

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



Solución de inteligencia de negocios basada en técnicas de minería de datos, para apoyar la toma de decisiones, en la Gerencia Regional de Agricultura-Lambayeque

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

AUTOR

Omar Erick Valderrama Mundaca

ASESOR

Huiler Juanito Mera Montenegro

<https://orcid.org/0000-0001-6830-5415>

Chiclayo, 2022

Solución de inteligencia de negocios basada en técnicas de minería de datos, para apoyar la toma de decisiones, en la Gerencia Regional de Agricultura-Lambayeque

PRESENTADA POR
Omar Erick Valderrama Mundaca

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

INGENIERO DE SISTEMAS Y COMPUTACIÓN

APROBADA POR

Héctor Miguel Zelada Valdivieso
PRESIDENTE

Segundo José Castillo Zumarán
SECRETARIO

Huiler Juanito Mera Montenegro
VOCAL

Dedicatoria

A Dios y a mis padres, por todo su esfuerzo, dedicación, paciencia y motivación para poder culminar satisfactoriamente esta etapa tan importante, como es el de ser un profesional y una persona de bien.

Agradecimientos

A toda mi familia por ser mi soporte y darme su apoyo incondicional desde siempre.

A mi asesor de tesis, Mg. Ing. Huilder Mera Montenegro, por la experiencia, confianza y dedicación para llevar adelante la culminación de este proyecto de investigación y brindarme su apoyo en todo momento.

A la Gerencia Regional de Agricultura – Lambayeque por darme la oportunidad de realizar este proyecto de investigación en la Dirección Ejecutiva de Información Agraria, al personal y amigos que me brindaron su apoyo incondicional.

Al Ing. Segundo José Castillo Zumarán, por el apoyo y conocimiento brindado en la realización de este proyecto de investigación, resolviendo mis dudas y brindándome sus recomendaciones.

Índice

| | |
|----------------------------------|-----------|
| Resumen | 10 |
| Abstract | 11 |
| I. Introducción..... | 12 |
| II. Marco teórico..... | 16 |
| <i>Antecedentes</i> | <i>16</i> |
| <i>Bases teóricas</i> | <i>20</i> |
| III. Metodología | 21 |
| IV. Resultados | 26 |
| V. Discusión | 48 |
| VI. Conclusiones | 50 |
| VII. Recomendaciones..... | 50 |
| VIII. Referencias | 51 |
| IX. Anexos | 54 |

Lista de tablas

| | |
|--|----|
| Tabla 1. Diseño de Contrastación de Hipótesis..... | 21 |
| Tabla 2. Métodos de Investigación | 22 |
| Tabla 3. Métodos y Técnicas de recolección de datos | 22 |
| Tabla 4. Matriz de Consistencia..... | 25 |
| Tabla 5. Matriz Bus del Proceso Siembra..... | 28 |
| Tabla 6. Matriz Bus del Proceso Cosecha..... | 28 |
| Tabla 7. Dimensiones: sus atributos y jerarquías..... | 29 |
| Tabla 8. Hecho: sus medidas y atributos..... | 30 |
| Tabla 9. Hecho: sus medidas y atributos..... | 30 |
| Tabla 10. Modelo Físico Tabla dim_Cultivo | 32 |
| Tabla 11. Modelo Físico Tabla dim_Campaña | 33 |
| Tabla 12. Modelo Físico Tabla dim_Valle..... | 33 |
| Tabla 13. Modelo Físico Tabla dim_Consumo..... | 34 |
| Tabla 14. Modelo Físico Tabla hecho_Siembra..... | 34 |
| Tabla 15. Modelo Físico Tabla hecho_Cosecha | 35 |
| Tabla 16. Medidas de precisión del modelo ARIMA (0,0,1) (0,1,0) [12] | 39 |
| Tabla 17. Medidas de precisión del modelo de Regresión lineal con ajuste estacional..... | 40 |
| Tabla 18. Planificación de la Implementación de la solución de BI..... | 54 |
| Tabla 19. Estadísticos descriptivos de la variable Producción (t.) de Arroz cáscara..... | 79 |
| Tabla 20. Prueba de Levene para la Homogeneidad de la varianza (Centro = Mediana)..... | 80 |
| Tabla 21. Resultados de Box-Ljung para Naive estacional | 83 |
| Tabla 22. Resultados de la Prueba Box-Ljung en R | 84 |
| Tabla 23. Fórmulas de Errores de Pronóstico | 85 |
| Tabla 24. Resultados obtenidos de la función accuracy de R [Train]..... | 86 |
| Tabla 25. Resultados obtenidos de los errores de c/Modelo predictivo..... | 86 |
| Tabla 26. Proceso ETL para la Dimensión Valle..... | 87 |

| | |
|--|----|
| Tabla 27. Proceso ETL para la Dimensión Campaña | 87 |
| Tabla 28. Proceso ETL para la Dimensión Cultivo | 88 |
| Tabla 29. Proceso ETL para la Dimensión Tipo de Consumo..... | 89 |
| Tabla 30. Proceso ETL para el Hecho Siembra y Cosecha..... | 90 |
| Tabla 31. Encuesta de Validación..... | 91 |

Lista de figuras

| | |
|--|----|
| Figura 1. Diseño lógico del DataMart DM_DEIA..... | 31 |
| Figura 2. Proceso ETL con la herramienta SQL Server Data Tools..... | 36 |
| Figura 3. Inicio de Sesión de la Plataforma..... | 37 |
| Figura 4. Navegación del Sistema..... | 37 |
| Figura 5. Reporte estadístico y visualización GIS..... | 38 |
| Figura 6. Módulo de pronóstico..... | 38 |
| Figura 7. Pronóstico de producción (t.) de Arroz cáscara para el período 2017..... | 40 |
| Figura 8. Arquitectura Técnica para el desarrollo de la Solución BI..... | 41 |
| Figura 9. Formulario para el Módulo de Pronóstico..... | 42 |
| Figura 10. Resultados del Pronóstico de Producción (t.) de cultivo seleccionado..... | 43 |
| Figura 11. Comprensibilidad de la información..... | 43 |
| Figura 12. Información actualizada..... | 44 |
| Figura 13. Completitud de la información..... | 44 |
| Figura 14. Relevancia de la información..... | 45 |
| Figura 15. Síntesis de la información..... | 45 |
| Figura 16. Confianza en la información solicitada..... | 46 |
| Figura 17. Eficiencia en la información..... | 46 |
| Figura 18. Eficiencia en la información solicitada..... | 46 |
| Figura 19. Disfrute en la información solicitada..... | 47 |
| Figura 20. Satisfacción general con la información solicitada..... | 47 |
| Figura 21. Diagrama de Grantt de la Solución BI..... | 56 |
| Figura 22. Diagrama de Grantt de Seguimiento de la Solución BI..... | 56 |
| Figura 23. Calendario de Actividades de la Solución BI..... | 57 |
| Figura 24. Archivos en extensión “XLS” encontrados del Proceso de GIA..... | 58 |
| Figura 25. Archivos en extensión “XLS” de la nueva Base de Datos estandarizada..... | 58 |

| | |
|--|----|
| Figura 26. Modelo dimensional de alto nivel Hecho_siembra..... | 59 |
| Figura 27. Modelo dimensional de alto nivel Hecho_cosecha..... | 59 |
| Figura 28. Determinación de Requerimientos: STAR NET | 60 |
| Figura 29. Determinación de Requerimientos: STAR NET | 61 |
| Figura 30. Proceso ETL para la Dimensión Valle | 62 |
| Figura 31. Relación entre campos de entrada y destino para la Dimensión Valle..... | 62 |
| Figura 32. Proceso ETL para la Dimensión Cultivo | 62 |
| Figura 33. Relación entre campos de entrada y destino para la Dimensión Cultivo..... | 63 |
| Figura 34. Proceso ETL para la Dimensión Tipo de Consumo | 63 |
| Figura 35. Relación entre campos de entrada y destino para la Dimensión Tipo de Consumo..... | 63 |
| Figura 36. Proceso ETL para la Dimensión Campaña | 64 |
| Figura 37. Relación entre campos de entrada y destino para la Dimensión Campaña..... | 64 |
| Figura 38. Proceso ETL para el Hecho Siembra..... | 64 |
| Figura 39. Relación entre campos de entrada y destino para el Hecho Siembra | 65 |
| Figura 40. Proceso ETL para el Hecho Cosecha..... | 65 |
| Figura 41. Relación entre campos de entrada y destino para el Hecho Cosecha | 66 |
| Figura 42. Dimensión Valle | 66 |
| Figura 43. Dimensión Tiempo | 66 |
| Figura 44. Dimensión Cultivo..... | 67 |
| Figura 45. Dimensión Tipo de Consumo | 67 |
| Figura 46. Dimensión Campaña..... | 67 |
| Figura 47. Hecho Siembra..... | 67 |
| Figura 48. Hecho Cosecha | 68 |
| Figura 49. Inicio de Sesión de la Plataforma | 68 |
| Figura 50. Registro de usuarios..... | 69 |
| Figura 51. Registro de cuenta de usuario | 69 |
| Figura 52. Navegación del Sistema..... | 70 |

| | |
|---|----|
| Figura 53. Reporte analítico de Siembra y Cosecha..... | 70 |
| Figura 54. Reporte de Consolidado de Siembra – DEIA | 71 |
| Figura 55. Tabulares, métricas y gráficas de Áreas instaladas de Cultivos | 71 |
| Figura 56. Reporte de variaciones y comparativos anuales | 72 |
| Figura 57. Estadística agraria de Cultivos..... | 72 |
| Figura 58. Reporte estadístico y visualización GIS | 73 |
| Figura 59. Calendario de siembra y cosecha..... | 73 |
| Figura 60. Semaforización de siembra, cosecha y producción | 74 |
| Figura 61. Perfil productivo de Cultivos transitorios..... | 74 |
| Figura 62. Reporte analítico de Sup. en Verde de Cultivos permanentes | 75 |
| Figura 63. Reporte analítico de Sup. en Producción,..... | 75 |
| Figura 64. Módulo de pronóstico | 76 |
| Figura 65. Consulta al hecho hecho_cosecha..... | 76 |
| Figura 66. Consulta al hecho hecho_cosecha..... | 77 |
| Figura 67. Producción (t.) de cultivos por años | 77 |
| Figura 68. Gráfico Estacional: Producción de F. Palo Grano Verde (t.)..... | 78 |
| Figura 69. Gráfico Estacional: Producción de Maíz Amiláceo..... | 78 |
| Figura 70. Gráfico Estacional: Producción de F. Palo Grano Seco (t.)..... | 78 |
| Figura 71. Resultados de la Prueba de Dickey-Fuller Aumentada en R..... | 79 |
| Figura 72. Prueba de Dickey-Fuller Aumentada..... | 80 |
| Figura 73. Gráfica de Descomposición de la Serie de Tiempo..... | 81 |
| Figura 74. Gráfica de la Serie de Tiempo con su Tendencia | 81 |
| Figura 75. Modelos predictivos en R | 82 |
| Figura 76. Resultados del método Naive estacional en R..... | 82 |
| Figura 77. Gráfica de Auto correlación de residuales..... | 83 |
| Figura 78. Valores observados y ajuste con Naive estacional | 84 |
| Figura 79. Valores observados y ajuste con Naive estacional | 85 |

Resumen

La relación que exige los procesos tecnológicos en el universo actualmente es de mucha trascendencia para el individuo y su colectividad. En corto tiempo el ser humano se ha profundizado en el uso de la tecnología para beneficio de sus ocupaciones habituales como investigativas, industriales o comerciales. Esta evolución tecnológica sirve de soporte a las operaciones y a la gestión de las instituciones, por esta relevancia en el progreso evolutivo, para apoyar a la Dirección Ejecutiva de Información Agraria de la Gerencia Regional de Agricultura de Lambayeque, propongo, el desarrollo de una solución de Inteligencia de negocios basado en técnicas de minería de datos como herramienta para potenciar sus actividades de gestión de información. Resumiendo, el trabajo consistió, primero, en estudiar las principales metodologías de desarrollo existentes: este caso se decidió llevar a cabo un híbrido de metodologías Ralph Kimball con el objeto de implementar almacenes de datos y CRISP-DM para el procesamiento electrónico de datos, elección y aplicación de métodos predictivos. La aplicación de la dimensión sistemática de alto nivel nos permitirá medir el hecho-siembra y hecho-cosecha tales como: valle, tiempo, tipo de consumo, cultivo, campaña agrícola. Los resultados demuestran que el modelo propuesto es de Regresión lineal con ajuste estacional, puesto que ostentó un error cuadrático menor a comparación con el enfoque Naive estacional y el modelo ARIMA (0,0,1) (0,1,0) [12], además de tener un valor MAPE ligeramente superior al valor obtenido por los demás.

Palabras clave: Procesos tecnológicos, minería de datos, solución de Inteligencia de negocios.

Abstract

The relationship that exempts the technological processes in the universe today is of great importance for the individual and his community. In a short time, the human being has deepened in the use of technology for the benefit of their occupations, such as investigative, industrial or commercial. This technological evolution serves as support to the operations and management of the institutions, due to this relevance in the evolutionary progress, to support the Executive Directorate of Agrarian Information of the Regional Management of Agriculture of Lambayeque, I propose, the development of a solution of Business intelligence based on data mining techniques as a tool to enhance your information management activities. In summary, the work consisted, first, of studying the main existing development methodologies: in this case it was decided to carry out a hybrid of Ralph Kimball methodologies in order to implement data warehouses and CRISP-DM for electronic data processing, choice and application of predictive methods. The application of the high-level systematic dimension will allow us to measure the fact-sowing and fact-harvest such as: valley, time, type of consumption, crop, agricultural season. The results show that the proposed model is Linear Regression with seasonal adjustment, since it exhibited a lower quadratic error compared to the seasonal Naive approach and the ARIMA model $(0,0,1) (0,1,0)$ [12], in addition to having a MAPE value slightly higher than the value obtained by the others.

Keywords: Technological processes, data mining, solution of Business intelligence.

I. Introducción

Gracias al progreso continuo de las tecnologías de la información y comunicación (TIC's) que van adquiriendo las compañías tanto en el entorno público como privado a largo del tiempo; hoy en día se cuenta con grandes volúmenes de datos personales, económicos, médicos, financieros, entre otros. Como resultado, los mercados han tomado una rigurosidad y competitividad en cada paso y decisión que toman con respecto a las estrategias que adoptan frente a sus competidores; que desarrollan y explotan al máximo sus recursos; y en donde la información se convirtió en un activo importante y concluyente tanto como en el óptimo funcionamiento de numerosas empresas como además en el crecimiento socioeconómico y político de nuestra sociedad.

El interés por organizar y administrar eficientemente el creciente auge de la información, ha pasado a la aparición de un nuevo concepto, denominado Gestión de la información. De acuerdo a Bustelo y Amarilla [1] lo conceptualizan como una serie de actividades que se realizan con la finalidad de un control, almacenamiento y seguidamente, la recuperación de datos generados, obtenidos o retenidos por las organizaciones en crecimiento de su actividad. No obstante, depende mucho de cómo las organizaciones puedan aprovechar este conocimiento que han ido adquiriendo y almacenando; de modo que represente una ventaja competitiva.

Ciertamente, en las actividades, es primordial ajustar cada exigencia de datos a la respuesta de la misma. Es por ello que, Suarez et. al. [2] exponen la premisa de que el desarrollo de la automatización y las comunicaciones (TIC), conforma una reforma estructural, en la perspectiva económica, profesional, sociopolítico, educativo, y configuran el surgimiento de nuevas comprensiones de la cultura.

En este sentido, en las empresas privadas se vienen desarrollando iniciativas en cuanto a la gestión de información, sin embargo, nos encontramos con una brecha tecnológica bastante notoria en los organismos públicos de nuestro ámbito; en donde pocas de ellas tienen una proyección e inversión en cuanto a tecnologías de información, aunado al nivel de su capital humano en torno a las prestaciones que brindan a sus beneficiarios.

Al presente, la modernización del estado y fortalecimiento de las instituciones gubernamentales son temas ya muy recurrentes, pues se ha visto viable que con una correcta y eficiente administración de las tecnologías de información genere un mejor desempeño de sus procesos y diligencias en pro a la búsqueda de satisfacer las necesidades tanto de su personal como de la ciudadanía.

El éxito de la filosofía de una estructura depende de la capacidad de la administración tiene para anticipar el futuro y elaborar estrategias adecuadas. La revolución digital ha contribuido las capturas, procesamiento, almacenamiento, distribuciones y transmisiones de datos digitales. [3].

Obtener conocimiento sobre esta enorme cantidad de información es un desafío en sí mismo. La minería de datos (MD) es un intento de comprender las explosiones de datos que en la actualidad se pueden almacenar [4]. Como resultado, poder comprender y analizar datos históricos para identificar tasas de crecimiento o variaciones con respecto a periodos anteriores, explorar minuciosamente nuestra variable de estudio (producción en t. en nuestro caso) para determinar si son o no estacionales, la presencia de tendencia o de datos erróneos o atípicos que puedan afectar el comportamiento de nuestros datos; nos permite finalmente efectuar un mejor modelamiento de predicción esperado.

A lo largo del tiempo, se ha desarrollado una diversidad de técnicas pronósticas, tanto de forma cualitativa como cuantitativa, con su variable relacionada a los comportamientos de información [5]. En esta investigación, se revisó distintas técnicas y algoritmos como medias móviles, suavizado exponencial, modelos de regresión lineal o la creación de la red neuronal para la analítica y exploración de patrones de datos.

De esta manera llegamos a la Dirección Ejecutiva de Información Agraria (DEIA), área perteneciente a la Gerencia Regional de Agricultura Lambayeque (GRA), tiene años de experiencia en los sectores agropecuarios, se ocupa del diseño, implementación y generación de estadística y metodología en la producción de datos confiables sobre cada variable principal agroeconómica, por el que cuenta con un personal calificado para el acopio y procesamiento de información estadística, distribuidos en el ámbito de las tres Agencias Agrarias, definidas por sus características agrarias y medioambientales, correspondiente a las provincias: Chiclayo, Lambayeque y Ferreñafe; y por ende la necesidad de información es fundamental para dicha labor.

En la mayoría de cada institución, no cuenta con el recurso ni la capacidad para la optimización de sus procesamientos y mucho menos una articulación entre los sistemas administrativos y sus mecanismos de coordinación y comunicación. Esta realidad no es ajena en la GRA, puesto que cuenta con un sitio web, el mismo que proporciona información general a los usuarios, como: boletines informativos, producción de cultivos, análisis de actividad agrícola, siembras y cosechas y rendimientos anuales; sin embargo, con el gran desarrollo que está logrando la actividad agrícola, el avance tecnológico y con miras a mejorar la atención de la población lambayecana; recibe notoriedad la necesidad de personalizar la información por medio de indicadores, proyecciones, interacciones dinámicas, con el fin de facilitar su usabilidad y reducir tiempos, esfuerzos tanto al personal como al usuario final. (Anexo 2, Pregunta 3).

La institución no cuenta con un sistema que le brinde soporte a la información que se manipula, ni las herramientas adecuadas para la elaboración de reportes de rendimiento y producción por área, sectorización de siembra y cosecha por cultivo específico, de rendimientos anuales, entre otros procesos que les permita presentar de manera gráfica y detallada el comportamiento de sembríos y cosechas de cultivos, es por ello que las decisiones que se toman son basados en la propia experiencia, pudiéndosele pasar ciertas cosas importantes. (Anexo 1, Pregunta 3 y 6; Anexo 2, Pregunta 3).

La principal desventaja que surge es la cuantiosa información que se manipula y la necesidad de contar con un repositorio o una base de datos que almacene toda la información que se genera, garantizando el respaldo y la seguridad de la misma. Es por esta razón, que uno de los factores que conlleva a la realización de este proyecto, es el crecimiento que está teniendo la actividad agrícola y por ende el soporte a la información es vital para su posterior tratamiento, de este modo, la tecnología que emplean no les facilita el desempeño óptimo de sus actividades, surgiendo una limitación en el proceso.

El tratamiento de la información requiere altos costos de tiempo, valor y esfuerzo; este proceso recopila cédulas con información estadística mensual de todo lo correspondiente al sembrío y cosecha de cereales, tubérculos, menestras (legumbres), etc.; hectáreas sembradas, producción por área, rendimientos, entre otros aspectos; por parte de técnicos especializados de las Agencias Agrarias (Sedes), lo que ocasiona grandes bancos de información y aunado a ello; no existe una política o práctica de uso y conservación de datos históricos, y cuando tienen la necesidad de comparar algún proceso en el tiempo; la demora se hace notar o se dan con la sorpresa que nadie sabe de la localización de dichos archivos. (Anexo 1, Pregunta 8).

Otra dificultad que tienen es que no hay un control del tiempo que demoran en brindar alguna información solicitada, ni la frecuencia con que vienen a solicitarla; es por ello que reconocen esa debilidad y pretenden mejorarla. Estos usuarios solicitan información ocasionalmente, de entre 5 a 7 personas a la semana y un tiempo aproximado de atención de entre 30 a 45 min. dependiendo la carga de trabajo. (Anexo 1, Pregunta 4 y 5; Anexo 2, Pregunta 2 y 4).

