

# Tecnologías de Inteligencia de Negocios y Minería de datos para el análisis de la producción y comercialización de cacao

## Business Intelligence and Data Mining Technologies for the analysis of cocoa production and commercialization

Bertha MAZÓN-Olivo [1](#); Marcia JARAMILLO-Paredes [2](#); Oscar ROMERO-Hidalgo [3](#); Amarilis BORJA-Herrera [4](#); Martha AGUIRRE-Benalcazar [5](#); María CONTENTO-Segarra [6](#)

Recibido: 27/02/2018 • Aprobado: 20/04/2018

### Contenido

- [1. Introducción](#)
- [2. Metodología](#)
- [3. Resultados](#)
- [4. Conclusiones](#)

[Referencias bibliográficas](#)

#### RESUMEN:

El propósito de este trabajo es identificar los indicadores clave de rendimiento (KPI) de producción y comercialización de una empresa de cacao; la metodología CRISP-DM se utilizó para orientar el desarrollo e implementación de un tablero de control (dashboard) de análisis de datos y apoyo a la decisión a través de tecnologías de inteligencia de negocios (BI) y minería de datos (DM). Los resultados más significativos son un dashboard BI-DM con interfaz web que accede a un dataware house, visualiza información estadística, detecta patrones de datos y predice el rendimiento y ventas de la cosecha de cacao mediante series temporales, método de Holt-Winters con un  $R^2 > 0.9$ .

**Palabras-Clave:** cacao, comercialización, dashboard, inteligencia de negocios, minería de datos, producción

#### ABSTRACT:

The purpose of this paper is to identify the key performance indicators (KPI) of production and commercialization of a cocoa company; the CRISP-DM methodology was used to guide the development and implementation of a dashboard for data analysis and decision support through business intelligence (BI) and data mining (DM) technologies. The most significant results are a dashboard BI-DM with web interface that accesses a dataware house, visualizes statistical information, detects data patterns and predicts the yield and sales of the cocoa crop through time series, Holt-Winters method with  $R^2 > 0.9$ .

**Keywords:** business intelligence, cocoa, commercialization, dashboard, data mining, production

## 1. Introducción

En la actualidad las empresas de cacao se han convertido en principales motores de la economía en varios países; Shavez, Ahmad, Jan, & Bashir (2017) mencionan, que la demanda de este producto aumenta a nivel mundial, en este sentido en Sudamérica según

la International Cocoa Organization (2017) es una de las principales regiones productoras de cacao. En el Ecuador, específicamente en la provincia de El Oro, existen varias organizaciones que se dedican a esta actividad desde hace varios años, según los reportes del Instituto Nacional de Estadística y Censos (2014); empresas que con el paso del tiempo y los cambios tecnológicos, se han visto en la necesidad de implementar mejoras, no sólo en el ámbito operativo, sino también en el manejo de la información.

Para analizar la información que puede ser clave en una empresa cacaotera, es necesario comprender el proceso de producción de cacao. Torres (2012) describe las siguientes fases: selección y preparación del suelo, siembra, riego, fertilización, control de plagas y malezas, poda, cosecha, quiebra, fermentación, secado, limpieza, selección del grano y almacenamiento; sin embargo, Utomo et al. (2016), destacan que los procesos efectuados, pueden variar y depender del lugar de producción del cacao.

Este trabajo se llevó a cabo en la empresa Agrisolsa S.A. radicada en el cantón Machala-Ecuador; su objeto social es la producción y comercialización de cacao. Esta empresa registra a diario múltiples transacciones en las que parcialmente se detallan cuentas inherentes a su actividad principal. Obtener información a partir de los datos en bruto de procesos operativos, puede ser útil para diagnosticar problemas, detectar patrones o identificar tendencias o simplemente generar ventajas frente a la competencia. En este sentido, Nguyen et al. (2017) explican que las diferentes fases productivas y de comercialización de cualquier empresa, generan grandes volúmenes de datos; si estos datos se procesan para obtener información organizada y presentada a la persona indicada en el momento oportuno, pueden ofrecer beneficios económicos a las organizaciones.

El objetivo de este artículo es identificar los indicadores clave de rendimiento (KPI) de producción y comercialización de una empresa de cacao, a través de herramientas y técnicas de Inteligencia de Negocios y Minería de Datos (Power BI, PostgreSQL, Open Talend Studio, y el lenguaje de programación R + librerías Shiny) para el apoyo a las decisiones. Se utilizó la metodología Cross Industry Standard Process for Data Mining (CRISP-DM) para el análisis, diseño e implementación un tablero de control (dashboard BI-DM), con interfaz web, mediante herramientas de inteligencia de negocios y minería de datos como: visualiza información estadística relevante del estado actual de la organización, detecta patrones y predice tendencias.

## **1.1. Trabajos relacionados de Inteligencia de Negocio y Minería de datos**

En toda empresa la información es importante para tomar decisiones y establecer rutas a seguir, es sinónimo de comprender su propia realidad y poder actuar a tiempo. Según Stuart et al. (2017), si la información está correctamente procesada y analizada, se constituye un factor estratégico para la toma de decisiones. En este contexto aparece la Inteligencia de Negocios (BI: Business Intelligence), como parte de una estrategia empresarial, que actúa como una tecnología relevante para mejorar el tratamiento de la información (Rosado G. & Rico B., 2010).

Curto Díaz (2017) señala que BI abarca metodologías, prácticas, y aplicaciones enfocadas a la creación, análisis y gestión de la información, que involucra identificar cuáles son aquellos indicadores claves que permiten medir el progreso de una organización. Por otra parte Lakew et al. (2017) argumenta que BI incide en los resultados trascendentales de la empresa, como alertas que permiten prevenir y actuar antes de que ocurran los problemas. Los llamados datos históricos de la empresa, han provocado el desarrollen nuevas herramientas o soluciones que mejoran la gestión de la información, como lo menciona Gandhi & Armstrong (2016).

Como una disciplina complementaria a la Inteligencia de Negocios para el análisis de datos, se encuentra la Minería de Datos (DM: Data Mining). DM se encuentra inmersa como una fase del proceso de descubrimiento de información en bases de datos (KDD: Knowledge Discovery in Databases). KDD es una metodología que agrupa procesos no comunes de identificación de patrones válidos, novedosos, potencialmente útiles y finalmente

comprensibles (Gorbea & Madera, 2017). DM se fundamenta en la Estadística, la Matemática y las Ciencias Computacionales; comprende una serie de técnicas para explorar grandes volúmenes de información, describir y/o diagnosticar patrones, identificar relaciones y clasificar, y predecir tendencias que pueden ocurrir a futuro en una organización (Hazen, Boone, Ezell, & Jones-Farmer, 2014; Niether et al., 2017). La minería de datos se encarga del proceso exploratorio y de análisis de información, a través de medios automáticos o semiautomáticos; e involucra técnicas para visualizar/describir patrones ocultos en datos (segmentación, reglas de asociación, secuencia de patrones, análisis de correlaciones y distribuciones, entre otras) o para identificar tendencias que sirven para predecir comportamientos futuros (regresiones, series temporales, arboles de decisiones, redes neuronales, etc.) (National Academi of Science, 2017; Provost & Fawcett, 2013; Sivarajah, 2017).

González, Lorés, & Granollers (2008) recomiendan que, antes de escoger un modelo o técnica de minería de datos se debe tener muy en claro el objetivo y direccionamiento de la investigación.

## 2. Metodología

La metodología CRISP-DM es utilizada debido al enfoque y tipo de investigación. Es un ciclo de vida, similar a los de ingeniería de software, creada para proyectos de análisis de datos, con la diferencia que presenta un proceso más normalizado y racionalizado (Macas, Lagla, Fuertes, Guerrero, & Toulkeridis, 2017). A continuación se describe cada fase de la metodología aplicada.

