

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



Facultad de Ingeniería en Electricidad y Computación

“MODELAMIENTO DEL PROCESO DE NEGOCIO Y USO DE
TECNOLOGÍAS DE INTELIGENCIA DE NEGOCIOS Y
MINERÍA DE DATOS PARA LA PRODUCCIÓN Y
COMERCIALIZACIÓN DE ARROZ”

TRABAJO DE TITULACIÓN

PREVIA A LA OBTENCIÓN DEL TÍTULO DE:

MAGISTER EN SISTEMAS DE INFORMACIÓN GERENCIAL

Presentada por:

LEYLA ESTHER MATÍAS RODRÍGUEZ

DAVID ANDRÉS LEÓN PEÑAFIEL

GUAYAQUIL - ECUADOR

AÑO 2021

AGRADECIMIENTOS

Agradecemos a Dios por la salud en estos tiempos difíciles, por cuidarnos y mantenernos con vida para alcanzar este objetivo.

Agradecemos a nuestras familias, amigos, colegas, profesores y compañeros.

DEDICATORIA

Dedico este trabajo a mi mamá quien en todo momento me ha dado su apoyo y ejemplo para lograr mis metas.



Leyla Matías Rodríguez

Dedico este trabajo a Lissete, mi esposa, amiga y apoyo constante. A mi familia por su esfuerzo y paciencia y a la vida, por permitirme llegar hasta donde estoy ahora.



David León Peñafiel

TRIBUNAL DE SUSTENTACIÓN

MGS. Lenin Freire Cobo

DIRECTOR MSIG

MGS. Lenin Freire Cobo

DIRECTOR DE TRABAJO TITULACIÓN

MGS. Juan Carlos García

MIEMBRO DEL TRIBUNAL

RESUMEN

El propósito de este trabajo es diseñar e implementar un sistema de soporte a la decisión que permita visualizar la información a partir de la definición de los principales indicadores de desempeño que sean relevantes para la empresa.

La metodología CRISP-DM fue la seleccionada para elaborar los tableros de control que incluirá un apartado de proyección de ventas mediante el uso de minería de datos con la ayuda de tecnologías de inteligencia de negocios.

Los resultados serán obtenidos de un almacén de datos que contará con un proceso de extracción, transformación y carga de datos y que se usará para armar data marts que contarán con las dimensiones que serán obtenidas del análisis de requerimientos. Esto se visualizará en una interfaz web ya sea local o publicada en internet.

Finalmente, se acudirá al estándar BPMN para realizar un análisis previo del proceso de negocio actual y se realizará un rediseño del mismo.

ÍNDICE GENERAL

AGRADECIMIENTOS.....	2
DEDICATORIA	3
TRIBUNAL DE SUSTENTACIÓN	4
RESUMEN.....	5
ÍNDICE GENERAL.....	6
ÍNDICE DE FIGURAS.....	10
ÍNDICE DE TABLAS	13
INTRODUCCIÓN.....	14
CAPÍTULO 1	1
GENERALIDADES	1
1.1 Antecedentes.....	1
1.2 Descripción del problema	6
1.3 Solución propuesta.....	7
1.4 Alcance de la propuesta	9
1.5 Objetivo General.....	10
1.6 Objetivos Específicos.....	10
CAPÍTULO 2.....	12
MARCO TEÓRICO.....	12

2.1 DSS Sistema de Apoyo de decisiones	12
2.1.1 Definición.....	12
2.1.2 Propósito	13
2.1.3 Estructura, Elementos y Componente	13
2.2 Metodología Cross-Industry Standar Process para Data Mining	14
2.2.1 Fase I: Business Understanding. Definición de necesidades del cliente (comprensión del negocio).....	16
2.2.2 Fase II: Data Understanding. Estudio y comprensión de los datos	17
2.2.3 Fase III. Data Preparation. Análisis de los datos y selección de características	18
2.2.4 Fase IV. Modeling. Modelado.....	19
2.2.5 Fase V. Evaluation. Evaluación (obtención de resultados).....	21
2.2.6 Fase VI. Deployment. Despliegue (puesta en producción).....	21
2.3 Modelado gráfico de procesos de negocio: BPMN	22
2.3.1 BPM (BUSINESS PROCESS MANAGEMENT)	23
2.3.2 Ciclo de vida BPM	23
2.3.3 BPMN (BUSINESS PROCESS MODEL AND NOTATION)	26
2.4 Inteligencia de Negocios.....	26
2.4.1 Business Intelligence BI	26

2.4.2 Herramientas de Business Intelligence	27
2.5 Data Mining – Python	28
2.5.1 Técnicas de Data Mining	29
2.5.2 Python	30
CAPÍTULO 3.....	32
DEFINICIÓN DE LA SITUACIÓN ACTUAL Y DEFINICIÓN DE REQUERIMIENTOS.....	32
3.1 Definición de la situación actual.....	32
3.2 Levantamiento de información de los procesos y actividades actuales	35
3.3 Levantamiento de Requerimientos Funcionales y no funcionales	39
3.4 Definición de principales indicadores de desempeño del negocio.....	42
3.5 Definición de la mejora continua del proceso de negocio.....	49
3.6 Definición del alcance de la solución propuesta	52
CAPÍTULO 4.....	55
ANÁLISIS, DISEÑO E IMPLEMENTACIÓN	55
4.1 Cálculo de KPIs	55
4.2 Diseño de proceso TO-BE.....	66
4.3 Arquitectura en capas e implementación del sistema.....	69

4.4 Implementación de Dashboard BI-DM.....	78
4.5 Implementación del modelo de proyección.....	82
4.5.1 Descripción de las variables.....	82
4.5.2 Ajuste del modelo.....	84
4.5.3 Interpretación	87
CAPÍTULO 5.....	89
PRUEBAS Y ANÁLISIS DE RESULTADOS.....	89
5.1 Descripción del entorno de usuario	89
5.2 Casos de Uso	92
5.3 Cumplimiento de sistema	95
5.4 Resultados de la implementación.....	107
CONCLUSIONES Y RECOMENDACIONES.....	108
BIBLIOGRAFÍA.....	111
ANEXO 1	117
ANEXO 2	119

ÍNDICE DE FIGURAS

Figura 1.1 Organigrama Empresa EVARROZ	4
Figura 1.2: Cadena de producción del arroz.....	10
Figura 2.1 Esquema de los 4 niveles de CRISP – CM.....	15
Figura 2.2 Modelo del proceso CRISP-DM.....	16
Figura 2.3 Elementos claves de un proceso de Negocio	22
Figura 2.4 Ciclo de vida del proceso BPM	23
Figura 3.3: Proceso de Generar Información Agrónomo	51
Figura 3.4: Proceso Generar Información Ventas.....	52
Figura 4.1: Proceso de negocio TO-BE EVARROZ.....	67
Figura 4.2: Arquitectura de la red de datos.....	69
Figura 4.3: Arquitectura de 3 capas EVARROZ.....	70
Figura 4.4: Job “Obtener de FTP	71
Figura 4.5: Validar conexión	71
Figura 4.6: Armado de los data marts a partir de las hojas de cálculo.....	72
Figura 4.8: Esquema copo de nieve EVARROZ	73
Figura 4.9: Armado de la dimensión temporal del cubo.....	74
Figura 4.10: Calculo de indicadores: Fase siembra	75
Figura 4.11: Calculo de indicadores: Fase Cosecha.....	75

Figura 4.12: Calculo de indicadores: Fase post cosecha.....	76
Figura 4.13: Calculo de indicadores: Fase mantenimiento	76
Figura 4.14: Calculo de indicadores: Fase ventas	77
Figura 4.15: Procedimiento completo Capa Fuente de Datos y Extracción, Capa de almacenamiento	77
Figura 4.16: Dashboard principa.....	79
Figura 4.17: Dashboard “SIEMBR	79
Figura 4.18: Dashboard “COSECHA	80
Figura 4.19: Dashboard “POSTCOSECHA.....	80
Figura 4.20: Dashboard “MANTENIMIENTO”	81
Figura 4.21: Dashboard “VENTAS.....	81
Figura 4.22: Código Python: correlación lineal entre dos variables	83
Figura 4.23: Resultado de correlación lineal entre dos variables.....	83
Figura 4.24: Código Python: División de los datasets.....	85
Figura 4.25: Código Python: Creación del modelo utilizando matrices	85
Figura 4.26: Resultados de la regresión	86
Figura 4.27: Ajuste de coeficientes, intervalo de confianza y predicciones ..	87
Figura 4.28: Gráfico de proyecciones integrado en el Dashboard BI-DM.....	87
Figura 5.1: Pantalla de inicio “Sistema EVARROZ”	90

Figura 5.2: Pantalla de opciones: dashboard o subir datos	90
Figura 5.3: Descarga de plantilla	91
Figura 5.4: Búsqueda local y carga de datos	91
Figura 5.5: Visualización de dashboard Gerencial EVARROZ	92
Figura 5.6: Validación de credenciales	93
Figura 5.7: Carga de información no coherente	94
Figura 5.9: Actualización de información en el dashboard gerencial	95
Figura 5.10: Estadística de consulta	96
Figura 5.11: Estadística de consulta 2	97
Figura 5.12: Estadística de consulta 3	98
Figura 5.13: Estadística de consulta 4	99
Figura 5.14: Estadística de consulta	100
Figura 5.15: Estadística de consulta 6	101
Figura 5.16: Estadística de consulta 7	102
Figura 5.17: Estadística de consulta 8	103
Figura 5.18: Estadística de consulta 9	104
Figura 5.19: Estadística de consulta 10	105
Figura 5.20: Estadística de consulta 11	106

ÍNDICE DE TABLAS

Tabla 1: Requerimientos funcionales	40
Tabla 2: Requerimientos funcionales	41
Tabla 3: Principales indicadores de desempeño y su relación con preguntas del negocio	43
Tabla 4: Fase Siembra.....	55
Tabla 5: Fase Cosecha.....	57
Tabla 6 Fase: Post Cosecha.....	58
Tabla 7 Fase: Mantenimiento.....	59
Tabla 8 Fase: Ventas.....	60
Tabla 9: Dashboard Principal.....	61
Tabla 10: Dashboards Secundarios.....	62
Tabla 11: Plantilla “Siembra”.....	117
Tabla 12: Plantilla “Cosecha”.....	117
Tabla 13: Plantilla “Post Cosecha”.....	117
Tabla 14: Plantilla “Mantenimiento”.....	118
Tabla 15: Plantilla “Ventas”.....	118
Tabla 16: Encuesta de satisfacción de usuario.....	119

INTRODUCCIÓN

El presente documento tiene como propósito describir el modelamiento del proceso de negocio e implementación de un sistema de toma de decisiones para la producción y comercialización de arroz. Se detallan los objetivos generales y específicos del trabajo. Además, se define el alcance y las herramientas que se utilizaran para desarrollar la solución.

El trabajo se apoya en un marco de trabajo donde se investigaron los conceptos y definiciones relacionadas a sistemas de apoyo de decisiones, modelado de procesos (BPMN), etc. que fueron piezas fundamentales para lograr el alcance propuesto.

Se realizó el levantamiento del proceso de negocio actual, se identificaron las necesidades de mejora, se definieron los requerimientos funcionales e indicadores del sistema y se planteó la solución propuesta para cumplir con el alcance definido.

Posterior a determinar los requerimientos y plantear una solución para lograr el alcance definido, se procede a analizar la información para definir el cálculo de KPIs, se define el nuevo modelo de negocio, se implementa los dashboard que forman parte del sistema de decisiones.

Finalmente, se realizaron las pruebas y análisis de resultados del trabajo realizado.

CAPÍTULO 1

GENERALIDADES

En el capítulo presente se indicarán los antecedentes de la empresa, los problemas que enfrentan actualmente, el alcance y la solución que se propone. De igual forma, se define el objetivo general y los objetivos específicos del proyecto.

1.1 Antecedentes

El arroz constituye el principal alimento de los ecuatorianos, especialmente formando parte de la dieta de los habitantes de la región Costa [1] . De hecho, el consumo anual por habitante llega a los 53.2 kg [2], lo cual remarca la magnitud de su importancia.

En el Ecuador el cultivo del arroz se realiza tanto en el invierno como en el verano donde se depende en su mayoría del agua de riego. En el 2018, la

superficie sembrada de arroz a nivel nacional fue de 301.853 hectáreas. La producción se concentra en la provincia del Guayas con el 72,7% [3].

Fundamentalmente en las provincias del Guayas y Los Ríos. Las zonas arroceras del país presentan un amplio rango en la distribución de los factores climáticos que varía desde el trópico húmedo hasta el trópico seco, con temperaturas de 20° a 30 °C, precipitaciones máximas de 2500 mm y mínimas de 500 mm por año con humedad relativa generalmente alta. Estas zonas son fértiles y su mayor limitante es la inadecuada disponibilidad de agua, factor que en extensas zonas de secano es mínimo, sujeto a las lluvias. [4]

A pesar de que el precio de productor de arroz ha decaído desde enero 2015 de \$415,00/tonelada a \$306,00/tonelada a enero del 2018, esto no ha repercutido en el precio internacional del cereal, el cual se cotiza en \$425,00 por tonelada. [4]. Sin embargo, el precio en el mercado internacional se estima que cotizará entre \$409/tonelada y \$424/tonelada en los próximos 5 años.

Domésticamente, las ventas de cultivo de arroz pasaron de 30.02 millones USD en el 2013 a 27.12 millones USD en el 2016, representando una disminución del 9%. [5].

En el cantón Salitre, provincia del Guayas, “EVARROZ” es una pequeña empresa que se dedica a la producción de arroz y cuenta con 10 hectáreas

de terreno para sembrío. Esta empresa pertenece a una conocida familia de Las Ramas, cabecera cantonal del Salitre. La junta de directiva está conformada por la Sra. Shirley Lorena (dueña de las tierras), el Sr. Germán Jiménez (Gerente General) y Marcos Jiménez (Representante legal).

La Hacienda "Shirley Lorena" está ubicada en la parroquia Junquillal, a 30,2 Kms de Las Ramas. Este lugar es apto para realizar las actividades tanto de la siembra como el proceso de pilación del cereal. Además del terreno, la hacienda cuenta con las instalaciones adecuadas para la producción:

- Tendal: área de 50m² para secar el arroz
- Maquinaria y equipo de trabajo: tractores, cosechadora, palas, recogedores, cosedora, picos, mangueras, bombas de riego, piladora y secadora de granos.
- Oficinas administrativas
- Bodegas para almacenaje

Los canales de distribución son principalmente tiendas de barrio, supermercados locales y entregas a domicilio bajo pedido, siempre y cuando sean dentro del cantón; además de las ventas que se realicen a distribuidores externos o pedidos que vengan fuera del cantón.

El organigrama de la empresa que se muestra en la figura 1.1, consta del gerente general, administrador, especialista de calidad y personal operativo.



Figura 1.1 Organigrama Empresa EVARROZ

Fuente: Autores

El Gerente General se encarga de abrir el mercado, negociar con clientes los valores de venta y tomar decisiones sobre del negocio. A él le responden directamente el contador general que elabora informes financieros mensuales, estado de pérdidas y ganancias, declara impuestos, control y pago de la nómina y realiza el control de crédito y cobranzas. Y el equipo de ventas que se encargan de buscar nuevos clientes, distribuir el producto y realizar campañas publicitarias.

El Administrador realiza revisiones periódicas de las maquinarias, gestiona el mantenimiento de los equipos e instalaciones. A su cargo están los que realizan la labor operativa: Sus actividades son:

- Cuadrilleros 1:
 - Siembra de la semilla
 - Aplica el abono y fertilizantes correspondientes durante el proceso de crecimiento de la siembra.
 - Estiba y moviliza el producto para cada proceso.
- Cuadrilleros 2:
 - Apoya al operador de maquinaria en el secado y pilación.
 - Estiba y moviliza el producto para cada proceso.
 - Empaca el producto.
- Cuadrilleros 3:
 - Limpieza diaria de las instalaciones.
 - Empacado del producto.
 - Coser sacos.

El Supervisor de Seguridad coordina los turnos de la guardianía privada, registra incidentes de seguridad y elabora los informes de control para gerencia.

Conociendo los antecedentes, el lugar donde se encuentra, tanto la hacienda productora como las oficinas y el funcionamiento del negocio, podremos describir la problemática que afecta a la empresa que atañe directamente al deficiente uso de la información en sus actividades y toma de decisiones.

1.2 Descripción del problema

Como parte de las actividades de control, la gerencia solicita regularmente informes de los ingresos y gastos, en donde, se contempla información sobre las ventas y de la producción. El enfoque gerencial de estos informes recae sobre el presupuesto general, costos asociados al proceso productivo, los ingreso por ventas e indicadores de gestión. No existen en la actualidad, algún sistema especializado en “EVARROZ” que los ayude a realizar estos informes de manera rápida y confiable, predecir cambios en el mercado y adaptarse a ellos ni decidir el mejor rumbo para la empresa.

Para cumplir con los informes solicitados por la gerencia, el contador general consolida los datos de ingresos y gastos de forma manual. Recopila los datos de las diferentes áreas del negocio y las registra en tablas de Excel. Este proceso no está definido y la responsabilidad del contador por realizar esta actividad es auto percibida. No existe un proceso establecido para registrar dicha data.

Con el fin de proporcionar la información requerida, se exportan los reportes del procesador de datos directamente. Al ser un proceso manual donde hay poca o nula contrastación de datos, los informes no son completamente confiables y pueden existir inconsistencias. Y a esto, se suma el tiempo que toma realizar los informes ya que se debe considerar

la disponibilidad de cada área para remitir los datos solicitados. Agregando horas de trabajo a los involucrados que se no solo deben preocuparse por cumplir sus funciones sino también dedicar tiempo y esfuerzo a la unificación de los datos.

Recopilar los datos puede llevar varias horas, incluso días para tener listo el informe. Lo cual, genera malestar e inconformidad en la gerencia y la junta directiva, al tener que esperar un tiempo considerable para tener información de la empresa, que, no es en tiempo real.

Luego de esclarecer los problemas que aquejan a la empresa en relación al manejo eficiente y oportuno de la información, se propondrá una solución para solventar dichos problemas.

1.3 Solución propuesta

Se propone la implementación de un Sistema de Soporte a la Decisión (Data-Driven System DSS, por sus siglas en inglés). Este sistema, entre sus ventajas, se menciona:

- Mejora la eficiencia personal
- Incrementa el control organizacional
- Facilita la comunicación interpersonal
- Genera nuevas pruebas en apoyo de una decisión
- Reducción de costos en labores que requieran decisiones

- Ayuda a automatizar los procesos de gestión

Con la correcta implementación del sistema, la empresa podrá realizar el análisis de los datos con mayor facilidad, con el propósito de tener información útil y confiable en cualquier momento.

La implementación de este sistema empieza con la carga de los datos. Como se mencionó anteriormente, estos se tienen en tablas de Excel que son digitalizadas a partir de comprobantes físicos. Posteriormente, se deberá continuar con el proceso de transformación e integración de datos (Extract, Transfor and Load ETL, por sus siglas en inglés). Luego, se continuará con el diseño dimensional de los Data Marts que trabajarán conjuntamente con el Data Warehouse que contendrá la data histórica de la empresa.

