuación que halla el valor de la variable **a**:

$$a = \bar{Y} - b \bar{X} = 9.969.334,87 - (475.956 * 2.5) = 8.779.444,74$$

Ya tenemos los valores de las variables **a** y **b** que deben ser reemplazados en la fórmula de regresión lineal.

$$Y = a + b X = (8.779.444,74) + (475.956)X$$

Solo resta reemplazar 'X' por los meses a los que deseemos proyectar la venta. En nuestro caso, debido a que esperamos conocer los ingresos del establecimiento para el segundo semestre de 2013, los meses de actividad universitaria en el sector serán los 4 siguientes a los contenidos en la muestra. Por cuanto las proyecciones de venta quedarían así:

Extracción de conocimiento.

Ahora reemplazando la X en la fórmula obtenida por los meses 5,6,7 y 8 (que para el caso de estudio serán los meses agosto, septiembre, octubre y noviembre de 2013 respectivamente), nuestras ventas proyectadas serán las siguientes.

X (mes)	Y (ventas)	Y (proyeccion de ventas)
1	\$9,207,600	
2	\$9,711,289	
3	\$10,390,850	
4	\$10,567,600	
5	$Y = 8.779.445 + (475.956 * 5)$	\$11,159,225
6	$Y = 8.779.445 + (475.956 * 6)$	\$11,635,181
7	$Y = 8.779.445 + (475.956 * 7)$	\$12,111,137
8	$Y = 8.779.445 + (475.956 * 8)$	\$12,587,093

➤ **Validación, comunicación, de los resultados obtenidos.**

Para validar que los datos obtenidos son totalmente fiables y podemos confiar plenamente en ellos, la técnica de Regresión Lineal Simple brinda la posibilidad de comprobarlo hallando el denominado Coeficiente de correlación "R".

Para hallarlo tenemos que despejar la siguiente ecuación:

$$R^2 = \frac{Cov(X,Y)}{s_x s_y} = \frac{\sum_{i=1}^n (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} * \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \leq +1$$

Debido a que en ecuaciones anteriores tenemos los datos solicitados para para despejar el interrogante, reemplazamos y obtenemos el resultado de R²

$$R^2 = 0.96$$

La teoría dice que entre más cercano esté R² a 1 (uno) más exacta es al predicción y más confiables seran los datos resultantes. En cambio, entre más cercano este el resultado de la variable R² a 0 (cero) la proyección es menos fiable y

los datos resultantes serán menos confiables.

Por tanto, se observa que el estudio y los resultados proyectados son de alta confiabilidad, ya que el coeficiente de correlación se aproxima bastante a 1 (uno).

Interpretación y evaluación de datos.

La siguiente grafica combina los puntos de dispersión (datos de venta), en combinación con la línea de tendencia dibujada a partir de la formula obtenida.



Grafica de dispersión, línea de tendencia y formula de la recta.

Analizando los datos se concluye que las ventas para el segundo semestre de 2013 serán mayores que las reportadas en el primer semestre de 2013. Se evidencia una tendencia de incremento mensual de aproximadamente \$ 475.956 lo cual equivale a un 6% aproximadamente. Este estudio demuestra que las políticas de atención al cliente, presentación de productos, aseo entre otros aspectos a los cuales se les presta mayor interés en el establecimiento evaluado, están teniendo efectos positivos en él. Por tanto, se continuaran aplicando aun con más control las políticas previstas por el establecimiento, pues el objetivo de este es superar esta proyección de ventas para el segundo semestre de 2013 en al menos un 4%.

BIBLIOGRAFÍA

[1]Inteligencia Negocios - Procibernetica.
<http://www.procibernetica.com/inteligencia-negocios.html>.

[2]«sofi_mineria.dvi -
Mineria_Datos_Vallejos.pdf». [En línea].:
http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/Mineria_Datos_Vallejos.pdf

[3]«Microsoft Word - estado del arte-def.doc - 308.pdf». [En línea].
Disponble en:
<http://cabrillo.lsi.uned.es:8080/aepia/Uploads/29/308.pdf>

[4]«Inteligencia de negocios». <http://www.slideshare.net/Alevsk/inteligencia-de-negocios-9667502>.

[5]«Datos, información, conocimiento». http://sinnexus.es/business_intelligence/piramide_negocio.aspx

[6]«DATA MINING PARA AGRONOMIA», *Ingeniería en Sistemas 2009 UNL.*
<http://sistemas2009unl.wordpress.com/data-mining-para-agronomia/>.

[7]«Análisis de regresión lineal - 18reglin_SPSS.pdf». [En línea].
http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/material/es/analisis_datosyMultivariable/18reglin_SPSS.pdf.

[8]«MDX: Implementar recta de regresión lineal simple». [En línea]. Disponible en: <http://blogs.solidq.com/bicorner/post.aspx?id=75&title=mdx%3A+implementar+recta+de+regresi%C3%B3n+lineal+simple>