### 2.1. Comprensión del Negocio

Agrisolsa S.A. se dedica a la producción y comercialización de cacao seco y en lata; cuenta con dos predios: "Don Chago" ubicado el cantón el Guabo y "El Dorado" en el cantón Arenillas, ambos pertenecientes a la provincia de El Oro – Ecuador; los predios tienen un total de 11 y 21 hectáreas respectivamente. Esta empresa almacena su información en sistemas transaccionales, hojas de cálculo y agendas, lo que dificulta el acceso, presentación, administración y análisis de forma oportuna; los directivos están preocupados porque en los últimos años no han obtenido resultados favorables de su actividad económica. Es importante mencionar, que los datos de sus transacciones diarias son almacenados en archivos o bases de datos, pero no siempre brindan información necesaria, lo que genera, en muchas ocasiones, posibles errores en la toma de decisiones. Debido a estos problemas surge la necesidad de crear un sistema que permita tener un control de la información de sus procesos; a fin de detectar fallas a tiempo para plantear acciones preventivas, correctivas y predictivas que puedan contribuir con el desarrollo empresarial.

En la Tabla 1, se aprecian las etapas que forman parte de los procesos de producción y comercialización de la empresa objeto de estudio, que se obtuvieron en base a preguntas del negocio aplicadas a los directivos de mandos tácticos y estratégicos.

**Tabla 1**  
KPIs y su relación con preguntas del negocio

Etapas	Indicadores	Preguntas de negocio
Siembra	Número de hectáreas sembradas	¿Cuántas hectáreas sembradas tienen cada predio organizado por años?
		¿Cuál es el número hectáreas sembradas por variedad?
	Número de plantas sembradas	¿Cuál es el número de plantas sembradas por predio y variedad?

	Costos de siembra	¿En qué predio se empleó mayores costos para la siembra por variedad de cacao que se sembró?
		¿Cuáles son los costos indirectos, materia prima y mano de obra en el proceso de siembra de cacao por años?
Cosecha	Número de quintales	¿Cuántos quintales de cacao fueron cosechados por años?
	Número de especie cosechadas	¿Cuál es la variedad de cacao que más se ha cosechado por año y semestre?
	Costos de cosecha	¿Cuáles son los costos indirectos y mano de obra en proceso de cosecha de cacao por años?
	Porcentaje de plantas no productoras	¿Cuál es el porcentaje de plantas no productoras por años y por predios?
Post-cosecha	Número de quintales	¿Cuántos quintales de cacao seco resultaron en la fase de pos cosecha por años?
	Número de latas	¿Cuántos quintales de cacao resultaron del proceso de fermentación por predio, año y semestre?
	Promedio de temperatura	¿Cuál es promedio de temperatura que ha obtenido el cacao en el proceso de fermentación por años, semestre y por predios?
	Costos de post-cosecha	¿Cuáles son los costos indirectos y mano de obra en el proceso de pos-cosecha de cacao por años?
Mantenimiento	Costo Total	¿Cuál es el costo total de mantenimiento por predio, tipo de control y año?
	Costo de Materiales	¿Cuánto es el gasto total por compra de materiales y herramientas para el control de cultivo por predio, por año y por tipo de control?
	Costo Mano de Obra	¿Cuánto es el gasto total por mano de obra, por predio, por mes entre los dos últimos años?
Ventas	Número de quintales	¿Cuántos quintales de cacao seco se vendieron por año y por predio?
	Número de latas	¿Cuántas latas de cacao en baba se vendieron por año y por predio?
	Total de ventas	¿Cuál es el total en ventas de cacao seco y en baba por años?

Se exploró a fondo todas las áreas relacionadas con el problema a solucionar para recolectar los datos necesarios que alimentan un data warehouse, mismos que fueron obtenidos a partir de hojas de cálculo y la base de datos de la empresa. Las preguntas del negocio sirvieron para determinar indicadores clave de rendimiento (KPI: Key Performance Indicator) agrupados por perspectivas o dimensiones. Las dimensiones ayudan a organizar y examinar los KPI.

Para identificar los KPIs de los procesos de producción y comercialización de cacao, se revisaron varios estudios empíricos y aportes científicos; los trabajos realizados por Lawrence (2003) aporta con fórmulas para el proceso de ventas; Loo Veloz & Espinoza Lema (2012) describen como un indicador el porcentaje de plantas no productoras; a través de la investigación de Franklin E. (2007) se concretó información sobre el cacao rechazado, para el promedio de cajas elaboradas por trabajador; Reyes E. (2008) presentan indicadores sobre el nivel de plántulas no adaptadas, el porcentaje de efectividad de la cosecha y promedio de mazorcas por saco. En la Tabla 2, se resumen tanto los indicadores propuestos por la comunidad científica como los propuestos por nuestro equipo de trabajo; indicadores que se identificaron analizando los procesos y la data de la organización y la información recabada mediante entrevistas y diálogos con los directivos de mandos tácticos y estratégicos.

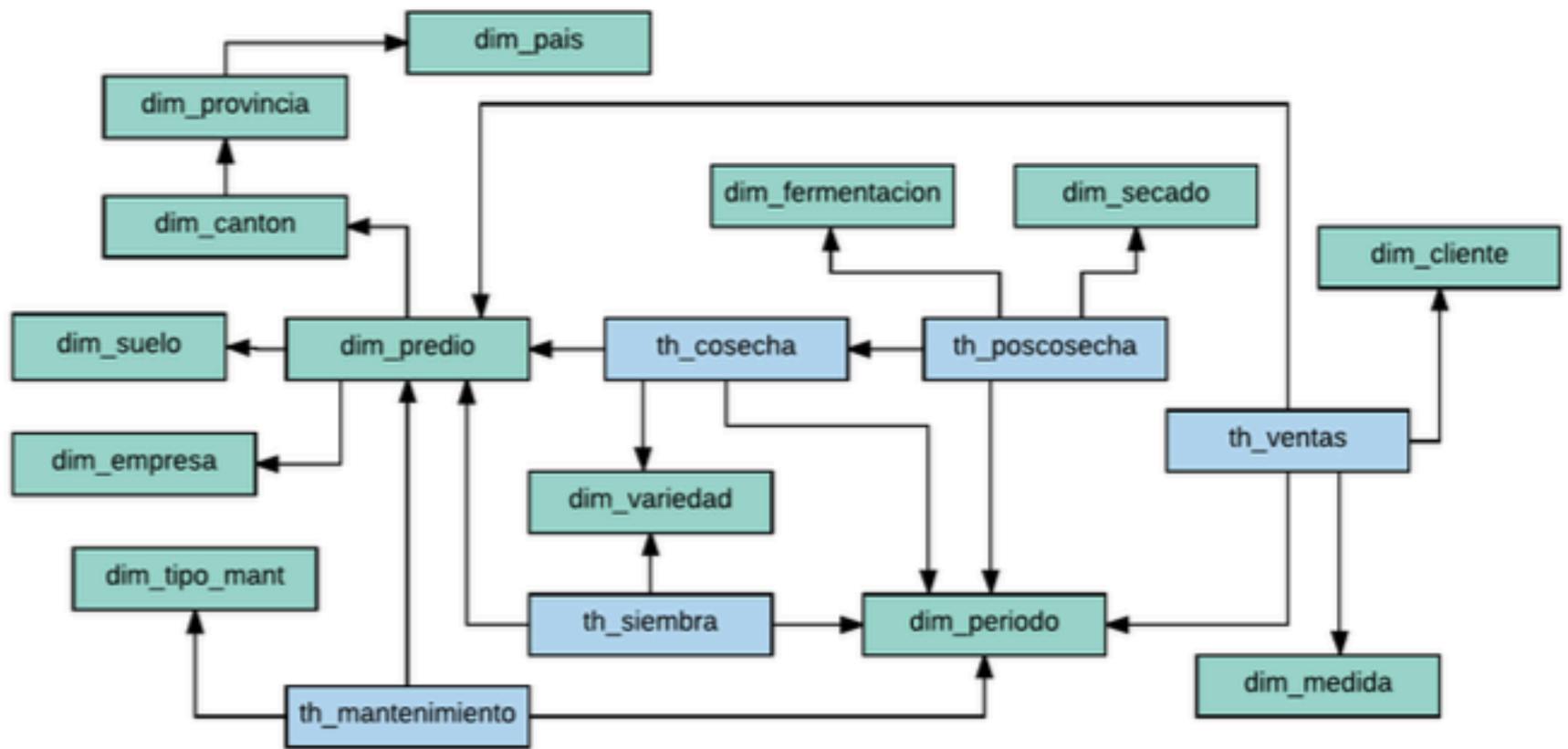
**Tabla 2**  
KPIs de los procesos de producción y comercialización de cacao