Hecho esto, para la analítica se usarán técnicas de minería de datos con el fin de explorar, clasificar y analizar información de cada fase de la producción del arroz. Las técnicas descriptivas visualizarán la información estadística mediante diagramas de barras, de cajas, histogramas, estadísticos básicos, etc., para una mayor comprensión del negocio. Los modelos predictivos empleados serán series temporales y de regresión para pronosticar tendencias (valores futuros) del rendimiento de la cosecha del arroz e ingresos por ventas. Para la visualización de la información en tableros interactivos se usará la herramienta Power BI.

Finalmente, para lograr una mayor eficiencia dentro de la empresa, el proceso completo estará detallando mediante el uso de la metodología BPMN. Donde, se realizará el diagrama AS-IS y luego de realizar la solución se presentarán los cambios en el proceso con el diagrama TO-BE.

Ya indicado la forma de proceder con la solución propuesta, se define el alcance de la propuesta.

1.4 Alcance de la propuesta

La presente propuesta contempla únicamente la fase de producción, manejo y almacenamiento y venta de la cosecha. Las demás fases no se tomarán en cuenta para el presente análisis, debido a que, considerar toda la cadena de valor de la producción de arroz tal como indica la figura 1.2, demandaría mayor cantidad de tiempo y recursos.



Figura 1.2: Cadena de producción del arroz

Fuente: elmundo.com [6]

1.5 Objetivo General

Crear un sistema que le permita a la empresa productora “EVARROZ” tener el control de la información de sus actividades, mediante el modelamiento BPMN (Business Process Management Notation) del proceso de negocio y utilizando entre sus actividades una herramienta de minería de datos, con el propósito de detectar errores a tiempo y plantear acciones preventivas, correctivas y predictivas para mejorar su actividad comercial.

1.6 Objetivos Específicos

- Levantar el proceso de producción de arroz tal como se lo lleva actualmente, utilizando la metodología BPMN.
- Analizar y diseñar la información actual de la empresa a partir de los indicadores claves de desempeño (KPIs) previo a su automatización.
- Automatizar el análisis y diseño de la información utilizando Power BI y Data Mining para incorporarlo dentro del nuevo proceso siguiendo la metodología BPMN.
- Realizar las pruebas y análisis de resultados del nuevo proceso.

Luego de conocer los antecedentes, los problemas que enfrentan, proponer una solución a estos problemas, enunciar tanto el objetivo general como los objetivos específicos, en el siguiente capítulo se revisará el marco teórico de las diferentes metodologías, herramientas y tecnologías a usarse en la solución propuesta.

CAPÍTULO 2

MARCO TEÓRICO

El presente capítulo contiene las definiciones técnicas en lo que se basa el desarrollo e implementación de este trabajo.

2.1 DSS Sistema de Apoyo de decisiones

2.1.1 Definición

Los DSS (Sistema de apoyo de decisiones) son aquellos sistemas orientados al apoyo en toma de decisiones, son herramientas útiles para la inteligencia empresarial.

Los DSS juntan los recursos intelectuales de los individuos con las capacidades del computador para mejorar la calidad de las decisiones. Es un sistema de soporte basado en computadoras para

la gerencia de decisiones para aquellos que lidian con problemas semiestructurados. [7]

2.1.2 Propósito

El propósito esencial de un DSS es el de simular soluciones, a fin de dar alternativas a las decisiones que se tomen a determinados problemas. Dada a la creciente demanda por la información en los tiempos actuales, y proporcionalmente a la capacidad de cómputo de las máquinas, es útil pensar que estos sistemas proveen una gran solución a la hora de tomar decisiones con bastantes alternativas y que requieren de muchos cálculos matemáticos. [8]

2.1.3 Estructura, Elementos y Componente

Subsistema de gestión de Datos

Es la espina dorsal del sistema DSS que contiene la información esencial para trabajar, sin ella no habría datos que procesar. Este subsistema trabaja con un manejador de base de datos, que sirve de medio con dicha base para hacer consultas de los datos, modificarla, borrarla, entre otras acciones. [8]

Subsistema de gestión de modelado

Los modelos son representaciones matemáticas de la realidad, que generalmente sirven para describir acciones o parámetros, es este caso, de una organización, y luego usarlos efectivamente. [8]

Subsistema de gestión de diálogos

Es la interfaz que se muestra al usuario, generalmente compuesta con íconos y menús para comunicarse efectivamente sin necesidad de hacer un proceso introductorio del sistema o de asistencia técnica. Permite llevar además un control de las actividades que se realizan en forma de historial. [8]

2.2 Metodología Cross-Industry Standar Process para Data Mining

Las técnicas de Data Science o Data Analytics, que tanto interés despiertan hoy en día, en realidad surgieron en la década de los 90, cuando se usaba el término KDD (Knowledge Discovery in Databases) para referirse al concepto de hallar conocimiento en los datos [9], son muchos los modelos de procesos que han sido propuestos para el desarrollo de proyectos Data Mining tales como CRISP-DM (Cross Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify, Model, and Assess), DMAMC (Definir, Medir, Analizar, Mejorar, Controlar), sin embargo uno de los modelos más utilizados en es CRISPDM. [9]

CRISP-DM Está dividida en 4 niveles de abstracción organizados de forma jerárquica en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases (figura 2.1):

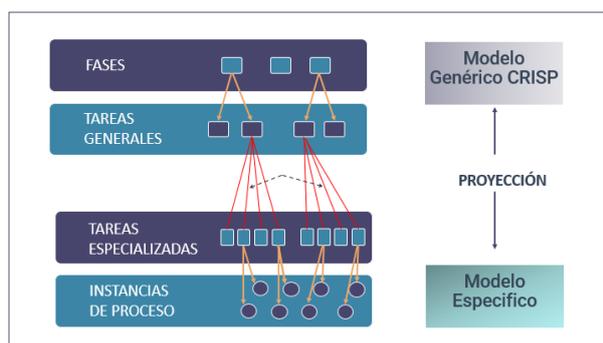


Figura 2.1 Esquema de los 4 niveles de CRISP – CM

Fuente: Metodología para el Desarrollo de proyectos en minería [10]

La secuencia de las fases no es rígida: se permite movimiento hacia adelante y hacia atrás entre diferentes fases. El resultado de cada fase determina qué fase, o qué tarea particular de una fase, hay que hacer después. Las flechas indican las dependencias más importantes y frecuentes. [10]

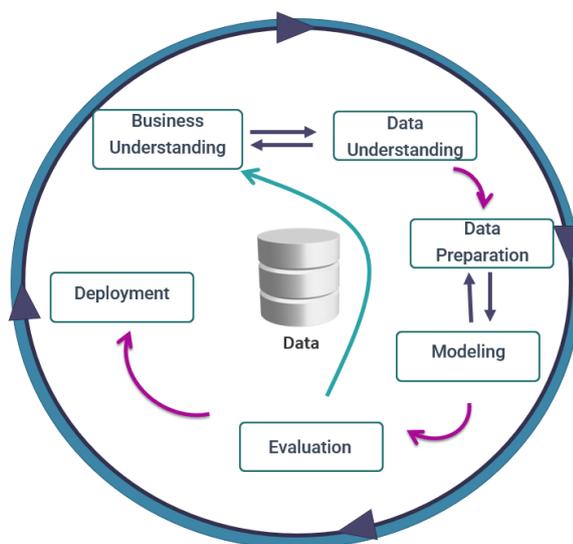


Figura 2.2 Modelo del proceso CRISP-DM

Fuente: Metodología para el Desarrollo de proyectos en minería [10]

A continuación, se describen cada una de las fases:

2.2.1 Fase I: Business Understanding. Definición de necesidades del cliente (comprensión del negocio)

La primera fase denominada fase de comprensión del negocio Figura 2.2, es la más importante y aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto. En esta fase, es muy importante la capacidad de poder convertir el conocimiento adquirido del negocio, en un

problema de Data Mining y en un plan preliminar cuya meta sea el alcanzar los objetivos del negocio. [11]

Descripción de cada una de las principales tareas que componen esta fase:

- Determinar los objetivos del negocio: Se determina el problema y definir criterios de éxito.
- Evaluación de la situación: Evaluar el estado de la situación para iniciar el proceso de Data Mining.
- Determinación de los objetivos de DM: Se definen las metas del proyecto de DM.

2.2.2 Fase II: Data Understanding. Estudio y comprensión de los datos

La segunda fase Figura 2.2, fase de estudio y comprensión de los datos, comprende la recolección inicial de datos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis. Esta fase junto a las próximas dos fases, son las que demandan el mayor esfuerzo y tiempo en un proyecto de DM. Por lo general si la organización cuenta con una base de datos corporativa, es deseable crear una nueva base de datos ad-hoc al proyecto de DM,

pues durante el desarrollo del proyecto, es posible que se generen frecuentes y abundantes accesos a la base de datos al objeto de realizar consultas y probablemente modificaciones, lo cual podría generar muchos problemas. [11]

Descripción de cada una de las principales tareas que componen esta fase:

- Descripción de datos: Se establecen los registros, campos de registros, identificación y descripción de los datos adquiridos.
- Exploración de datos: Aplicación de pruebas estadísticas.
- Verificación de calidad de los datos: Se validan errores.

2.2.3 Fase III. Data Preparation. Análisis de los datos y selección de características

En esta fase y una vez efectuada la recolección inicial de datos, se procede a su preparación para adaptarlos a las técnicas de Data Mining que se utilicen posteriormente, tales como técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para exploración de los datos. La preparación de datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato. [11]

Esta fase se encuentra relacionada con la fase de modelado, puesto que, en función de la técnica de modelado elegida, los datos requieren ser procesados de diferentes formas. Es así que las fases de preparación y modelado interactúan de forma permanente. Una descripción de las tareas involucradas en esta fase es la siguiente:

Selección de datos. En esta etapa, se selecciona un subconjunto de los datos adquiridos en la fase anterior, apoyándose en criterios previamente establecidos en las fases anteriores: calidad de los datos en cuanto a completitud y corrección de los datos y limitaciones en el volumen o en los tipos de datos que están relacionadas con las técnicas de DM seleccionadas.

Limpieza de los datos. Esta tarea complementa a la anterior, y es una de las que más tiempo y esfuerzo consume, debido a la diversidad de técnicas que pueden aplicarse para optimizar la calidad de los datos con el objeto de prepararlos para la fase de modelación. Algunas de las técnicas a utilizar para este propósito son: normalización de los datos, discretización de campos numéricos, tratamiento de valores ausentes, reducción del volumen de datos, etc. [11]

2.2.4 Fase IV. Modeling. Modelado

En esta fase de CRISP-DM, se seleccionan las técnicas de modelado más apropiadas para el proyecto de Data Mining específico. Las

técnicas para utilizar en esta fase se eligen en función de los siguientes criterios:

- Ser apropiada al problema.
- Disponer de datos adecuados.
- Cumplir los requisitos del problema.
- Tiempo adecuado para obtener un modelo.
- Conocimiento de la técnica

Previamente al modelado de los datos, se debe determinar un método de evaluación de los modelos que permita establecer el grado de bondad de ellos. Después de concluir estas tareas genéricas, se procede a la generación y evaluación del modelo. Los parámetros utilizados en la generación del modelo dependen de las características de los datos y de las características de precisión que se quieran lograr con el modelo. [11]

Descripción de las principales tareas:

- Selección de la técnica de modelado: Se selecciona la técnica de Data Mining.
- Generación del plan de prueba: Generación del plan de prueba.
- Construcción del modelo: Ejecución de la técnica DM en los datos.
- Evaluación del modelo: Se interpretan los modelos.

2.2.5 Fase V. Evaluation. Evaluación (obtención de resultados)

En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse, además, que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis. Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para poder repetir algún paso anterior, en el que se haya posiblemente cometido algún error. [10]

Descripción de las principales tareas:

Evaluación de los resultados: Se evalúa el modelo en relación con los objetivos del negocio.

Proceso de revisión: Se califica el proceso entero y se busca mejoras.

2.2.6 Fase VI. Deployment. Despliegue (puesta en producción)

En esta fase (figura 4), y una vez que el modelo ha sido construido y validado, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio, ya sea que el analista recomiende acciones basadas en la observación del modelo y sus resultados, ya sea aplicando el modelo a diferentes conjuntos de datos o como parte del proceso. Generalmente un proyecto de Data Mining no concluye en

la implantación del modelo, pues se deben documentar y presentar los resultados de manera comprensible para el usuario, con el objetivo de lograr un incremento del conocimiento. Por otra parte, en la fase de explotación se debe asegurar el mantenimiento de la aplicación y la posible difusión de los resultados.

2.3 Modelado gráfico de procesos de negocio: BPMN

Según Henry J. Johansson “un proceso de negocio es un conjunto de actividades relacionadas que permiten crear un producto o servicio final a través de la transformación de uno o varios productos o servicios iniciales”. [12]

En la Figura 2.3 se muestran los elementos claves que deben ser tomados en cuenta para el modelaje.



Figura 2.3 Elementos claves de un proceso de Negocio

Fuente: www.dspace.uclv.edu.cu [12]

2.3.1 BPM (BUSINESS PROCESS MANAGEMENT)

Los procesos de negocio deben tener dos características principales: ser eficientes y efectivos, Para lograr estas dos características se ha propuesto un enfoque de gestión llamado BPM (del inglés Business Process Management) que tiene como objetivo mejorar el funcionamiento de los procesos de negocio de una organización.

BPM se lo puede definir como “Un conjunto de herramientas, tecnologías, técnicas, métodos y disciplinas de gestión para la identificación, modelización, análisis, ejecución, control y mejora de los procesos de negocio”. [13]

2.3.2 Ciclo de vida BPM

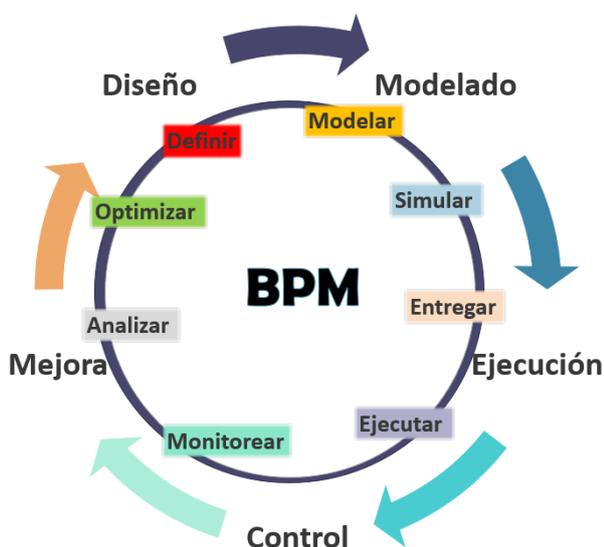


Figura 2.4 Ciclo de vida del proceso BPM

Fuente: Artículo Gestión de procesos con BPM [14]

1. Diseño

Es tal vez la parte más importante del proceso, aquí se definen los objetivos y estrategias con miras a dar cumplimiento a la dimensión de valor habiendo entendido previamente los procesos actuales de la empresa y las reglas del negocio. [14]

2. Modelamiento

En esta etapa se definen los usuarios, roles y reglas para la ejecución del paso a paso, así, por ejemplo, se deben cumplir unas condiciones mínimas para poder ejecutar el paso siguiente del proceso. [14]

Se descubren los procesos de negocio, se define el AS-IS (cómo es actualmente el proceso) y TO-BE (cómo debe ser), se analiza la mejora o el rediseño del proceso.

Una vez modelado el proceso se deben hacer las respectivas pruebas que corroboren que el proceso efectivamente soluciona la necesidad por la que fue concebido, para ello es importante simular el proceso en las herramientas de modelado antes de aplicarlo en ambiente real. [14]

3. Ejecución

Después de corroborar que el proceso funciona de acuerdo a las necesidades, se debe hacer un proceso de entrega donde se da a los usuarios involucrados la información necesaria para paso a producción del proceso. [14]

La etapa de ejecución es de vital importancia porque es aquí donde se reemplaza el proceso antiguo por el optimizado, en caso de que ya existiera el proceso, o se pone en marcha el nuevo proceso automatizando las tareas necesarias siempre con miras al cumplimiento de los objetivos del negocio. [14]

4. Control

Después del paso a producción del proceso, es importante monitorear y controlar que el proceso se esté ejecutando correctamente y analizar los fallos que se llegaran a presentar para proponer mejoras en los mismos. [14]

5. Optimización

Los hallazgos de la etapa de control serán el punto de partida para la optimización de los procesos, pues se identificarán rutas críticas, flujos a mejorar, cuellos de botella, entre otros, que deben ser analizados, de ellos se generarán cambios, eliminación de pasos que sean innecesarios, modificación

2.3.3 BPMN (BUSINESS PROCESS MODEL AND NOTATION)

El proceso de mejora de producción de arroz será modelado utilizando la notación estándar BPMN. La notación BPMN fue creada con el fin de modelar los procesos de negocio de las organizaciones, haciendo uso de gráficos que permiten comunicar los procesos de forma clara, completa y eficiente.

2.4 Inteligencia de Negocios

2.4.1 Business Intelligence BI

Business Intelligence (BI) es un término paraguas que abarca los procesos, las herramientas, y las tecnologías para convertir datos en información, información en conocimiento y planes para conducir de forma eficaz las actividades de los negocios. [15]

BI abarca las tecnologías de datawarehousing los procesos en el “back end”, consultas, informes, análisis y las herramientas para mostrar información (estas son las herramientas de BI) y los procesos en el “front end”. [15]

Data Warehouse

Un Data Warehouse es una base de datos corporativa en la que se integra información depurada de las diversas fuentes que hay en la organización. Dicha información debe ser homogénea y fiable, se

almacena de forma que permita su análisis desde muy diversas perspectivas, y que a su vez dé unos tiempos de respuesta óptimos. Para ello la información se encuentra altamente desnormalizada y modelada de una forma bastante diferente a los sistemas transaccionales, principalmente se utilizan los modelos en estrella (star schema) y en copo de nieve (snowflake schema) que estudiaremos más adelante. [16]

Data mart

Un Data Mart es una copia de las transacciones específicamente estructurada para la consulta y el análisis.