Sin embargo, en la actualidad coexisten soluciones como la de Inteligencia de Negocio (BI) que nos proporcionarían un panorama más intuitivo e innovador, necesaria para poder colaborar con las mejoras en cuanto a información se trate a través de herramientas, técnicas o modelos para poder brindar información oportuna, útil y confiable para su posterior utilización y toma de decisiones.

En lo que corresponde a la información con la que trabajan, no hay una integración o un consolidado de las mismas, pues existen cuadros por cada cultivo, producción por provincia, por departamento, exportaciones, etc. todo por separado, lo cual genera una gran dificultad para poder analizarlo y tomar decisiones. (Anexo 1, Pregunta 8).

De esta manera, llegamos a la conclusión de que existen estos problemas en cuanto al desarrollo tecnológico y las necesidades de gestión y aseguramiento de calidad de la información; y hoy en día, se precisa la implementación de un elemento catalizador e integrador. Con el propósito de poder brindar apoyo a la DEIA; área perteneciente a la GRA, se desarrollará una solución tecnológica que le permita gestionar toda la información recolectada a lo largo de los últimos años y así, potenciar su prospectiva y tácticas operacionales de desarrollo en torno al ámbito tan competitivo de hoy en día.

Con el objetivo de contribuir con lo ya expuesto, surgió la siguiente interrogante: ¿De qué manera se puede apoyar la toma de decisiones en la Gerencia Regional de Agricultura - Lambayeque? Ante lo cual se propuso implementar una solución de Inteligencia de negocios basada en técnicas de Minería de Datos, logrando que de manera oportuna, confiable y útil consiga acceder a dicha información para su posterior beneficio.

Como objetivo general se propuso apoyar la gestión en la DEIA mediante la solución de BI basada técnicas de Minería de Datos, mientras que como objetivos se establecieron los siguientes: (A) Determinar la técnica de Minería de datos que más se ajuste a los datos históricos de la variable en estudio: Producción (t.), (B) Implementar la técnica de Minería de datos que permita predecir la Producción de cultivos, (C) Implementar informes analíticos en base a las necesidades de información requeridas, y finalmente; (D) Lograr una validación positiva de la Solución de BI por parte de los usuarios.

Desde el punto de vista social se fundamentó porque la implementación de la herramienta propuesta beneficia a los directivos de la GRA y personal en cuestión, productores y empresas asociadas a proporcionar información oportuna y útil en sus procesos con mayor detalle y tomar decisiones acordados, a tiempo. Además, se permitirá disminuir tiempos de procesamiento, costos, satisfacción del usuario tanto externo como interno; pero principalmente será una herramienta de apoyo para el área pues a través de data valiosa que poseen, se podrá crear información útil para tomar decisiones acertadas. Así como también ser modelo para investigaciones futuras en otras actividades empresariales.

En cuanto al aspecto tecnológico, la exploración se espaldó porque con la solución que se construyó quise demostrar que la información histórica sirve como fuente para la toma de decisiones pues les permite a los directivos tener una visión más amplia de la información que

se gestiona. Se trató de una herramienta que se generó a partir de una suite de inteligencia de negocio de Microsoft, la cual ofrece una amplia gama de recursos en cuanto al proceso de Extracción, Transformación y Carga, generación del cubo multidimensional y sobre todo la elaboración de reportes personalizados para los usuarios finales los cuales, para un acceso más amigable, serán mostradas en una plataforma web que funcionara como apoyo a la gestión de información. Basándose en técnicas de minería de datos buscaremos tipos específicos de patrones o tendencias que definirán criterios óptimos para la inventar un software con el fin de extraer patrones factibles y estadísticas detalladas en medidas que puedan continuarse en el futuro.

Finalmente, para lograr el rigor científico, ejecuto el diseño de investigación pre-test y post-test y se aplicó a los miembros y directivos de la Institución para evaluar la manera en que se gestiona la información actualmente y como el sistema de Inteligencia de negocio los apoyó después. Más a allá de eso, este proyecto interesa como antecedentes para investigaciones futuras pertinentes con la problemática.

II. Marco teórico

Antecedentes

Rudnichenko et al. [6], el estudio se concentra en las características del procedimiento de respaldo a la toma de medidas para la elección de métodos de aprendizaje automático en el desarrollo de la minería de BIG DATA. La investigación circunscribe las resultados del estudio del problema del proceso científico y análisis de BIG DATA por fases. En ella, se puntualiza propuestas formuladas para valerse de metadatos como base para la formación de una valoración metódica para evaluar los métodos de aprendizaje automático; basándose en este caso de un estudio experimental de un conjunto de datos de Diabetes conformado por 442 registros y 10 atributos con valores reales. El instrumento muestra los resultados del formulación y uso de un procedimiento de respaldo a la toma de medidas para evaluar las distintas metodologías de aprendizaje; para ello se aplicó y probó varias comparaciones de modelos experimentales con métodos de regresión lineal, árbol de decisión, random forest, método SVM (para clasificación y regresión), xgboost, regresión logística y clasificación ingenua de Bayes. En cuanto a las métricas de evaluación utilizadas para la calidad de los modelos se utilizaron AUC, F1, precisión y sensibilidad para clasificación; y MSE con MAE para regresión. En caso de las comparaciones experimentales; utilizaron el Método de Validación cruzada. El sistema desarrollado para tomar medidas de solución, nos permite disminuir el tiempo de análisis de métodos apropiados para solucionar inconvenientes de aprendizaje automático por un analista de ciencia de datos, teniendo en cuenta las particularidades definidas en las matrices de datos de entrada, sus volúmenes, estructura y otros metadatos. Finalmente, los autores concluyeron que el nivel de precisión del Sistema DSS alcanza el 70-75% y si bien no es siempre precisa en tareas con requisitos críticos, es aconsejable realizar un análisis exploratorio adicional de datos. Asimismo, se debe tener en cuenta los casos que presenten inconsistencia de datos; en donde la capacidad de correctitud y completitud utilizando varios algoritmos es una de las opciones para el futuro desarrollo y actualización del sistema.

Chougule, Kumar y Mukhopadhyay [7], realizaron la indagación en base a algunas proximidades de la cuantía de abonos y el tipo de cultivo a sembrar o injertar. Para ellos, es trascendental prosperar en el agro, por lo que consideran que la agricultura mucho depende de la naturaleza del suelo y el clima. Según el instrumento usado se plantea el proceso de un sistema de recomendación fundamentado en Ontologías para la idoneidad de cultivos y abonos. El sistema pronostica cultivos apropiados para el campo bajo consideración según la región del estado de Maharashtra en la India y el tipo de suelo. Igualmente aporta una recomendación apropiada de abonos a los sembradores, esta se fundamentó en el complemento de nitrógeno, fósforo y potasio (NPK) de la tierra y utilizando datos de investigación de años precedentes que se recogen del Departamento de agricultura del Gobierno de Maharashtra. Web Ontology Language (OWL) se utilizó para la representación de dichas ontologías. Prosiguiendo con el sistema de recomendación de abonos, éste está basado en Android conectada a un servidor el cual provee además conjeturas sobre la idoneidad de cultivos en zonas específicas. Dicho sistema maneja un algoritmo de Random forest utilizando la base de conocimientos en Ontología como conjunto de aprendizaje para la recomendación de cultivos y el de agrupación en clústeres de k-medias para así descubrir abonos con NPK contenido más cercano a los requisitos para el cultivo especificado. Finalmente, los autores concluyeron que el objetivo de este sistema fue aumentar la producción agrícola recomendando cultivos y fertilizantes correctos; mostrando que la precisión del sistema desarrollado es razonablemente alta. En el futuro, la aplicación de Android se desarrollará en idioma regional.

Khanian, Mohamed y Naz'ri [8] según su investigación los sistemas de recomendación han adquirido un papel más específico con la publicación del primer concepto de investigación sobre filtrado colaborativo (FC) en la década de 1990. CF pronostica las ventajas de un usuario dinámico apoyándose en los informes de consumidores con intereses semejantes. Se puede realizar un examen crítico luego de conseguir la referencia sobre la predilección de los consumidores por un grupo de elementos y evaluar el desempeño de las tecnologías y algoritmos del sistema de recomendación. Es por ello, que el presente estudio aprovecha el análisis crítico de 131 artículos en el área de la FQ de 36 revistas propagadas entre los años 2010 y 2016. Esta observación considera a la encuesta que gravita y motiva a la colectividad de estudiosos y competitivos, la investigación determina el uso de las diligencias de las acciones de los usuarios y las sistemáticas de inteligencia artificial y exploración de datos en los sistemas de propuesta de CF. Igualmente, facilita un ordenamiento del lenguaje sobre bases de datos pedagógicas en concordancia con las bancos de datos de puntos de referencia, retroalimentación de dos usuarios (retroalimentación explícita e implícita) que muestra sus acciones y naturalezas de métodos de informática, de inteligencia y listados de datos. En último lugar, los autores concluyeron que este estudio aporta una hoja de ruta para ubicar la trayectoria futura en la búsqueda de sistemas de recomendación y proporciona el conocimiento acumulado y derivada sobre su aplicación en sistemas de recomendación de CF. Además de comprender las aplicaciones de los comentarios de los usuarios, la inteligencia artificial y las técnicas de minería de datos en dichos sistemas de recomendación mediante el examen de los artículos publicados y brindar a la comunidad de investigadores y profesionales información y orientación futura sobre el tema investigado.

Mazon-Olivo et al. [9], la intención de esta investigación es reconocer los itinerarios claves de rendimiento (KPI) de productividad y mercantilización de una asociación de cacao; la metodología CRISP-DM que se uso es para alinear el perfeccionamiento y ejecución de un tablero de control (Dashboard) de examen de información y apoyado por conjunto de técnicas de inteligencia de negocios (BI) y minería de datos (DM). Los autores refieren que otros investigadores han aportado en los estudios de referencias, estableciendo o empleando técnicas de BI y DM; los aportes han contribuido en optimizar las artes de productividad y de mercadeo; esto, no se demuestra en la producción de cacao. Hay que tener en consideración que las ventajas competitivas se dan por la buena y adecuada información para la elección de medidas en tal sentido los aportes para identificar KPI's y el empleo de técnicas de BI y DM; no describen la situación del negocio ni predice las tendencias futuras. Los resultados son significativos porque permiten visualizar información estadística, detectar patrones de datos y predecir el rendimiento y comercialización de la cosecha de cacao mediante series temporales, método de Holt-Winters con un Chi cuadrado mayor a 0.9. Finalmente, los autores concluyeron que la inteligencia de negocios y minería de datos, incluyen técnicas que permiten analizar los datos históricos de una organización, buscando información útil que describa la situación em el pasado o presente de la empresa, o prediciendo tendencias futuras que generen ventaja competitiva favorable.

Según Merchán, Mero y Castro [10], ante el riesgo probable de tener inconstancias en el momento de analizar la progresiva diversidad de datos y fuentes generadas en las actividades diarias los investigadores consideran que los sistemas de visualización de datos contribuyen de forma rápida y sencilla el examen de dicha información. Los resultados precisan que los sistemas BI es la herramienta adecuada para obtener conocimiento, garantizando una óptima toma de decisiones en las instituciones por su rapidez de proceso. Así mismo su facilidad de uso permite aplicarla a cualquier espacio para el análisis de información en tiempo real. En ese sentido el uso de los sistemas propuestos son una alternativa factible. Finalmente, los autores

concluyeron que los sistemas de BI con una herramienta útil para la obtención de conocimiento, lo que acredita que se tomen mejores decisiones en la organización a través de la velocidad de procesamiento. Con los DW se pueden integrar diferentes bases de datos, los CMI permiten el control de planes estratégicos, la DM aplica métodos inteligentes para predecir o describir información y los Sistema Expertos manejan conocimientos imprecisos, contradictorios o incompletos.

Améstica, Liendo y Luna [11], los autores argumentan que la Pyme Agroindustrial referida en el estudio está centrada en comercialización de agroquímicos, abonos y semillas, de igual manera de alimentos balanceados en el departamento de San Justo. Los variados retos que afronta la organización por el crecimiento del mercado, son fortalecidos por la problemática en los vigentes sistemas de gestión informáticos ya obsoletos, como lo es COBOL. Ante la necesidad de perfeccionar la gestión diaria de pedidos y el control exhaustivo del stock con un sistema de información con estrategias modernas, multiplataforma, flexible y escalable, se propone de forma estructurada el estudio, con proceso e implementación de un sistema Web, manejando para esto el conjunto de técnicas del framework .NET, base de datos SQL Server y Tableau como Suite de apoyo a medidas. El proyecto, además, de ser configurable, permite asignar otras capacidades de los sistemas de gestión, relacionados al rol. Finalmente, los autores concluyeron que, en el Predio Agropecuario, se viene utilizando y verificando el correcto funcionamiento del Sistema y, gracias a ello, permitiéndole al personal realizar pedidos a través de dispositivos móviles (teléfonos y tabletas). Mediante a la aplicación web, el dueño del Predio puede controlar todo el inventario por intermedio de diversas funciones del sistema, lo que le permite visualizar la causa de las diferencias, si las hubiera, y mantener una estadística de las mismas, así como también realizar informes instantáneos de información general del negocio.

Santa Cruz [12], el autor apuesta por un instrumento practico como Bussines Intelligence (BI) para moderar o eliminar riesgos que otorguen beneficios de productividad del presente estudio. Garantizar la competitividad aporta visibilidad luego de optimizar las estrategias que se pueden gestionar y medir mediante el uso de un sistema de BI. En esta indagación se proyectó exponer que el uso de las tecnologías de información y concretamente la inteligencia empresarial (BI) cree impacto efectivo en las métricas empresariales, de tal forma que los trabajadores de la compañía consigan acceder de forma sistematizada y pertinente a la información generada en el área de produccion de cultivos de palma y cacao a fin de contribuir a las decisiones de los gerentes directamente a los resultados organizacionales en el breve, mediano y largo plazo. La ejecución de esta medida fue un resultado explicativo que concluyentemente marcó una brecha significativa entre el entorno en la que se hallaba produciendo anteriormente y posteriormente de la obtención de la medida ofrecida. Por último, consintió reconocer y demostrar cada una de los ciclos del plan y crear, igualmente, el manual del sistema. Finalmente, el autor concluyo en primera instancia que, como resultado de la implementación del sistema se lograron agilizar las decisiones, mejor la velocidad de acceso a la información y aumentar la calidad de los informes empleados en el ámbito gerencial. En segunda instancia, se consiguió una gran experiencia en el manejo de la plataforma Pentaho para soportar el sistema de BI de manera satisfactoria y a partir de los resultados obtenidos, se confirmó que la decision tomada de invertir en BI y manejar software libre es una combinación ideal para un desarrollo y competitividad favorable. Como último punto, el impacto positivo que tuvo la metodología Hefesto en la realización del proyecto.

Castro y Díaz [13], para los investigadores los instrumentos de Inteligencia permiten proyectar tácticas para trazar, efectuar, gestionar, controlar y amparar las decisiones de BI. En ese ambiente, el objetivo de la investigación pretende implementar esos procesos para implicar a todo el personal que labora TI. Los métodos han dejado de ser solo un apoyo y creadora de

precio, a ser una revolución para cobijar la escasez vital. La problemática observada fue la falta de reportes e información desarticulada que genera insatisfacciones y pérdidas de tiempo al momento de presentar la realidad de la empresa y de tomar decisiones. Es aquí donde ingresa el conjunto de instrumentos al área de TI: Business Intelligence, para amparar a la organización a obtener una superior comprensión. Finalmente, los autores concluyeron que los resultados del estudio expresan que los materiales de procesos de recolección, posicionará a la organización generándole mayor valor a las líneas de negocios existentes, del mismo modo accederá predecir nuevas conformidades; para ayudar a disminuir los ciclos de mejora de productos, acelerar operaciones, armonizar maniobras de marketing, optimizar y ampliar relaciones con clientes a periodo largo, llevando a la disminución de costos y aumento de utilidades.

Arbieto [14] en su tesis investigativa hace referencia a las instituciones encargadas del acopio de información del clima, la cual la realizan a través de sensores que permiten medir y comprender los fenómenos de los métodos acostumbrados en varias escalas numerarias así se conoce la intensidad, frecuencia, volumen, etc. En su estudio pretende demostrar la eficiencia de técnicas de recolección de información ampliable sobre sucesiones estacionales meteorológicas de 36 dimensiones, con el objeto de mejorar el proceso en período informático y aprovecharse de las nuevas propuestas de hardware históricos, tratando de aligerar el procedimiento de recolección de información a través del uso de instrucciones KNN (búsqueda del vecino más cercano) en sistemas multiprocesador 1 en GPU (Unidad de procesamiento gráfico). Los resultados logrados demostraron que incrementando medidas de alto trabajo manejando el prototipo de programación GPU se logra optimizar el ejercicio de estos en diversos ritmos de dimensión. Por los antecedentes este trabajo es una solución aplicable porque contribuye a reducir los tiempos y acelera la búsqueda y la reconstrucción de tareas de minería de datos. Finalmente, el autor concluyo que se deben enfatizar tres aspectos de este trabajo. El primero fue una reconstrucción de una base de datos climática, como segundo punto fue un estudio de los algoritmos de búsqueda por similitud, orientado en los algoritmos aproximados. y el tercero fue la adaptación de algoritmos para arquitecturas paralelas en CPU-GPU y su comparación al respecto.

Fernández [15], en su publicación, se aprecia las tecnologías de minería de datos, como un instrumento computacional para pronosticar el cultivo de espárragos para fines de la industria agrícola en distrito de Jayanca - Lambayeque. Las tecnologías de comportamiento futuros tanteadas fueron: ARIMA, Holt-Winters, Redes Neuronales, el examen y comparación de los patrones en series temporales. Los resultados observados determinan cotejos demostrativos de las predicciones de las inventivas trabajadas, con los antecedentes visualizados. Finalmente, el autor concluyo que las capacidades de reciprocidad más altos se alcanzaron con la práctica ARIMA en comparación a Holt-Winters y Red Neuronal autorregresiva, esto se revela en los productos alcanzados al computar el conocimiento (valor computado entre el valor real y el valor predicho para establecer el nivel de relación que obra entre estos), consiguiendo para Holt Winters unos 80.52 % de confiabilidad, contra un Red Neuronal Autoregresiva con 85.29 % y un ARIMA con un 86.86 % que lo ubica como el dominante de los tres algoritmos. Por ello, se determinó la viabilidad de la usanza de técnicas de predicción aplicables a la obtención de espárragos en la región de Lambayeque, por cuanto el diseño del aplicativo web también permitió visualizar información del prototipo, en un dispositivo informático, donde PHP es el lenguaje servidor, HTML 5 es la perspectiva y R Project es el motor ordenado; el manejo de efectos se manejó el motor con SQL Server.

Bases teóricas

Gestión de la Información (GI)

Es la designación convenida de métodos contiguos por los que se examina el tiempo de duración de la investigación, desde su preparación (por espacio o conquista), hasta su práctica final (su registro o separación). Estos métodos están complementados por la procedencia, composición, purificación y repartimiento del examen a los comprometidos. Su objetivo es avalar la entereza, reserva y privacidad de la información. Concluyendo, White [16] la llama como la coherencia eficaz y vigorosa de las búsquedas nativas de orígenes ocultas y exteriores.

Bajo definición de Woodman [17], la gestión de información es la recolección de antecedentes considerada en forma ordenada, para el sujeto indicado, al mejor costo, en el momento y lugar conveniente para realizar la gestión precisa.

La gestión de la información según Ponjuán [18] es el despliegue de un sistema de información (si la intención del método es obtener salidas de información) y lo especifica como: el proceso para manipular antecedentes dentro y fuera de la sociedad donde está inmersa, porque en ella se obtienen, desarrolla o manejan recursos elementales (económicos, corporales, humanos, materiales). Al tener unidades especializadas las unidades de información sus recursos son intensivos y tiene como componente principal la gestión del tiempo de vida de este medio que sucede en cualquier estructura. La Gestión de información posee los siguientes objetivos de acuerdo a Ponjuán [19]:

- La Extender la totalidad de los beneficios provenientes de la usanza de la indagación.
- Reducir el precio de compra, proceso y usanza de los antecedentes.
- Establecer compromisos para la usanza real, eficaz y módica de información.
- Afirmar un abastecimiento perenne de la información.

Bajo este argumento de Páez, los antecedentes son creídos un medio, un fruto y un activo; la fichas como recurso tiene costo y por consiguiente rendimiento, por lo tanto, habrá que exigirle. Así mismo, como activo involucra a la organización a poseer, gestionar y utilizar la información [20].

Para Huber, una estructura experimenta si por la ruta del proceso de fichas, cambia el procedimiento de sus estándares permitidos [21]. Por ello Choo, refiere para fortalecer sus medios de información y las potencialidades con el fin de que la estructura experimente y se acomode a su contexto es importante la gestión de información [22].