Etapa	Nombre	Fórmula
Siembra	Nivel de plántulas no adaptadas (Reyes, 2008)	$\frac{\text{Núm. de plántula no adaptadas}}{\text{tot. plántulas sembradas}} \times 100$
Cosecha	Cacao rechazado (Franklin, 2007)	$\frac{\text{Quintales de cacao rechazado}}{\text{Quintales de cacao producido}} \times 100$
	Porcentaje de efectividad de la cosecha (Reyes, 2008)	$\frac{\text{Núm. de quintales cosechados}}{\text{Núm. de plantas}} \times 100$
	Porcentaje de plantas no productoras (Loo & Espinoza, 2012)	$\frac{\text{Núm. de plantas N. P.}}{\text{Total de semillas plantadas}} \times 100$
Post-cosecha	Promedio de mazorcas por saco (Reyes, 2008)	$\frac{\text{Num. de mazorcas cosechadas}}{\text{total de sacos producidos}}$
Mantenimiento	Promedio de cajas elaboradas por trabajador	$\frac{\text{total de sacos producidos}}{\text{Núm. de trabajadores}}$
Ventas (Lawrence, 2003)	Porcentaje de la cantidad de sacos vendidos	$\frac{\text{Num. de de sacos vendidos}}{\text{Núm. de sacos producidos}} \times 100$
	Margen de utilidad	$\frac{\text{Utilidad Neta}}{\text{Ventas Netas}} \times 100$
	Costo de producción	$\frac{\text{Costo total de producción}}{\text{Cantidad de quintales producidas}}$

## 2.2. Comprensión del Negocio

En la Figura 1 se representan el esquema datos que se implementó en el gestor de base de datos PostgreSQL. Las dimensiones (rectángulos con prefijo dim\_) y tablas de hechos (rectángulos con prefijo th\_) forman parte de los data marts del data warehouse que sirve de repositorio para la aplicación web de analítica de datos (dashboard BI-DM). El diagrama está representado en el modelo de estructura de tipo constelación debido a que se encuentran acoplados los datamarts, conformando el data warehouse final.

**Figura 1**  
Diseño del Data Warehouse



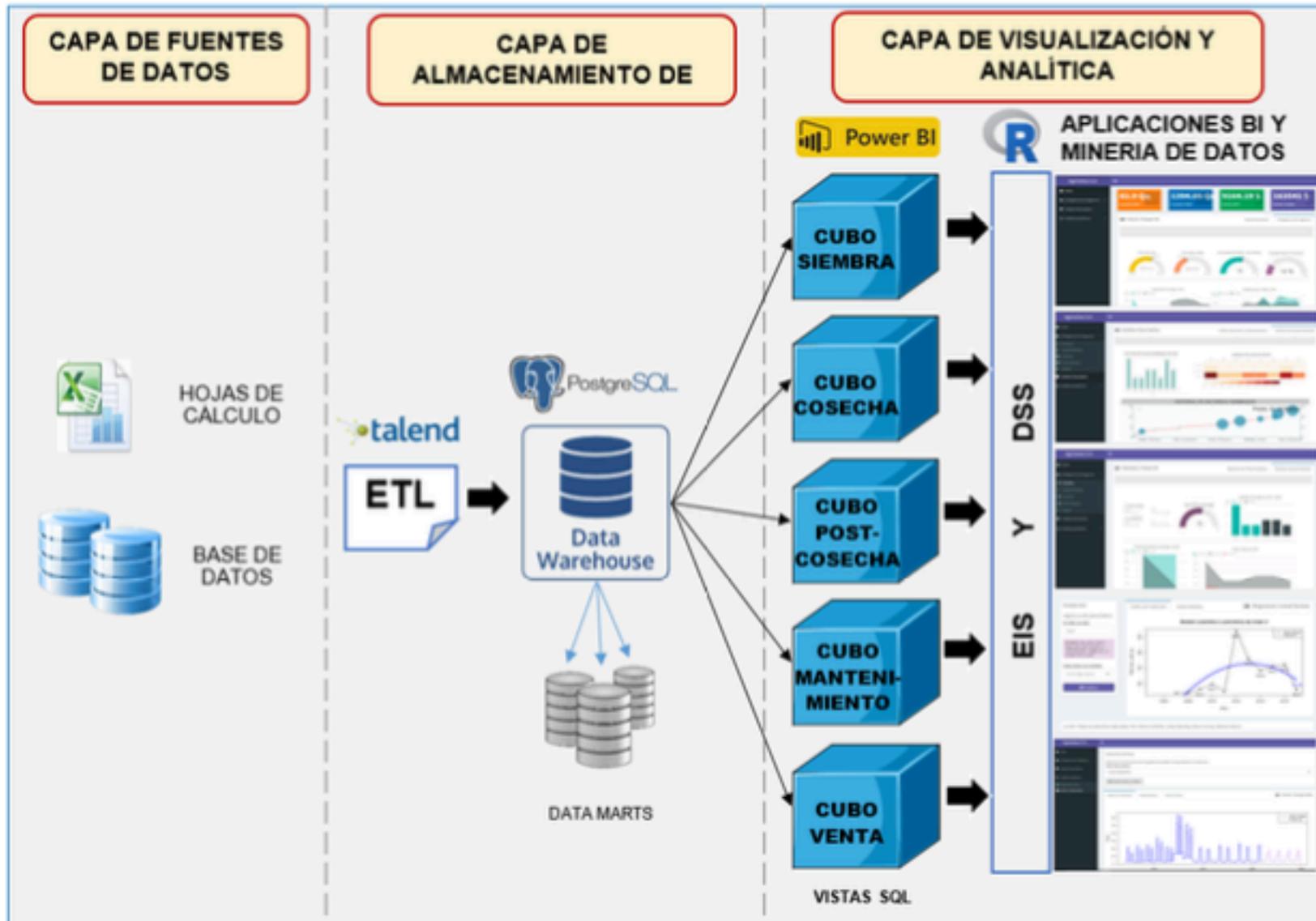
### 2.3. Preparación de los datos

Esta fase es una de las de mayor trabajo, porque los datos brutos (sin procesar) pasan por un proceso de extracción, transformación y carga (ETL) (Astriani & Trisminingsih, 2016) o pre-procesamiento (Provost & Fawcett, 2013); por ejemplo: limpieza o corrección de datos erróneos, eliminación de registros incompletos, atribución de nuevos valores a faltantes, creación de nuevos campos como rangos en datos numéricos, etc. Todos estos cambios son necesarios y se deben realizar con mucho cuidado para una migración exitosa al data warehouse, teniendo como origen los datos fuentes transaccionales de la empresa. El proceso ETL y el pre-procesamiento de datos, en nuestro caso se realizó en las herramientas Talend y R respectivamente. La creación de vistas por data mart a nivel del data warehouse, facilitan las consultas desde las aplicaciones de analítica de datos.

### 2.4. Arquitectura del sistema BI y DM

La Figura 2, representa la arquitectura de la solución de análisis de datos implementada mediante tecnologías de inteligencia de negocios y minería de datos. A continuación se describen las capas de la arquitectura: 1) La capa de fuente de datos proporciona los insumos (datos brutos) sin procesar, 2) mediante ETL y técnicas de pre-procesamiento se cargan los datos a la capa de almacenamiento. 3) en la capa de almacenamiento se ubica el repositorio de data histórica (data warehouse) organizado por temas (data marts) que sirve para analítica. 4) la capa de visualización y analítica representada por la aplicación web tipo tablero de control, implementa las distintas técnicas de BI y DM, y accede a la capa de almacenamiento a través de cubos de Procesamiento Analítico en Línea (MDX-OLAP: Multidimensional Expressions - Online Analytical Processing) y sentencias expresadas en lenguaje estructurado de consultas (SQL). La implementación del componente BI en tablero de control se efectuó mediante la herramienta Power BI.

**Figura 2**  
Arquitectura del Data Warehouse



## 2.5. Modelo de minería de datos

Se aplicaron modelos descriptivos de minería de datos con el fin de explorar, clasificar y analizar información de cada fase de la producción de cacao. Las técnicas descriptivas visualizan la información estadística mediante diagramas de barras, de cajas, histogramas, estadísticos básicos, etc., para una mayor comprensión del negocio.

Los modelos predictivos empleados son las series temporales y de regresión para pronosticar tendencias (valores futuros) del rendimiento de la cosecha de cacao e ingresos por ventas.