Según Ralph Kimball, “Un Data Mart es un conjunto de datos flexible, idealmente basado en el nivel de granularidad mayor que sea posible, presentado en un modelo dimensional que es capaz de comportarse bien ante cualquier consulta del usuario. En su definición más sencilla, un data mart representa un único proceso de negocio”. [16]

2.4.2 Herramientas de Business Intelligence

Definición

Una herramienta de inteligencia de negocios es un software de aplicación creado para ayudar con los procesos de inteligencia de negocios en organizaciones. Estas herramientas cumplen las

funciones de extracción, transformación y carga (ETL), análisis, y presentación de datos, siendo las funciones de ETL las que menos comúnmente se ve en este tipo de software ya que existen otros softwares destinados específicamente a este fin y que pese a formar parte del proceso de implementación no son considerados inteligencia de negocios como tal. [18]

2.5 Data Mining – Python

La Minería de Datos es el conjunto de técnicas y tecnologías que permiten analizar grandes cantidades de información depositados en bases de datos; mediante la extracción y depuración de datos se pueden identificar patrones, clases, grupos, tendencias y asociaciones que expliquen su comportamiento en un contexto y periodo de tiempo determinado. [19]

El proceso de minería de datos empieza con la selección y filtrado de las fuentes y el tipo de dato. Utilizar un filtrado correcto es fundamental para evitar datos ambiguos, desconocidos e innecesarios que tomen más tiempo de análisis y expresen patrones poco precisos. El resultado es una estructura de datos adecuada a la que aplicaremos la fase de modelado y normalización. Con esto se logra obtener un modelo final con patrones válidos para interpretar y evaluar los resultados obtenidos. [19]

2.5.1 Técnicas de Data Mining

- **Redes Neuronales artificiales:** Son sistemas de aprendizaje que en base a resultados construyen procesos. Son principalmente útiles en la minería de datos ya que al conocer los resultados esperados pueden encontrar patrones que conduzcan a los objetivos planteados. Por ejemplo: ¿Qué cliente comprará más en el mes 3? [17]
- **Árboles de decisión:** De forma visual representan las decisiones históricas de los clientes, de manera que se pueden visualizar patrones y tendencias. [17]
- **Agrupamiento:** Al agrupar datos se buscan elementos cercanos que formen grupos, y a partir de este grupo ver sus similitudes y encontrar tendencias. [17]
- **Algoritmo Jerárquico:** Se calcula la distancia entre pares de datos cercanos, se repite esta acción hasta que no quedan más elementos de comparación. [17]
- **Regla de inducción:** Consiste en generar reglas para clasificar casos, estas reglas se aplican sobre arboles de decisión y se contrastan con patrones para encontrar coincidencias. [17]

2.5.2 Python

Python es un lenguaje de programación creado por Guido van Rossum a principios de los años 90, su nombre está inspirado en el grupo de cómicos ingleses “Monty Python”, es un lenguaje similar a Perl, pero, con una sintaxis muy limpia y que favorece un código legible. Entre sus principales características están:

Ser un lenguaje interpretado o de Script, Tipado Dinámico, Fuertemente Tipado, Multiplataforma (Windows, Linux, MAC) y orientado a objetos, es uno de los lenguajes más utilizados en el desarrollo de aplicaciones de ciencia de datos porque combina en gran forma su potencia, con una sintaxis muy clara lo que es atractivo para los programadores principiantes o personas que han dejado de programar por algún tiempo.

Sus librerías nos permiten una gran variedad de funcionalidades generales y específicas para: carga y visualización de datos, estadísticas, procesamiento de lenguaje natural y de imágenes, entre otras; y una gran comunidad de usuarios dedicada a fomentar el uso del lenguaje, ayudando a los desarrolladores con cualquier tipo de dudas.

Sus bibliotecas de manejo de tareas relacionadas con grandes volúmenes de datos son: numpy y pandas, éstas incluyen muchas de las capacidades del software R y MATLAB, pero son más intuitivas.

Estas características hacen de Python un lenguaje superior o a la par de los lenguajes disponibles, y es por esto, que actualmente se lo elija con más frecuencia para el desarrollo de un gran número de aplicaciones para análisis de datos.

CAPÍTULO 3

DEFINICIÓN DE LA SITUACIÓN ACTUAL Y DEFINICIÓN DE REQUERIMIENTOS

En el presente capítulo se detalla la situación actual de la empresa productora de arroz, así como se muestra el diagrama AS-IS usando la metodología BPMN. Se identifican los requerimientos deseados por parte de la empresa con el propósito de mejorar la capacidad de toma de decisiones.

3.1 Definición de la situación actual

Debido al giro de negocio de EVARROZ, producción y comercialización de arroz, el flujo de dinero que ingresa no es constante. Las ventas se hacen luego del proceso productivo, el cual toma entre dos a tres meses, y este se realiza una vez por cada estacione climática del año, invierno y verano.

Lo cual, deja un amplio margen de tiempo donde no se recopila ni mucho menos se consolida la información que se genera en cada etapa del negocio.

El departamento de ventas son los encargados de la promoción, distribución y ventas del arroz producido. Por lo cual, los convierte en el área de mayor relevancia dentro de la empresa, por lo tanto, es imprescindible que cuenten con información confiable y a tiempo para elaborar metas reales que proporcionen la rentabilidad que los dueños desean.

Como se mencionó con anterioridad, EVARROZ no cuenta con un sistema que les permita visualizar la información de su actividad productiva y comercial previa. Inclusive, no disponen de ningún sistema especializado para ordenar, almacenar y muchos menos generar reportes. Lo cual, ocasiona severos problemas al momento de tomar decisiones ya que, no tienen visibilidad ni trazabilidad del negocio.

Actualmente, el encargado de recopilar y consolidar la información de cada área del proceso productivo es el contador, el cual, solicita verbalmente a las áreas involucradas la información necesaria para generar los informes que le piden. Esta información es enviada al contador en tablas de Excel, con formatos no estandarizados y que suelen incluir datos no relevantes y en ciertas ocasiones presentan incoherencias.

Estas tablas contienen la cantidad de semillas usadas para la siembra, cantidad de fertilizantes y fitosanitarios, costos directos e indirectos relacionados con la siembra, cosecha, entre otros. Sin embargo, la forma en la que se tratan a los datos no permite realizar un análisis a profundidad de la situación actual de la empresa, ni mucho menos predecir cambios en el sector arrocero, realizar una planificación de la producción ni una estrategia de ventas acorde a la situación actual del mercado.

Cada área aporta los datos que el contador solicita, el área de agronomía, conformada por el administrador y cuadrilleros, proporcionan datos sobre la cosecha, siembra y post cosecha. A su vez, el contador, el cual lleva el control de compras y gastos, proporciona los datos de mantenimiento y finalmente, el departamento de ventas aporta datos del tipo de producto vendido y la cantidad de los mismos. La información es almacenada en el computador del contador, el cual, previamente se encargó de ordenar los datos, corregir las inconsistencias y generar el reporte solicitado.

Este proceso de recopilación y consolidación de información es un trabajo que demanda mucho tiempo y está grandemente expuesto a errores, tanto por la transcripción de los datos como por la excesiva interacción humana. Así mismo, luego de generado el reporte, el contador debe validar la veracidad de la información y la coherencia de estos, tarea que incrementa el tiempo demandado.

A pesar de ser un proceso crítico para la toma de decisiones oportunas y acertadas, los dueños no han considerado implementar un sistema de soporte a la toma de decisiones lo cual deriva en una significativa desventaja competitiva de la empresa.

3.2 Levantamiento de información de los procesos y actividades actuales

Para realizar el levantamiento de información, se procedió a entrevistar a cada uno de los involucrados claves en el proceso de negocio, estos usuarios son responsables de enviar informes en base a su trabajo realizado en cada periodo de producción. Estos usuarios tienen amplio conocimiento de sus respectivas áreas de trabajo y cuentan con la experiencia en el sector de la producción de arroz.

En las entrevistas realizadas, los usuarios dieron a conocer como elaboran los informes sobre los reportes de cada periodo, el análisis de calidad requerido para poder lanzar su producto al mercado, la clasificación por tipo de producto final y las ventas. Además, conversamos con la gerencia quien nos mostró los informes que actualmente disponen y nos hizo conocer la necesidad de tener información disponible y confiable para tomar nuevas decisiones.

Dentro del proceso de negocio, se identificaron los siguientes subprocesos y actividades:

Actividad de Planificación de siembra: En este proceso el gerente con apoyo del administrador planifica las hectáreas de arroz que van a sembrar, se realiza la lista de productos que debe adquirirse como semillas, fertilizantes, entre otros. Estos datos son enviados al contador, la comunicación es verbal.

Actividad de adquisición de insumos y equipos: En este proceso, el contador se encarga de comprar los suministros que se detallaron en el proceso de planificación de siembra. El contador comunica la recepción de las compras vía verbal.

Proceso de siembra: Este proceso tiene un tiempo aproximado de tres meses, donde el grupo de agronomía realiza la preparación de tierra, siembra de semillas, continua fertilización y cuidado de plagas, para posterior pasar a cosechar.

Proceso de Cosecha: En este proceso se recogen las plantas de arroz posterior a drenar el agua para luego pasar a un proceso de secado para luego pasar al proceso post cosecha.

Proceso post cosecha: En este proceso se procede a almacenar el arroz.

Actividad de control de calidad de producto: Se toma una muestra de la cosecha del periodo y se procede a enviar a una empresa externa para

que se evalúe la calidad del arroz, los resultados son enviados al área contable quienes realizan la comunicación por mail al departamento agrónomo, para que se clasifique el producto.

Actividad de empaquetado: Se procede a dividir el arroz en tres productos finales:

- Grupo 1: arroz que se pila para venta por saco,
- Grupo 2: arroz que no se pila y pasa a empaquetado para producto de arroz integral.
- Grupo 3: utilizado para derivados.

La cantidad de sacos que se obtienen según su calidad y clasificación es comunicada al departamento de ventas, para que procedan a la venta de estos.

Proceso de Venta: En este proceso el grupo de trabajo del área realiza un estudio de mercado y en base a la calidad del producto definen el precio de venta, se encarga de la recepción de pedidos del cliente, el registro de los datos referentes al pedido en un documento de Excel, tales como: detalles de productos, cantidades, fechas de entrega y otras consideraciones, posteriormente se comunica telefónicamente al cliente y demás áreas involucradas (distribución de producto y contabilidad).

Actividad de distribución de producto: Posterior a que el producto se encuentra almacenado en las bodegas, el personal de distribución revisa la información enviada por ventas vía mail, donde se detallan las cantidades, direcciones, clientes y fechas donde se realizarán las entregas del producto.

Proceso contable de actividad: Posterior al registro de entrega, se procede a realizar la facturación de los diferentes despachos, se realiza utilizando como guía la información enviada por ventas vía mail.

Proceso de generación de Informe: En este proceso el contador solicita verbalmente a los vendedores y al administrador los informes de ventas y la producción de arroz respectivamente.

- **Generación Informe Administrador:** Los datos son registrados manualmente en una bitácora del encargado, quien entrega la información impresa al contador.
- **Generación Informe Ventas:** El personal de ventas envía los detalles en Excel al contador, para su análisis.
- **Validación de información:** Posterior a recibir la información el personal contable evalúa los datos, en caso de existir errores, procede a devolverlos y realizar una evaluación hasta tener una aprobación final.
- **Generación de Informe:** Luego de contar con la información requerida, se procede a generar un informe con el detalle de gastos vs ingresos

durante el periodo de producción, el resultado final es entregado impreso a la gerencia.

3.3 Levantamiento de Requerimientos Funcionales y no funcionales

Previo diseño de la solución es necesario conocer lo que los usuarios esperan del sistema y que necesidades desean que este satisfaga. Para esto, es preciso definir qué es un requerimiento.

Un requerimiento funcional es una declaración de los servicios que proveerá el sistema, de manera que éste reaccionará en situaciones particulares. A su vez, un requerimiento no funcional es una restricción de los servicios o funciones ofrecidos por el sistema. [18]

Es decir, los requerimientos funcionales son las tareas que el sistema debe hacer y los requerimientos no funcionales indican qué tan bien debe hacer esas tareas.

Definido esto, se procedió a realizar entrevistas a los principales usuarios de la herramienta a implementar. Se consideraron las perspectivas de producción como la perspectiva comercial.

En la Tabla 3.1 se detallan los requerimientos funcionales obtenidos de las entrevistas que se les realizaron a los usuarios principales.

REQUISITO	REQUERIMIENTO FUNCIONAL
Tener la totalidad de ventas realizadas durante cada periodo, separado por tipo de producto, cantidad y dinero.	El sistema debe mostrar el detalle de las ventas realizadas ordenados por gráficos y poder generar el reporte de las ventas realizadas y compararla con otros periodos y otros años.
Saber la cantidad de hectáreas sembradas tipo de semilla usada, calidad de la semilla y número de kilos de semilla sembrada por periodo.	El sistema debe mostrar el detalle de la cantidad de hectáreas sembradas y no sembradas ordenados por gráficos y poder generar el reporte donde se compara con otros periodos y otros años.
Totalizar los costos directos e indirectos por cada fase de producción separados por periodo.	El sistema debe mostrar el detalle de los costos totales incurridos en la producción ordenados por gráficos y poder generar el reporte donde se compara con otros periodos y otros años.

Conocer la ganancia neta por cada periodo.	El sistema debe mostrar el detalle de la utilidad neta de la actividad comercial ordenados por gráficos y poder generar el reporte donde se compara con otros periodos y otros años.
--------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Tabla 1: Requerimientos funcionales

En la Tabla 3.2 se detallan los requerimientos no funcionales que fueron obtenidos a partir de entrevistas que se les realizaron a los usuarios principales.

Tabla 2: Requerimientos no funcionales

REQUISITO	REQUERIMIENTO FUNCIONAL
El sistema debe ser útil.	El sistema debe garantizar que la información mostrada sea relevante para los usuarios.
El sistema debe ser confiable.	El sistema debe proporcionar garantías para que este sea

REQUISITO	REQUERIMIENTO FUNCIONAL
	accesible en cualquier momento de forma rápida y segura.
El sistema debe ser fiable.	El sistema debe contener y mostrar información valida y no sesgada.
El sistema debe ser comprensible.	El sistema debe mostrar la información de forma ordenada y comprensible a los usuarios.

Identificados las principales necesidades de los usuarios y habiendo detallado tanto los requerimientos funcionales como no funcionales, en el siguiente punto se indica los indicadores de medición.

3.4 Definición de principales indicadores de desempeño del negocio

Para identificar los KPIs de los procesos de producción y comercialización del arroz, se analizaron los requerimientos funcionales previamente obtenidos, además de realizar una serie de entrevistas a los involucrados,

donde, el objetivo fue encontrar similitudes en las necesidades de cada área.

Para esto, se exploró a fondo los departamentos y personas relacionadas con el problema a solucionar para recolectar los datos necesarios.

En la tabla 3.3 se muestran las fases del negocio tanto del proceso productivo como comercial, los indicadores a mostrar y como estos se relacionan con el negocio.

Tabla 3: Principales indicadores de desempeño y su relación con preguntas del negocio

FASE DEL NEGOCIO	INDICADOR	PREGUNTA DEL NEGOCIO
SIEMBRA	Número de hectáreas sembradas	¿Cuántas hectáreas sembradas tienen cada periodo organizado por años?

FASE DEL NEGOCIO	INDICADOR	PREGUNTA DEL NEGOCIO
	Porcentaje de pureza de la semilla	¿Cuál es el porcentaje de pureza de la semilla sembrada?
	Número de semillas sembradas	¿Cuál es el número de semillas sembradas por periodo y año?
	Costos directos de siembra	¿Cuáles son los costos directos por periodo y año que se sembró?
COSECHA	Número de sacos	¿Cuántos sacos de arroz fueron

FASE DEL NEGOCIO	INDICADOR	PREGUNTA DEL NEGOCIO
		cosechados por años y periodo?
	Tipo de grano cosechado	¿Cuál es la variedad de arroz que más se ha cosechado por año y periodo?
	Costos directos de cosecha	¿Cuáles son los costos directos y mano de obra en proceso de cosecha de arroz por años y periodos?
	Porcentaje de suelo no productor	¿Cuál es el porcentaje de suelo no productor por años y por

FASE DEL NEGOCIO	INDICADOR	PREGUNTA DEL NEGOCIO
		periodo?
POST COSECHA	Promedio de calidad	¿Cuál fue el porcentaje de sacos que cumplieron y no cumplieron con el estándar de calidad?
	Número de sacos	¿Cuántos sacos de arroz blanco y arroz integral se obtuvieron?
	Número de sacos	¿Cuántos sacos de polvillo y arrocillo se obtuvieron?
	Costos de la post-cosecha	¿Cuál es el costo directo por post

FASE DEL NEGOCIO	INDICADOR	PREGUNTA DEL NEGOCIO
		cosecha y empaquetado?
MANTENIMIENTO	Costo Total	¿Cuál es el costo indirectos total por producción por periodo y año?
	Costo de materiales	¿Cuál es el gasto por materiales y maquinaria por la operación por periodo y año?
	Gasto en salarios	¿Cuál es el gasto total de salarios por la operación?

FASE DEL NEGOCIO	INDICADOR	PREGUNTA DEL NEGOCIO
	Gastos administrativos adicionales	¿Cuál es el gasto administrativo adicionales total por la operación?
VENTAS	Número de sacos	¿Cuántos sacos de arroz blanco se vendieron?
	Número de sacos	¿Cuántos sacos de arroz integral se vendieron?
	Número de sacos	¿Cuántos sacos de arrocillo y polvillo se vendieron?
	Total de ventas	¿Cuál es el total de arroz blanco, integral y

FASE DEL NEGOCIO	INDICADOR	PREGUNTA DEL NEGOCIO
		derivados ventas por periodo y año?

3.5 Definición de la mejora continua del proceso de negocio

Definiendo de acuerdo al enfoque BPM, “la mejora continua es un método por el cual las empresas realizan optimizaciones a pequeña escala de forma continuada. Este proceso mejora la calidad de los productos, los procesos y los servicios a largo plazo.” [19]

Las oportunidades de mejora surgen cuando los resultados parciales o finales de un proceso no alcanzan la meta establecida y provocan insatisfacción en los involucrados o dueños del proceso. Por lo cual, es imprescindible que para mejorar un proceso se conozcan las actividades o tareas que realizan en él.

En la figura 3.1 se muestra el proceso de negocio actual; modelos AS-IS, en el cual se observan todas las actividades y subprocesos anteriormente detallados.

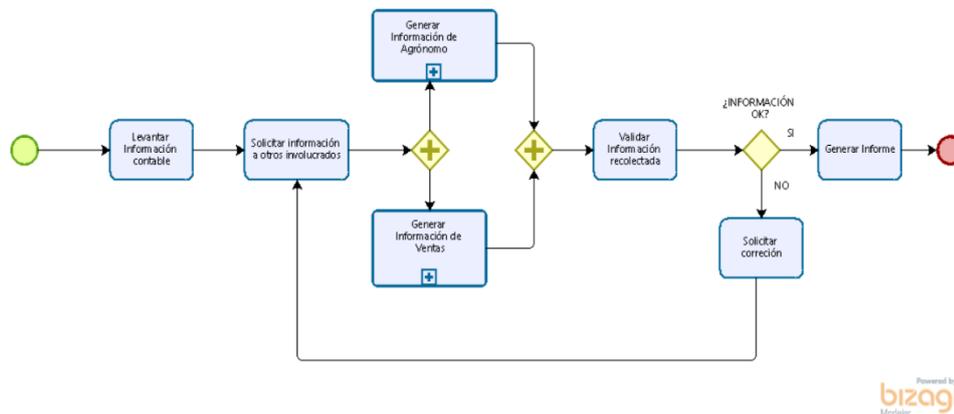


Figura 3.2: Proceso de Generación de Informes

Fuente: Autores

Finalmente, en las figuras 3.3 y 3.4 se detallan las actividades que se realizan para enviar la información de Agronomía y Ventas, respectivamente.