Minería de Datos

La minería de datos descrita por [23], es como un paso de hallazgos de acontecimientos y explicativas relaciones, de esquemas para explorar inmensos volúmenes de información. Por otro lado [24], explica que la minería de datos se define como el paso de sustracción de la investigación de beneficio al dividir los datos, argumentando que solo la noción es de beneficio perpetuo y que sea siempre innovador.

Según [25], la minería de datos es el hallazgo de información ventajosa en inmensas cantidades de antecedentes. Se conoce como una voluntad entre las personas y los ordenadores. Este concepto lo amplían Witten y Frank, [26] que dicen que la minería de datos es la actividad de desagregar elementos favorables y visibles, inicialmente desconocidos desde enormes valores de antecedentes acumulados en variados espacios. Mientras que Hernández [27] nos explica

que los datos de los resultados como historiales son favorables que hay que utilizar como mecanismos para tener el real beneficio dispuesto.

Teniendo en cuenta los antecedentes, la actuación de la Minería de datos es un escenario de actividades que implica concordar pilotos o establecer estándares partiendo del acopio de enormes volúmenes de investigación ordenados por diferentes tipos de historiales no perennes, correctos y perfectos como se anhelaría. Beltrán [28] revela que el empleo de esquemas regularmente es de tipo detallado, lo que acepta un indiscutible ruido o desacierto internamente. Los modelos implantados son regularmente de tipo determinístico o probabilístico, esto hace que los caminos para su cumplimiento no sufran cambios, y otorguen independencia de la práctica delimitada de origen de los conocimientos utilizada.

III. Metodología

1. Tipo de investigación

El tipo de investigación según el propósito pretendido, es el de una investigación tecnológica aplicada, dado que busca desplegar un metódico conjunto de mecanismos, pasos y técnicas procedimentales reflexivas que se aplican para el uso, transformación o construcción de una nueva tecnología. En este caso sería una solución de BI basado en cada técnica de minería de datos para el pronóstico de producción agrícola [29].

2. Nivel de investigación

En los niveles de investigación, fue Cuantitativa preexperimental.

3. Diseño de investigación

La investigación tiene un diseño que se desarrolló con preprueba/posprueba con un solo grupo [30]: G_1 o G_2 que pertenece al nivel de investigación pre-experimental, puesto que se evaluó un antes y un después a causa de que se pretendió saber cómo esta solución de inteligencia de negocio apoyará en la mejora de la Gestión de Información de los Directivos o usuarios de la Gerencia Regional de Agricultura Lambayeque.

El diseño de contrastación que se utilizó se muestra a continuación:

*Tabla 1.
Diseño de Contrastación de Hipótesis*

| Variable dependiente | Solución BI (X) | Variable dependiente |
|---|--|---|
| O1 | X | O2 |
| Toma de decisiones en la Dirección Ejecutiva de Información Agraria de la Gerencia Regional de Agricultura, Lambayeque. | Solución de Inteligencia de negocios en la DEIA. | Toma de decisiones en la Dirección Ejecutiva de Información Agraria de la Gerencia Regional de Agricultura, Lambayeque. |
| A través de entrevistas con el responsable del área de Información Agraria y de Sistemas de Información de la GRA, se pudo obtener información del estado situacional y procesos involucrados | | A través de un cuestionario con escala de Likert de validación con los responsables y directivos del área de Información Agraria, se pudo conseguir una valoración positiva de la Solución. |

4. Población y muestra

4.1. Población

La presente investigación tiene una población constituida por los directivos y miembros de la Dirección Ejecutiva de Información Agraria y el encargado del Centro de Sistemas de Información de la GRA.

Muestra de estudio:

- Miembros de la Dirección Ejecutiva de Información Agraria: En total son seis trabajadores, por ser pequeña, será igual a la población.
- Encargado del Centro de Sistemas de Información: Es un empleado. La muestra, por ser pequeña, será igual a la población.

4.2. Métodos de investigación

El método de estudio empleado es:

Tabla 2.
Métodos de Investigación

| Método | Descripción |
|----------------|---|
| Descriptivo | Revisión y comparación de modelos de Minería de datos más idóneas con el objeto de resolver la problemática de investigación en cuestión. |
| Analítico | Estudio y análisis de los procesos, actividades y herramientas para la comprensión de requerimientos en la Dirección Ejecutiva. |
| Deductivo | Plan y estrategia para proponer la solución propuesta al problema. |
| Implementación | Sistema Web para el análisis y gestión de Información en la Dirección Ejecutiva. |

4.3. Técnicas e instrumentos de recolección de datos

La Tabla III muestra los métodos, técnicas e instrumentos de recolección de datos que se emplearon en la presente investigación.

Tabla 3.
Métodos y Técnicas de recolección de datos

| Método | Técnicas e instrumentos | Elementos de la población |
|------------------------|--------------------------------|--|
| Entrevista | Guía de entrevista | Director del Área de Información Agraria. Encargado del Centro de Sistemas de Información. |
| Encuesta | Cuestionario | Directora y miembros del Área de Información Agraria. |
| Observación | Observación | Se observó el movimiento de la gestión de información en la GRA. Recopilación, Selección, Tratamiento, Utilización, Distribución, etc. |
| Observación documental | Documento | Redacción de la documentación prevista en el Desarrollo e Implementación de la misma. |

4.4. Procedimientos

1. Metodología de desarrollo

Asimismo, se relata cada actividad realizada en la interacción del método a seguir, como en el presente caso, se decidió llevar a cabo un híbrido de metodologías Ralph Kimball con el objeto de implementar almacenes de datos y CRISP-DM para la minería de datos, seleccionando y aplicando el algoritmo inteligente.

El primer método se ejecutará cada fase, y el soportar de cada modelo no cubierto por esta se aplicó CRISP-DM, donde se considera la cuarta y quinta fase del modelado y evaluación.

Seguidamente, se enumera cada actividad realizada por cada fase de la hibridación entre las metodologías descritas:

1. Fase #1: Planificación del Proyecto (RALPH KIMBALL)
En esta fase se desarrollaron las siguientes actividades:
 - ✓ Definir el alcance del Proyecto.
 - ✓ Definir el propósito.
 - ✓ Presentación de la Entidad en estudio.
 - ✓ Desarrollo del Proyecto.
2. Fase #2: Definición de requerimientos del Negocio (RALPH KIMBALL)
En esta fase se desarrollaron las siguientes actividades:
 - ✓ Análisis de requerimientos del Proceso.
 - ✓ Construcción de la Matriz Bus
 - ✓ Análisis de las fuentes de datos internas.
3. Fase #3: Modelado dimensional (RALPH KIMBALL)
En esta fase se desarrollaron las siguientes actividades:
 - ✓ Elaboración del modelo dimensional de alto nivel.
 - ✓ Establecimiento del nivel de granularidad del proceso modelado.
 - ✓ Elección de las dimensiones y sus atributos.
 - ✓ Identificación de la tabla hecho y sus medidas.
 - ✓ Documentación del modelo dimensional.
4. Fase #4: Diseño físico (RALPH KIMBALL)
Esta fase involucra el tamaño que tendrá el Proyecto, la plataforma tecnología a requerir, configuraciones, servidores, tablas relacionales, entre otras.
5. Fase #5: Diseño e implementación del subsistema de Extracción, Transformación y Carga (ETL) (RALPH KIMBALL)
En esta fase se desarrollaron las siguientes actividades:
 - ✓ Diseño del proceso de Extracción, Transformación y Carga.
 - ✓ Limpieza de los datos sucios de las fuentes internas obtenidas en el paso I.
 - ✓ Desarrollo del proceso ETL para los Hechos y dimensiones.
 - ✓ Validación y revisión de la carga correcta de datos.
6. Fase #6: Implementación (RALPH KIMBALL)
En esta fase se comprende las capacitaciones y soporte técnico que se pudieran necesitar, así como también las diferentes comunicaciones y estrategias de feedback futuras.

7. Fase #7: Mantenimiento y Crecimiento (RALPH KIMBALL)
En esta fase se documentará todo el Plan de desarrollo para futuros flujos de retroalimentación y oportunidades de crecimiento y evolución del Proyecto.
8. Fase #8: Especificación de aplicaciones de BI (RALPH KIMBALL)
En esta fase se desarrollaron las siguientes actividades:
 - ✓ Informes y aplicaciones de BI.
 - ✓ Herramientas de análisis.
 - ✓ Aplicaciones analíticas.
9. Fase #9: Modelado y desarrollo de aplicaciones analíticas (CRISP-DM)
En esta fase se desarrollaron las siguientes actividades:
 - ✓ Selección de la técnica de modelado.
 - ✓ Generación de la prueba de diseño
 - ✓ Construcción y evaluación del modelo.
10. Fase #10: Diseño de la Arquitectura Técnica (RALPH KIMBALL)
En esta fase se desarrollaron las siguientes actividades:
 - ✓ Establecer diseño Back Room (Adquisición de datos).
 - ✓ Establecer diseño Front Room (Acceso de datos).

2. Producto acreditable

1. Interfaces

Se construyeron las interfaces del sistema haciendo uso de la suite de SQL Server 2012, Power BI Desktop y services, así como también el lenguaje R para el análisis de datos; las mismas que se presentan en el *ítem 4.6. Fase #06: Implementación, en el Capítulo IV. Resultados.*

2. Arquitectura

Se diseñó una arquitectura idónea para el desempeño de la Solución de Business Intelligence, el cual se detalla en el *ítem 4.10. Fase #10: Diseño de la Arquitectura técnica, en el Capítulo IV. Resultados.*

3. Infraestructura tecnológica

Tomando en cuenta la arquitectura previamente descrita, se definen las particularidades de cada uno de sus componentes en el *ítem 4.4. Fase #04: Diseño físico, en el Capítulo IV. Resultados.*

4. Manual de usuario

Se elaboró un manual de usuario con el fin de contribuir a los interesados en la utilización de la Solución de BI que se implementó.

5. Técnicas de procesamiento de datos.

Luego de recolectar la información utilizando las herramientas descritas, analizamos los datos de la investigación en relación a las variables de estudio. Para ello se utilizaron herramientas estadísticas de la suite de MICROSOFT EXCEL.

Tabla 4.
Matriz de Consistencia

| <u>FORMULACIÓN DEL PROBLEMA</u> | | <u>MÉTODOLÓGIA DE INVESTIGACIÓN</u> | | | |
|---|---|---|---|---|--|
| ¿De qué manera se puede apoyar la toma de decisiones en la Gerencia Regional de Agricultura - Lambayeque? | | <u>TIPO DE INVESTIGACIÓN</u> | | | |
| | | Investigación Cuantitativa Pre-experimental | | | |
| <u>OBJETIVO GENERAL</u> | <u>MÉTODO</u> | <u>DESCRIPCIÓN</u> | | | |
| Implementar una Solución de Inteligencia de negocios basada en técnicas de Minería de datos, para apoyar la toma de decisiones en la Gerencia Regional de Agricultura - Lambayeque. | Análítico | Estudio y análisis de los procesos, actividades y herramientas para la comprensión de requerimientos en la Dirección Ejecutiva. | | | |
| | Deductivo | Plan y estrategia para el planteamiento de la propuesta de solución ante el problema. | | | |
| | Implementación | Solución de BI para el análisis y toma de decisiones en la Dirección Ejecutiva. | | | |
| | <u>TÉCNICAS</u> | <u>INSTRUMENTOS</u> | <u>ELEMENTOS DE LA POBLACIÓN</u> | <u>PROPÓSITO</u> | |
| | Observación Documental | Documento | Redacción de la documentación prevista en el Desarrollo e Implementación de la misma. | Documentar Solución. | |
| | Entrevista | Guía de entrevista | * Director del Área de Información Agraria. * Encargado del Centro de Sistemas de Información. | Análisis Operacional. | |
| | Observación | Observación | Se observó el entorno de trabajo y actividades realizadas en el análisis y criterios de información requeridas por los usuarios para la toma de decisiones. | Estado Situacional de la Información. | |
| | Encuesta | Cuestionario | Actores (usuarios) de las Cadenas Agroproductivas. | Validar Solución. | |
| <u>OBJETIVOS ESPECÍFICOS</u> | <u>DESCRIPCIÓN DEL LOGRO DE LOS OBJETIVOS ESPECÍFICOS</u> | | <u>INDICADORES</u> | | |
| 1. Determinar la técnica de minería de datos que más se ajuste a los datos históricos de la variable en estudio: Producción (t.). | 1. | Técnica de Minería de datos cuyo pronostico sea el más cercano a los datos históricos de la variable en estudio. | 1.1. | Número de técnicas de Minería de datos evaluadas | |
| 2. Implementar la técnica de minería de datos que permita predecir la Producción de los Cultivos. | 2. | Predicción de la producción de los cultivos mediante la técnica de Minería de datos implementada. | 1.2. | Medida absoluta: MSE – RMSE | |
| 3. Implementar informes analíticos en base a las necesidades de información requeridas. | 3. | Informes analíticos implementados según las necesidades de información requeridas. | 1.3. | Medida porcentual: MAPE – sMAPE | |
| 4. Lograr una validación positiva de la Solución de BI por parte de los usuarios. | 4.1. | Calidad de información reportada por la solución de BI. | 1.4. | Medida relativa: MASE | |
| | 4.2. | Satisfacción de los usuarios con respecto a la información requerida. | 2. | Reporte de Predicción de Cultivos. | |
| | | | 3. | Número de informes analíticos implementados en base a las necesidades de información requerida. | |
| | | | 4.1. | Nivel de calidad de la información reportada por la Solución de BI. | |
| | | | 4.2. | Nivel de satisfacción del usuario con respecto a la información requerida. | |

5. Consideraciones éticas

Posteriormente, se enumeran los aspectos que se han tomado en consideración para la protección y el bienestar de los copartícipes de esta investigación:

- Las políticas internas de administración de datos dentro de la Gerencia han sido tomadas en consideración por la Ley N° 27806, Ley de Transparencia y Acceso a la Información Pública y decreto legislativo N° 1353 que crea la Autoridad Nacional de Transparencia y acceso a la Información Pública, fortalece el régimen de protección de datos personales y la regulación de la Gestión de Interés.
- Las disposiciones para el desarrollo de la Solución se han dado conforme con la información declarada por la Dirección Ejecutiva aseverando la veracidad de la investigación.
- Privacidad y conformidad de cada una de las encuestas y entrevistas aplicadas tanto en el pre-test como en el post-test, así como también en la integridad para el análisis de los resultados obtenidos.
- Limitación mediante accesos para los interesados finales que utilizarán la plataforma.
- Este proyecto de investigación es original en su totalidad y se han citado de carácter oportuna los trabajos e ideas de otros autores.
- Las pruebas que se efectuaron para su posterior valoración de la usabilidad y desempeño del sistema se realizaron con el consentimiento previo tanto del departamento como de la población muestral estudiada.

IV. Resultados

En base a la metodología utilizada

1.1 Planificación del Proyecto (RALPH KIMBALL)

1.1.1 Alcance

El proceso de gestión de la información en la Dirección Ejecutiva de Información Agraria de la Gerencia Regional de Agricultura localizada en la Ciudad de Chiclayo.

1.1.2 Propósito

Proporcionar a los miembros de la Dirección Ejecutiva de Información Agraria en estudio una herramienta que les permita analizar la realidad del proceso de gestión de la información de diferentes perspectivas empresariales. Asimismo, hemos tratado de brindarle una disponibilidad diligente a la información siempre y cuando sea necesario, ya que estarán disponibles a cualquier hora del día para alegar a sus preguntas de manera específica, visto que pueden formular las combinaciones que deseen en función de sus necesidades para satisfacer sus necesidades. Todo ello tenía como finalidad servir de soporte para gestionar esa información en post de tomar decisiones adecuadas.

1.1.3 Presentación de la Dirección Ejecutiva DEIA

La GRA Lambayeque, a través de la Dirección Ejecutiva de Información Agraria, para la generación, unificación, análisis y transmisión de información pertinente en el ámbito agrario, cuenta con personal calificado para el acopio de la información estadística, distribuida en el ámbito de las tres Agencias, definidas por sus características agrarias y medioambientales, correspondiente a las provincias: Chiclayo, Lambayeque y Ferreñafe.

1.1.4 Desarrollo del Proyecto

Una planeación completa del proyecto se asignará, se determinará y detallará en cada fase, así como los entregables y documentación para contribuir a cada una de ellas.

1.2 Definición de requerimientos del Negocio (RALPH KIMBALL)

a) Requerimientos funcionales:

- ✓ Reportes de indicadores de producción que incluya:
 - Variación de producción y rendimiento de cultivos por año, mes y semana y evaluadas a través de las perspectivas de Valle, Tiempo, Tipo de Consumo, Cultivo y Campaña.
 - Avance de Cosecha (ha) de Cultivos por tipo, según Valles y/o Provincias de una determinada Campaña Agrícola o Periodo de Tiempo.
 - Avance de Siembras (ha) de Cultivos según Valles o Provincias de una determinada Campaña Agrícola o Periodo de Tiempo.
 - Avance de Hectáreas Instaladas (Sembradas) (ha) por Provincia según Cultivo de una determinada Campaña Agrícola y/o Periodo de Tiempo.
 - Distribución porcentual de las Áreas Instaladas (ha) de los Cultivos según el Tipo de la Región en una determinada Campaña Agrícola y/o Periodo de Tiempo.
 - Series Históricas de Siembra (ha) según Cultivo de una determinada Campaña Agrícola y/o Periodo de Tiempo.
 - Evolución de las Áreas Instaladas según Cultivo en la Región de una determinada Campaña Agrícola y/o Periodo de Tiempo.
 - Acumulados de Producción en (t) y Cosechas (ha) de Cultivos de una determinada Campaña Agrícola y/o Periodo de Tiempo.
 - Semaforización de Producción y Rendimiento de Cultivos por Valle y/o Cultivo de una determinada Campaña Agrícola y/o Periodo de Tiempo.
 - Mapeo de Zonas de Sembríos y Cosechas de la Región de Lambayeque.
- ✓ Reporte de tendencias que incluya: Predicción de la variable de Producción de Cultivos en la Región de Lambayeque.
- ✓ Reporte multidimensional que incluya: Evaluación de la Producción y Rendimientos a través de las perspectivas del Negocio requeridas por los dueños del mismo: Valle, Tiempo, Tipo de Consumo, Cultivo y Campaña.

b) Requerimientos no funcionales:

- ✓ Disponibilidad las 24 horas del día.
- ✓ Celeridad en la generación de consultas.

- ✓ Manejo único por usuarios específicos de la Gerencia (Ingenieros de sistemas y alta dirección).
- ✓ Sea fácil de emplear para generar informes.
- ✓ Asumir costos de realización y mantenimiento bajos.
- ✓ Ser escalable en el tiempo (Mediante un posible crecimiento del DataMart hacia un Data Warehouse).

c) Matriz bus (proceso/dimensiones)

La matriz bus se usó para detectar qué perspectivas guardaban interacción con el proceso empresarial a evaluar.

*Tabla 5.
Matriz Bus del Proceso Siembra*

| PROCESO DEL NEGOCIO | DIMENSIONES | | | | |
|------------------------|-------------|--------|--------------|---------|---------|
| | Valle | Tiempo | Tipo_Consumo | Cultivo | Campaña |
| Hecho_Siembra | X | X | X | X | X |

*Tabla 6.
Matriz Bus del Proceso Cosecha*

| PROCESO DEL NEGOCIO | DIMENSIONES | | | | |
|------------------------|-------------|--------|--------------|---------|---------|
| | Valle | Tiempo | Tipo_Consumo | Cultivo | Campaña |
| Hecho_Cosecha | X | X | X | X | X |

Mostramos las dimensiones de análisis del proceso en estudio y su respectiva granularidad:

- Dimensión Valle: Nombre del Valle
- Dimensión Cultivo: Nombre del Cultivo
- Dimensión Campaña: Descripción de la Campaña
- Dimensión Tipo-Consumo: Nombre del Tipo de Consumo
- Dimensión Tiempo: Fecha

d) Dimensiones y sus atributos

Cada dimensión tiene atributos especiales que son relevantes para el proceso de gestión de información agrícola de la unidad en estudio. La Tabla VII nos muestra la distribución de ésta y las jerarquías existentes, lo que nos permitió evaluar la realidad examinada utilizando relaciones uno a muchos análogas a las de bases de datos transaccionales, pero aquí se realizaron entre los atributos de las dimensiones.

*Tabla 7.
Dimensiones: sus atributos y jerarquías*

| Dimensión | Atributos | Jerarquías |
|------------------|-----------------------|---------------------|
| Cultivo | ✓ Key_cultivo | Jerarquía 1: |
| | ✓ codigo_oltp | . valle |
| | ✓ nombre_cultivo | .. provincia |
| | ✓ nombre_tipo_cultivo | ... distrito |
| | | nombre_cultivo |
| Tiempo | | tipo |
| | | Campaña |
| | ✓ key_tiempo | Jerarquía 1: |
| | ✓ fecha | . año |
| | ✓ mes | .. mes |
| Valle | ✓ año | ... fecha |
| | ✓ key_valle | No tiene |
| | ✓ codigo_oltp | |
| | ✓ nombre_valle | |
| | ✓ distrito | |
| Campaña | ✓ provincia | |
| | ✓ key_campaña | No tiene |
| | ✓ descripcion_campaña | |
| Tipo_Consumo | ✓ key_tipo | No tiene |
| | ✓ codigo_oltp | |
| | ✓ tipo_consumo | |

e) Hecho y sus medidas

La tabla de hechos o *fact table* agrupó las medidas que fueron utilizadas para evaluar el proceso de gestión de información de la entidad en estudio. La Tabla VIII y IX nos muestra la distribución de los atributos del hecho Inf. Agrícola.