El lenguaje R fue la herramienta seleccionada para programar los algoritmos de las técnicas de minería de datos y el framework Shiny se utilizó para crear la aplicación web que integra el dashboard BI-DM.

### Descripción de las variables

Para los modelos de predicción de las ventas, se utilizó datos de ventas mensuales a partir del año 2014 hasta el mes de julio del año 2017; estos valores fueron empleados para predecir las ventas de los próximos meses y años. A continuación se expresan las variables utilizadas:

Variable dependiente (Y) = Total en ventas (\$)

Variable independiente (X) = mes-año

Para los modelos de predicción del rendimiento de la cosecha se obtuvieron los valores de la fórmula de rendimiento en quintales a partir del año 2010; información que se empleó para predecir el rendimiento de los próximos meses y años.

Variable dependiente (Y) = Rendimiento (quintales)

Variable independiente (X) = mes-año

Modelado y pronóstico de tendencia de ventas y rendimiento de la cosecha

Una de las técnicas empleadas son las series temporales, usada para secuenciar los datos (variable Y) y ordenarlos de manera cronológica (variable X). El modelo multiplicativo estacional de Holt-Winters se utilizó para suavizar la serie de tiempo, debido a su fácil adaptación a medida que los datos aumentan y a los repentinos cambios de tendencia (Tratar & Strmčnik, 2016). Por otro lado, los modelos de regresión, pueden ser alternativas viables de predicción. En la tabla 3 se visualizan los modelos elegidos y que deben ser probados para ver si son óptimos (López, 2012).

**Tabla 3**  
Posibles modelos de regresión

Tipo	Función
Regresión Lineal	$y = ax + b$
Regresión Polinomial 2°	$y = ax^2 + bx + c$
Regresión Polinomial 3°	$y = ax^3 + cx^2 + bx + c$
Series temporales Holt-Winter	$s_t = \delta \frac{y_t}{M_t} + (1 - \delta)s_{t-12}$

Fuente: Elaboración propia

Según Abreu et al. (2016), para determinar si un modelo es óptimo, se considera el coeficiente de determinación R<sup>2</sup>, que indica la proporción de variación de la variable dependiente (Y) explicada por la variable independiente (X); y además, el valor de p (p-value) representa el nivel de significación del modelo, al menos debe ser p < 0.05; es decir, un margen de error del 5% y un nivel de significancia del 95% (Keat & Young, 2004).

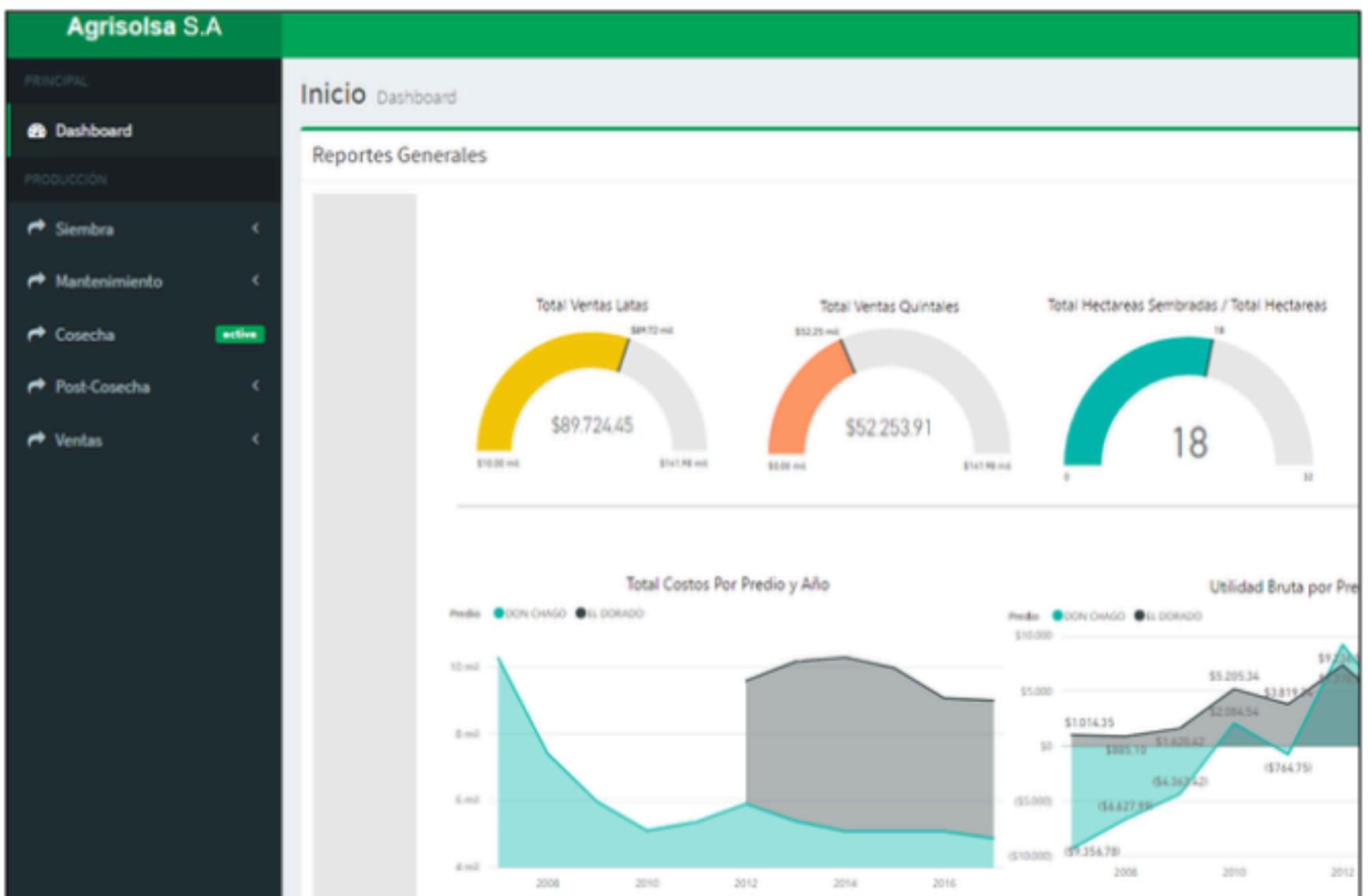
### 3. Resultados

En primera instancia se presenta la evidencia de la aplicación web dashboard BI-DM desarrollado y luego se describen los resultados de los modelos de minería de datos.

En la Figura 3, se muestra el interfaz del dashboard BI-DM, donde se presentan los datos generales de la empresa objeto de estudio. En la primera sección de la gráfica se muestra un conjunto de medidores acerca del total de ventas en latas y quintales, el total de hectáreas sembradas y el porcentaje de plantas no productoras. En la segunda sección se diseñó gráficos de áreas apiladas (stacked area chart) con el objetivo de representar el total de costo y la utilidad bruta de la empresa; estos resultados fueron organizados por predio y los años respectivos en los que fueron cultivados.