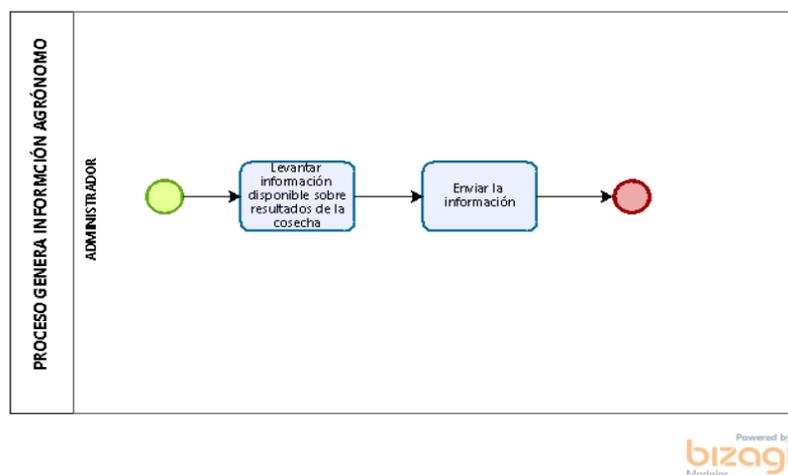


Figura 3.3: Proceso de Generar Información Agrónomo

Fuente: Autores

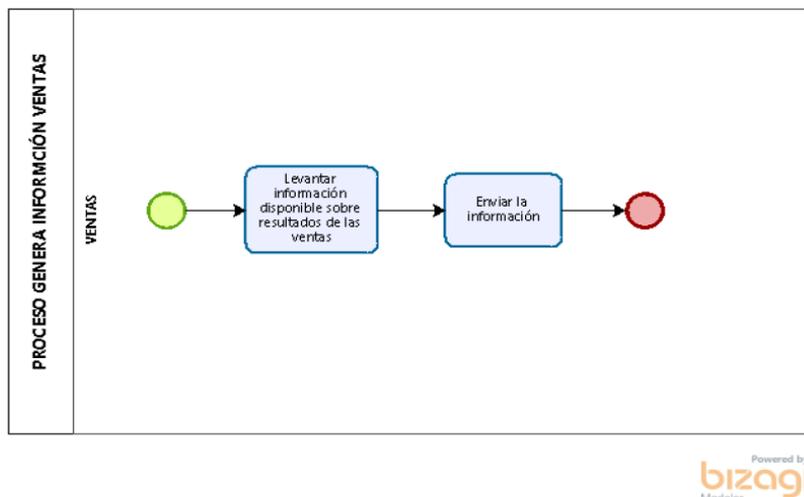


Figura 3.4: Proceso Generar Información Ventas

Fuente: Autores

En el siguiente capítulo, se detallarán las mejoras en el proceso de Generación de Informes, así como los diferentes automatismos que se introducirán a partir de la implementación de un sistema de generación de Informes.

3.6 Definición del alcance de la solución propuesta

La solución sugerida se basa en la creación de un sistema de soporte a la decisión que permitirá el análisis de los datos desde las perspectivas requeridas. De acuerdo al negocio de EVARROZ, las perspectivas son: siembra, cosecha, post cosecha, mantenimiento y ventas. Además de poder visualizar en gráficos comparativos por años y periodos, el sistema permitirá predecir el volumen de ventas que se tendrán en años

posteriores, a partir del uso de herramientas de minería de datos usando los datos almacenados.

El alcance del sistema se detalla a continuación:

- Se creará una aplicación web donde se podrán visualizar de los datos.
- Se creará un almacén de datos donde se recopile la información de las 5 etapas del proceso productivo: siembra, cosecha, post cosecha, mantenimiento y ventas.
- Se podrá descargar desde la aplicación web un formato en Excel para que los usuarios del sistema pueden alimentar continuamente el almacén de datos.
- Se podrán visualizar todos los indicadores de desempeño anteriormente mencionados por año y periodo productivo ordenados por vistas las diferentes etapas del proceso productivo
- Se podrá visualizar un resumen con los principales indicadores de desempeño solicitados por la gerencia: ventas totales, porcentaje de hectáreas sembradas, gastos totales y utilidad neta. Todo ordenado por periodo y año.
- Se podrá visualizar en una vista diferenciada el pronóstico de ventas para los años siguientes.

- El sistema a implementar no es un ERP (Enterprise Resource Planning), por lo cual, no se considera la creación de una sección contable o de control de inventarios.

Los entregables del sistema son:

- Diagramas de base de datos relacionales, dimensiones y tablas de hechos.
- Proceso de extracción ETL de los datos cargados al sistema.
- Cubos de cada etapa del proceso productivo.
- Dashboards de cada etapa del proceso productivo.
- Dashboard principal con principales indicadores de desempeño.
- Reportes generales de la actividad comercial.
- Presentación de indicadores mediante gráficos comparativos.

Ahora que se ha definido el alcance de la solución, se procede a diseñar el sistema propuesto.

CAPÍTULO 4

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN

4.1 Cálculo de KPIs

Posterior a la definición de los indicadores de negocio, ahora se establecerá como se medirán dichos indicadores. Pero antes, se deben especificar las variables que intervienen en el cálculo. A su vez, se indica, que persona es la encargada de ingresar el dato al sistema y que tipo de dato es el ingresado.

Tabla 4: Fase Siembra

VARIABLE	INGRESO AL SISTEMA	FORMATO DE DATO
Total de hectáreas disponibles	Administrador	Numérico

VARIABLE	INGRESO AL SISTEMA	FORMATO DE DATO
Número de hectáreas no sembradas	Administrador	Numérico
Pureza real del laboratorio	Administrador	Porcentaje
Pureza mínima aceptable	Administrador	Porcentaje
Número de hectáreas sembradas	Administrador	Numérico
Número de sacos usados por hectárea	Administrador	Numérico
Costo de semillas usadas	Contador	Moneda
Costo de fertilizantes usados	Contador	Moneda
Costo de Fitosanitarios	Contador	Moneda
Costo de jornales usados	Contador	Moneda
Jornales en la siembra	Administrador	Numérico

Tabla 5: Fase Cosecha

VARIABLE	INGRESO AL SISTEMA	FORMATO DE DATO
Número de libras cosechadas	Administrador	Numérico
Total de sacos cosechados	Administrador	Numérico
Libras de granos largos cosechados	Administrador	Peso
Libras de granos totales cosechados	Contador	Peso
Costo de cosecha con maquinaria	Contador	Moneda
Costo de transporte a la piladora	Contador	Moneda
Número de sacos cosechados	Administrador	Numérico
Número de sacos esperados	Administrador	Numérico

Tabla 6 Fase: Post Cosecha

VARIABLE	INGRESO AL SISTEMA	FORMATO DE DATO
Número de sacos sin cumplimiento de calidad	Administrador	Numérico
Total de sacos cosechados	Administrador	Numérico
Número de sacos con cumplimiento de calidad	Administrador	Numérico
Porcentaje destinado para arroz blanco	Contador	Porcentaje
Número de sacos con cumplimiento de calidad	Administrador	Numérico

VARIABLE	INGRESO AL SISTEMA	FORMATO DE DATO
Porcentaje destinado para polvillo	Contador	Porcentaje
Número de sacos cosechado	Administrador	Numérico
Costo unitario por ensacado de arroz	Contador	Moneda

Tabla 7 Fase: Mantenimiento

VARIABLE	INGRESO AL SISTEMA	FORMATO DE DATO
Costo indirecto siembra	Contador	Moneda
Costo indirecto cosecha	Contador	Moneda
Costo indirecto post cosecha	Contador	Moneda
Costo de mantenimiento de maquinaria	Contador	Moneda

Costo de operación de maquinaria	Contador	Moneda
Gasto de salarios agrónomo	Contador	Moneda
Gasto salarios administrativos	Contador	Moneda
Gasto salarios ventas	Contador	Moneda
Gasto de mantenimiento de instalaciones administrativas	Contador	Moneda
Gastos de insumos de consumo de personal	Contador	Moneda

Tabla 8 Fase: Ventas

VARIABLE	INGRESO AL SISTEMA	FORMATO DE DATO
Cantidad de sacos de arroz blanco vendidos	Vendedor	Numérico
Cantidad de sacos de arroz integral vendidos	Vendedor	Numérico

Cantidad de sacos de derivados	Vendedor	Numérico
Valor de saco de arroz blanco	Gerente	Moneda
Valor de saco de arroz integral	Gerente	Moneda
Valor de saco de derivados	Gerente	Moneda

Luego de establecer las variables que intervienen, se define la forma de calcular cada indicador. Se estableció una variable interna “CÓDIGO” que facilita la forma de identificar cada indicador y variable.

Tabla 9: Dashboard Principal

INDICADOR	CÓDIGO	FÓRMULA
Total ventas	DAS1	$\text{DAS1} = (\text{Cantidad de sacos de arroz blanco vendido} * \text{Valor de saco de arroz blanco}) + (\text{Cantidad de sacos de arroz integral vendido} * \text{Valor de saco de arroz integral}) + (\text{Cantidad de sacos de derivados} * \text{Valor de saco de derivados})$

INDICADOR	CÓDIGO	FÓRMULA
Porcentaje de hectáreas sembradas	DAS2	$DAS = (\text{Número de hectáreas sembradas} / \text{Total de hectáreas disponibles}) * 100$
Gastos totales	DAS3	$DAS3 = \text{Gasto directo de siembra} + \text{Gasto directo de cosecha} + \text{Gasto directo de la post cosecha} + \text{Gasto indirecto total} + \text{Gasto de materiales y maquinaria} + \text{Gasto salarios} + \text{Gastos administrativos}$
Utilidad neta	DAS4	$DAS4 = \text{Total ventas} - \text{Gastos totales}$
Margen de utilidad	DAS5	$DAS5 = (\text{Utilidad neta} / \text{Total ventas}) * 100$

Tabla 10: Dashboards Secundarios

FASE	INDICADOR	CÓDIGO	FÓRMULA
SIEMB RA	Número de hectáreas sembradas	SIE1	$SIE1 = \text{Total de hectáreas disponibles} - \text{Número de hectáreas no sembradas}$
	Porcentaje de pureza de la semilla	SIE2	$SIE2 = (\text{Pureza real de laboratorio} / \text{Pureza mínima aceptable}) * 100$
	Número de semillas sembradas	SIE3	$SIE3 = \text{Número de hectáreas sembradas} * \text{Número de sacos usados por hectárea}$
	Gasto directo de siembra	SIE4	$SIE4 = \text{Costo de semillas usadas} + \text{Costo de fertilizantes usados} + \text{Costo de Fitosanitarios} + (\text{Costo de jornales usados} * \text{Jornales en la siembra})$
COSE CHA	Promedio de libras por sacos	COS1	$COS1 = (\text{Número de libras cosechadas} / \text{Total de sacos cosechados})$

FASE	INDICADOR	CÓDIGO	FÓRMULA
	Porcentaje de granos largos cosechados	COS2	$COS2 = (\text{Libras de granos largos cosechados} / \text{Libras de granos totales cosechados}) * 100$
	Gasto directo de cosecha	COS3	$COS3 = \text{Costo de cosecha con maquinaria} + \text{costo de transporte a la piladora}$
	Rendimiento porcentual de la cosecha	COS4	$COS4 = (\text{Número de sacos cosechados} / \text{Número de sacos esperados}) * 100$
POST COSE CHA	Promedio de calidad	POS1	$POS1 = [(\text{Número de sacos sin cumplimiento de calidad} - \text{Total de sacos cosechados}) / \text{Total de sacos cosechados}] * 100$
	Número de sacos de arroz blanco	POS2	$POS2 = \text{Número de sacos con cumplimiento de calidad} * \text{Porcentaje destinado para arroz blanco}$
	Número de sacos de polvillo	POS3	$POS3 = \text{Número de sacos con cumplimiento de calidad} * \text{Porcentaje destinado para polvillo}$

FASE	INDICADOR	CÓDIGO	FÓRMULA
	Gasto directo de la post cosecha	POS4	$POS4 = \text{Número de sacos cosechado} * \text{Costo unitario por ensacado de arroz}$
MANTENIMIENTO	Gasto indirecto total	MAN1	$MAN1 = \text{Costo indirecto siembra} + \text{Costo indirecto cosecha} + \text{Costo indirecto post cosecha}$
	Gasto de materiales y maquinaria	MAN2	$MAN2 = \text{Costo de mantenimiento de maquinaria} + \text{Costo de operación de maquinaria}$
	Gasto salarios	MAN3	$MAN3 = \text{Gasto de salarios agrónomo} + \text{gasto salarios salarios administrativos} + \text{gasto salarios ventas}$
	Gastos administrativos	MAN4	$MAN4 = \text{Gasto de mantenimiento de instalaciones administrativas} + \text{gastos de insumos de consumo de personal}$
VENTAS	Número de sacos arroz blanco	VEN1	$VEN1 = \text{Cantidad de sacos de arroz blanco vendidos}$

FASE	INDICADOR	CÓDIGO	FÓRMULA
	Número de sacos arroz integral	VEN2	VEN2=Cantidad de sacos de arroz integral vendidos
	Número de sacos de derivados	VEN3	VEN3=Cantidad de sacos de derivados
	Total ventas	VEN4	VEN4= (Cantidad de sacos de arroz blanco vendido * Valor de saco de arroz blanco) + (Cantidad de sacos de arroz integral vendido * Valor de saco de arroz integral) + (Cantidad de sacos de derivados * Valor de saco de derivados)

4.2 Diseño de proceso TO-BE

En esta sección se muestra el modelado “TO-BE”, el cual se obtuvo mediante el análisis del modelado “AS-IS”, así como también se consideró las peticiones de los interesados, nuestra mejora se enfoca en rediseñar el proceso de negocio de cosecha de arroz, para que permita registrar información en cada fase y obtener informes desde una fuente unificada,

podrá proyectar las ganancias en base a la información histórica registrada y tomar una decisión sobre la inversión de la presente cosecha.

Actividad registro compra en el sistema: Si antes no existía un registro de información en alguna base compartida, con la mejora del proceso, se permitirá al usuario registrar los datos de la compra en el sistema, optimizando la comunicación y manteniendo los datos disponibles para la generación del informe final.

Actividad de control de calidad de producto: El propósito de esta actividad es conocer los resultados de la calidad del producto cosechado, en el modelo AS – IS, la información era conocida únicamente por el contador hasta proporcionar el informe final a la gerencia. Con el rediseño del proceso se busca compartir esta información a todos los interesados, ya que se podrá registrar en el sistema y realizar proyecciones sobre las ganancias basadas en la clasificación de ventas del producto (sacos de arroz blanco, integral y polvillo).

Actividad de registro de ventas: Esta actividad se adicionó al flujo, para que el área pueda registrar en tiempo real las ventas de la cosecha.

Actividad generación de Informe: Con el rediseño del proceso de negocio se cambió la generación de informe de un proceso a una actividad, debido a que la información se va obteniendo en cada actividad del

proceso general, en lugar de concentrar la información en registros separados por cada área y no conectados.

4.3 Arquitectura en capas e implementación del sistema

Se realizó una evaluación costo/beneficio de la infraestructura, servicios y componentes a utilizar y se decidió implementar el sistema en infraestructura propia, migrar de un plan corporativo de acceso a internet a acceso dedicado y mejorar la infraestructura de redes de datos y acceso inalámbrico. Esto con la finalidad de crecer e implementar nuevos sistemas que ayuden a mejorar los demás procesos.

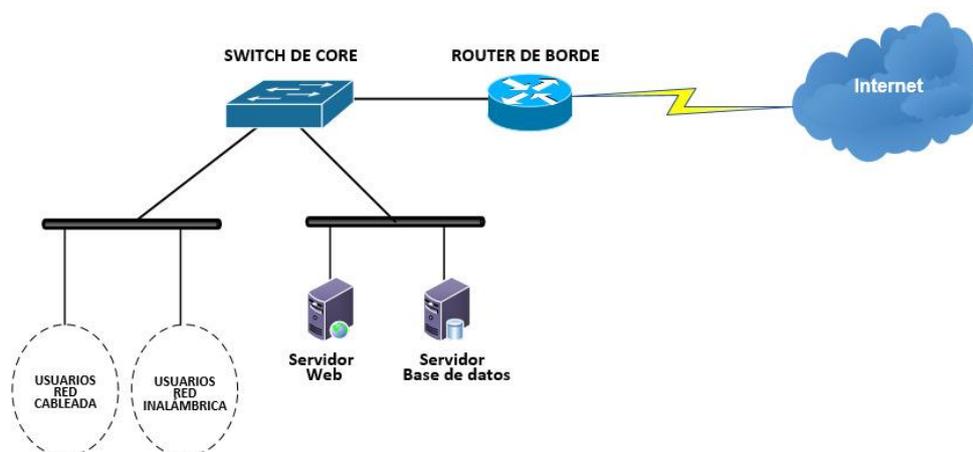


Figura 4.2: Arquitectura de la red de datos

Fuente: Autores

El sistema EVARROZ implementado es un modelo de tres capas. Una arquitectura en capas permite dividir y segmentar los roles que cada capa debe ejecutar en el sistema. A pesar de que cada capa alimenta a otra, esto no implica que el funcionamiento de una capa afecte a la otra.

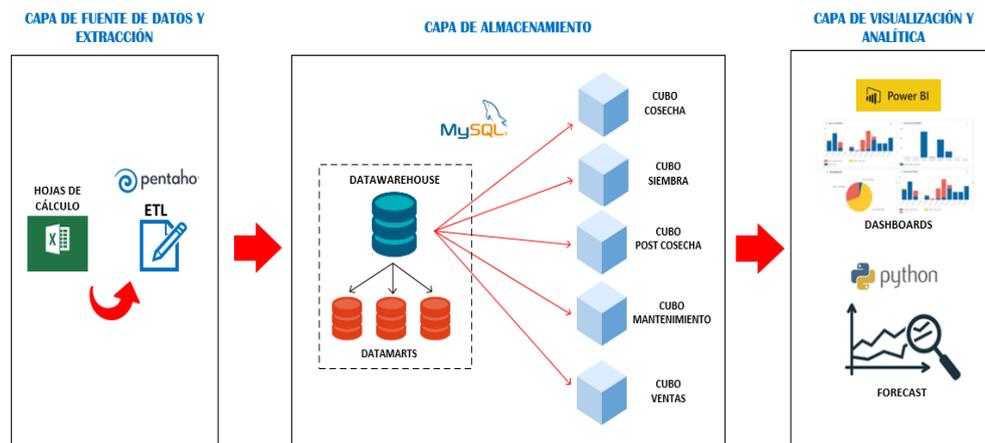


Figura 4.3: Arquitectura de 3 capas EVARROZ

Fuente: Autores

De acuerdo a la figura 4.3, la solución de análisis de datos implementada se divide en 3 capas que se relacionan entre sí. La capa de fuente de datos y extracción, capa de almacenamiento y la capa de visualización y analítica.

La capa de fuente de datos contiene el origen de datos en bruto, la cual se obtuvo a partir de la digitalización de la data que se tenía físicamente. Además, antes de pasar por el proceso ETL, se diseñó una plantilla para un registro estructurado de los datos.

En el proceso ETL, se valida que el cliente tenga conexión con el servidor FTP para extraer los archivos que se hayan cargado a dicho servidor. En el caso de que no exista tal conexión, el proceso ETL se suspenderá. Si, al contrario, la conexión se confirma como exitosa, el proceso ETL obtiene los archivos que están almacenados en el servidor FTP y los mueve a un directorio local para continuar con el siguiente paso. Este proceso es una tarea programada que se ejecuta cada 8 horas (este valor es modificable).