Tabla 8.
Hecho: sus medidas y atributos

| Hecho | Atributos | Medidas |
|----------------------|---------------------|--------------------|
| Hecho_Siembra | ✓ key_hecho_siembra | ✓ cantidad_siembra |
| | ✓ key_valle | |
| | ✓ key_tiempo | |
| | ✓ key_consumo | |
| | ✓ key_cultivo | |
| | ✓ key_campaña | |

Tabla 9.
Hecho: sus medidas y atributos

| Hecho | Atributos | Medidas |
|----------------------|---------------------|-----------------------|
| Hecho_Cosecha | ✓ key_hecho_cosecha | ✓ cantidad_cosecha |
| | ✓ key_valle | ✓ precio_chacra_kg |
| | ✓ key_tiempo | ✓ rendimiento |
| | ✓ key_consumo | ✓ producción |
| | ✓ key_cultivo | ✓ sup. en producción |
| | | ✓ sup. en crecimiento |
| | ✓ sup. perdida | |
| | ✓ sup. verde | |

Finalmente, en la Figura 1 defino el diseño lógico de cada una de las dimensiones y del hecho utilizado en el desarrollo del proyecto.

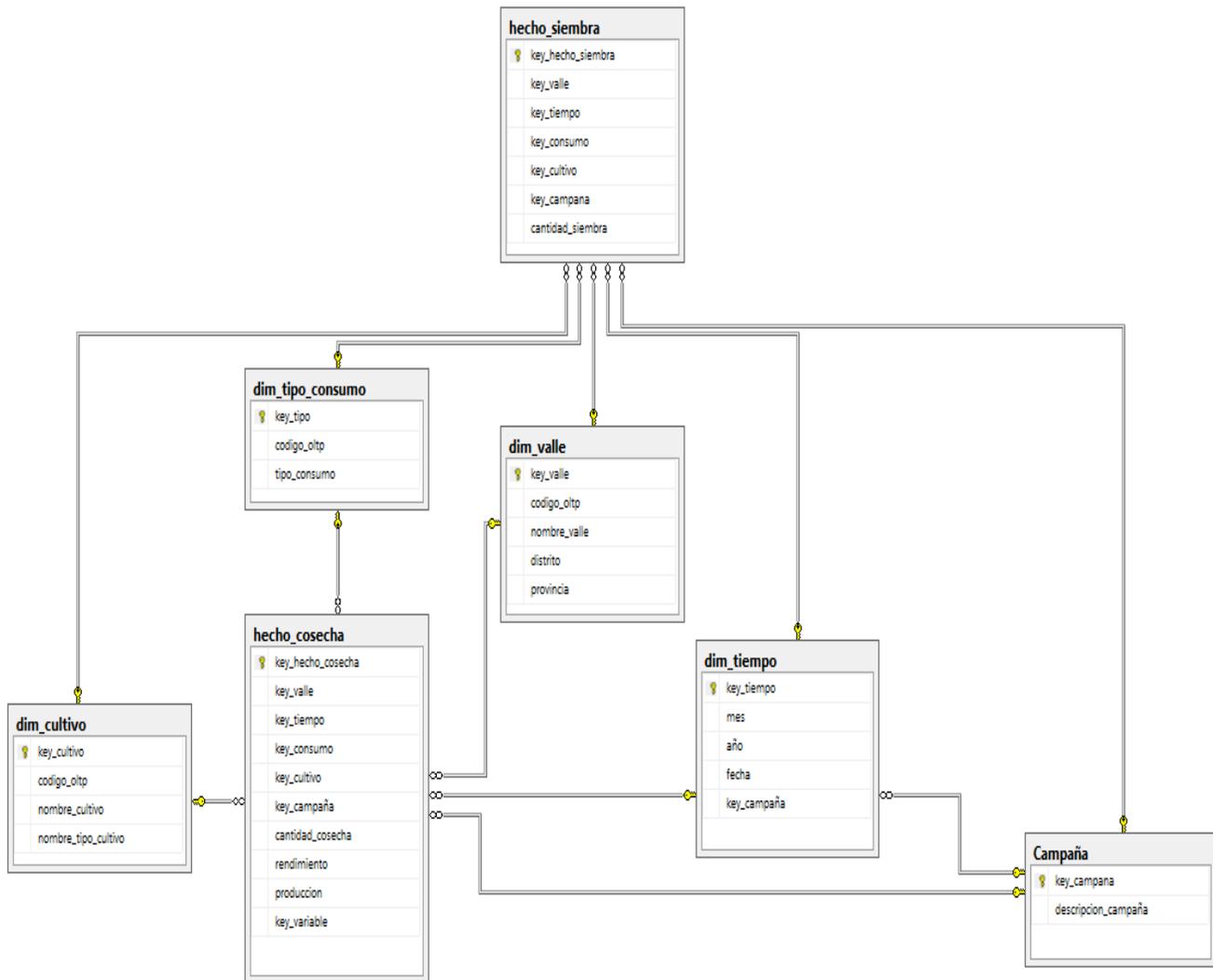


Figura 1. Diseño lógico del DataMart DM_DEIA

1.3 Diseño físico (RALPH KIMBALL)

A continuación, se muestra el modelo físico de la base de datos multidimensional, que apoyó el diseño lógico; en otras palabras, la dependencia entre las dimensiones y el hecho (de uno a muchos), la relación entre dimensiones (de uno a muchos) cuando se trata de esquemas copo de nieve, especificaciones de restricciones en los campos de las dimensiones y del hecho (primary key, foreign, etc.), los estándares de nomenclatura que se definen para los campos de las tablas, entre otras cosas. Las Tablas de la X a la XV muestra modelos físicos tanto para las dimensiones como para los hechos de siembra y cosecha.

Tabla 10.
Modelo Físico Tabla dim_Cultivo

| Dimensión | Atributo | Tipo Dato | Longitud | Valor Aceptados | Nulo | Autoincremental | Restricción |
|--------------------|----------------------------|-----------------|----------|---------------------------|----------|-----------------|-------------|
| | key_cultivo | Int | | | No | Si | PK |
| dim_Cultivo | Nombre | varchar | 100 | | No | No | |
| | Tipo | varchar | 100 | TRANSITORIO PERMANENTE | No No | No | |
| | Origen de datos BD antigua | Documento Excel | | Tablas | | | - |
| | Origen de datos BD nueva | dm_agricultura | | Tablas | | | dim_cultivo |

Tabla 11.
Modelo Físico Tabla dim_Campaña

| Dimensión | Atributo | Tipo Dato | Longitud | Valor Aceptados | Nulo | Autoincremental | Restricción |
|--------------------|----------------------------|-----------|----------|-----------------|--------|-----------------|-------------|
| dim_Campaña | Key_campaña | int | | | No | Si | PK |
| | descripcion_campaña | varchar | 9 | | No | No | |
| | Origen de datos BD antigua | | | Documento Excel | Tablas | | - |
| | Origen de datos BD nueva | | | dm_agricultura | Tablas | | Campaña |

Tabla 12.
Modelo Físico Tabla dim_Valle

| Dimensión | Atributo | Tipo Dato | Longitud | Valor Aceptados | Nulo | Autoincremental | Restricción |
|-------------------|----------------------------|-----------|----------|-----------------|--------|-----------------|-------------|
| dim_Ubigeo | Key_valle | int | | | No | Si | PK |
| | nombre_valle | varchar | 100 | | No | No | |
| | Distrito | varchar | 100 | | No | No | |
| | Provincia | Varchar | 100 | | No | No | |
| | Origen de datos BD antigua | | | Documento Excel | Tablas | | - |
| | Origen de datos BD nueva | | | dm_agricultura | Tablas | | dim_valle |

Tabla 13.
Modelo Físico Tabla dim_Consumo

| Dimensión | Atributo | Tipo Dato | Longitud | Valor Aceptados | Nulo | Autoincremental | Restricción |
|--------------------|----------------------------|-----------------|----------|-----------------|------|------------------|-------------|
| dim_Consumo | Key_tipo | Int | | | No | Si | PK |
| | tipo_consumo | Varchar | 100 | | No | No | |
| | Origen de datos BD antigua | Documento Excel | | Tablas | | - | |
| | Origen de datos BD nueva | dm_agricultura | | Tablas | | dim_tipo_consumo | |

Tabla 14.
Modelo Físico Tabla hecho_Siembra

| Dimensión | Atributo | Tipo Dato | Longitud | Valor Aceptados | Nulo | Autoincremental | Restricción |
|--------------------------|----------------------------|----------------|----------|-----------------|--------|-----------------|-------------|
| hecho_siembra | key_hecho_siembra | int | | | No | Si | PK |
| | key_valle | int | | | No | No | PK + FK |
| | key_tiempo | int | | | No | No | PK + FK |
| | key_consumo | int | | | No | No | PK + FK |
| | key_cultivo | int | | | No | No | PK + FK |
| | key_campaña | int | | | No | No | PK + FK |
| | Cantidad_siembra | Numeric(18, 2) | | | No | No | |
| | Intención_siembra | Numeric(18, 2) | | | No | No | |
| | Origen de datos BD antigua | | - | | Tablas | | - |
| Origen de datos BD nueva | | dm_agricultura | | Tablas | | hecho_siembra | |

Tabla 15.
Modelo Físico Tabla hecho_Cosecha

| Dimensión | Atributo | Tipo Dato | Longitud | Valor Aceptados | Nulo | Autoincremental | Restricción |
|----------------------------|-------------------------|----------------|----------|-----------------|------|-----------------|-------------|
| | key_hecho_cosecha | int | | | No | Si | PK |
| | key_valle | int | | | No | No | PK + FK |
| | key_tiempo | int | | | No | No | PK + FK |
| | key_consumo | int | | | No | No | PK + FK |
| | key_cultivo | int | | | No | No | PK + FK |
| hecho_cosecha | Cantidad_cosecha | Numeric(18, 2) | | | No | No | |
| | Precio_chacra_kg | Numeric(18, 2) | | | No | No | |
| | Rendimiento | Numeric(18, 2) | | | No | No | |
| | Producción | Numeric(18, 2) | | | No | No | |
| | Sup_producción | Numeric(18, 2) | | | No | No | |
| | Sup_crecimiento | Numeric(18, 2) | | | No | No | |
| | Sup_perdida | Numeric(18, 2) | | | No | No | |
| | Sup_verde | Numeric(18, 2) | | | No | No | |
| Origen de datos BD antigua | | - | | Tablas | | - | |
| Origen de datos BD nueva | | dm_agricultura | | Tablas | | hecho_cosecha | |

1.4 Diseño e implementación del subsistema ETL (RALPH KIMBALL)

a) Diseño y desarrollo de presentación de datos

Aquí se ha realizado el proceso de extracción, transformación y carga (ETL), con el objeto de poblar en paralelo de las dimensiones para que al final se logre hacer lo mismo en el hecho siembra y cosecha. De esta manera la información es suficiente para obtener una estandarización de ésta (información) en el DataMart. Se han realizados los procesos para cumplir con el poblamiento del DataMart de la siguiente manera y mostrados además en la Figura 2:

- ✓ Poblamiento de dimensión cultivo.
- ✓ Poblamiento de dimensión tiempo.
- ✓ Poblamiento de dimensión campaña.
- ✓ Poblamiento de dimensión valle.
- ✓ Poblamiento de dimensión tipo_consumo.
- ✓ Poblamiento de hecho siembra.
- ✓ Poblamiento de hecho cosecha.

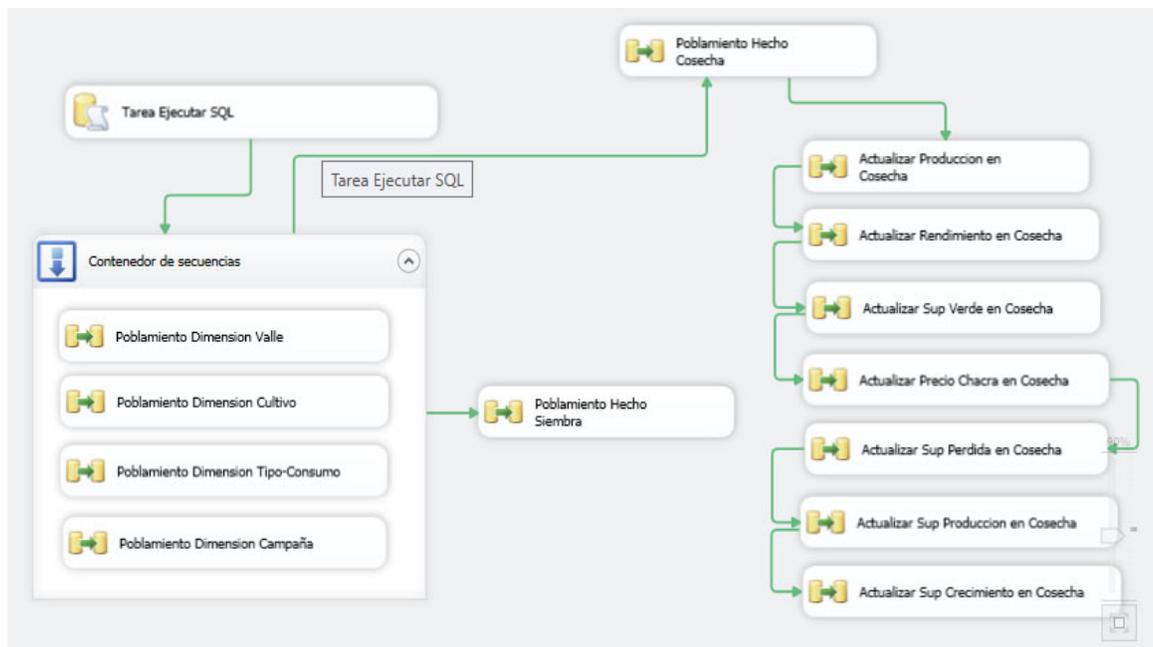


Figura 2. Proceso ETL con la herramienta SQL Server Data Tools

1.5 Implementación (RALPH KIMBALL)

Habiendo culminado con el proceso de ETL e implementado el DataMart, se procedió a la elaboración de Dashboard y reportes analíticos requeridos por la Dirección. La implementación de las interfaces se ha ajustado a los arquetipos que son confirmados por los interesados involucrados, y para ello se utilizó la Suite de Power BI para su posterior realización.

Subsiguientemente, se han descrito algunas de las interfaces requeridas para permitirles un mejor análisis y comportamiento de la información obtenida por el área en estudio, mejorando así una adecuada gestión y toma de decisiones. Esta información brindada podrá ser consultada por todos los actores de la cadena agro productivas, sin embargo; se consideró inicios de Sesión para un mayor control de acceso y validación de usuarios.

Además, tendremos usuarios AD HOC que tendrán la labor de diseñar nuevos reportes en base a necesidades según las dimensiones consideradas o actualizaciones de información posterior, entre otros. El acceso se hizo a través de la siguiente interfaz con un módulo para el registro de nuevos usuarios:

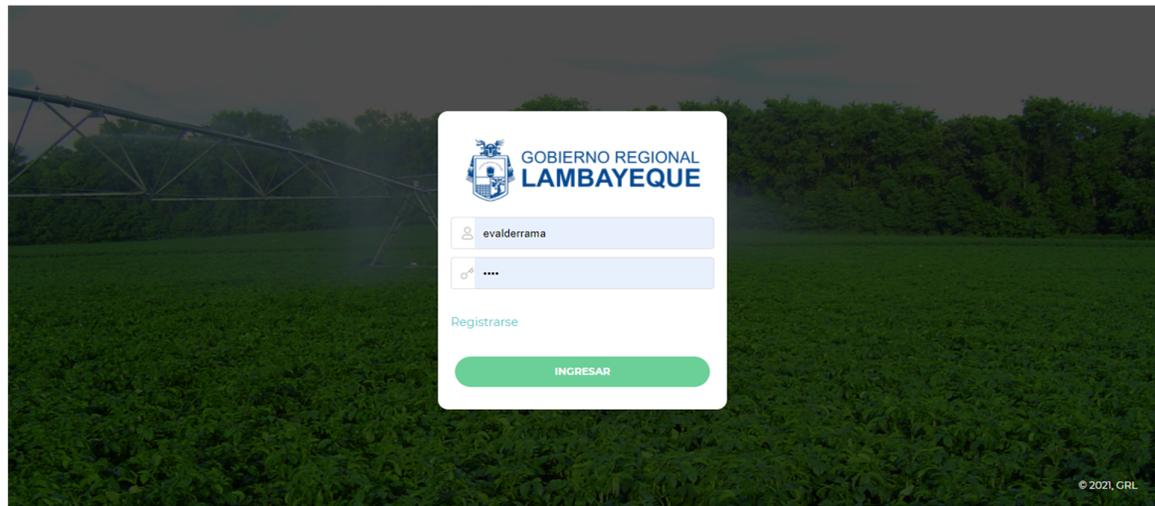


Figura 3. Inicio de Sesión de la Plataforma

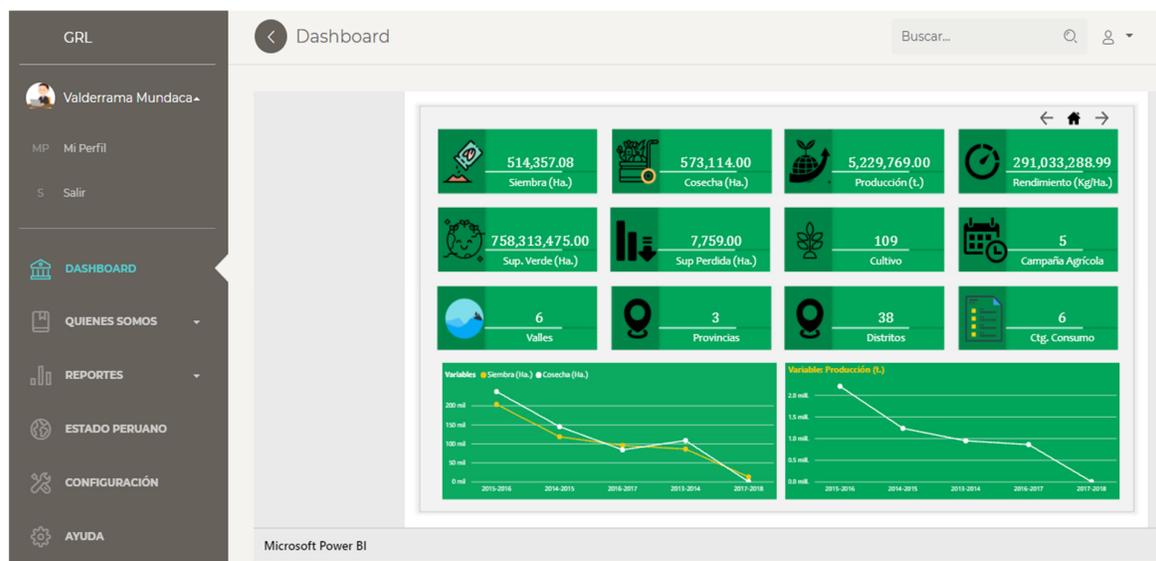


Figura 4. Navegación del Sistema

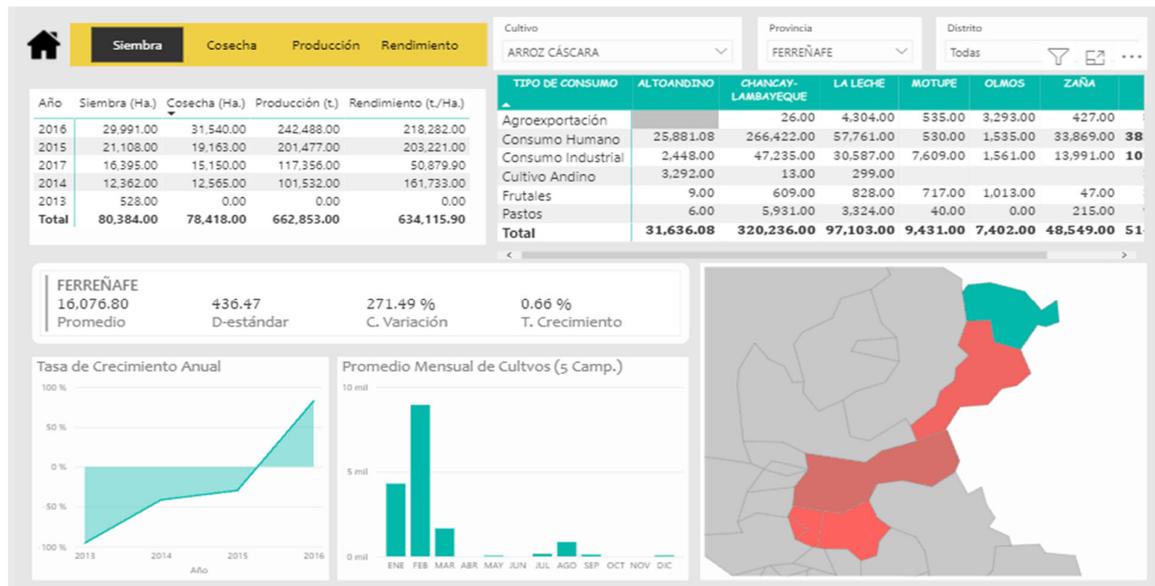


Figura 5. Reporte estadístico y visualización GIS

En el módulo de Pronóstico de la Figura 6, donde tendremos la opción de predecir la producción de cultivos según el plazo a pronosticar; nos mostrara los resultados en una tabla con los ítems pronosticados.