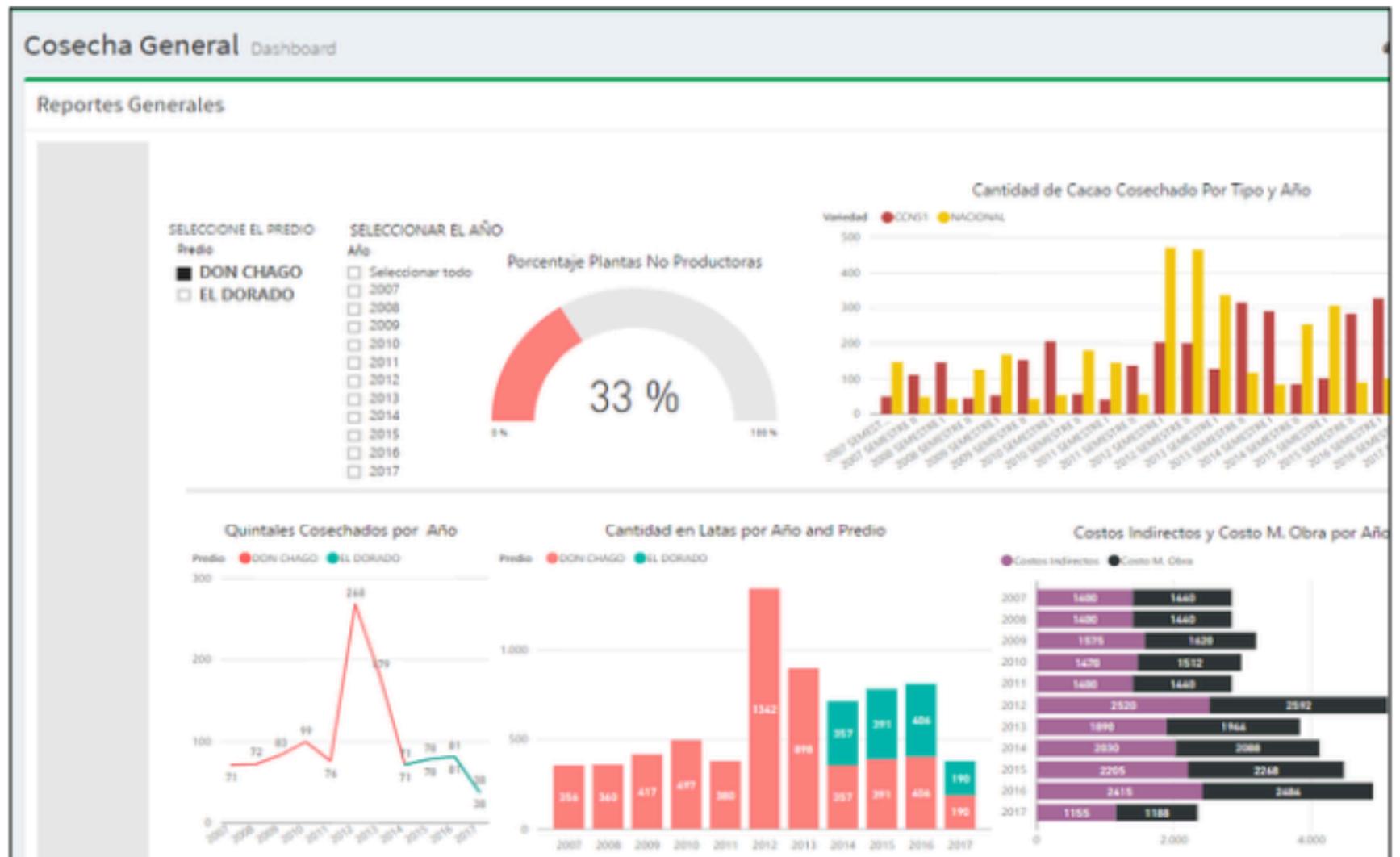
En la Figura 5, se proporciona información del proceso de cosecha como la cantidad de cacao recogido, mediante un gráfico de columnas apilado (Stacked column chart); además, se proporciona ítems al lado izquierdo para filtrar por año y/o predio. Uno de los indicadores hace referencia al costo de mano de obra y costos indirectos empleados para el proceso de cosecha, como se observa en el gráfico de barras apilado (Stacked bar chart).

**Figura 3**  
Dashboard BI-DM que presenta datos generales de la empresa



-----

**Figura 4**  
Dashboard del proceso de cosecha



En la tabla 4 se detallan los valores de  $R^2$  y P de cada uno de los modelos predictivos de regresiones y serie temporal.

**Tabla 4**  
Valor de R<sup>2</sup> y p para cada modelo de predicción

Modelo	Rendimiento de Cosecha		Ventas	
	R <sup>2</sup>	p	R <sup>2</sup>	p
Regresión. Lineal	0.0325	0.1521	0.4097	0.0333
Regresión Polinomial 2°	0.3697	0.2543	0.6937	0.0087
Regresión Polinomial 3°	0.4258	0.3572	0.8492	0.0029
Series temporales Holt-Winter	0.9369	4.2x10 <sup>-5</sup>	0.9999	2.2x10 <sup>-16</sup>

Fuente: Datos obtenidos de la herramienta R.

A través de los coeficientes obtenidos de R y P-value se pudo construir la ecuación matemática del modelo de series temporales que presentan valores de 0.9369 para el rendimiento de cosecha y 0.9999 en caso del proceso de ventas, estos valores se encuentran en un nivel de ponderación óptimo.

La figura 5 presenta los parámetros estadísticos para un modelo de regresión cúbica y calculada en la herramienta R.

**Figura 5**  
Coeficientes de la regresión y R<sup>2</sup> de ventas

```

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      14867      1438  10.338 1.72e-05 ***
poly(v_indp_peranio, 3, raw = FALSE)1  20801      4770   4.361 0.00331 **
poly(v_indp_peranio, 3, raw = FALSE)2 -17318      4770  -3.631 0.00839 **
poly(v_indp_peranio, 3, raw = FALSE)3 -12814      4770  -2.687 0.03124 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4769 on 7 degrees of freedom
Multiple R-squared:  0.8492,    Adjusted R-squared:  0.7846
F-statistic: 13.14 on 3 and 7 DF,  p-value: 0.002912
    
```

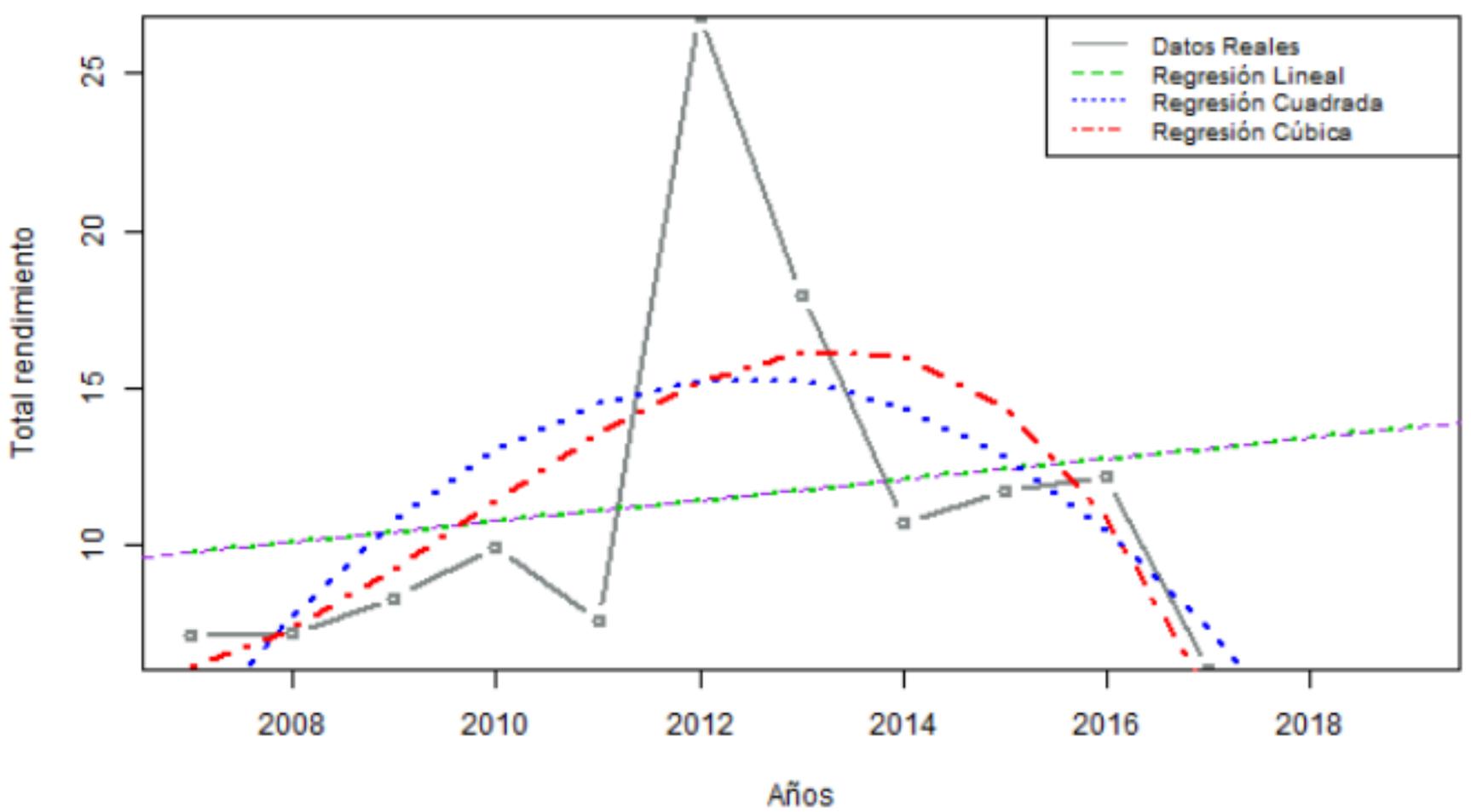
Obteniendo como resultado la siguiente la ecuación:

$$y_i = 14867 + 20801x - 17318x^2 - 12814x^3 \quad (1)$$

Sin embargo, el modelo predictivo de series temporales fue escogido tanto para el proceso de ventas y rendimiento de cosecha, esto se debe a que presentan valores de R<sup>2</sup> y P-value en un nivel aceptable; por otro lado, el modelo de regresión polinómico de tercer orden para ventas presenta un valor de 0.8492, es decir que se encuentra en la categoría de aceptable, siendo también una alternativa de predicción.