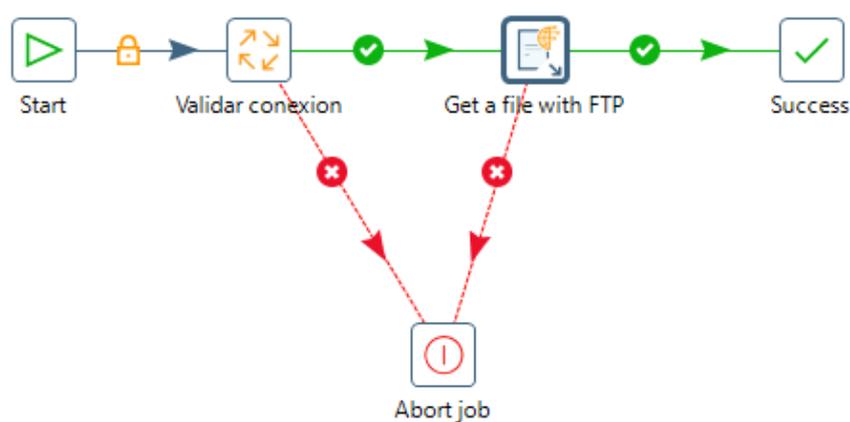


Figura 4.4: Job “Obtener de FTP”

Fuente: Autores

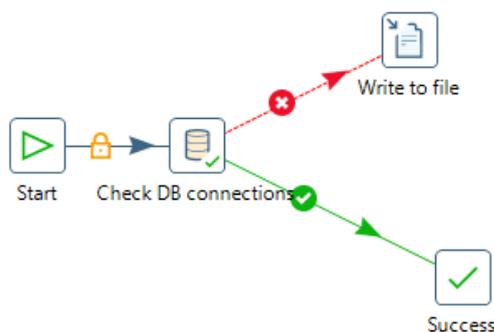


Figura 4.5: Validar conexión

Fuente: Autores

Luego de mover los archivos al directorio local, se procesan uno a uno hasta que el directorio quede vacío. En este proceso se seleccionan los datos de la plantilla alojados en el directorio local, se renombran las cabeceras de las tablas, esto para cumplir la estructura de una base de datos y armar el modelo dimensional que se trata en la siguiente capa.

En el caso de que los archivos tengan alteraciones en la estructura definida en la plantilla, se cancela el job ETL. El diseño de la plantilla se adjunta en el anexo 1.

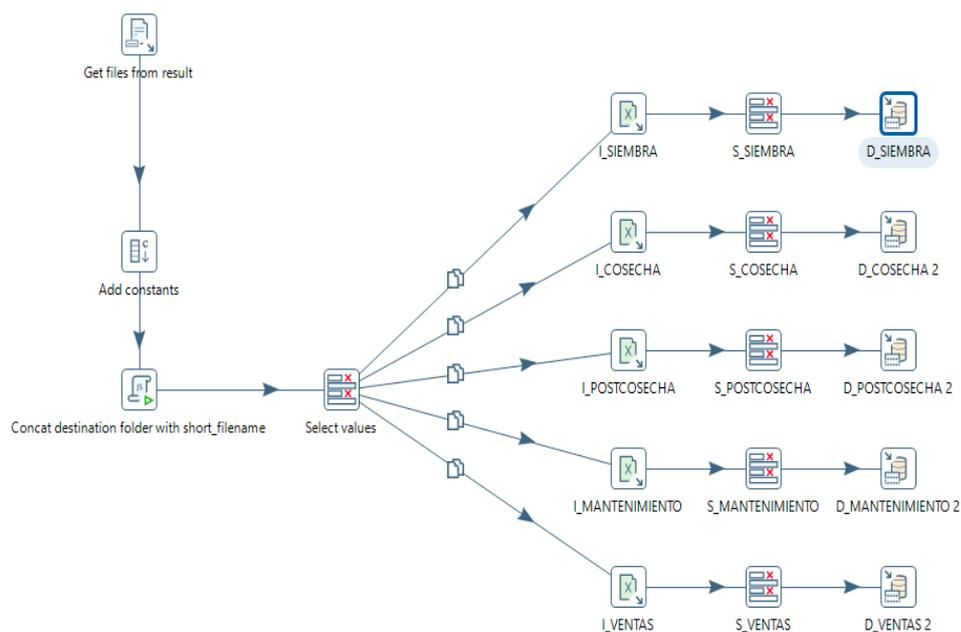


Figura 4.6: Armado de los data marts a partir de las hojas de cálculo

Fuente: Autores

En la capa de almacenamiento, luego de haber extraído cada hoja de la plantilla pre establecida por medio de procesos de transformación en el software de Data Integration de Pentaho, se obtuvieron las dimensiones de tiempo de los data marts para formar la tabla de dimensión correspondiente.

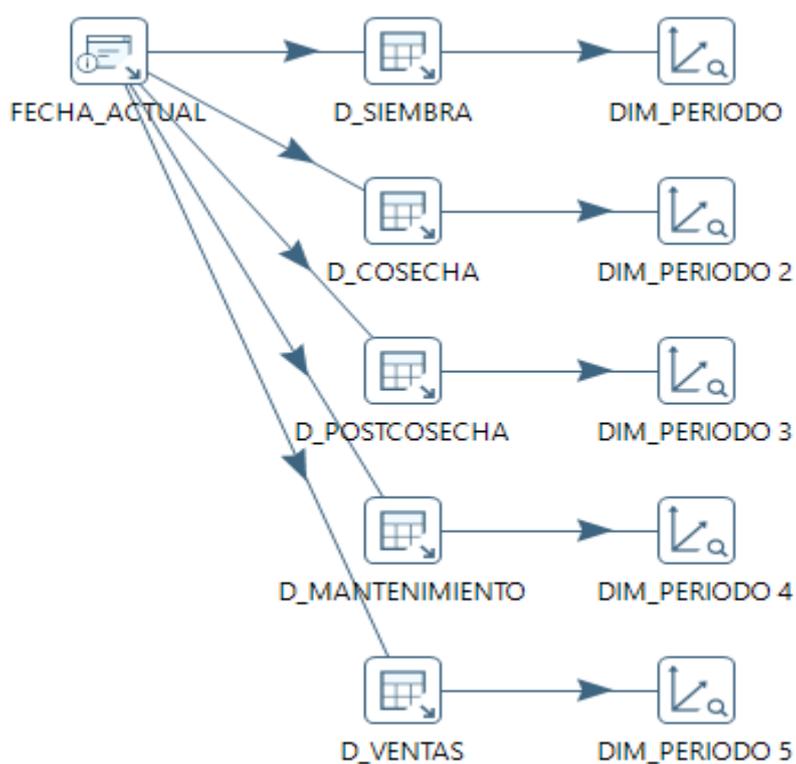


Figura 4.9: Armado de la dimensión temporal del cubo

Fuente: Autores

Luego de haber establecido la tabla de dimensión temporal, se relaciona las medidas (indicador de desempeño) con la dimensión correspondiente en la tabla de hechos.

Pentaho nos permite realizar el cálculo de los indicadores de manera automática posterior a la carga de los datos.

El cálculo de cada indicador, se realiza en base a las tablas de indicadores que se definieron anteriormente.

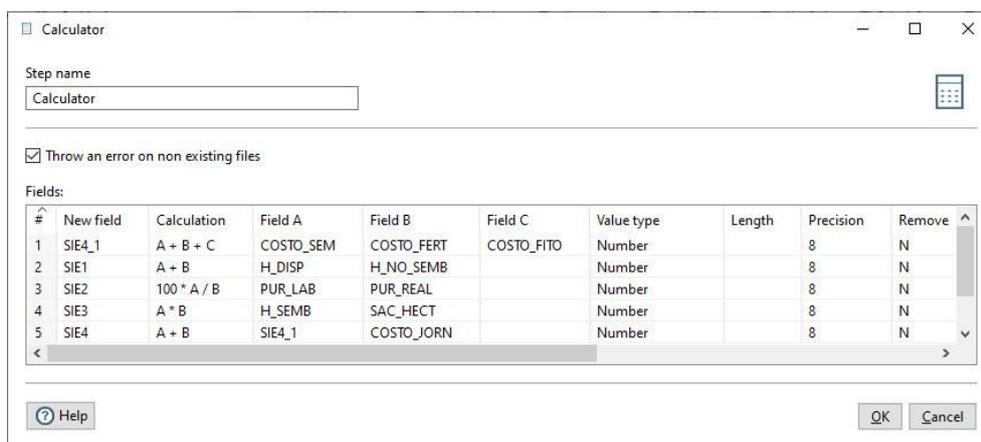


Figura 4.10: Calculo de indicadores: Fase siembra

Fuente: Autores

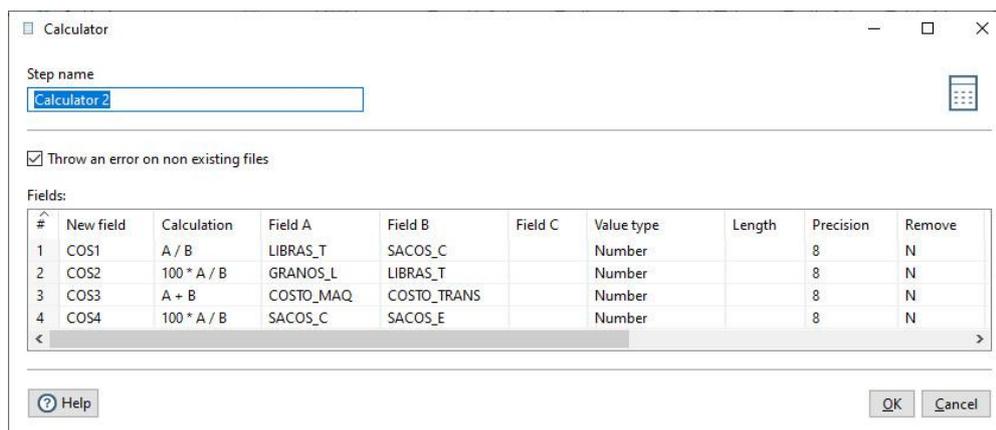


Figura 4.11: Calculo de indicadores: Fase Cosecha

Fuente: Autores

Calculator

Step name
Calculator 3

Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove
1	POS2	A * B	P_AB	SACO_CCC		Number		8	N
2	POS3	A * B	P_AI	SACO_CCC		Number		8	N
3	POS4	A * B	P_P	SACO_SCC		Number		8	N
4	POS5	Create a copy of field A	COSTO_ENS			Number		8	N

Help OK Cancel

Figura 4.12: Calculo de indicadores: Fase post cosecha

Fuente: Autores

Calculator

Step name
Calculator 4

Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove
1	MAN1	A + B + C	COSTO_I_C	COSTO_I_PC	COSTO_I_S	Number		8	N
2	MAN2	A + B	COSTO_MAQ	COSTO_OPER		Number		8	N
3	MAN3	A + B + C	COSTO_AGRO	COSTO_ADM	COSTO_VEN	Number		8	N
4	MAN4	A + B	COSTO_INS	COSTO_MAN		Number		8	N

Help OK Cancel

Figura 4.13: Calculo de indicadores: Fase mantenimiento

Fuente: Autores

Calculator

Step name: Calculator 5

Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove
1	VEN1	Create a copy of field A	SV_BLANCO			Number	8		N
2	VEN2	Create a copy of field A	SV_DERIVADOS			Number	8		N
3	VEN3	Create a copy of field A	SV_INTEGRAL			Number	8		N
4	VEN4_1	A * B	P_BLANCO	SV_BLANCO		Number	8		N
5	VEN4_2	A * B	P_DERIVADOS	SV_DERIVADOS		Number	8		N
6	VEN4_3	A * B	P_INTEGRAL	SV_INTEGRAL		Number	8		N
7	VEN4	A + B + C	VEN4_1	VEN4_2	VEN4_3	Number	8		N

Help OK Cancel

Figura 4.14: Calculo de indicadores: Fase ventas

Fuente: Autores

Finalmente, con los cubos ya formados con la dimensión de tiempo, estos sirven de fuente de datos para que se puede integrar con la siguiente capa.

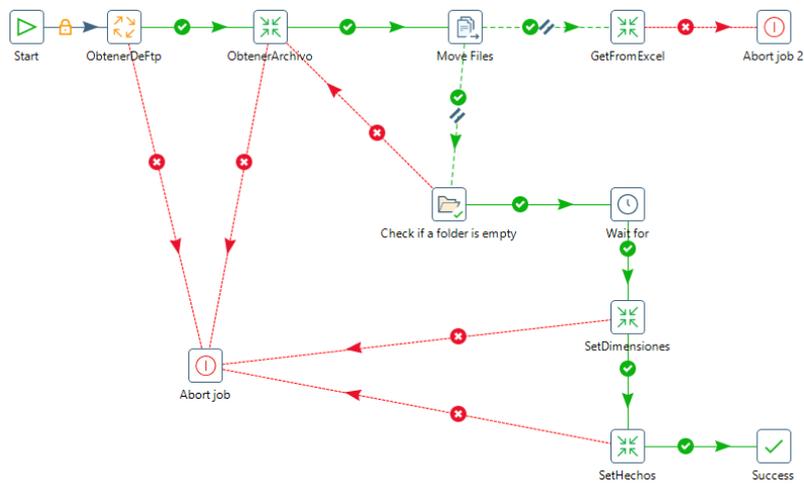


Figura 4.15: Procedimiento completo Capa Fuente de Datos y Extracción, Capa de almacenamiento

Fuente: Autores

En la capa de Visualización y Analítica, Power Bi establece conexión con la base de datos MySQL donde se encuentran las tablas de hechos y dimensiones. En la interfaz de Power Bi, se diseñan los cuadros de mando con la información de cada tabla de hechos y se realiza una proyección de ventas del arroz en dólares versus las hectáreas sembradas por año.

4.4 Implementación de Dashboard BI-DM

De acuerdo a la metodología CRISP-DM incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. [20]

El dashboard principal contiene la información de los KPIs en forma gráfica que se definieron anteriormente. Se muestran las ventas totales, los gastos totales, el margen de utilidad acumulado y la utilidad acumulada por cada periodo y año. Además, el porcentaje de hectáreas sembradas, los sacos cosechados ves los vendidos de acuerdo a las hectáreas disponibles y finalmente, se muestra un resumen de las ventas totales.

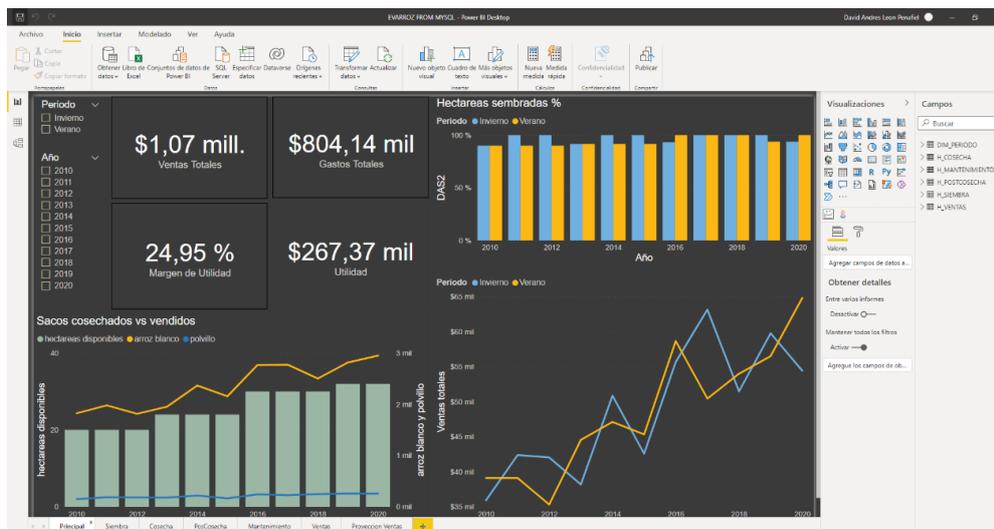


Figura 4.16: Dashboard principal

Fuente: Autores

En las demás dashboards, se muestra una representación gráfica de los KPIs de cada fase del proceso de negocio que se definieron anteriormente.



Figura 4.17: Dashboard "SIEMBRA"

Fuente: Autores

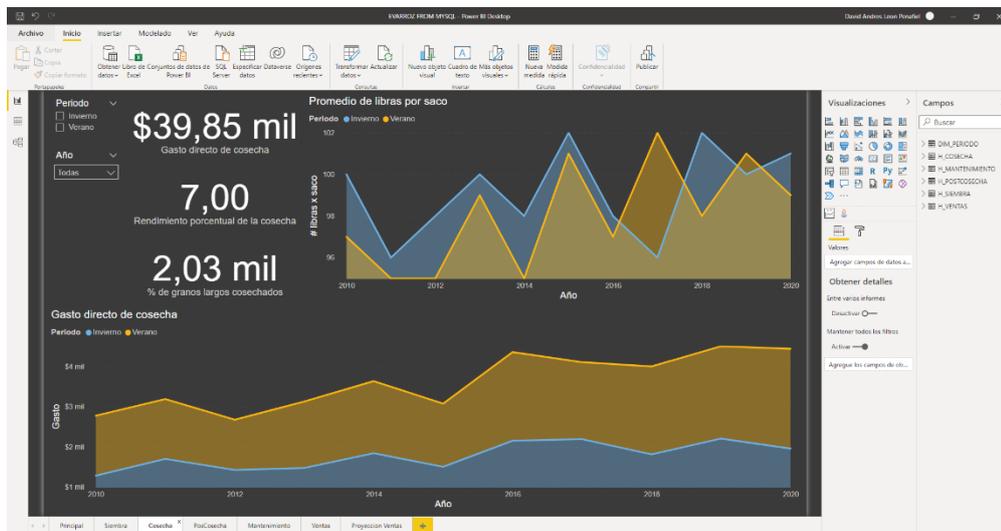


Figura 4.18: Dashboard “COSECHA”

Fuente: Autores

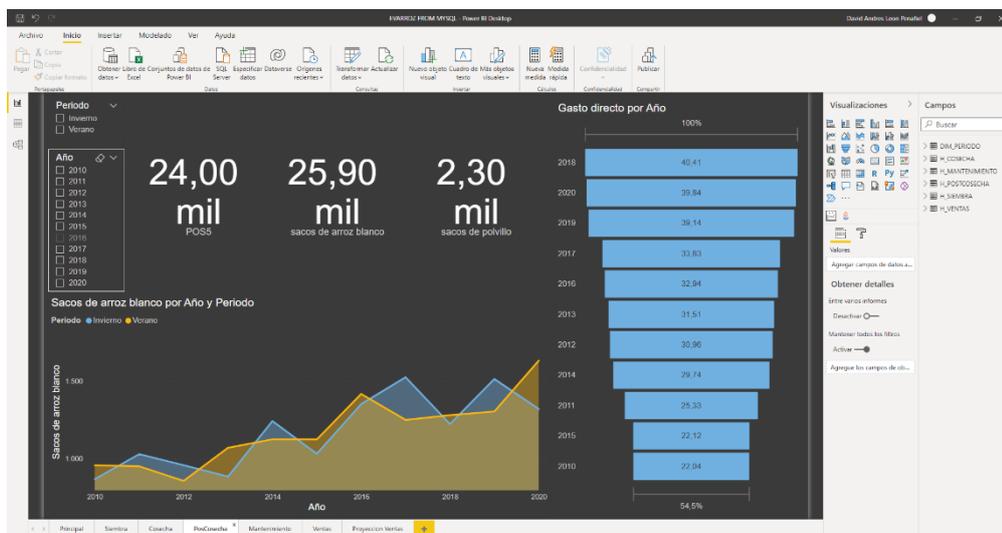


Figura 4.19: Dashboard “POSTCOSECHA”