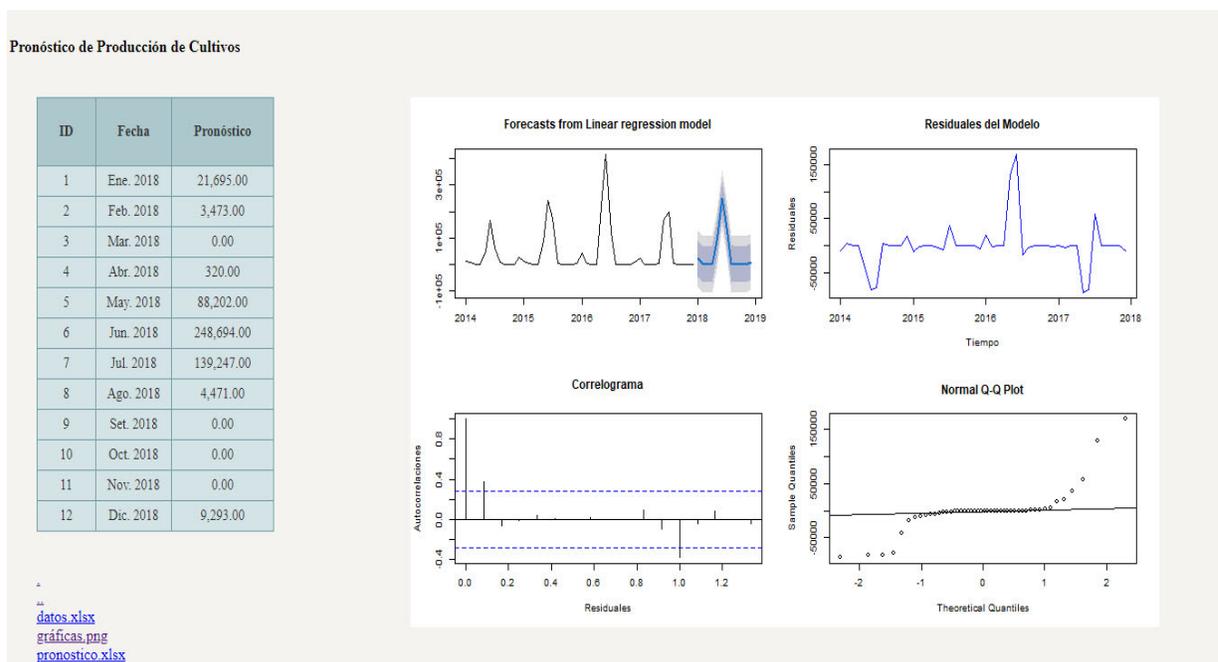


Figura 6. Módulo de pronóstico

1.6 Modelado y desarrollo de aplicaciones analíticas (CRISP-DM)

R Studio presenta diferentes librerías que tienen implementados modelos y técnicas de predicción de Series de tiempo; como métodos simples, regresión lineal, métodos de suavizamiento (Holt y Holt-Winters), promedios móviles (ARIMA), redes neuronales autorregresivas; entre otros. Para el análisis de Minería de datos se realizará la proyección de la producción de cultivos para los 12 últimos meses del valor histórico y se comparará con los valores reales.

Si bien la diferencia entre el ajuste y el grado de error de los modelos es mínima, para efectuar los pronósticos de producción mensual de cultivos, se utilizará el de Regresión lineal con ajuste estacional, pues entre los demás es el que mejor modela los datos en estudio y posee menor error y mayor ajuste. En las siguientes Tablas XVI Y XVII se presentan los resultados para c/modelo.

Tabla 16.
Medidas de precisión del modelo ARIMA (0,0,1) (0,1,0) [12]

| MODELO: ARIMA (0,0,1) (0,1,0) [12] | | | | | | | | |
|---|-------------------|-------------------|----------------|----------------|----------------------|----------------|-------------|--------------|
| Periodo | Cant. Real | Pronóstico | CFE | MAD | MSE | MASE | MAPE | SMAPE |
| 1 | 22,480 | 17,830.00 | 4,650 | 4,650 | 21,622,500 | 0.3729 | 21% | 23% |
| 2 | 2,680 | 2,688.00 | -8 | 8 | 64 | 0.0006 | 0% | 0% |
| 3 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 4 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 5 | 31,790 | 62,075.00 | -30,285 | 30,285 | 917,181,225 | 2.4290 | 95% | 65% |
| 6 | 167,628 | 231,188.00 | -63,560 | 63,560 | 4,039,873,600 | 5.0977 | 38% | 32% |
| 7 | 197,617 | 151,843.00 | 45,774 | 45,774 | 2,095,259,076 | 3.6712 | 23% | 26% |
| 8 | 7,190 | 7,497.00 | -307 | 307 | 94,249 | 0.0246 | 4% | 4% |
| 9 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 10 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 11 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 12 | 27,280 | 22,245.00 | 5,035 | 5,035 | 25,351,225 | 0.4038 | 18% | 20% |
| Suma de errores | | | -38,701 | 149,619 | 7,099,381,939 | 12.0000 | 200% | 170% |
| | | | -38,701 | 12,468 | 591,615,162 | 1.0000 | 17% | 14% |

En la Figura 7, se observa el gráfico de la serie original y los pronósticos para el periodo 2017 con sus respectivos intervalos de confianza. Estos pronósticos se observan que mantienen el patrón estacional y la tendencia creciente de la serie de datos.

Tabla 17.
Medidas de precisión del modelo de Regresión lineal con ajuste estacional

| MODELO: REGRESIÓN LINEAL CON AJUSTE ESTACIONAL | | | | | | | | |
|---|-------------------|-------------------|----------------|----------------|----------------------|----------------|-------------|--------------|
| Periodo | Cant. Real | Pronóstico | CFE | MAD | MSE | MASE | MAPE | SMAPE |
| 1 | 22,480 | 15,766.00 | 6,714 | 6,714 | 45,077,796 | 0.6604 | 30% | 35% |
| 2 | 2,680 | 2,737.00 | -57 | 57 | 3,249 | 0.0056 | 2% | 2% |
| 3 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 4 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 5 | 31,790 | 57,006.00 | -25,216 | 25,216 | 635,846,656 | 2.4803 | 79% | 57% |
| 6 | 167,628 | 209,049.00 | -41,421 | 41,421 | 1,715,699,241 | 4.0742 | 25% | 22% |
| 7 | 197,617 | 153,123.00 | 44,494 | 44,494 | 1,979,716,036 | 4.3765 | 23% | 25% |
| 8 | 7,190 | 7,571.00 | -381 | 381 | 145,161 | 0.0375 | 5% | 5% |
| 9 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 10 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 11 | 0 | 0.00 | 0 | 0 | 0 | 0.0000 | 0% | 0% |
| 12 | 27,280 | 23,564.00 | 3,716 | 3,716 | 13,808,656 | 0.3655 | 14% | 15% |
| Suma de errores | | | -12,151 | 121,999 | 4,390,296,795 | 12.0000 | 177% | 161% |
| | | | -12,151 | 10,167 | 365,858,066 | 1.0000 | 15% | 13% |

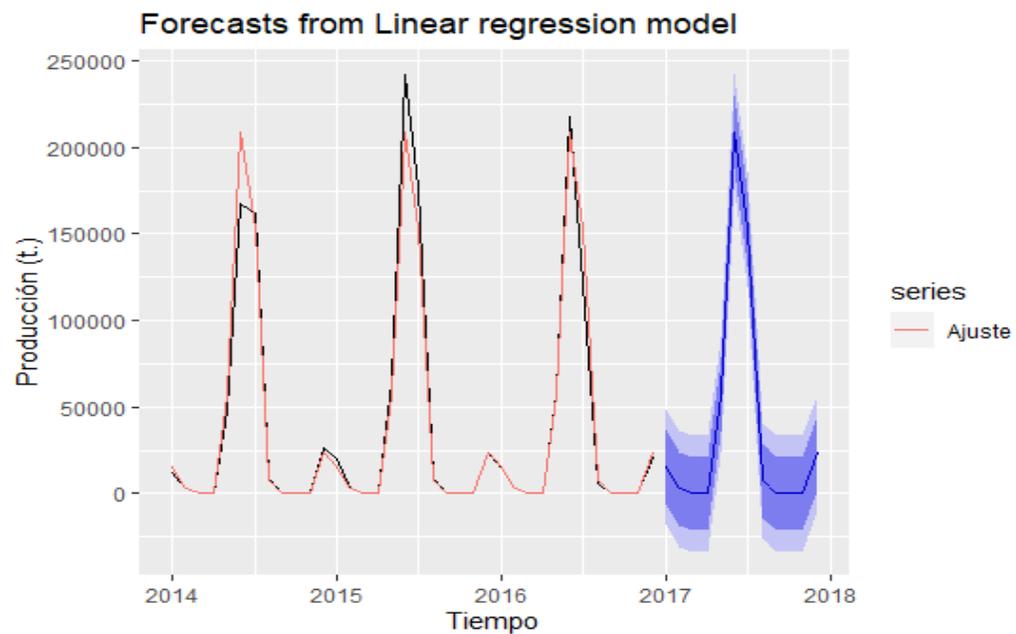


Figura 7. Pronóstico de producción (t.) de Arroz cáscara para el periodo 2017

1.7 Diseño de la Arquitectura Técnica (RALPH KIMBALL)

Para comprender el proceso realizado para la elaboración del Modelo dimensional y los informes detallados, en la Fig. 75 se muestra el flujo para el desarrollo de la Solución de BI, y como estas se relacionan con las distintas herramientas.

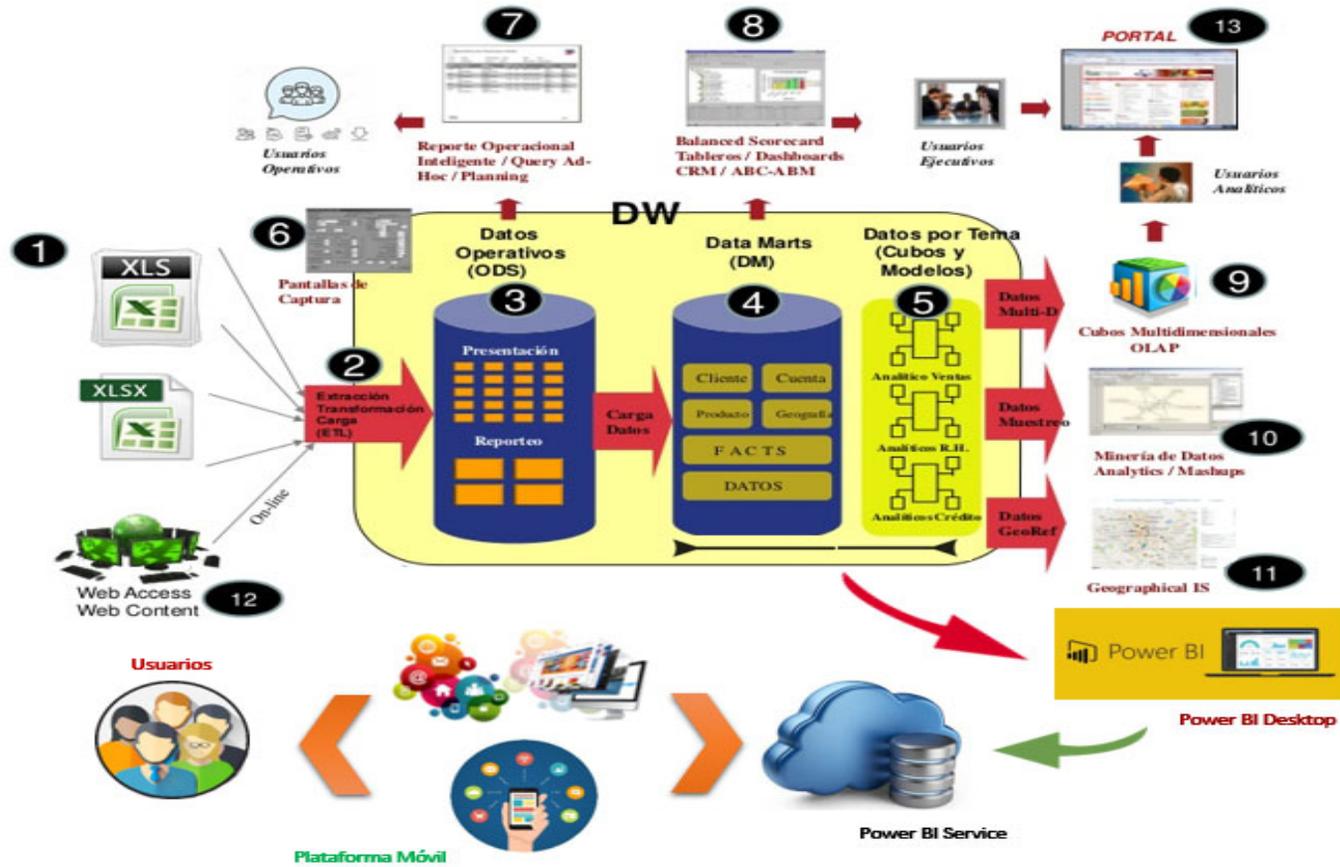


Figura 8. Arquitectura Técnica para el desarrollo de la Solución BI

En base a los objetivos de la investigación:

1. Determinar la técnica de Minería de datos que más se ajuste a los datos históricos de la variable en estudio: Producción (t.)

Previo a la aplicación de minería de datos se realizó un análisis y descripción del comportamiento de la Producción agrícola basada en la data histórica desde los años 2013 al 2017 registradas por la DEIA de Gerencia Regional de Agricultura-Lambayeque. El cumplimiento de este objetivo se detalla en el ítem **4.9. Fase #09: Modelado y desarrollo de aplicaciones analíticas, sección Construcción y evaluación del modelo, en el Capítulo IV. Resultados.**

Tras la aplicación de minería de datos y encontrar patrones de comportamiento que determinen valores futuros de la variable en estudio, se basó en el modelo de Regresión lineal con ajuste estacional previamente analizada en la fase de análisis de la metodología y validar el modelo con el cálculo de los errores de pronóstico, se exportaron los resultados mediante un script en lenguaje R consultados mediante argumentos para ser leído por la aplicación.

2. Implementar la técnica de Minería de Datos que permita predecir la Producción de Cultivos.

Teniendo los resultados exportados en formato Excel y graficas de la predicción con la técnica seleccionada, procederemos a cargarla y mostrarla en la Solución. Para ello se creó un formulario web conformado por un combo-box para seleccionar el cultivo que queremos predecir, una caja de texto en la contendrá el número de periodos a considerar y finalmente, un botón que me muestre los resultados de la consulta como visualizaremos en la sgte. figura.

Figura 9. Formulario para el Módulo de Pronóstico

Como resultado nos mostrará la siguiente información, una tabla con los periodos y resultados del pronóstico del cultivo seleccionado, una gráfica mostrando el comportamiento de los datos y el pronóstico final conjuntamente con los residuales, correlograma y normalidad del Modelo.

Asimismo, visualizaremos en la parte lateral izquierda inferior un directorio de archivos en donde podremos descargar la información que se utilizó para el pronóstico como también los resultados del mismo en extensión .xlsx y guardarlo para su posterior uso o estudio.

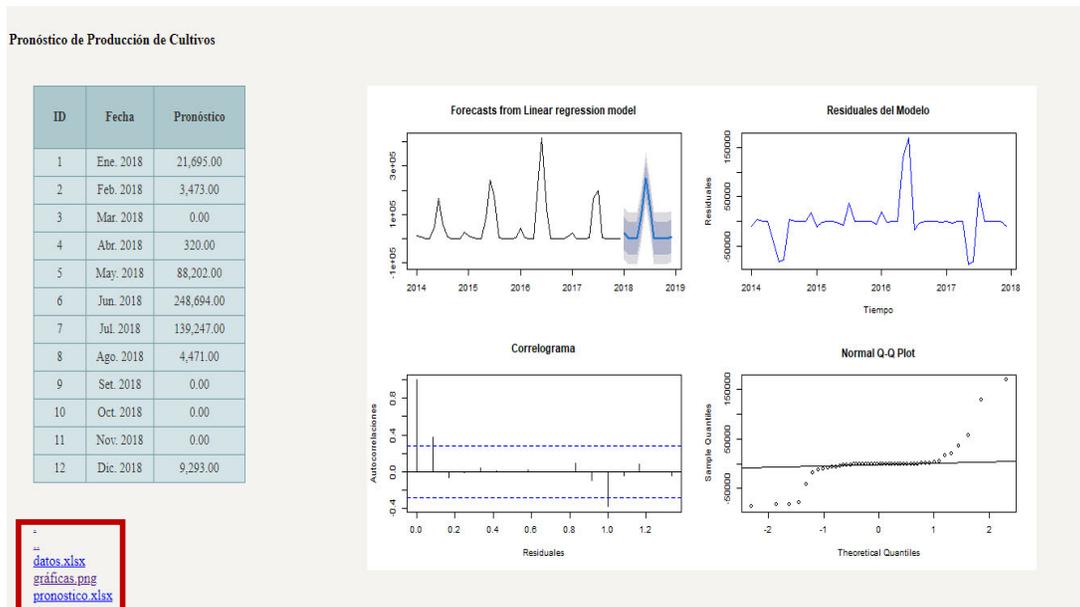


Figura 10. Resultados del Pronóstico de Producción (t.) de cultivo seleccionado

3. Implementar informes analíticos en base a las necesidades de información requeridas.

En base al cumplimiento de la implementación de los DataMart y cubos multidimensionales, se diseñaron prototipos de salida de información que nos facilitaron el desarrollo de la solución. En esta fase, se construyeron reportes analíticos y Dashboard acorde a las necesidades y requisitos de los interesados que les permita tener conocimiento del comportamiento de las distintas variables agroeconómicas mejorando su toma de decisiones en base a dicha información analítica. El cumplimiento de este objetivo se detalla en el ítem 4.6. **Fase #06: Implementación, en el Capítulo IV. Resultados.**

4. Lograr una validación positiva de la Solución de Inteligencia de negocios por parte de los usuarios.

Para el indicador correspondiente a dar respuesta al nivel de calidad de la información entregada, se formularon las preguntas siguientes: ¿El sistema provee reportes de rápida comprensión? Obteniendo como resultado de que el 67% de la población está de acuerdo con la medida de comprensibilidad de la información proporcionada por la Solución de Inteligencia de negocios y un 33%, totalmente de acuerdo.

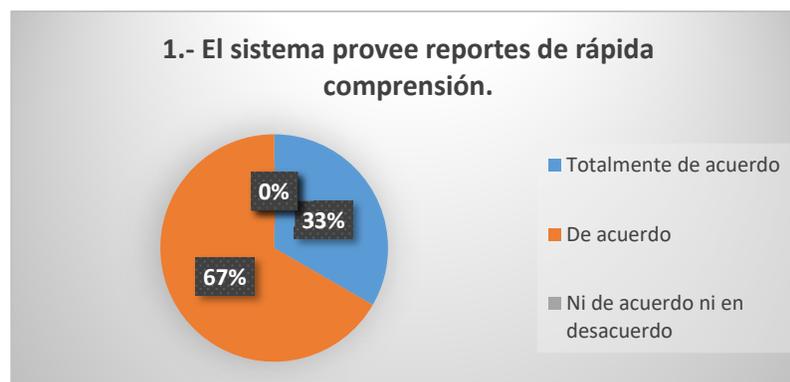


Figura 11. Comprensibilidad de la información

Continuando con la siguiente pregunta, se les realizó a las partes interesadas si: ¿El sistema provee reportes con información actualizada? Y en base a las respuestas obtenidas un 67% ha indicado que están de acuerdo con que la Solución brinda información actualizada conforme a los datos recolectados en las fases previas de implementación y otro 33% están totalmente de acuerdo.