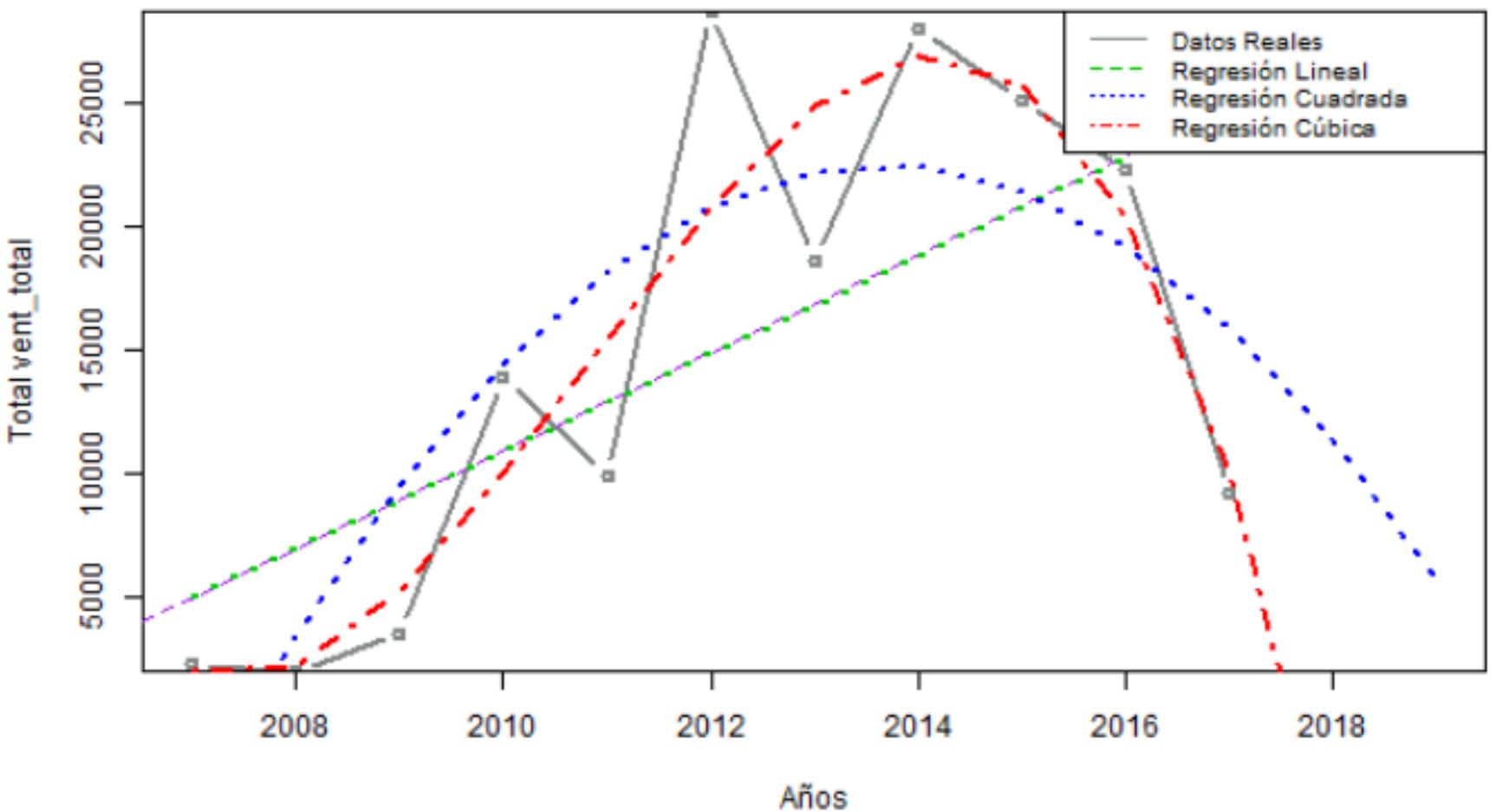
Los modelos de regresión del rendimiento de la cosecha presentados en la tabla 4 fueron graficados por medio de la herramienta R y son mostrados las Figuras 6 y 7.

**Figura 6**  
Modelos de regresión para predicción del rendimiento de cosecha



**Figura 7**

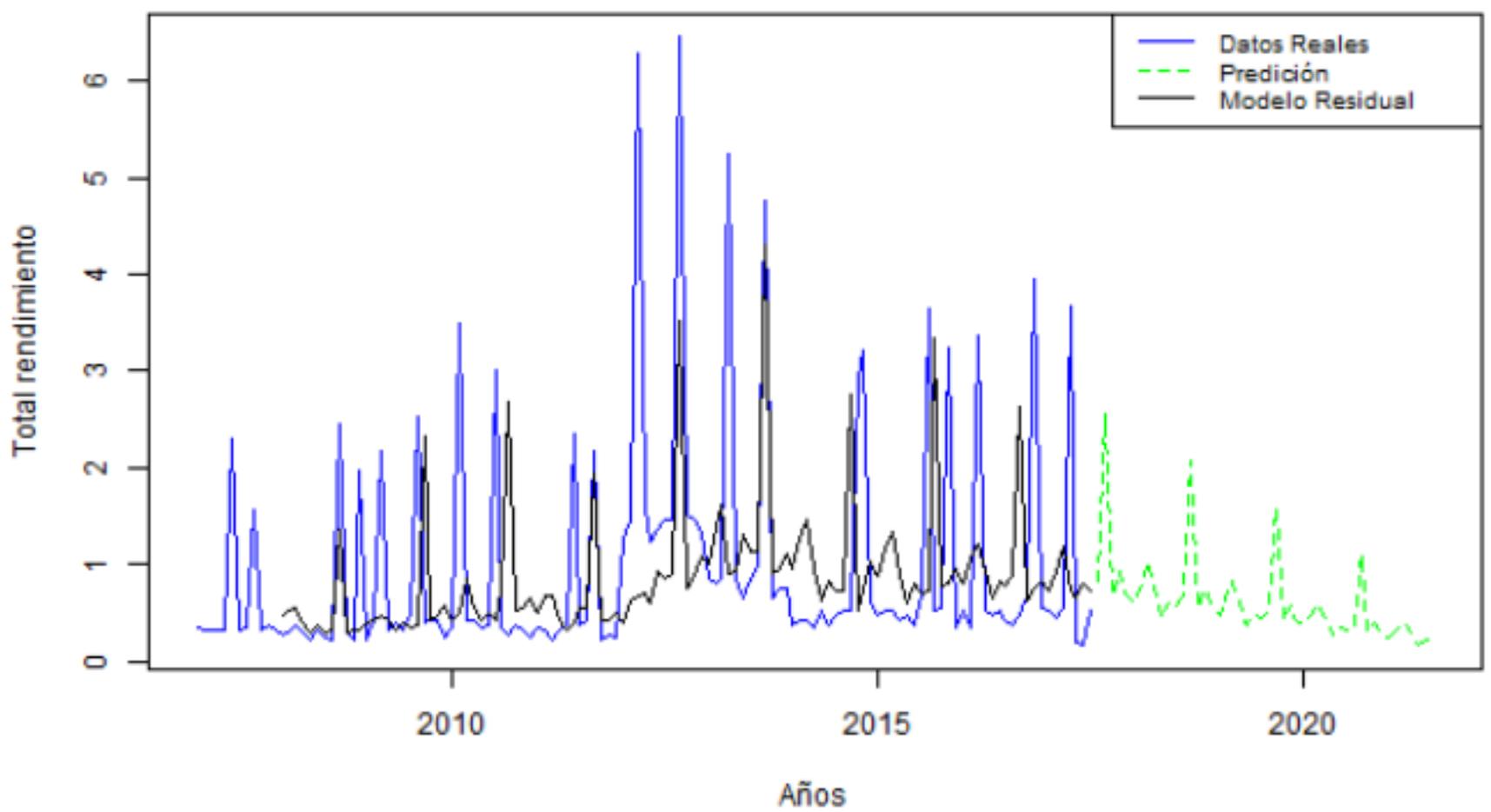
Modelos de regresión para predicción de ventas