Fuente: Autores

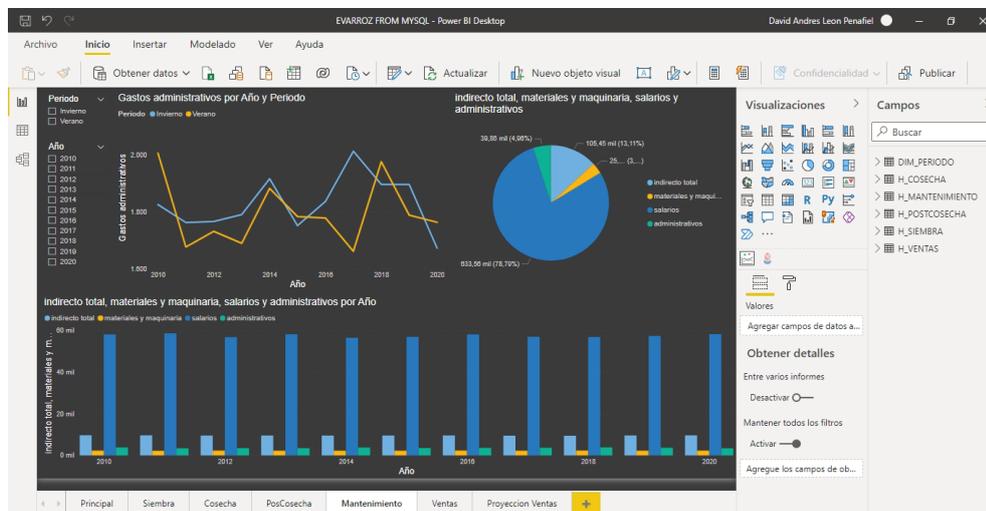


Figura 4.20: Dashboard “MANTENIMIENTO”

Fuente: Autores

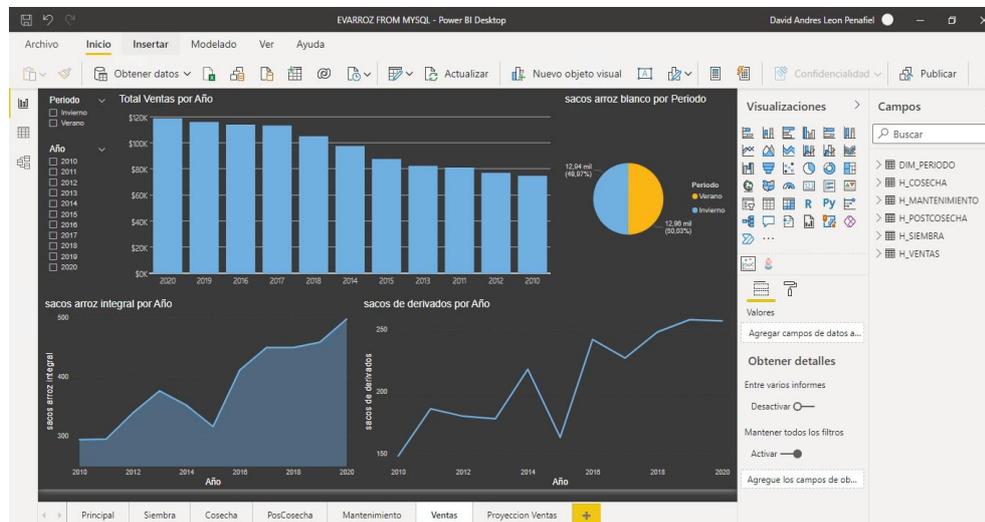


Figura 4.21: Dashboard “VENTAS”

Fuente: Autores

4.5 Implementación del modelo de proyección

Para la predicción de ventas a partir de los datos se usaron técnicas descriptivas de minería a partir de un modelo predictivo empleando regresión lineal para el pronóstico de ventas de arroz blanco versus el número de hectáreas sembradas.

Se usó Python para codificar los algoritmos de datamining y se integró al tablero de control a través de un editor de scripts de Python directamente en el Power Bi.

Para la implementación del modelo de proyección se usaron datos del número de sacos de arroz blanco vendidos y las hectáreas sembradas a partir del 2010. Se decidió optar por estas variables debido a que, por decisión gerencial, se considera que los principales indicadores para determinar buenas ventas es la cantidad de hectáreas sembradas ya que, si la proyección de ventas no es la deseada no se incurriría en un sembrío extenso.

4.5.1 Descripción de las variables

Según de regresión lineal es $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$, donde:

β_0 : es la ordenada en el origen, se corresponde con el valor promedio de la variable respuesta y cuando todos los predictores son cero.

β_j : es el efecto promedio que tiene sobre la variable respuesta el incremento en una unidad de la variable predictor x_j , manteniéndose constantes el resto de variables. Se conocen como coeficientes parciales de regresión.

E: es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo. Recoge el efecto de todas aquellas variables que influyen en y pero que no se incluyen en el modelo como predictores. [26]

Para saber con certeza de que las variables propuestas tienen correlación entre si se usó el test de correlación de Pearson que se calculó de la siguiente forma:

```
# Correlación lineal entre las dos variables
# -----
corr_test = pearsonr(x = dataset['SIE1'], y = dataset['VEN1'])
print("Coeficiente de correlación de Pearson: ", corr_test[0])
print("P-value: ", corr_test[1])
```

Figura 4.22: Código Python: correlación lineal entre dos variables

Fuente: Autores

```
Coeficiente de correlación de Pearson: 0.9528676359813453
P-value: 5.858434677699133e-06
```

Figura 4.23: Resultado de correlación lineal entre dos variables

Fuente: Autores

El test de correlación señala una relación lineal ($r = 0.95$) bastante aceptable ($p\text{-value} = 0.000339$). Se procede con ello generar un modelo de regresión lineal para predecir el número de sacos de arroz blanco en función de las hectáreas sembradas.

4.5.2 Ajuste del modelo

Se ajusta un modelo empleando como variable respuesta el número de sacos de arroz blanco y como predictor las hectáreas sembradas.

La variable dependiente (y) son los sacos de arroz blanco vendidos

La variable independiente (x) son el número de hectáreas sembradas

En el caso del estudio predictivo, es importante cuantificar su capacidad para predecir nuevas observaciones. Para ello se deben dividir los datos en dos grupos uno de test y otro de entrenamiento.

```

# División de los dataset en train y test
# =====
X = dataset['SIE1']
y = dataset['VEN1']

X_train, X_test, y_train, y_test = train_test_split(
    X.values.reshape(-1,1),
    y.values.reshape(-1,1),
    train_size = 0.8,
    random_state = 1234,
    shuffle = True
)

```

Figura 4.24: Código Python: División de los datasets

Fuente: Autores

La implementación de regresión lineal se realiza en Statsmodels, ya que nos permite calcular los test estadísticos y análisis, los cuales son necesarios para analizar que se cumplan las condiciones del modelo.

```

# Creación del modelo utilizando matrices como en scikitlearn
# =====
# A la matriz de predictores se le tiene que añadir una columna de 1s para el intercept del modelo
X_train = sm.add_constant(X_train, prepend=True)
modelo = sm.OLS(endog=y_train, exog=X_train,)
modelo = modelo.fit()

```

Figura 4.25: Código Python: Creación del modelo utilizando matrices

Fuente: Autores

```

=====
OLS Regression Results
=====
Dep. Variable:          y      R-squared:                0.957
Model:                  OLS    Adj. R-squared:           0.950
Method:                 Least Squares  F-statistic:              132.8
Date:                   Mon, 29 Nov 2021  Prob (F-statistic):       2.57e-05
Time:                   14:36:14    Log-Likelihood:          -46.457
No. Observations:      8          AIC:                     96.91
Df Residuals:          6          BIC:                     97.07
Df Model:              1
Covariance Type:       nonrobust
=====
               coef    std err          t      P>|t|      [0.025    0.975]
-----
const         107.7638    203.172     0.530     0.615    -389.380    604.907
x1            163.5831     14.194    11.524     0.000     128.850    198.316
=====
Omnibus:                 2.073    Durbin-Watson:           2.535
Prob(Omnibus):           0.355    Jarque-Bera (JB):        0.822
Skew:                   0.257    Prob(JB):                0.663
Kurtosis:                1.517    Cond. No.                88.9
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Figura 4.26: Resultados de la regresión

Fuente: Autores

Una vez entrenado el modelo, se pueden obtener predicciones para nuevos datos, statmodels nos devuelve las predicciones y los

```

# Intervalos de confianza para los coeficientes del modelo
# =====
modelo.conf_int(alpha=0.05)

# Predicciones con intervalo de confianza del 95%
# =====
predicciones = modelo.get_prediction(exog = X_train).summary_frame(alpha=0.05)
predicciones.head(4)