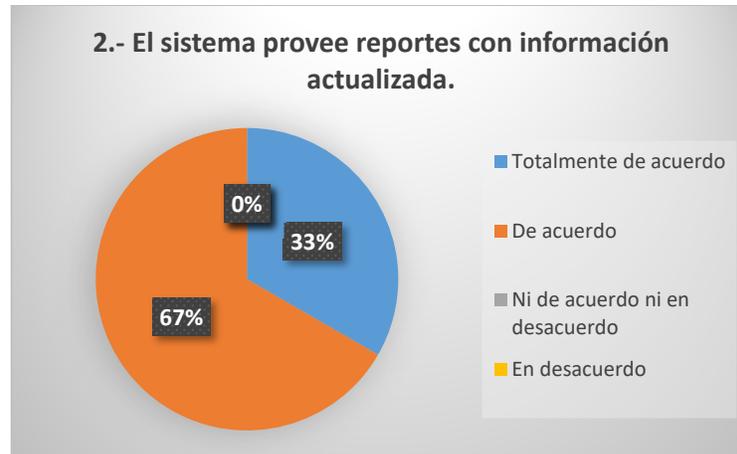


Figura 12. Información actualizada

En el ítem 03 correspondiente a la completitud de la información, tenemos la siguiente pregunta: ¿A través de los reportes obtengo información suficiente y necesaria para abarcar todos los factores críticos que influyen en la correcta ejecución de mis funciones?, teniéndose como respuesta que el 67% de los usuarios indicaron estar de acuerdo con la obtención de información suficiente y necesaria en post de una correcta ejecución de sus actividades y funciones; mientras que con 33%, usuarios totalmente de acuerdo.



Figura 13. Completitud de la información

Siguiendo con el cuestionario de validación, nos encontramos la siguiente pregunta: ¿Los reportes son consistentes, es decir, la información que brinda es congruente con todos los objetivos de negocio? En este caso un 17% no están de acuerdo ni en desacuerdo con la relevancia que tiene la información brindada en la Solución de Inteligencia de negocios, sin embargo, se manifiesta una opinión favorable respecto a las demás partes interesadas con un porcentaje del 83% de la población.



Figura 14. Relevancia de la información

Finalmente, el último ítem correspondiente a la percepción frente al nivel de calidad de la información entregada, nos formula la siguiente pregunta: ¿Los reportes son concisos, puntuales, selectivos y resumidos adecuadamente? Obteniendo como resultado de que el 33% de la población está de acuerdo con la medida de síntesis en la información proporcionada por la Solución de Inteligencia de negocios y un 67%, totalmente de acuerdo.

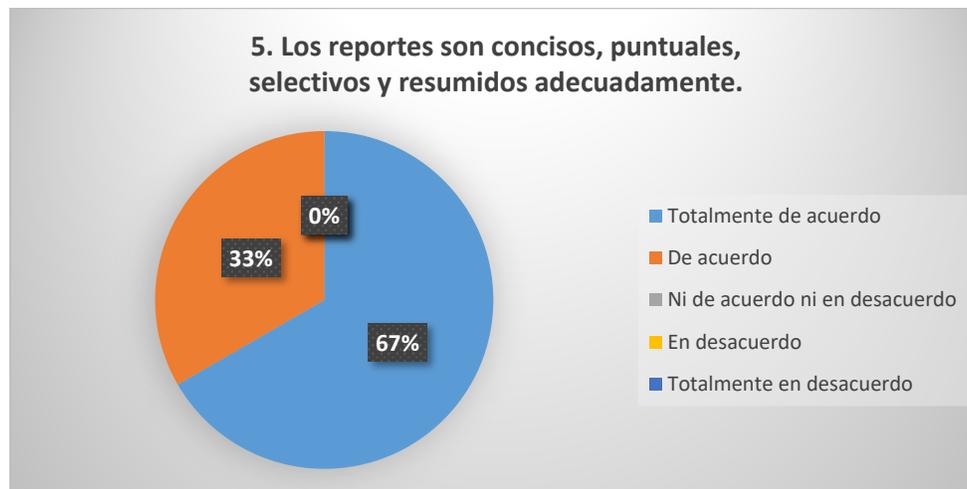


Figura 15. Síntesis de la información

Para el indicador correspondiente a dar respuesta al nivel de satisfacción de la información solicitada, se formularon las preguntas siguientes: ¿Confío plenamente en los reportes que me brinda el sistema? Se repartieron los resultados en fracciones equivalentes entre los usuarios que indicaron que estaban de acuerdo y totalmente de acuerdo con respecto a la confianza emitida de la información proporcionada por la Solución, obteniendo un 50% de la población.



Figura 16. Confianza en la información solicitada

En cuanto los ítems que corresponden a la Eficiencia en la información solicitada, tenemos las siguientes preguntas: ¿Si me fuera a trabajar a una empresa de la competencia, recomendaría que implanten un sistema igual por la eficiencia de sus reportes? y si, ¿Los reportes me permiten adquirir un conocimiento valioso que de otra forma me resultaría muy difícil o imposible de alcanzarla? Obteniendo resultados propicios y positivos de la población que manifestó estar de acuerdo y totalmente de acuerdo con la premisa formulada, presentadas en la Fig. 84 y 85 respectivamente.



Figura 17. Eficiencia en la información



Figura 18. Eficiencia en la información solicitada

En el ítem 09 correspondiente al disfrute en la información solicitada, tenemos la siguiente pregunta: ¿Cuándo estoy usando las funcionalidades de los reportes del sistema, mi estado es de disfrute?, teniéndose como respuesta que el 67% de los usuarios indicaron estar de acuerdo con este apartado; mientras que con un 33%, usuarios totalmente de acuerdo.

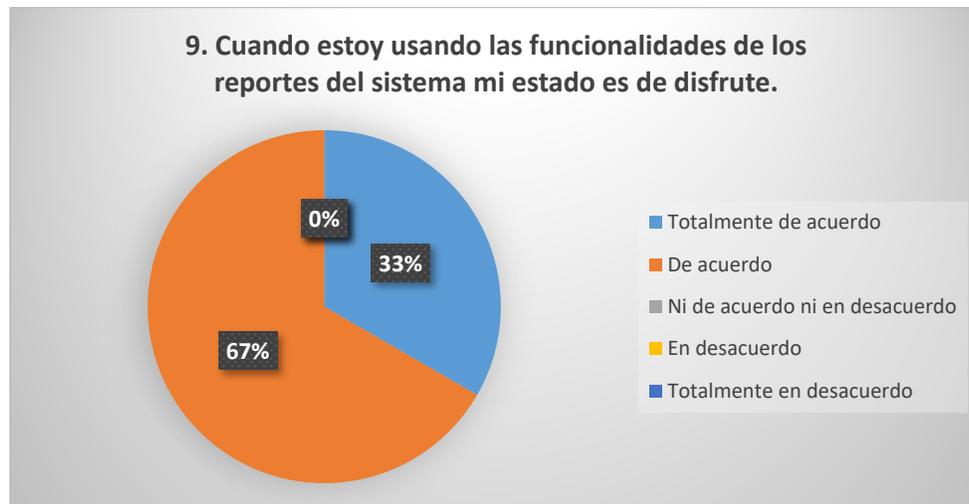


Figura 19. Disfrute en la información solicitada

Finalmente, el último ítem correspondiente a la percepción frente al nivel de satisfacción de la información solicitada, nos formula la siguiente pregunta: En términos generales, ¿Estoy satisfecho con el sistema de información? Obteniendo como resultado de que el 33% de la población está de acuerdo con la medida de satisfacción general con la información proporcionada por la Solución de Inteligencia de negocios y un 67%, totalmente de acuerdo.

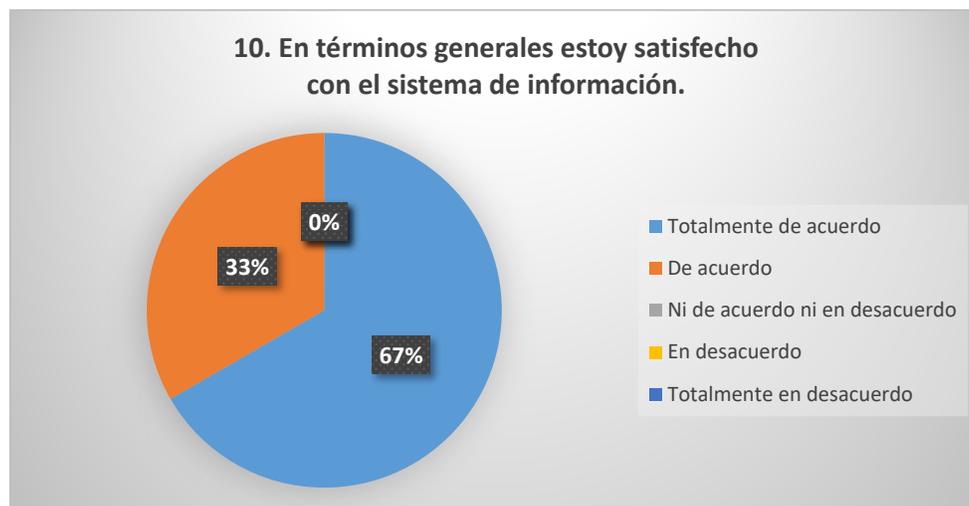


Figura 20. Satisfacción general con la información solicitada

Impactos esperados

1. Impactos económicos

El sistema desarrollado constituye no solo una herramienta para dar sustento al proceso de análisis y toma de decisiones, pues sumado a ello nos ayudará también a gestionar táctica y estratégicamente otras actividades en la DEIA al garantizar información asequible, eliminando algunos retrasos y redistribuyendo al personal hacia otras actividades que proporcionen un valor agregado a la Dirección ejecutiva. Asimismo, la plataforma concede una mejor articulación del flujo de información en los diferentes actores de las cadenas agro productivas.

2. Impactos sociales

La expectativa entorno a la implantación futura de la plataforma desarrollada permita una transición en los procesos internos en la Dirección ejecutiva, facilitando a un incremento en la productividad de los interesados y la innovación en subprocesos de análisis y gestión de datos, rediseñándose y formulando progresos que acrecienten la eficacia y eficiencia del proceso completo y la integración de otras herramientas tecnológicas que mejoren el ecosistema digital de la misma. Por otro lado, con la solución implementada, se tomarán decisiones oportunas en beneficio a disponer de información y almacenes de datos pertinentes para monitorear regularmente el desarrollo de las cadenas productivas y principales variables agroeconómicas, estrategias de comunicación efectiva, campañas para el fortalecimiento de la agricultura en nuestra Región, capacitaciones y asesorías basándose en los reportes analíticos brindados por la Solución, mecanismos y programas para mejorar la información de calidad proporcionada a las partes interesadas de la cadena productiva.

3. Impactos en tecnología

El aporte de la investigación se fundamenta en la innovación tecnológica para el análisis de datos y toma de decisiones adecuadas para las partes interesadas de bajo costo, pues la solución fue desarrollada a medida y fundamentado en las medidas y necesidades del entorno en estudio. Se espera que la implementación de esta solución y su futura integración con los sistemas utilizados por la DEIA representen un antecedente para próximas investigaciones en el uso de soluciones de BI basados en minería de datos.

V. Discusión

Los medios de soporte a la toma de decisiones son los instrumentos que emplean los directores para tomar medidas efectivas, fundamentándose en la hipótesis del resultado. Para poder proponer un modelo para el diseño e implementación de Minería de Datos que más se ajusten a la variable de estudio: Producción (t.), se comparó diferentes modelos predictivos con el fin de encontrar el que mayor se ajuste al comportamiento esperado y cumpla con el criterio de evaluación de los errores de pronóstico. En su investigación, Rudnichenko y Vychuzhanin [6] provee métricas de evaluación utilizadas para la calidad de los modelos regresivos como son el Error cuadráticos medio (MSE) y el Error absoluto medio (MAE) para los cuales fueron tomados en cuenta en esta investigación, similar a Fernández [15] en su examen y comparación de los modelos de ARIMA, Holt-Winters y redes neuronales exploradas para el pronóstico de espárragos con fines industriales en el distrito de Jayanca-Lambayeque. Se concluyó, que el modelo seleccionado es el de Regresión lineal con ajuste estacional, puesto que ostentó un error cuadrático menor a comparación con el enfoque Naive estacional y el modelo ARIMA (0,0,1) (0,1,0) [12], además de tener un valor MAPE, medida porcentual, ligeramente superior al valor obtenido por los demás.

Como segundo objetivo se tuvo que implementar una Solución de BI basada en el modelo de Minería de Datos seleccionado; para ello se tomó en cuenta una metodología híbrida de Ralph Kimball para proyectos de Inteligencia de negocios y la de CRISP-DM en el desarrollo e implementación de Minería de Datos como lo expone Mazon-Olivo et al. [9] al apoyarse de estas técnicas y herramientas para optimizar el proceso de identificación de indicadores claves de rendimiento (KPI) de la producción y comercialización de cacao en la empresa Agrisolsa S.A. en el cantón de Machala-Ecuador y en donde además sirvió como fundamento a la hora de diseñar e implantar los reportes analíticos y de pronóstico según las necesidades del usuario; herramientas como la suite de Power BI, repositorio de base de datos PostgreSQL, lenguaje de programación R + librerías que en el caso del estudio de Améstica y Liendo [11] manejaron la suite de Tableau como apoyo a la creación de indicadores y tableros de control.

Para dar cumplimiento con el dicho objetivo se implementaron reportes los cuales muestran información actualizada, haciendo uso de la suite de Power BI con el fin crear reportes analíticos e interactivos de las variables agroeconómicas en estudio y la técnica de análisis de series temporales [15] para el análisis de patrones de comportamiento y lograr predecir valores futuros en base a data histórica almacenada a lo largo del tiempo.

Finalmente, para cumplir con el último objetivo de lograr una validación positiva de la Solución de Inteligencia de negocios por parte de los usuarios, mediante un cuestionario que valora la calidad y satisfacción de la información solicitada propuesta por DeLone & McLean y otros autores; haciendo uso de una escala de Likert de 1-5 donde:

| 1 | 2 | 3 | 4 | 5 |
|--------------------------|---------------|--------------------------------|------------|-----------------------|
| Totalmente en desacuerdo | En desacuerdo | Ni de acuerdo ni en desacuerdo | De acuerdo | Totalmente de acuerdo |

Como resultado permitió evidenciar la presencia de una mejora y evaluación positiva por parte de los usuarios con la solución propuesta y desarrollada en beneficio de la Dirección ejecutiva, sus procesos y actividades concernientes con la gestión de la información y toma de decisiones acertadas.

VI. Conclusiones

1. La técnica seleccionada para pronosticar durante el periodo de enero a diciembre del 2018, es un modelo de Regresión lineal con ajuste estacional, puesto que ostentó un error cuadrático menor a comparación con el enfoque Naive estacional y el modelo ARIMA (0,0,1) (0,1,0) [12], además de tener un valor MAPE ligeramente superior al valor obtenido por los demás.
2. Se generó el pronóstico de la Producción de cultivos de acuerdo a los parámetros que elija el usuario en el módulo implementado en la Solución utilizando la técnica seleccionada en el objetivo anterior, ya que es la más óptima.
3. Se han implementado reportes relacionadas a: Avances, comparativos, consolidados y calendarios de siembra y cosecha (Ha.) de cultivos por ubigeo, tipo de consumo, cultivos y campañas agrícolas o periodos de tiempo; perfil productivo; mapeos de zonificación de cultivos; información estadística; y pronóstico de la producción agrícola, los cuales han sido implementados de acuerdo a las necesidades del usuario y tomando en cuenta la metodología y modelo de minería de datos seleccionado.
4. Se ha llevado a cabo una reunión con los interesados llegándoles a realizar un demostrativo de la Solución, quienes mediante un instrumento validado por el investigador han opinado favorablemente sobre las características del producto software que se ha implementado mediante esta investigación, bajo la perspectiva de la calidad de la información proporcionada y la satisfacción sobre la información solicitada.
5. Con la solución implementada, los miembros de la GRA tomarán decisiones oportunas en beneficio a disponer de información y almacenes de datos necesarias para monitorear regularmente el desarrollo de las cadenas productivas y principales variables agroeconómicas, estrategias de comunicación efectiva, campañas para el fortalecimiento de la agricultura en la Región, capacitaciones y asesorías basándose en los reportes analíticos brindados por la Solución, mecanismos y programas para mejorar la información de calidad.

VII. Recomendaciones

1. Para proyectos posteriores, es importante vincular los sistemas existentes con el producto acreditable propuesto y así alcanzar un mejor desempeño en pro al desarrollo y crecimiento de la Dirección Ejecutiva de Información Agraria.
2. Se recomienda implementar, si es necesario, nuevos reportes para el análisis de datos y toma de decisiones en el proceso de gestión de información agrícola, utilizando como base el diseño actual del DataMart de este proyecto.
3. Para posteriores investigaciones, el alumno deberá familiarizarse con otros modelos predictivos a aplicar que se ajusten con las necesidades y problemática que se aborde.
4. Evaluar y ponderar la posibilidad de escalar el proyecto e involucrar no sólo el área de información agrícola sino ir interconectando también las áreas de exportaciones e insumos agrícolas, precios de mercado, actividad pecuaria, apícola, ente otros.

VIII. Referencias

- [1] C. Bustelo Ruesta y R. Amarilla Iglesias, «Gestión del conocimiento y gestión de información.,» 2001. [En línea]. Available: DOI:10.33349/2001.34.1153. [Último acceso: 8 set 2021].
- [2] A. Suárez Alfonso, I. Cruz Rodríguez y Y. Pérez Macías, «La gestión de la información: herramienta esencial para el desarrollo de habilidades en la comunidad estudiantil universitaria.,» *Revista Universidad y Sociedad*, ago 2015. [En línea]. Available: <http://scielo.sld.cu/pdf/rus/v7n2/rus10215.pdf>. [Último acceso: 8 sep 2021].
- [3] H. F. Vallejo Ballesteros, E. Guevara Iñiguez y S. R. Medina Velasco, «Minería de Datos.,» *Revista Científica Mundo de la Investigación y el Conocimiento*, 2018. [En línea]. Available: <https://doi.org/10.26820/recimundo/2.esp.2018.339-349>. [Último acceso: 9 sep 2021].
- [4] J. C. Riquelme, R. Ruiz y K. Gilbert, «Minería de Datos: Conceptos y Tendencias.,» *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial [en linea]*, vol. X, n° 29, pp. 11-18, 2006.
- [5] A. A. Mora Bermúdez y J. A. Sancho Chacón, «Modelo de redes neuronales optimizadas con algoritmos genéticos. Una aplicación para proyecciones sobre pacientes con cáncer hospitalizados en la C.C.S.S.,» *Universidad Estatal a Distancia*, 2013. [En línea]. Available: <http://repositorio.uned.ac.cr/reuned/handle/120809/881>. [Último acceso: 9 sep 2021].
- [6] N. Rudnichenko, V. Vychuzhanin, I. Petrov y D. Shibaey, «“Decision support system for the machine learning methods selection in big data mining”.,» 2020. [En línea]. Available: <http://ceur-ws.org/Vol-2608/paper65.pdf>. [Último acceso: 6 set 2021].
- [7] A. Chougule, V. Jha y D. Mukhopaddyay, «Crop suitability and fertilizers recommendation using data mining techniques”.,» 2019. [En línea]. Available: <https://bit.ly/3mjebQS>. [Último acceso: 7 set 2021].
- [8] M. Najafabadi, A. Mohamed y M. Mahrin, «A survey on data mining techniques in recommender systems”.,» 2019. [En línea]. Available: doi.org/10.1007/s00500-017-2918-7. [Último acceso: 6 set 2021].

- [9] B. Mazon-Olivo, M. Jaramillo-Paredes, O. Romero-Hidalgo, A. Borja-Herrera, M. Aguirre-Benalcazar y M. Contenido-Segarra, «Tecnologías de Inteligencia de Negocios y Minería de datos para el análisis de la producción y comercialización de cacao,» 2018. [En línea]. Available: <http://www.revistaespacios.com/a18v39n32/a18v39n32p06.pdf>. [Último acceso: 5 oct 2021].
- [10] E. J. Merchán Carreño, K. V. Mero Suárez y Y. Castro Blanco, «Empleo de sistemas de inteligencia de negocio para la visualización y análisis eficiente de información en las organizaciones,» 2019. [En línea]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=7149452>. [Último acceso: 6 oct 2021].
- [11] G. Amestica, L. Liendo y F. Luna, «Sistema informático de apoyo a la gestión de stock y sistema de soporte a la toma de decisiones del predio agropecuario,» CRUC-IUA UNDEF - Córdoba – Argentina. , 2017. [En línea]. [Último acceso: 8 set 2021].
- [12] W. Santa Cruz, «Business intelligence aplicada al monitoreo de la producción en las empresas Agrícolas del Grupo Palmas,» 2018. [En línea]. Available: <http://repositorio.unsm.edu.pe/handle/11458/2912>. [Último acceso: 8 set 2021].
- [13] J. Castro y E. Díaz, «“Aplicación de inteligencia de negocios para optimizar el proceso de toma de decisiones en una empresa agrícola- caso empresa agrícola el rancho se desarrolló en la ciudad de Chíncha”, (20,» 2017. [En línea]. Available: <https://repositorio.unica.edu.pe/handle/UNICA/3069>. [Último acceso: 12 ago 2021].
- [14] C. E. Arbieta Batallanos, «Herramientas de Minería de datos de gran escala para apoyar investigaciones en cambio climático en la Amazonia Peruana,» 2018. [En línea]. Available: <http://repositorio.unsa.edu.pe/handle/UNSA/6601>. [Último acceso: 6 oct 2021].
- [15] L. Fernandez, «Aplicacion de tecnicas de mineria de datos para pronostico de produccion de espárragos,» 2019. [En línea]. Available: <https://repositorio.uss.edu.pe/handle/20.500.12802/5270>. [Último acceso: 12 ago 2021].
- [16] M. White, «Intelligence Management,» *Information Management: From strategies to action*, pp. 19-35, 1985.
- [17] L. Woodman, «Information Management in large Organizations,» *Information Management: From strategies to action*, pp. 95-114, 1985.
- [18] G. Ponjuán, *Gestión de la Información en las Organizaciones: Principios, conceptos y aplicaciones*, Santiago de Chile: CECAPI, 1998.