En las Figuras 8 y 9, se representan las series temporales tanto para el rendimiento de la cosecha ( $R^2=0.9369$ ) como para las ventas ( $R^2=0.9999$ ); los datos reales de la empresa (línea color azul), el modelo residual (línea color negro) y la predicción (línea color verde) representan los modelos de predicción basado en los datos históricos. Como se puede apreciar, la línea de tendencia sigue un patrón que predice valores no es muy alentadores para la empresa; justamente este patrón puede servir para que los propietarios o directivos tomen los correctivos adecuados.

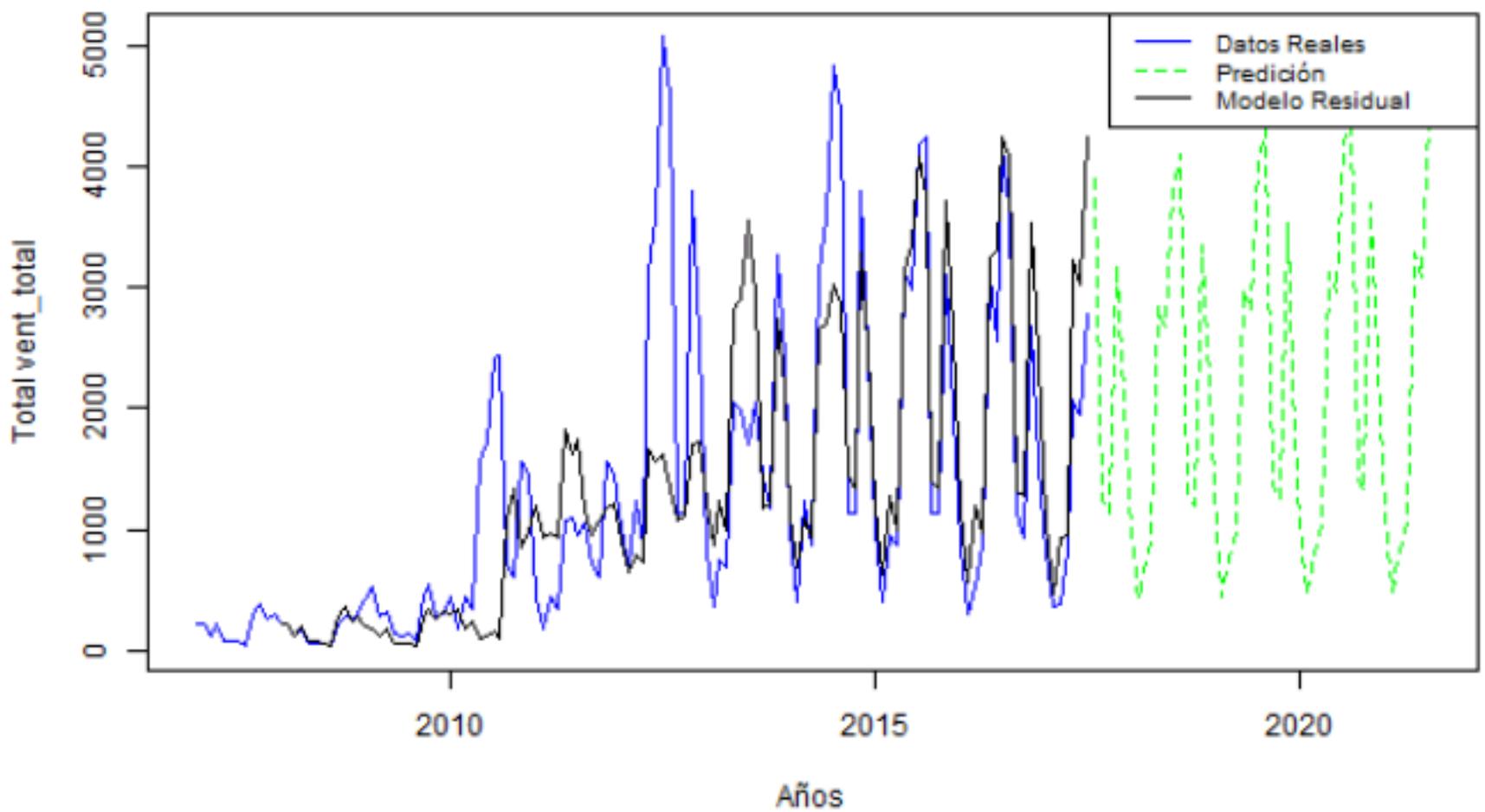
**Figura 8**

Modelo de serie temporal aplicado para predecir el rendimiento de las cosechas



**Figura 9**

Modelo de temporales aplicada en el proceso para predecir las ventas



**Discusión.** Varios aportes de la comunidad científica relacionados con la inteligencia de negocios y minería de datos fueron revisados. Un dashboard de soporte de decisiones para análisis de datos de los procesos de una empresa minera proponen (Mazon-Olivo et al., 2017). El caso de (Reuter & Brambring, 2016) emplea técnicas de minería de datos (métodos Bayesianos y de árboles de decisión) para conseguir un aumento de la consistencia en los controles de producción de empresas manufactureras. Otro caso es el trabajo de (Erkollar & Oberer, 2016), presentan un modelo scorecard y de dashboard basado en KPIs, métricas y marcos de tiempo para la visualización de datos. El trabajo de (Ng, Bandaru, & Frantzen, 2016) consiste en mejorar los sistema de producción con técnicas de optimización multi-objetivo y minería de datos. Muchos autores han realizado aportes en

el análisis de datos, creando o aplicando técnicas de BI y DM; incluso se han concentrado en mejorar los procesos productivos y de comercialización; sin embargo, en el ámbito la producción de cacao no se evidencian mayores aportes en la identificación de KPIs y aplicación de tecnologías de BI y DM para describir la situación del negocio y predecir tendencias futuras; información que puede ser útil para tomar decisiones. Una empresa bien informada de su propia realidad, tiene ventajas frente a la competencia.

---

## 4. Conclusiones

La inteligencia de negocios y minería de datos, incluyen técnicas que permiten analizar los datos históricos de una organización, en busca de información útil que describa la situación pasada o actual del negocio, o prediga tendencias futuras que generen ventajas frente a la competencia. Este trabajo consistió, en la identificación de los KIPs de los procesos de producción y comercialización para una empresa de cacao (caso Agrisolsa S.A); la metodología CRISP-DM, sirvió para guiar el desarrollo de un tablero de control (dashboard BI-DM) con interfaz web, que accede a un datawarehouse, visualiza información estadística, detecta patrones de datos y predice el rendimiento y ventas de la cosecha mediante series temporales y el método de Holt-Winters con un  $R^2 > 0.9$ .