# Predicciones con intervalo de confianza del 95%
# =====
predicciones = modelo.get_prediction(exog = X_train).summary_frame(alpha=0.05)
predicciones['x'] = X_train[:, 1]
predicciones['y'] = y_train
predicciones = predicciones.sort_values('x')

```

intervalos de confianza.

Figura 4.27: Ajuste de coeficientes, intervalo de confianza y predicciones

Fuente: Autores

En el gráfico, se incluye los límites superior e inferior del intervalo de confianza, para indicar el rango en el que se encontraría nuestra variable de respuesta.

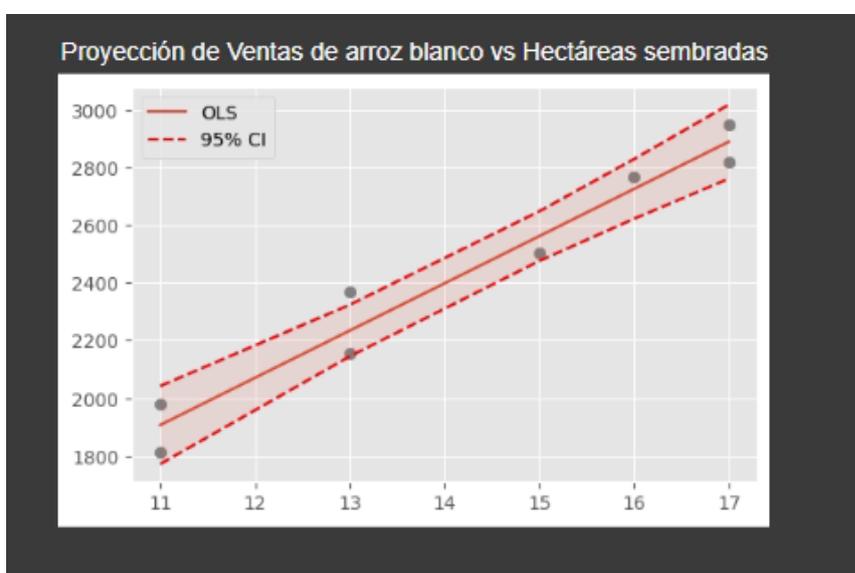


Figura 4.28: Gráfico de proyecciones integrado en el Dashboard BI-DM

Fuente: Autores

4.5.3 Interpretación

De acuerdo a la figura 4.26, la columna (coef) devuelve el valor estimado para los dos parámetros de la ecuación del modelo lineal (β^0 y β^1) que equivalen a la ordenada en el origen (intercept o const) y a la pendiente.

El valor de R-squared indica que el modelo es capaz de explicar el 95.7% de la variabilidad observada en la variable respuesta (ventas de arroz blanco).

El modelo de regresión lineal generado sigue la ecuación:

Número de arroz blanco = $107.7638 + 163.5831(\text{número de hectáreas sembradas})$. Por cada hectárea sembrada que se incrementa, el número de sacos de arroz blanco vendido aumenta en promedio de 163.5831.

CAPÍTULO 5

PRUEBAS Y ANÁLISIS DE RESULTADOS

En este capítulo se describirá la forma en la que el usuario interactúa con el sistema, errores del sistema que puedan ocurrir al momento de usar el sistema EVARROZ. Además, se medirá el nivel de satisfacción de los usuarios con respecto al uso del sistema y su utilidad. Finalmente, se enunciarán los resultados de la implementación y desempeño del sistema.

5.1 Descripción del entorno de usuario

Con el sistema ya implementado en un servidor local, al cual tendrán acceso sólo aquellas personas que deban subir la data y puedan visualizar la información. A estos, se les proporcionará un usuario y contraseña, los cuales deberán ingresar en la siguiente pantalla.

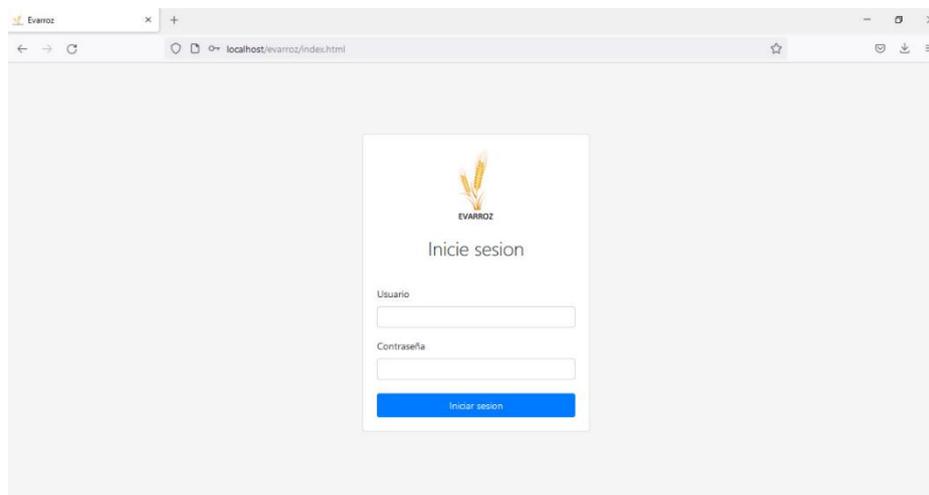


Figura 5.1: Pantalla de inicio “Sistema EVARROZ”

Fuente: Autores

Luego de haber ingresado las credenciales correctamente, el sistema direcciona a la siguiente pantalla donde el usuario tiene la opción de visualizar los dashboard o subir datos. Esta segunda opción permite subir la plantilla o cargar una plantilla nueva.

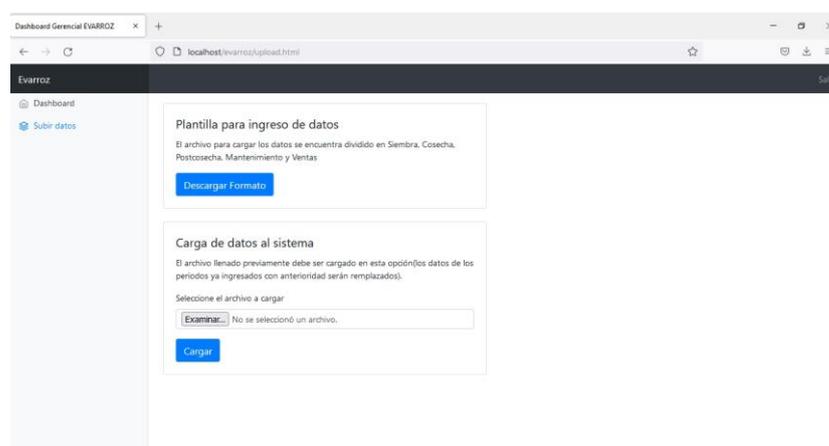


Figura 5.2: Pantalla de opciones: dashboard o subir datos

Fuente: Autores

Si se selecciona “Plantilla para ingreso de datos”, se descarga el archivo llamado “INPUT”.

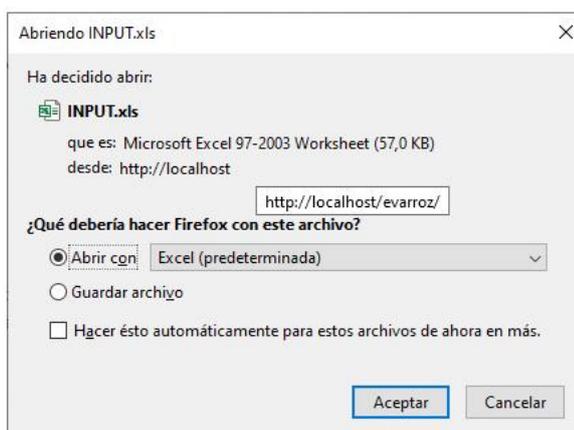


Figura 5.3: Descarga de plantilla

Fuente: Autores

Para cargar una nueva plantilla al sistema, se debe seleccionar la opción “Carga de datos al sistema”. Se debe seleccionar el archivo el archivo descargado. Se realizó la validación para que el nombre del archivo no tenga relevancia.

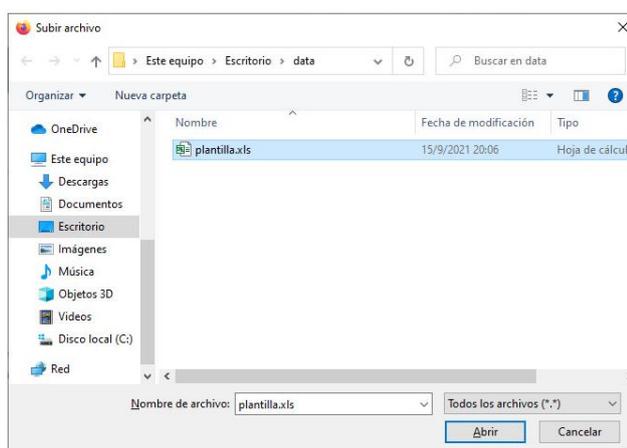


Figura 5.4: Búsqueda local y carga de datos

Fuente: Autores

Por el contrario, si se desea visualizar los dashboard, se debe seleccionar la opción “Dashboard”.

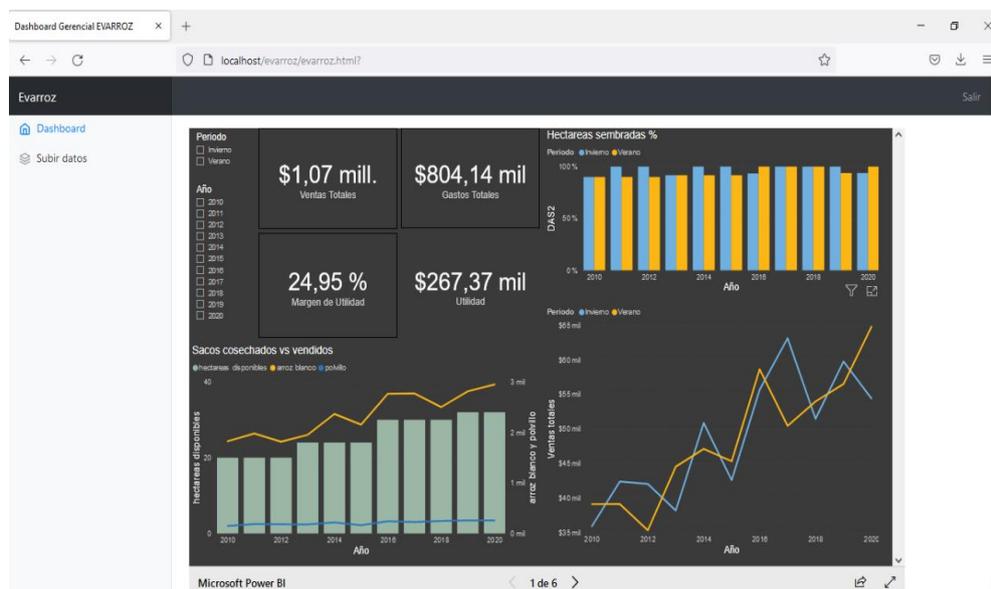


Figura 5.5: Visualización de dashboard Gerencial EVARROZ

Fuente: Autores

5.2 Casos de Uso

Para validar el correcto funcionamiento del sistema se realizaron junto a usuarios seleccionados varias pruebas de uso. El primer caso de uso fue la validación de las credenciales. En este caso, se intentó ingresar credenciales no válidas y el sistema envió un mensaje de alerta para este caso.

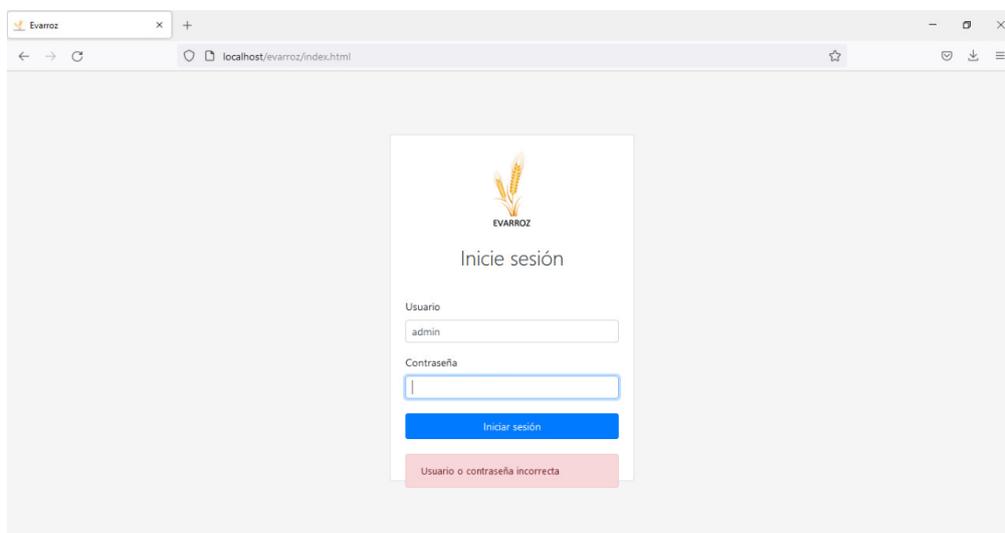


Figura 5.6: Validación de credenciales

Fuente: Autores

Un segundo caso de uso fue intentar la carga de un archivo que no es el apropiado. El sistema internamente; a través del proceso ETL, valida que el archivo cargado sea el apropiado y no permite la carga.

Otro caso de uso es subir data no correcta en una plantilla. En este caso, dado que no existe una herramienta que valide que la información sea coherente, se debería poder modificar subiendo nuevamente datos correctos del año y periodo correspondiente.

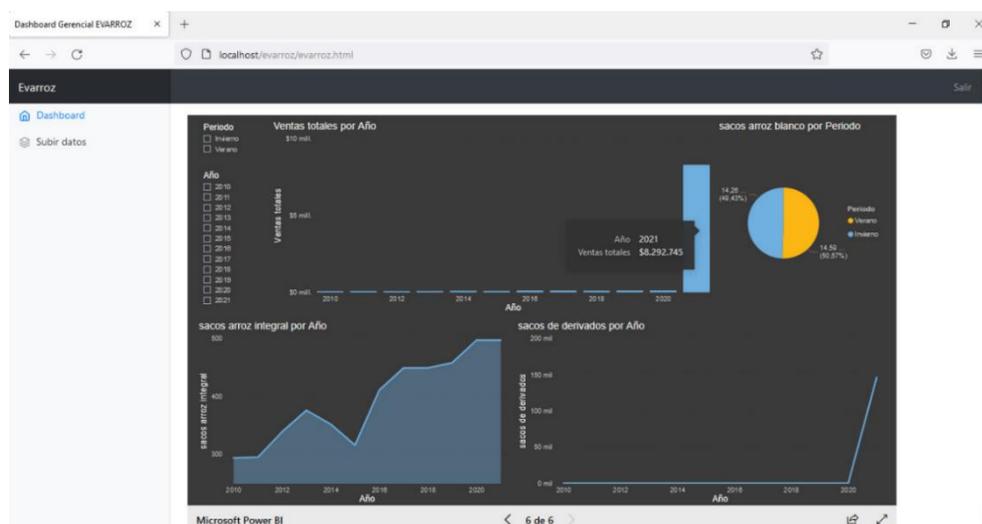


Figura 5.7: Carga de información no coherente

Fuente: Autores

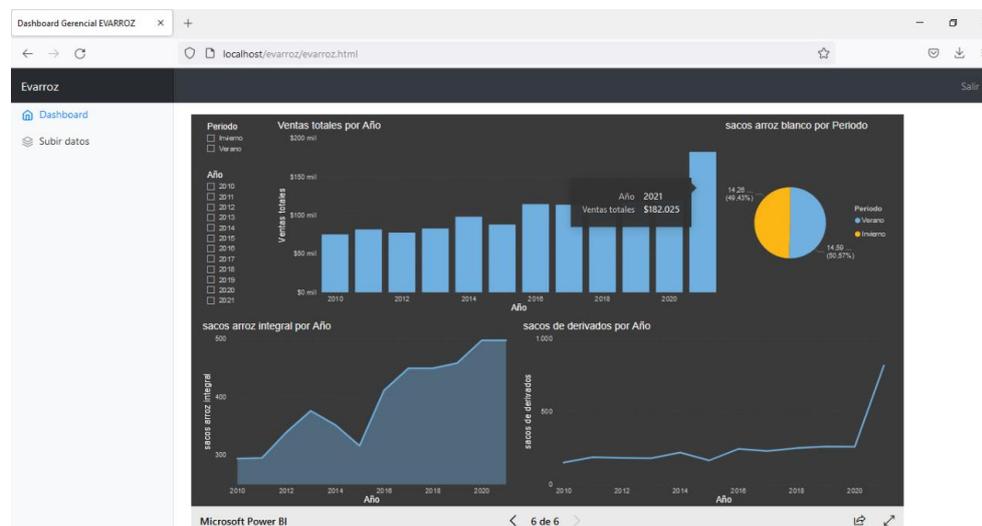


Figura 5.8: Modificación de información

Fuente: Autores

Finalmente, se validó que al subir data de otros años, el sistema actualice la información mostrada y se pueda visualizar en el dashboard lo cargado.

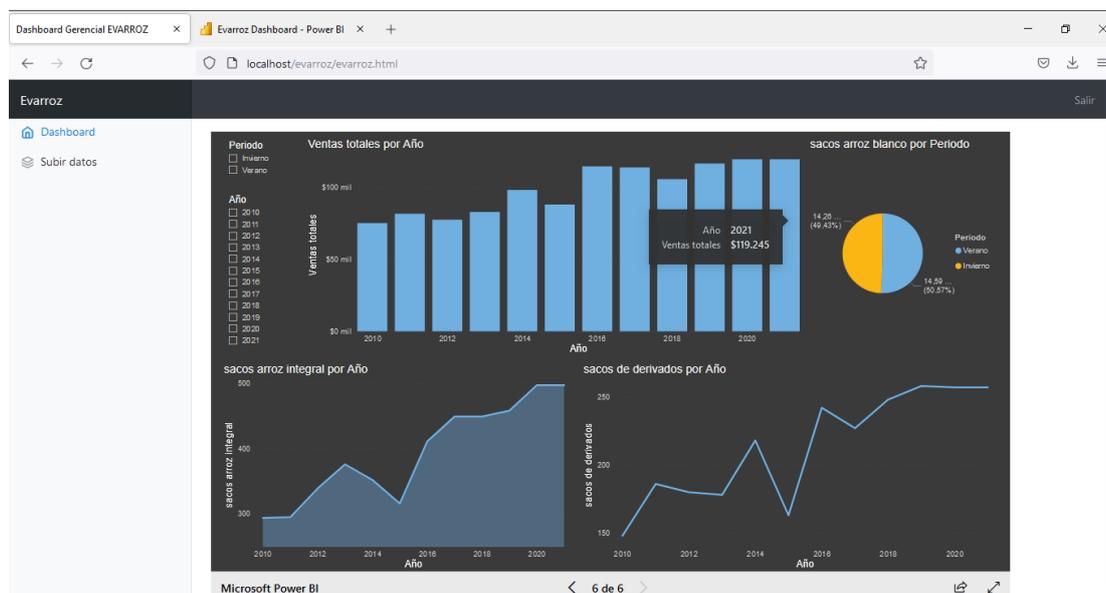


Figura 5.9: Actualización de información en el dashboard gerencial

Fuente: Autores

Para este caso se subió información de ventas del 2021.

5.3 Cumplimiento de sistema

Después de haber implementado el sistema, corregido errores y modificados parámetros, capacitado al personal que usará el sistema y monitorear el desempeño, se obtuvieron las siguientes mediciones obtenidas de la encuesta que se adjunta en el anexo 2, acerca del nivel de satisfacción del usuario con respecto a información general y uso del sistema.

Se realizaron once consultas donde se seleccionaron a ocho personas dentro de la empresa entre los cuales incluían directivos, gerencias y personal operativo.

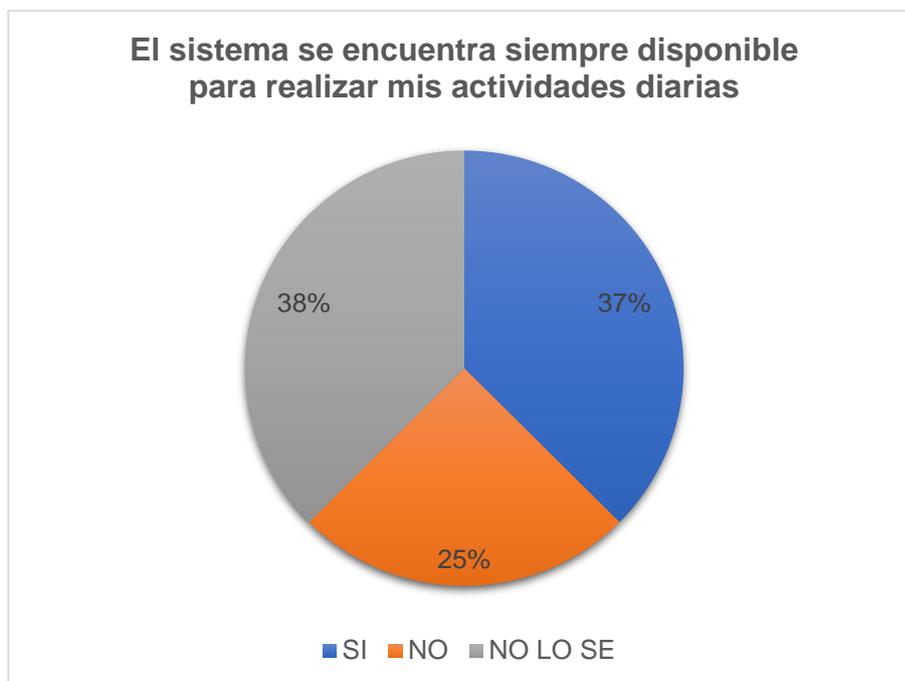


Figura 5.10: Estadística de consulta 1

Fuente: Autores

Como se muestra en la figura 5.10; el 37% de los participantes afirmaron que el sistema estaba siempre disponible. En contra punto, el 25% afirmó que no lo estaba. Esto se puede explicar a partir de las externalidades que pueden existir al momento de disponer del sistema. Como, por ejemplo, fallo en el flujo eléctrico en la empresa, fallas con el hardware que alberga el sistema, problemas humanos al momento de ejecutar el sistema, etc.

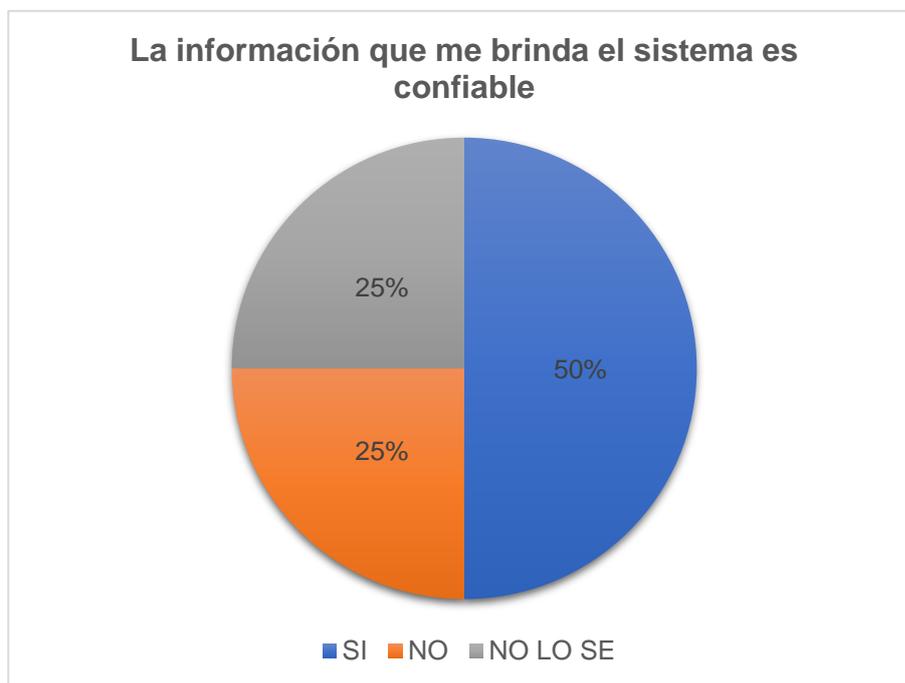


Figura 5.11: Estadística de consulta 2

Fuente: Autores

Como se muestra en la figura 5.11; el 50% de los participantes afirmaron que el sistema brinda información confiable. Se le consultó al 25% de los participantes que respondieron "NO LO SE" cuál fue la razón de su respuesta. Estos mencionaron que no sabían interpretar la información de los gráficos a pesar de haber recibido capacitación previa.

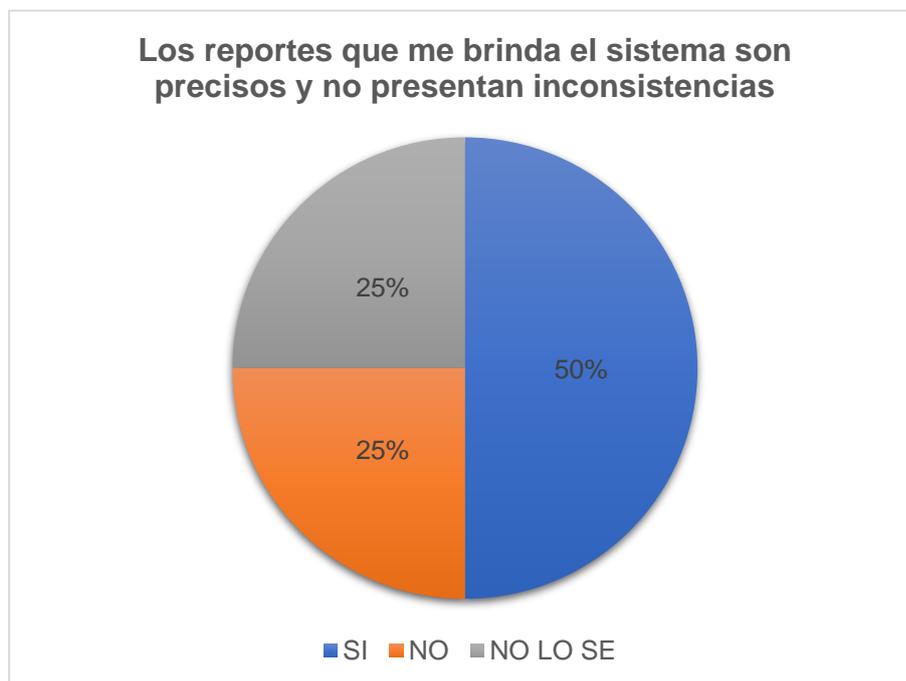


Figura 5.12: Estadística de consulta 3

Fuente: Autores

Como se muestra en la figura 5.12; el 50% de los participantes afirmaron que el sistema brinda reportes precisos y sin inconsistencias. La mitad restante no estaba segura de la fidelidad de la información ya que dudaban de que al momento de cargar esta, no estaba seguros de haberlo hecho correctamente.



Figura 5.13: Estadística de consulta 4

Fuente: Autores

Como se muestra en la figura 5.13; solo el 37% de los participantes afirmaron que el sistema ofrece las funciones necesarias de acuerdo a las necesidades de la empresa. Esta estadística es particularmente llamativa debido a que para elegir los KPIs se hizo un análisis de requerimientos.

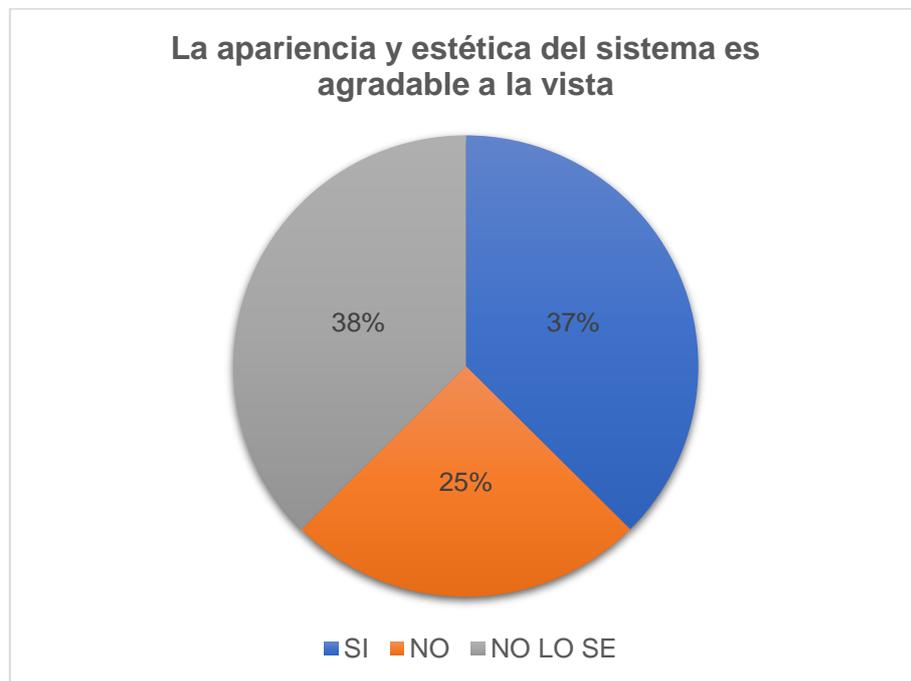


Figura 5.14: Estadística de consulta 5

Fuente: Autores

Como se muestra en la figura 5.14; el 50% de los participantes no estaban conformes con la apariencia de la interfaz del sistema. Aseguraron que con el uso del mismo se acostumbrarían a la estética del sistema.

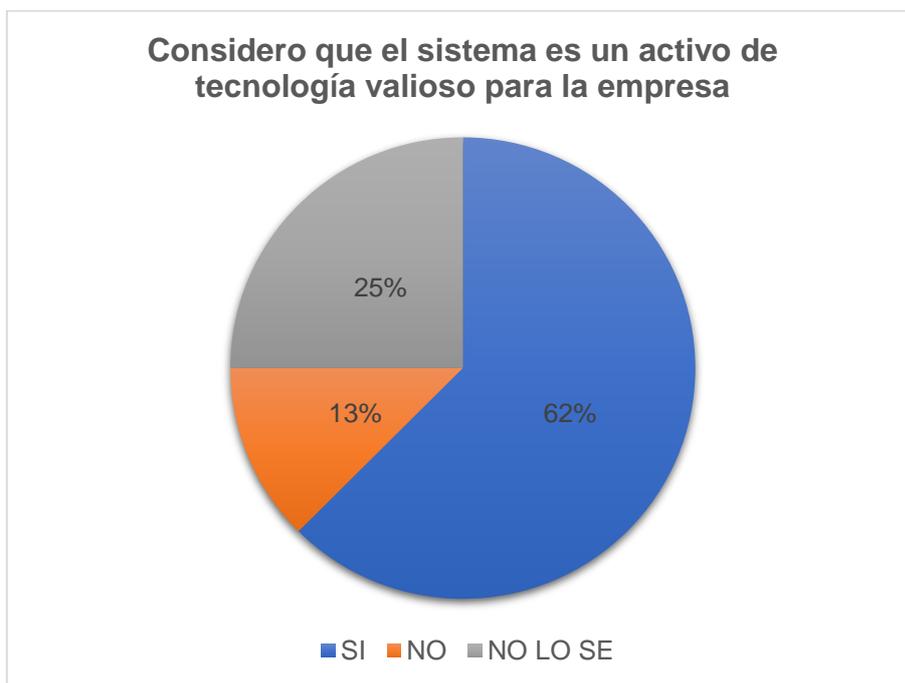


Figura 5.15: Estadística de consulta 6

Fuente: Autores

Como se muestra en la figura 5.15; el 62% de los participantes considera que el sistema es valioso para la empresa. Esta estadística es alentadora ya que se podría inferir que, en el futuro cercano, los usuarios podrían valorar de mejor forma el sistema.

Luego de interpretar las estadísticas relacionadas con la información general del sistema, se hará lo mismo con las estadísticas relacionadas con el uso del sistema.

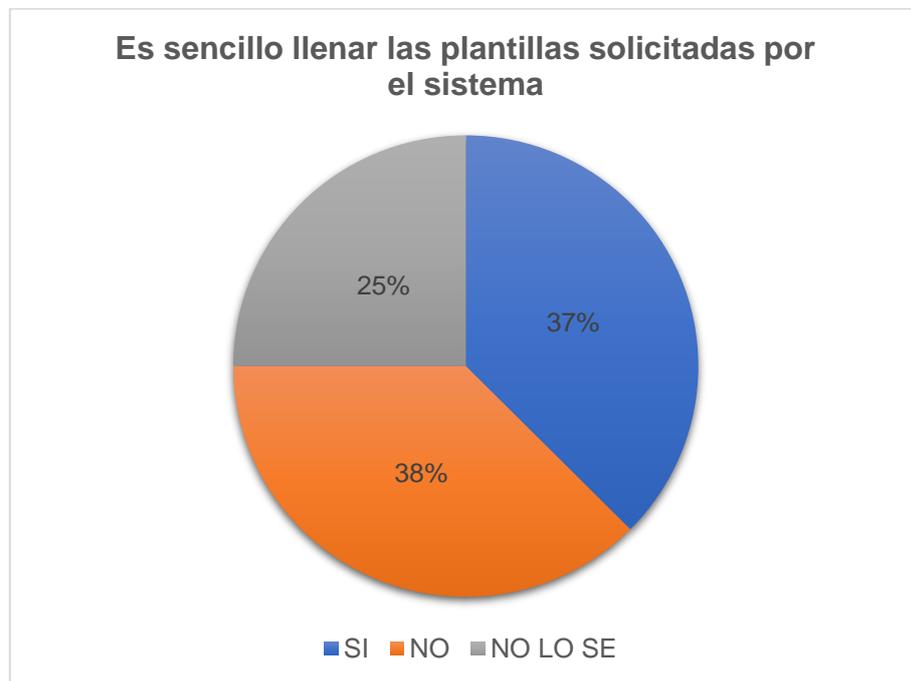


Figura 5.16: Estadística de consulta 7

Fuente: Autores

Como se muestra en la figura 5.16; el 38% de los participantes considera que es sencillo llenar las plantillas. Esto es similar a lo antes mencionado con el uso del sistema. Esto quiere decir que, con el uso prolongado se mejorará esta estadística.

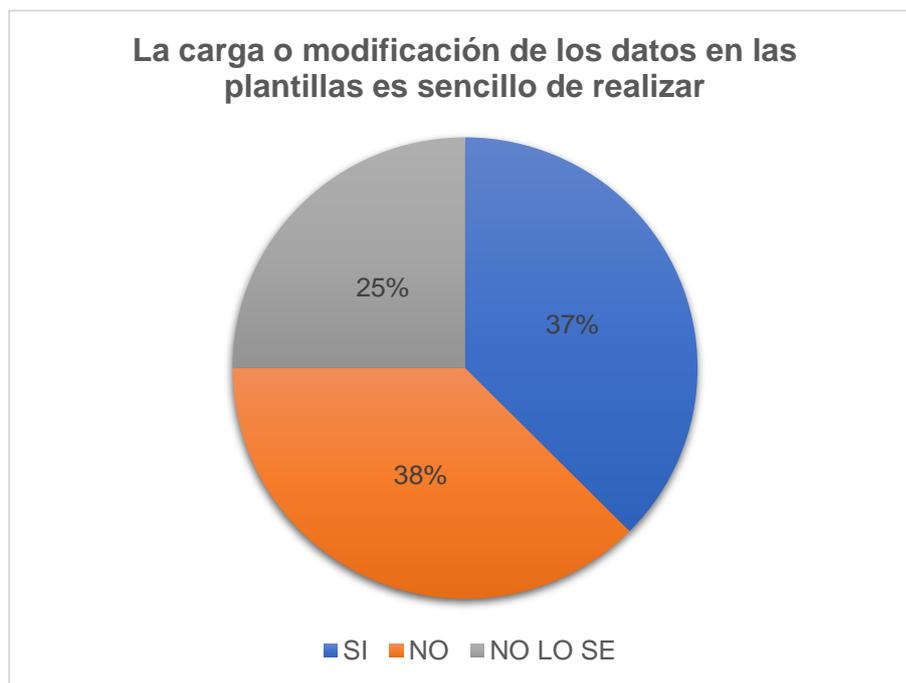


Figura 5.17: Estadística de consulta 8

Fuente: Autores

Similar a la estadística anterior, lo que muestra la figura 5.17 está relacionado con el poco uso del sistema. Con el tiempo esta estadística podría mejorar.

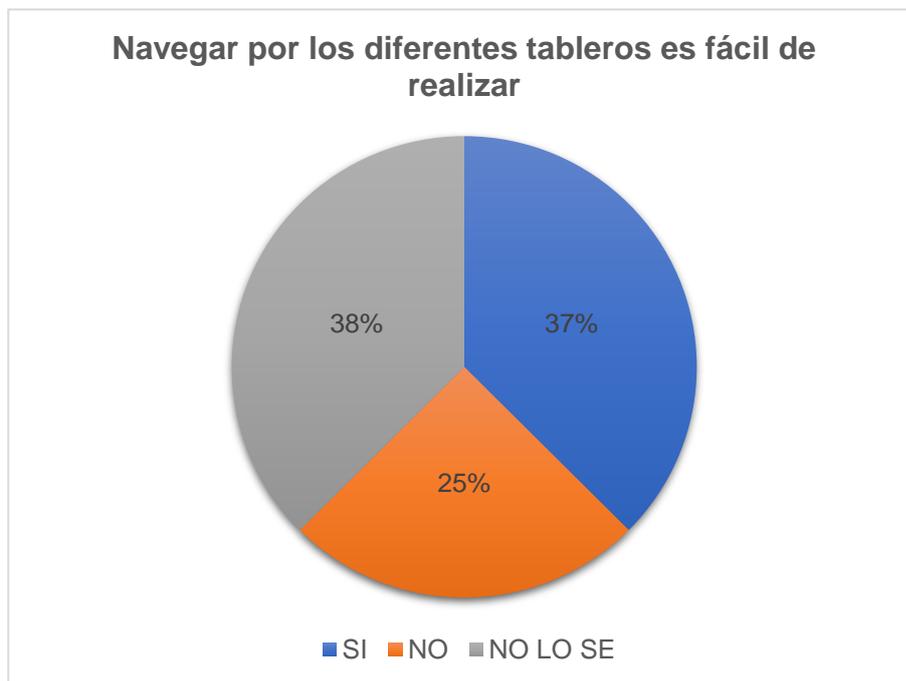


Figura 5.18: Estadística de consulta 9

Fuente: Autores

Nuevamente se evidencia lo antes indicado. La figura 5.18 muestra que solo el 37% de los participantes navega con facilidad por el tablero.



Figura 5.19: Estadística de consulta 10

Fuente: Autores

Como indica la figura 5.19, el 62% de los participantes está satisfecho con el tiempo de respuesta de la aplicación. No hubo participantes que renegaran por el tiempo de respuesta del sistema.



Figura 5.20: Estadística de consulta 11

Fuente: Autores

Como indica la figura 5.20, el 87% de los participantes cree necesario tener una persona que de soporte al sistema y a los usuarios. Esto es muy coherente con las estadísticas antes vistas. Claramente existe la necesidad de mayor capacitación en el uso del sistema y tiempo de acompañamiento