- [19] G. Ponjuán, *Gestión de la Información: Dimensiones e implementación para el éxito organizacional*, Rosario: Ediciones Nuevo Paradigma, 2004.
- [20] I. Páez Urdaneta, *Gestión de la inteligencia, aprendizaje tecnológico y modernización del trabajo. Retos y oportunidades.*, vol. 3, Caracas: Instituto de Estudios del Conocimiento de la Universidad Simón Bolívar, 1992, pp. 1-13.
- [21] G. P. Huber, «Organizational Learning: The contributing processes and the literature,» *Organization Science*, pp. 88-115, 1991.
- [22] E. A. y. C. Choo, *Managing Information for the Competitive Edge*, New York, NY: Neal Schuman, 1995.
- [23] C. Perez Lopez y D. Santin Gonzales , «Minería de datos. Técnicas y herramientas,» 23 mar 2007. [En línea]. [Último acceso: 3 set 2021].
- [24] R. Carrasco, 2011. [En línea]. [Último acceso: 2 set 2021].
- [25] S. M. Weiss y N. Indurkha, «Predictive Data Mining. A Practical Guide,» Morgan Kaufmann Publishers, San Francisco, USA, 1998.
- [26] I. H. Witten y E. Frank, *Data Minig. Practical Machine Learning Tools and Techniques*, San Francisco, California: Morgan Kaufmann Publishers, 2000.
- [27] M. R. y. C. F. J. Hernández, *Introducción a la Minería de Datos*, Madrid: Pearson Educación, 2004.
- [28] B. Beltrán, *Minería de Datos*, México: Benemérita Universidad Autónoma de Puebla, 2016.
- [29] H. Ñaupas Paitán, M. R. Valdivia Dueñas, J. J. Palacios Vilela y H. E. Romero Delgado, «Metodología de la investigación cuantitativa-cualitativa y redacción de la tesis. 5ta Ed.,» Ediciones de la U, sep 2018. [En línea]. Available: <https://bit.ly/3mz2DsW>. [Último acceso: 9 oct 2021].
- [30] R. Hernández Sampieri, C. Fernández Collado y M. d. P. Baptista Lucio, «Metodología de la investigación. 5ta Ed.,» McGRAW-HILL, 2010. [En línea]. Available: <https://bit.ly/2WYu06V>. [Último acceso: 10 oct 2021].
- [31] D. F. Betancourt, «Medición del error en pronósticos de demanda,» 07 mar 2016. [En línea]. Available: <https://www.ingenioempresa.com/medicion-error-pronostico/>. [Último acceso: 11 oct 2021].

IX. Anexos

Tabla 18.
Planificación de la Implementación de la solución de BI

| Ítem | Descripción / Fase | Duración (Días) | Comienzo | Fin | % Avance |
|-----------|---|-----------------|---------------------|---------------------|------------|
| 1 | PLANIFICACIÓN DEL PROYECTO | 2 | mié 21/03/18 | jue 22/03/18 | |
| 2 | DETERMINACIÓN DE REQUERIMIENTOS | 6 | vie 23/03/18 | jue 29/03/18 | |
| 3 | Análisis de requerimientos del Proceso de Gestión de Información Agraria | 2 | vie 23/03/18 | sáb 24/03/18 | |
| 4 | Construcción de la Matriz Bus (Proceso/Dimensiones) | 2 | lun 26/03/18 | mar 27/03/18 | |
| 5 | Análisis de las fuentes de datos internas | 2 | mié 28/03/18 | jue 29/03/18 | |
| 6 | MODELO DIMENSIONAL | 6 | vie 30/03/18 | jue 05/04/18 | |
| 7 | Elaboración del modelo dimensional de alto nivel | 1 | vie 30/03/18 | vie 30/03/18 | |
| 8 | Establecimiento del nivel de granularidad del proceso modelado | 1 | sáb 31/03/18 | sáb 31/03/18 | |
| 9 | Elección de las dimensiones y sus atributos | 1 | lun 02/04/18 | lun 02/04/18 | |
| 10 | Identificación de la tabla hecho y sus medidas | 1 | mar 03/04/18 | mar 03/04/18 | |
| 11 | Documentación del modelo dimensional de base de datos que se construirá | 2 | mié 04/04/18 | jue 05/04/18 | 50% |
| 12 | DISEÑO DE LA ARQUITECTURA TECNICA | 4 | vie 30/03/18 | mar 03/04/18 | |
| 13 | ESPECIFICACION DE APLICACIONES DE BI | 4 | vie 30/03/18 | mar 03/04/18 | |
| 14 | SELECCIÓN DE PRODUCTOS E IMPLEMENTACION | 4 | vie 06/04/18 | jue 12/04/18 | |
| 15 | DISEÑO FISICO | 6 | vie 13/04/18 | sáb 21/04/18 | |
| 16 | DESARROLLO DE APLICACIONES DE BI | 6 | lun 23/04/18 | mar 01/05/18 | |
| 17 | DISEÑO E IMPLEMENTACION DEL SUBSISTEMA ETL | 15 | mié 02/05/18 | vie 18/05/18 | |
| 18 | DESARROLLO DEL PROCESO ETL | 15 | mié 02/05/18 | vie 18/05/18 | |
| 19 | Diseño del proceso de Extracción, Transformación y Carga | 1 | mié 02/05/18 | mié 02/05/18 | |
| 20 | Limpieza de los datos sucios de las fuentes internas obtenidas en el paso I | 2 | jue 03/05/18 | vie 04/05/18 | |
| 21 | Desarrollo del proceso ETL para las dimensiones | 5 | sáb 05/05/18 | jue 10/05/18 | |
| 22 | Desarrollo del proceso ETL para la tabla hecho | 5 | vie 11/05/18 | mié 16/05/18 | |

| 23 | Validación y revisión de la carga correcta de datos en la base de datos dimensional | 2 | jue 17/05/18 | vie 18/05/18 | |
|------|--|-----------------|---------------------|---------------------|-------------|
| 24 | AVANCE AL 50% | 53 | mié 21/03/18 | vie 18/05/18 | |
| Ítem | Descripción / Fase | Duración (Días) | Comienzo | Fin | % Avance |
| 25 | DISEÑO E IMPLEMENTACIÓN DEL CUBO OLAP | 7 | sáb 19/05/18 | sáb 26/05/18 | |
| 26 | Implementación del cubo OLAP a partir del hecho en estudio | 5 | sáb 19/05/18 | jue 24/05/18 | |
| 27 | Validación y revisión de los datos transferidos en la implementación del cubo OLAP | 2 | vie 25/05/18 | sáb 26/05/18 | |
| 28 | EXPLOTACIÓN DEL DATA MART | 7 | lun 28/05/18 | lun 04/06/18 | |
| 29 | Generación de reportes tabulares multidimensionales según prototipos | 2 | lun 28/05/18 | mar 29/05/18 | |
| 30 | Generación de reportes gráficos multidimensionales según prototipos | 3 | mié 30/05/18 | vie 01/06/18 | 80% |
| 31 | Generación de reportes de semaforización sobre producción según prototipos | 3 | mié 30/05/18 | vie 01/06/18 | |
| 32 | Generación de reportes de indicadores de producción según prototipos | 3 | mié 30/05/18 | vie 01/06/18 | |
| 33 | Validación y revisión del correcto funcionamiento de los reportes | 2 | sáb 02/06/18 | lun 04/06/18 | |
| 34 | DISEÑO E IMPLEMENTACION DEL MODELO DE PREDICCIÓN | 14 | mar 05/06/18 | vie 22/06/18 | |
| 35 | AVANCE AL 80% | 28 | sáb 19/05/18 | vie 22/06/18 | |
| 36 | Generación de reporte de pronóstico de producción a partir del Algoritmo de Predicción | 2 | jue 21/06/18 | vie 22/06/18 | |
| 37 | IMPLEMENTACION | 5 | sáb 23/06/18 | vie 29/06/18 | |
| 38 | PRUEBAS | 4 | sáb 30/06/18 | mié 04/07/18 | 100% |
| 39 | Validación del funcionamiento en red de la Solución BI | 2 | mar 03/07/18 | mié 04/07/18 | |
| 40 | AVANCE AL 100% | 90 | mié 21/03/18 | mié 04/07/18 | |
| 41 | SUSTENTACION FINAL | 1 | | | 100% |

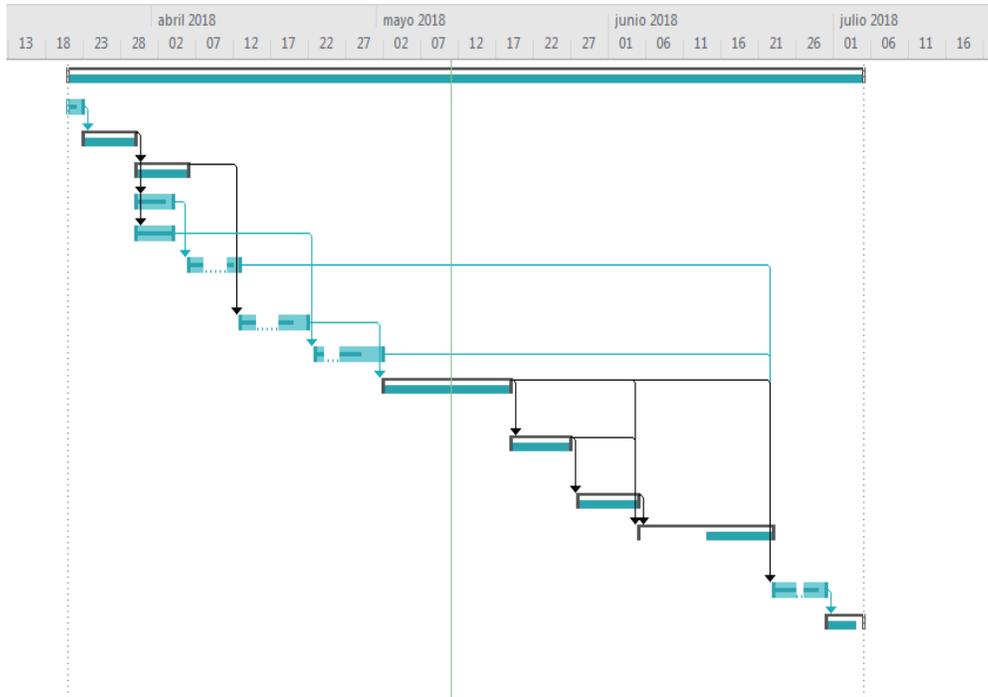


Figura 21. Diagrama de Grantt de la Solución BI

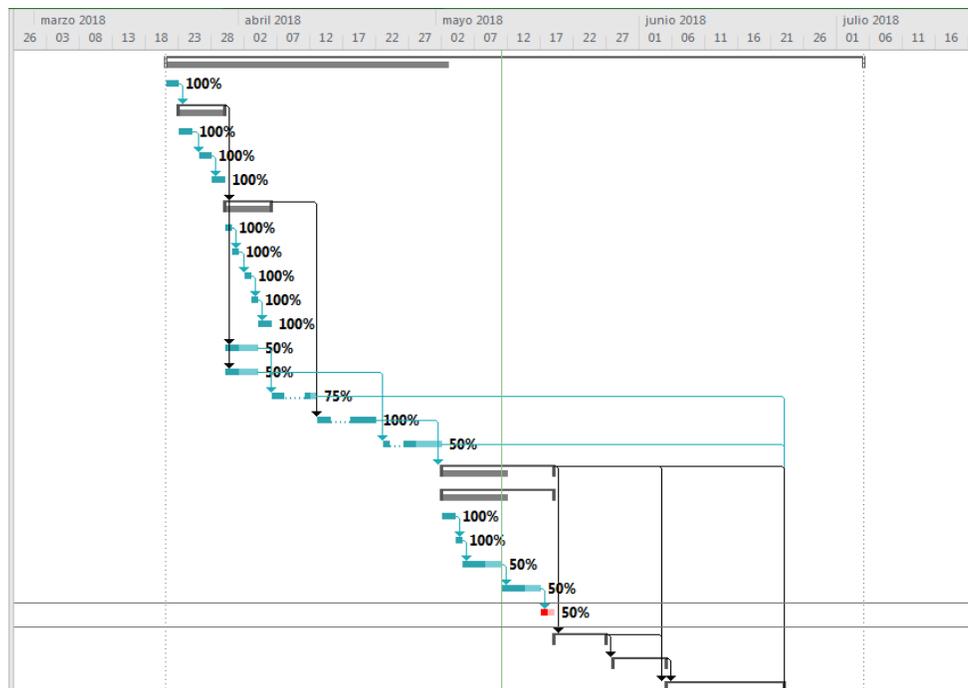


Figura 22. Diagrama de Grantt de Seguimiento de la Solución BI

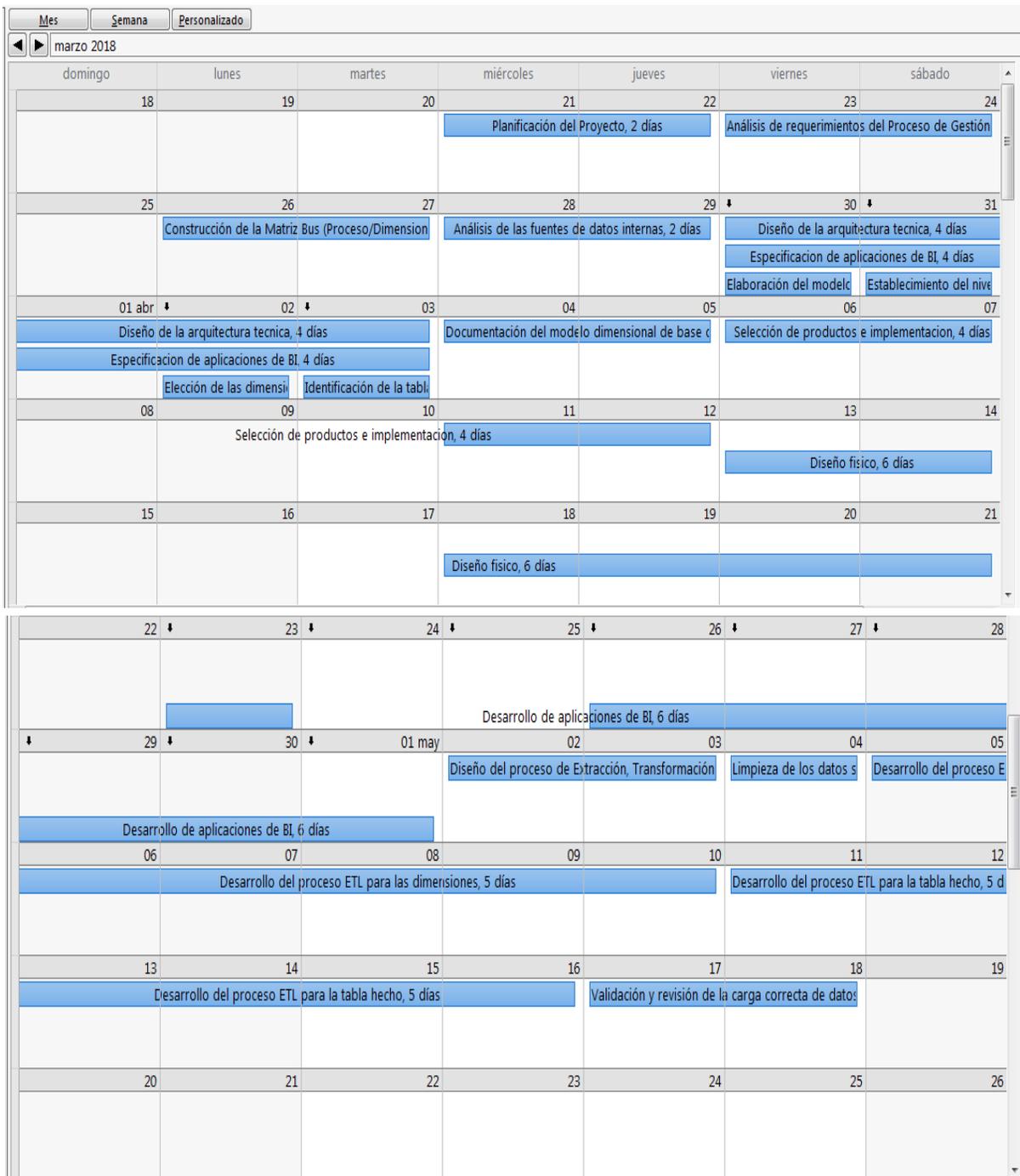


Figura 23. Calendario de Actividades de la Solución BI

Base agrícola últimas 3 campañas - Excel

Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista ¿Qué desea hacer? Compartir

Portapapeles Fuente Alineación Número Estilos Celdas Edición

AF133

| | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG | AH | AI | AJ | AK | AL | AM | AN |
|-----|------|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---------|---------|------------|---------------|-------------------|---------|----------------|-------|
| 1 | Nov. | Dic. | Ene | Feb | Mar | Abr | May | Jun | Jul | Ago | Set | Oct | Nov | Dic | Siembra | Cosecha | Produccion | siembras (ha) | cos (ha)/prod (t) | P. base | promedio/areas | |
| 2 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 3 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 4 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 5 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 6 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 7 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | | | | 108 | 144 | 0 | 12.00 | |
| 8 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 9 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 10 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | | | | 108 | 144 | 0 | 12.00 | |
| 39 | | | | | | | | | | | | 15 | | | | | 15 | | 15 | 0 | 0 | |
| 40 | | | | | | | | | | | | 0 | 0 | | | | | 0.15 | 0.45 | 0 | 0.15 | |
| 41 | | | | | 29 | | | | | | | 30 | 12 | | | | | 71 | 29 | 71 | 0.12 | 29.00 |
| 42 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 43 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 44 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | | | | 135 | 180 | 0 | 15.00 | |
| 45 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 46 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 47 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 | | | | 135 | 180 | 0 | 15.00 | |
| 104 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 105 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 106 | | | | | | | | | | | | | | | | | | 0 | 0 | 0.06 | 0 | |
| 107 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 108 | | | | | | | | | | | | | | | | | | 0 | 0 | 0 | 0 | |
| 109 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | 679 | | | | 5111 | 8148 | | 679.00 | |

Base Agrícola

Modo Filtrar

Figura 24. Archivos en extensión "XLS" encontrados del Proceso de GIA

datos_reales [Reparado] - Excel

Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista ¿Qué desea hacer? Compartir