---

## Referencias bibliográficas

- Abreu, A., Abreu, J., Iglesias, R., & Navarro, I. (2016). Interfaz gráfica en matlab para el cálculo de criterios de bondad de ajuste. *Revista Ingeniería, Matemáticas Y Ciencias de La Información*, 13–21.
- Astriani, W., & Trisminingsih, R. (2016). Extraction, Transformation, and Loading (ETL) Module for Hotspot Spatial Data Warehouse Using Geokettle. *Procedia Environmental Sciences*, 33, 626–634. <http://doi.org/10.1016/j.proenv.2016.03.117>
- Curto Díaz, J. (2017). *Introducción al Business Intelligence* (Universita). Barcelona. España.
- Erkollar, A., & Oberer, B. (2016). Multidimensional Dashboards for Evaluating Strategic brand Management Processes for Multi-brand Companies. *Procedia - Social and Behavioral Sciences*, 235(October), 505–513. <http://doi.org/10.1016/j.sbspro.2016.11.062>
- Franklin, E. (2007). *Auditoria administrativa: Gestión estratégica del cambio*. (Universidad Nacional Autónoma de México, Ed.). México: PEARSON.
- Gandhi, N., & Armstrong, L. J. (2016). A review of the application of data mining techniques for decision making in agriculture. *2016 2nd International Conference on Contemporary Computing and Informatics*, 1–6. <http://doi.org/10.1109/IC3I.2016.7917925>
- González, M. P., Lorés, J., & Granollers, A. (2008). Enhancing usability testing through datamining techniques: A novel approach to detecting usability problem patterns for a context of use. *Information and Software Technology*, 50(6), 547–568. <http://doi.org/10.1016/j.infsof.2007.06.001>
- Gorbea, S., & Madera, M. (2017). Diseño de un data warehouse para medir el desarrollo disciplinar en instituciones académicas. *Investigacion Bibliotecologica*, 31(72), 161–181. <http://doi.org/10.22201/iibi.0187358xp.2017.72.57828>
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *Int. J. Production Economics Journal*, 154, 72–80. <http://doi.org/10.1016/j.ijpe.2014.04.018>
- ICCO. (2017). Press release-quarterly bulletin of cocoa statistics. Retrieved February 28, 2018, from <https://www.icco.org/about-us/icco-news/380-quarterly-bulletin-of-cocoa-statistics-november-2017.html>
- INEC. (2014). Análisis del sector cacao y elaborados. Retrieved February 28, 2018, from <http://www.ecuadorencifras.gob.ec/estadisticas-sectoriales/>
- Keat, P., & Young, P. (2004). *Economía de empresa*. México: Pearson Educación.
- Lakew, E., Papadopoulos, A., Maggio, M., Klein, C., & Elmroth, E. (2017). KPI-Agnostic

- Control for Fine-Grained Vertical Elasticity. *IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 17(978-1-5090-6610-0), 589–598.  
<http://doi.org/10.1109/CCGRID.2017.71>
- Lawrence, J. G. (2003). *Principios de administración financiera*. México: PEARSON.
- Loor, K., & Espinoza, P. (2012). *Diseño de un sistema de control de procesos de producción y comercialización para una empresa productora de banano ubicada en la provincia de los ríos*. ESPOL. Retrieved from <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/38826>
- López, A. (2012). *Regresión Lineal y No Lineal en Contexto*. EAE.
- Macas, M., Lagla, L., Fuertes, W., Guerrero, G., & Toulkeridis, T. (2017). Data Mining model in the discovery of trends and patterns of intruder attacks on the data network as a public-sector innovation. *IEEE*, 55–62. <http://doi.org/10.1109/ICEDEG.2017.7962513>
- Mazon-Olivo, B., Rivas, W., Pinta, M., Mosquera, A., Astudillo, L., & Gallegos, H. (2017). Dashboard para el soporte de decisiones en una empresa del sector minero. *Conference Proceedings - Universidad Técnica de Machala, 1*, 1218–1229. Retrieved from <http://investigacion.utmachala.edu.ec/proceedings/index.php/utmach/article/view/219/191>
- National Academi of Science. (2017). Overview of Data Science Methods. In *Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions* (pp. 53–79). The National Academies Press. <http://doi.org/10.17226/23670>
- Ng, A., Bandaru, S., & Frantzen, M. (2016). Innovative Design and Analysis of Production Systems by Multi-objective Optimization and Data Mining. *Procedia CIRP*, 50, 665–671. <http://doi.org/10.1016/j.procir.2016.04.159>
- Nguyen, P. M., Haghverdi, A., de Pue, J., Botula, Y. D., Le, K. V., Waegeman, W., & Cornelis, W. M. (2017). Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils. *Biosystems Engineering*, 153, 12–27. <http://doi.org/10.1016/j.biosystemseng.2016.10.013>
- Niether, W., Schneidewind, U., Armengot, L., Adamtey, N., Schneider, M., & Gerold, G. (2017). Spatial-temporal soil moisture dynamics under different cocoa production systems. *Catena*, 158, 340–349. <http://doi.org/10.1016/j.catena.2017.07.011>
- Provost, F., & Fawcett, T. (2013). *Data Science for Business. What you need to know about Data Mining and Data-Analytic thinking*. O'Reilly Media.
- Reuter, C., & Brambring, F. (2016). Improving Data Consistency in Production Control. *Procedia CIRP*, 41, 51–56. <http://doi.org/10.1016/j.procir.2015.12.116>
- Reyes, E. (2008). *Contabilidad de costes*. México: LIMUSA.
- Rosado G., A. A., & Rico B., D. W. (2010). Inteligencia de Negocios: Estado del Arte. *Scientia Et Technica*, XVI, 321–326. Retrieved from <http://www.redalyc.org/articulo.oa?id=84917316060>
- Shavez, M., Ahmad, S., Jan, K., & Bashir, K. (2017). Status, supply chain and processing of cocoa - A review. *ScienceDirect*, 66, 108–116. <http://doi.org/https://doi.org/10.1016/j.tifs.2017.06.007>
- Sivarajah, U. et al. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <http://doi.org/10.1016/J.JBUSRES.2016.08.001>
- Stuart, M. L., Menéndez, D., Martínez, Y., Cordero, A., & Delgado, T. (2017). Experiencia en el diagnóstico de la Gestión de Información con enfoque de Arquitectura de Información Empresarial (Experience in the Diagnostic of Information Management with a Business Information Architecture Approach). *GECONTEC: Revista Internacional de Gestión Del Conocimiento Y La Tecnología*, 5(1), 1–16.
- Torres, L. A. (2012). *Manual de producción de cacao fino de aroma a través de manejo ecológico*. Universidad de Cuenca. Retrieved from <http://dspace.ucuenca.edu.ec/bitstream/123456789/3250/1/TESIS.pdf>
- Tratar, L., & Strmčnik, E. (2016). The comparison of Holt–Winters method and Multiple regression method: A case study. In *Energy* (pp. 266–276).

Utomo, B., Prawoto, A. A., Bonnet, S., Bangviwat, A., & Gheewala, S. H. (2016). Environmental performance of cocoa production from monoculture and agroforestry systems in Indonesia. *ScienceDirect*, 134, 583–591. <http://doi.org/10.1016/j.jclepro.2015.08.102>

---

1. Ingeniera en Sistemas. Magíster en Informática Aplicada. Profesora en la Unidad Académica de Ingeniería Civil, Universidad Técnica de Machala. [bmazon@utmachala.edu.ec](mailto:bmazon@utmachala.edu.ec)
  2. Ingeniera en Banca y Finanzas. Magíster en Administración de Empresas. Profesora en la Universidad Técnica de Machala. [mjaramillo@utmachala.edu.ec](mailto:mjaramillo@utmachala.edu.ec)
  3. Magíster en Dirección de Empresas. Doctor en Ciencias Administrativas. Universidad Nacional Mayor de San Marcos. Profesor en la Universidad Técnica de Machala. [oromero@utmachala.edu.ec](mailto:oromero@utmachala.edu.ec)
  4. Ingeniera Comercial. Doctora en Ciencias Contables y Empresariales. Universidad Nacional Mayor de San Marcos. Vicerrectora Académica de la Universidad Técnica de Machala. [lborja@utmachala.edu.ec](mailto:lborja@utmachala.edu.ec)
  5. Economista. Máster en Administración de Empresas. Doctora en Ciencias Administrativas. Universidad Nacional Mayor de San Marcos. Sub-decana de la Unidad Académica de Ciencias Empresariales de la Universidad Técnica de Machala. [maguirre@utmachala.edu.ec](mailto:maguirre@utmachala.edu.ec)
  6. Miembro del semillero de investigación de la carrera de Ingeniería de Sistemas. Unidad Académica de Ingeniería Civil, Universidad Técnica de Machala. [mjcontento\\_est@utmachala.edu.ec](mailto:mjcontento_est@utmachala.edu.ec)
- 

Revista ESPACIOS. ISSN 0798 1015  
Vol. 39 (Nº 32) Año 2018

[Índice]

[En caso de encontrar un error en esta página notificar a [webmaster](#)]

©2018. revistaESPACIOS.com • ®Derechos Reservados