5.4 Resultados de la implementación

En definitiva, la implementación del sistema EVARROZ se puede considerar un éxito parcial. Esto se vio corroborado en las encuestas de satisfacción y la nueva forma de cargar, actualizar datos y generar información relevante.

Tanto la gerencia, directivos como personal operativo notaron una mejora significativa al momento de disponer de la información y visualizarla.

Sin embargo, el proceso de generación de informes no pudo ser completamente automatizado debido a que, se requieren mayor cantidad de tiempo y recursos para implementar sistemas y herramientas que permitan obtener la información de todas las áreas de negocio.

A pesar de esto, el primer paso está dado hacia la completa automatización.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

1. En este trabajo se ha realizado la implementación de un sistema de soporte a decisión para una empresa de producción y comercialización de arroz. Dicho sistema cumple con las demandas y requerimientos de los usuarios de EVARROZ. El sistema brinda fácil acceso a los datos más importantes, de acuerdo a los parámetros establecidos a través de tableros de control y gráficos.
2. El sistema EVARROZ incluye un modelo de proyección a partir de regresión lineal, el cual fue construido a partir de la metodología CRISP-DM. Dicha metodología permitió trabajar los datos, estructurarlos y presentarlos de forma orgánica facilitando tanto en el desarrollo del sistema como en el desarrollo del proyecto.

3. A pesar de que se insistió muy enfáticamente en la adopción del nuevo proceso de negocio como política empresarial y que incluía una mejora en el proceso de generación de informes, se notó que, muy probablemente no se sigan los procesos establecidos en el diagrama TO-BE. Esto será debido a que, no se tiene una cultura empresarial que apoye el cambio que la empresa necesita a mediano plazo.
4. Finalmente, se lograron cumplir completamente dos de los cuatro objetivos específicos establecidos al inicio de este trabajo. Sin embargo, no se logró automatizar completamente el análisis y diseño de la información debido ni analizar los resultados del nuevo proceso debido a falta de tiempo y recursos. Por lo cual, podemos afirmar que estos dos objetivos específicos se lograron cumplir parcialmente.

RECOMENDACIONES

1. Concientizar a los trabajadores sobre las buenas prácticas en el cumplimiento de procesos, con el fin de aumentar la productividad de los mismos.
2. Crear una dimensión adicional en los data marts que contenga datos sobre los productos que se venden en EVARROZ. De esta forma, poder integrar esta información con las demás áreas del proceso de negocio y facilitar la visibilidad de la información.
3. Invertir en infraestructura tecnológica, tanto de hardware como de software. A pesar de que el giro de negocio de EVARROZ es la producción y comercialización, esta no está exenta a cambio de paradigma en la forma de hacer negocios. Una mayor injerencia de herramientas como un ERP (Sistema de planificación de recursos empresariales), ayudaría a cumplir completamente el objetivo de automatizar y analizar la obtención de información.
4. Incrementar la cantidad de datos que se carguen al sistema. Esto con la finalidad de aumentar la precisión del modelo de proyección de ventas y poder disponer otro modelo de proyección como regresión polinomial o series temporales que valide lo proyectado por el modelo de regresión lineal.

BIBLIOGRAFÍA

- [1] Garcés Fiallos, F. R., Aguirre Calderón, Ángel J., & Díaz Coronel, «<https://revistas.uteq.edu.ec>,» 23 1 2013. [En línea]. Available:<https://revistas.uteq.edu.ec/index.php/cyt/article/view/125>
- [2] M. Ing. Freddy Gavilánez Luna, «<http://archivo.uagraria.edu.ec>,» 8 1 2016. [En línea]. Available: http://archivo.uagraria.edu.ec/web/revistas_cientificas/10/030-2016.pdf.

[3] INEC, «www.ecuadorencifras.gob.ec,» 2018. [En línea]. Available: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_agropecuarias/espac/espac-2018/Presentacion%20de%20principales%20resultados.pdf.

[4] INIAP, «<http://tecnologia.iniap.gob.ec>,» 2019. [En línea]. Available: <http://tecnologia.iniap.gob.ec/index.php/explore-2/mcereal/rarroz>.

[5]CFN, «<https://www.cfn.fin.ec/wp-content/uploads/2018/04/Ficha-Sectorial-Arroz.pdf>,» CFN, Febrero 2018. [En línea].

[6]M.A.S.Crespo, «<https://www.elmundo.com/portal/pagina.general.impresion.php?idx=278178>,» El Mundo, 2016. [En línea].

[7]Wikipedia, https://es.wikipedia.org/wiki/Sistemas_de_soporte_a_decisiones,» Wikipedia, Julio 2021. [En línea].

[8] D. L. Cesar Hernandez, D. Lyon, A. Pérez, G. Nebiolo, J. García y L. Sosa, «<https://jmpovedar.files.wordpress.com>,» Universidad Simón Bolívar, Marzo 2014. [En línea]. Available:

<https://jmpovedar.files.wordpress.com/2014/03/tipos-de-sistemas-de-informacion.pdf>.

[9] A. y. Santos, «www.sngular.com,» 2008. [En línea]. Available: <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>.

[10] J. A. Gallardo, «<http://www.oldemarrodriguez.com>,» 2000. [En línea]. Available: http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037.

[11] J. A. Gallardo, «www.oa.upm.es,» Universidad UPM, 2009. [En línea]. Available: https://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf.

[12] T. Vilariño y J. M. De La Torre, «www.dspace.uclv.edu.cu,» 28 062014. [En línea]. Available: <https://dspace.uclv.edu.cu/handle/123456789/6621>.

[13] A.E.CALDERÓN, «<http://repositorio.puce.edu.ec/>,» PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR, Enero 2015. [En línea]. Available: <http://repositorio.puce.edu.ec/handle/22000/8840>.

[14] M. A. C. Riaño, «Gestión de procesos con BPM,» Tecnología, Investigación y Academia - Universidad Distrital Francisco José de Caldas, vol. 1, p. 12, 2016.

[15] R. F. Oltra Badenes, «<https://riunet.upv.es/>,» 2015. [En línea]. Available:<https://riunet.upv.es/bitstream/handle/10251/84471/Oltra%20%20Business%20Intelligence.%20Definici%C3%B3n.pdf?sequence=1>.

[16] S. Ramos, «<http://www.solidq.com/>,» 2016. [En línea]. Available:http://www.solidq.com/ebs/DataWarehouse_DataMarts_ModelosDimensionales_v2.pdf.

[17] F. J. Z. Barreno, «<http://repositorio.utn.edu.ec/>,» Agosto 2018. [En línea]. Available:<http://repositorio.utn.edu.ec/bitstream/123456789/8898/1/04%20ISC%20498%20TRABAJO%20DE%20GRADO.pdf>.

[18] M. I. Alfonso Galipienso, A. Botía Martínez, A. Mora Lizán y J. Tigreros Jover, Ingeniería de Software, Madrid, 2005.

[19] B. Hitpass, Business Process Mangement, 2017.

[20] IBM, «<https://www.ibm.com>,» 2021. [En línea]. Available: <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=dm-crisp-help-overview>.

[21] C. F. N. d. Ecuador, «<https://www.cfn.fin.ec>,» [En línea]. Available: <https://www.cfn.fin.ec/wpcontent/uploads/2018/04/Ficha-Sectorial-Arroz.pdf>.

[22] F. R. G. Fiallos, «www.uteq.edu.ec,» 23 Enero 2013. [En línea]. Available: <https://revistas.uteq.edu.ec/index.php/cyt/article/view/125>. [Último acceso: 06 2020].

[23] J. M. G. C. M. C. C. R. F. M. A. Freddy Gavilánez Luna, «www.uagraria.edu.ec,» 02 2016. [En línea]. Available: http://archivo.uagraria.edu.ec/web/revistas_cientificas/10/030-2016.pdf.

[24] INEC, «www.ecuadorencifras.gob.ec,» ABRIL 2019. [En línea]. Available: https://www.ecuadorencifras.gob.ec/documentos/web-inec/Estadisticas_agropecuarias/espac/espac-2018/Presentacion%20de%20principales%20resultados.pdf.

[25] INEC, «www.iniap.gob.ec,» 2020. [En línea]. Available: <http://tecnologia.iniap.gob.ec/index.php/explore-2/mcereal/rarroz>.

[26] IDS 702: Module 1.8, Transformations, Dr. Olanrewaju Michael Akande. <https://ids702-f21.olanrewajuakande.com/slides/1-8-MLR-transformations.pdf>

ANEXO 2

Tabla 16: Encuesta de satisfacción de usuario



EVARROZ

SISTEMA DE SOPORTE EN LA TOMA DE DECISIONES EVARROZ

Con el propósito de determinar si el sistema implementado es útil para la empresa pedimos su ayuda dando a conocer su punto de vista sobre el sistema y su uso. Le recordamos que la sinceridad sobre lo consultado es un factor clave para determinar el éxito de las pruebas.

		SI	NO	No lo sé
		1	2	3
1. INFORMACIÓN GENERAL				
1	El sistema se encuentra siempre disponible para realizar mis actividades diarias			
2	La información que me brinda el sistema es confiable			
3	Los reportes que me brinda el sistema son precisos y no presentan inconsistencias			
4	El sistema ofrece las funciones necesarias de acuerdo a las necesidades de la empresa			
5	La apariencia y estética del sistema es agradable a la vista			
6	Considero que el sistema es un activo de tecnología valioso para la empresa			
2. USO DEL SISTEMA				
		1	2	3
7	Es sencillo llenar las plantillas solicitadas por el sistema			
8	La carga o modificación de los datos en las plantillas es sencillo de realizar			
9	Navegar por los diferentes tableros es fácil de realizar			
10	El tiempo de respuesta del sistema es el apropiado			
11	Cree necesario tener una persona especializada para dar soporte a los usuarios y al sistema			