Portapapeles Fuente Alineación Número Estilos Celdas Edición

F54

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|-----------|--------------------|-------------|-----------|----------|--------------------|----------------------------|-------------------------|-------------|------|-------|---------|
| 1 | campana | Tipo consumo | tipo | Provincia | Distrito | Valle | cultivo | Variables | Int. Siemb. | año | mes | valores |
| 2 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Cosechas (ha) | | 2013 | 8.000 | |
| 3 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | P. Chacra (S./Kg) | | 2013 | 8.000 | |
| 4 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Producción (t) | | 2013 | 8.000 | |
| 5 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Rendimiento (kg/ha) | | 2013 | 8.000 | |
| 6 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Siembras (ha) | | 2013 | 8.000 | |
| 7 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Sup. en producción(Ha) | | 2013 | 8.000 | 12.000 |
| 8 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Sup. en crecimiento(Ha) | | 2013 | 8.000 | 0.000 |
| 9 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Sup. Perd. (ha) | | 2013 | 8.000 | |
| 10 | 2013-2014 | Consumo Industrial | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Caña de Azúcar para azúcar | Sup. Verde (ha) | | 2013 | 8.000 | 12.000 |
| 11 | 2013-2014 | Agroexportación | transitorio | Ferreñafe | Pítipo | Chancay-Lambayeque | Aji Paprika | Cosechas (ha) | | 2013 | 8.000 | |
| 12 | 2013-2014 | Agroexportación | transitorio | Ferreñafe | Pítipo | Chancay-Lambayeque | Aji Paprika | P. Chacra (S./Kg) | | 2013 | 8.000 | 0.000 |
| 13 | 2013-2014 | Agroexportación | transitorio | Ferreñafe | Pítipo | Chancay-Lambayeque | Aji Paprika | Producción (t) | | 2013 | 8.000 | |
| 14 | 2013-2014 | Agroexportación | transitorio | Ferreñafe | Pítipo | Chancay-Lambayeque | Aji Paprika | Rendimiento (kg/ha) | | 2013 | 8.000 | |
| 15 | 2013-2014 | Agroexportación | transitorio | Ferreñafe | Pítipo | Chancay-Lambayeque | Aji Paprika | Siembras (ha) | | 2013 | 8.000 | |
| 16 | 2013-2014 | Agroexportación | transitorio | Ferreñafe | Pítipo | Chancay-Lambayeque | Aji Paprika | Sup. Perd. (ha) | | 2013 | 8.000 | |
| 17 | 2013-2014 | Agroexportación | transitorio | Ferreñafe | Pítipo | Chancay-Lambayeque | Aji Paprika | Sup. Verde (ha) | | 2013 | 8.000 | 0.000 |
| 18 | 2013-2014 | Pastos | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Alfalfa | Cosechas (ha) | | 2013 | 8.000 | |
| 19 | 2013-2014 | Pastos | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Alfalfa | P. Chacra (S./Kg) | | 2013 | 8.000 | |
| 20 | 2013-2014 | Pastos | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Alfalfa | Producción (t) | | 2013 | 8.000 | |
| 21 | 2013-2014 | Pastos | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Alfalfa | Rendimiento (kg/ha) | | 2013 | 8.000 | |
| 22 | 2013-2014 | Pastos | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Alfalfa | Siembras (ha) | | 2013 | 8.000 | |
| 23 | 2013-2014 | Pastos | permanente | Ferreñafe | Pítipo | Chancay-Lambayeque | Alfalfa | Sup. en producción(Ha) | | 2013 | 8.000 | 15.000 |

datos_reales

Modo Filtrar

Figura 25. Archivos en extensión "XLS" de la nueva Base de Datos estandarizada

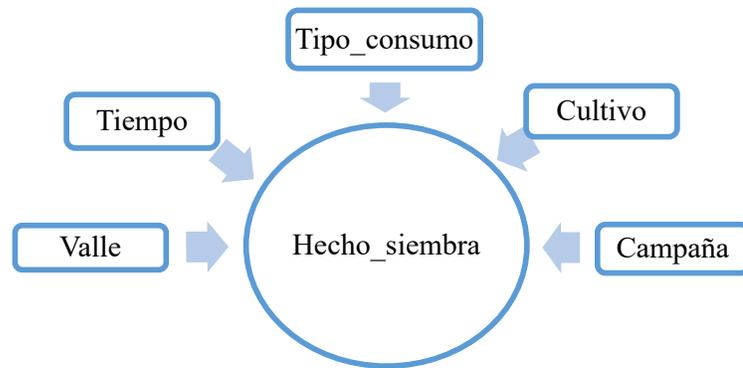


Figura 26. Modelo dimensional de alto nivel Hecho_siembra

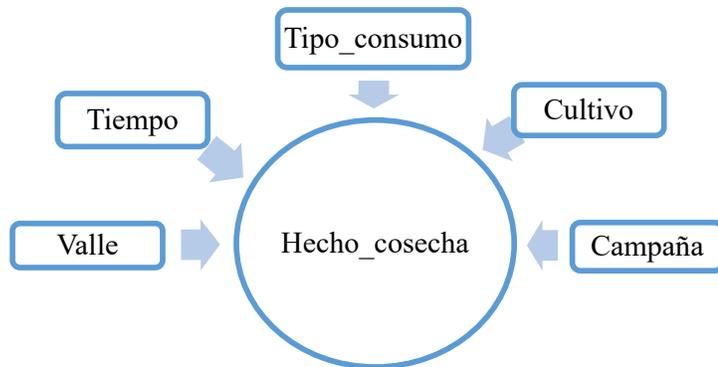


Figura 27. Modelo dimensional de alto nivel Hecho_cosecha

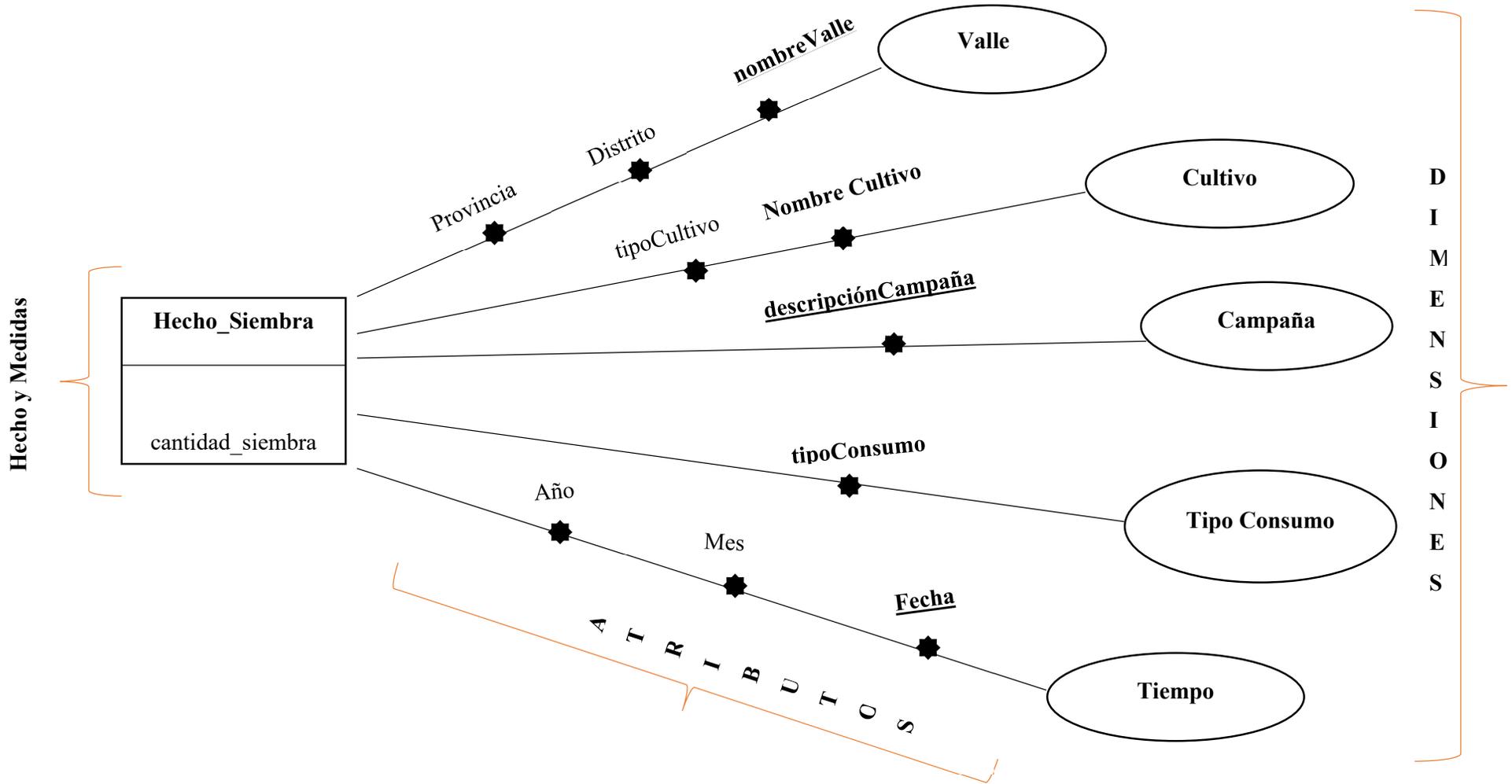


Figura 28. Determinación de Requerimientos: STAR NET

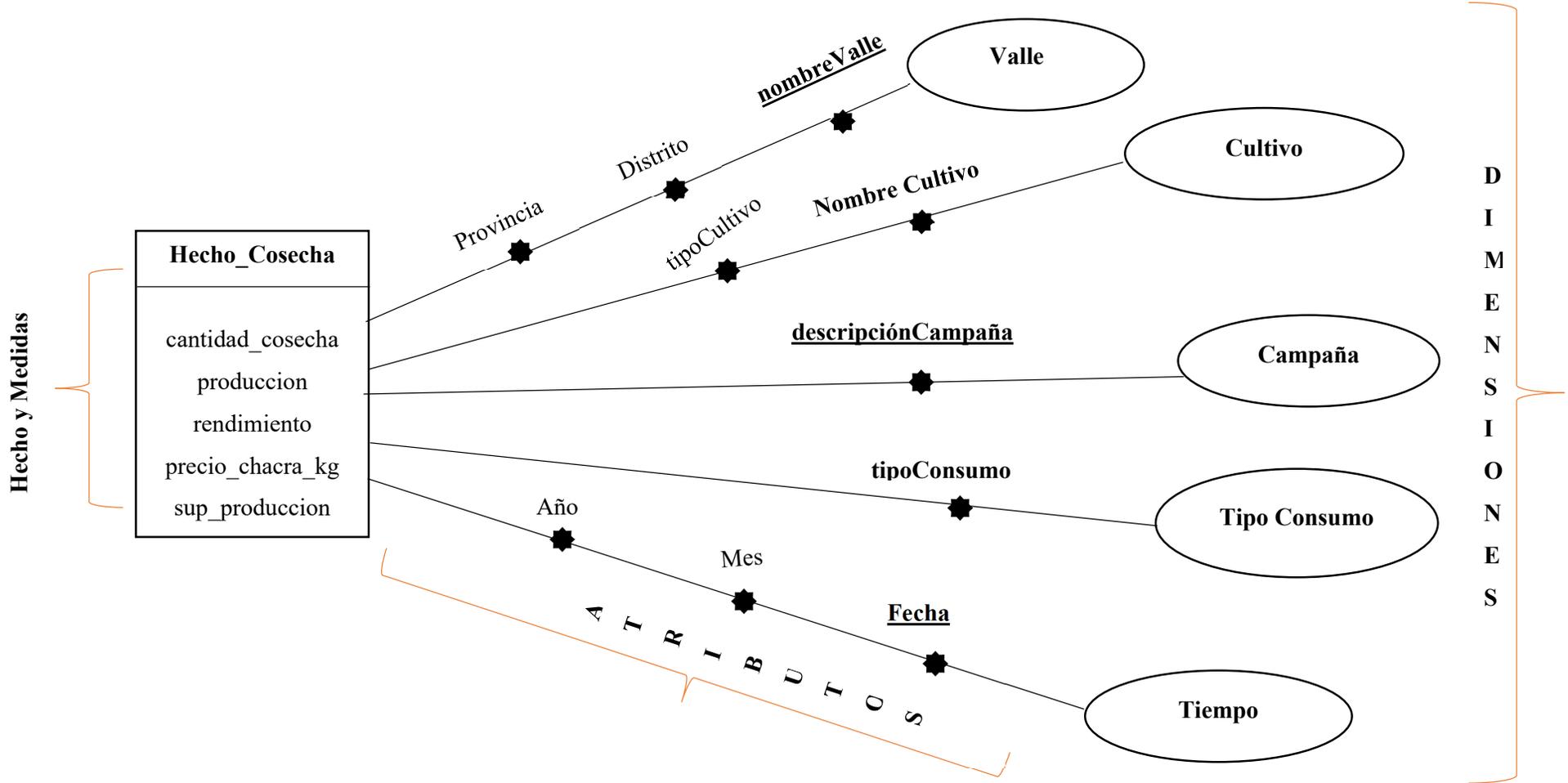


Figura 29. Determinación de Requerimientos: STAR NET



Figura 30. Proceso ETL para la Dimensión Valle

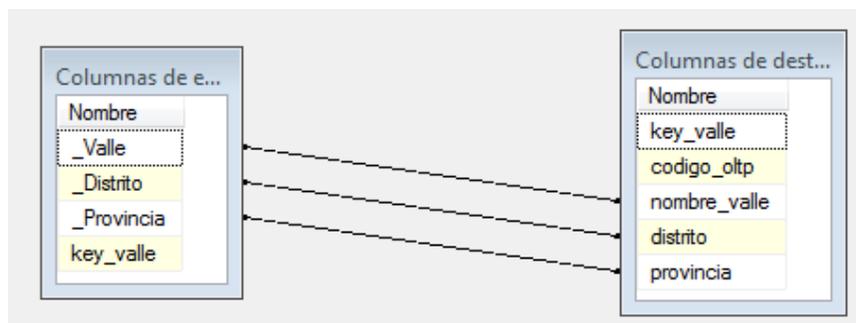


Figura 31. Relación entre campos de entrada y destino para la Dimensión Valle

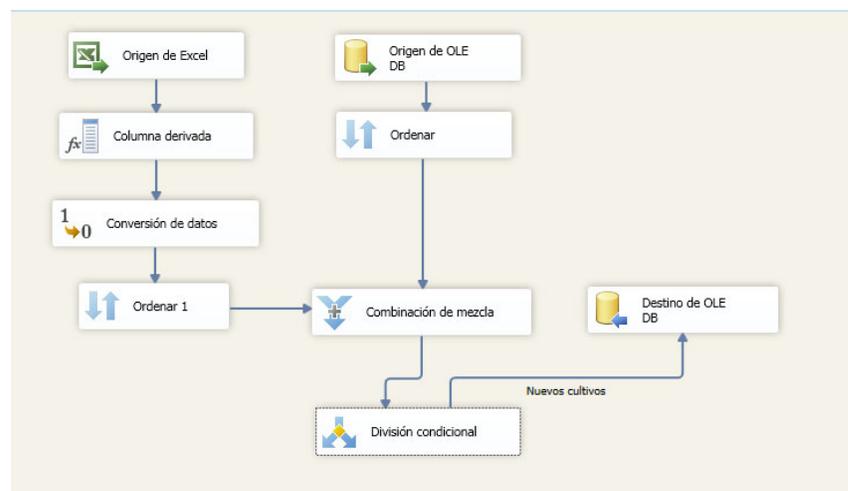


Figura 32. Proceso ETL para la Dimensión Cultivo

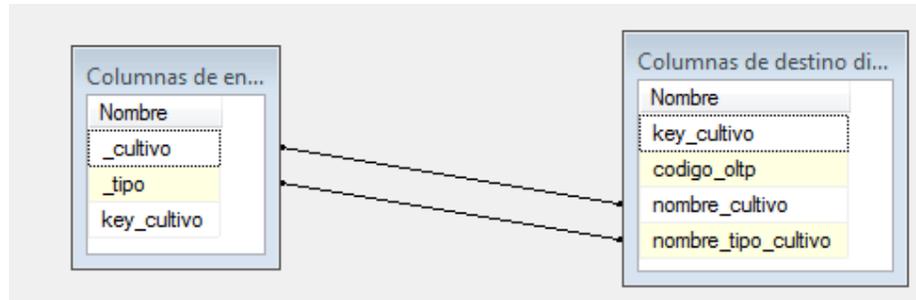


Figura 33. Relación entre campos de entrada y destino para la Dimensión Cultivo

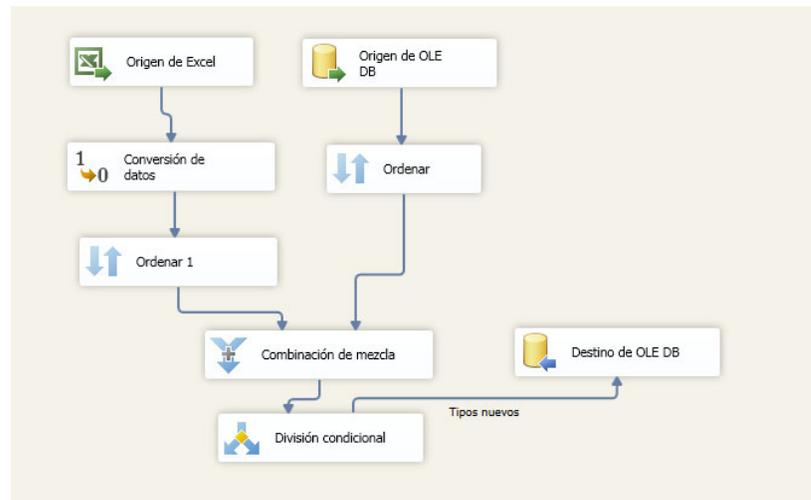


Figura 34. Proceso ETL para la Dimensión Tipo de Consumo

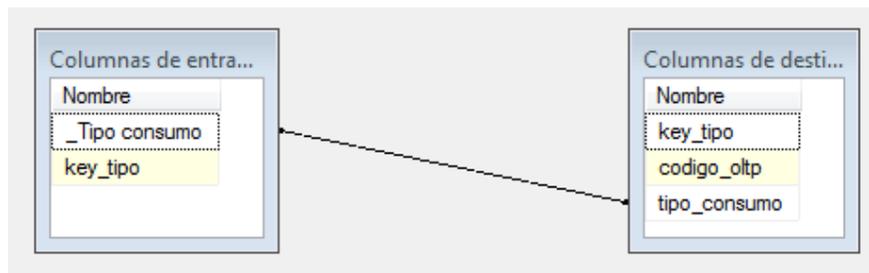


Figura 35. Relación entre campos de entrada y destino para la Dimensión Tipo de Consumo

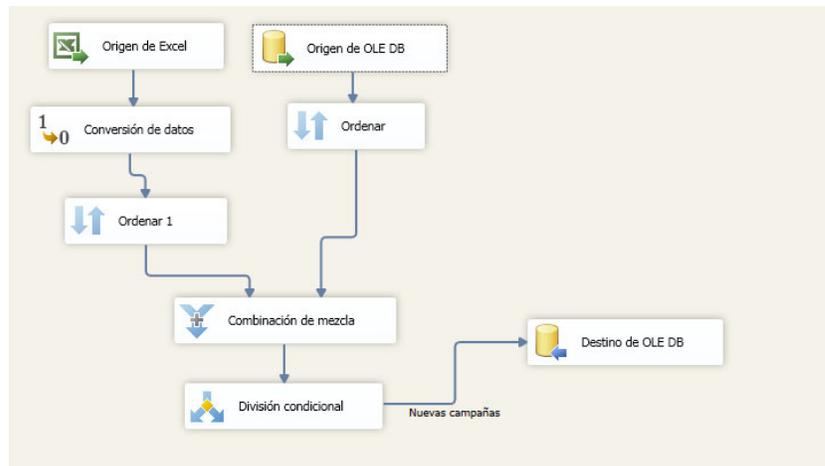


Figura 36. Proceso ETL para la Dimensión Campaña

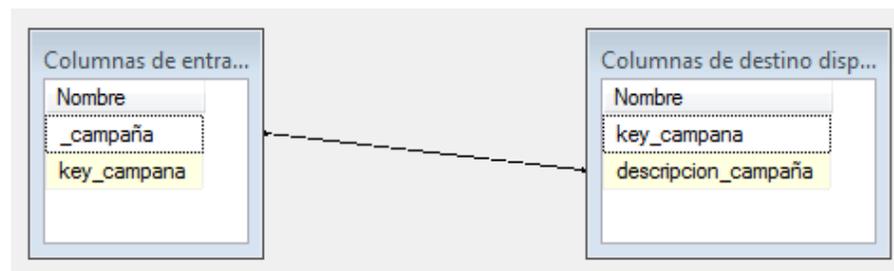


Figura 37. Relación entre campos de entrada y destino para la Dimensión Campaña

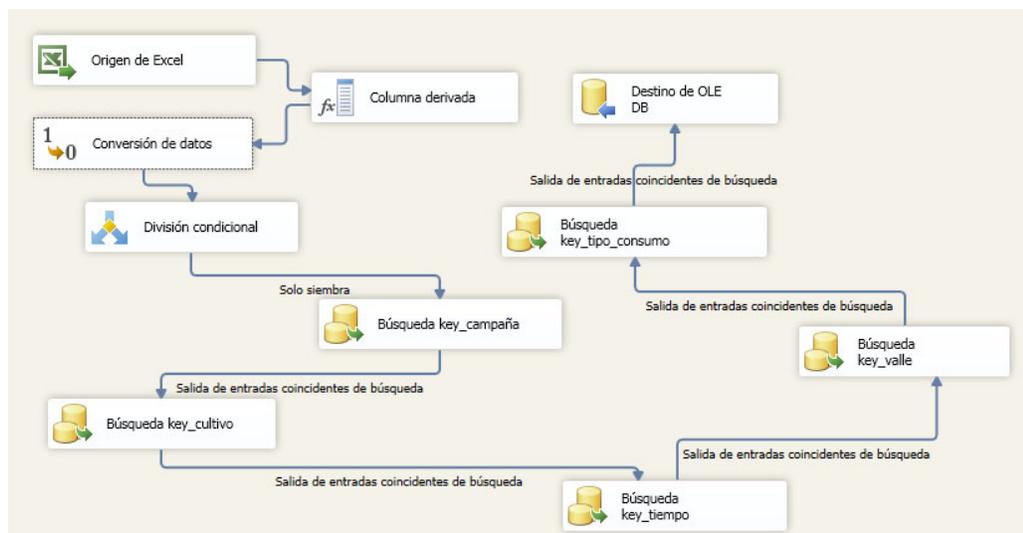


Figura 38. Proceso ETL para el Hecho Siembra

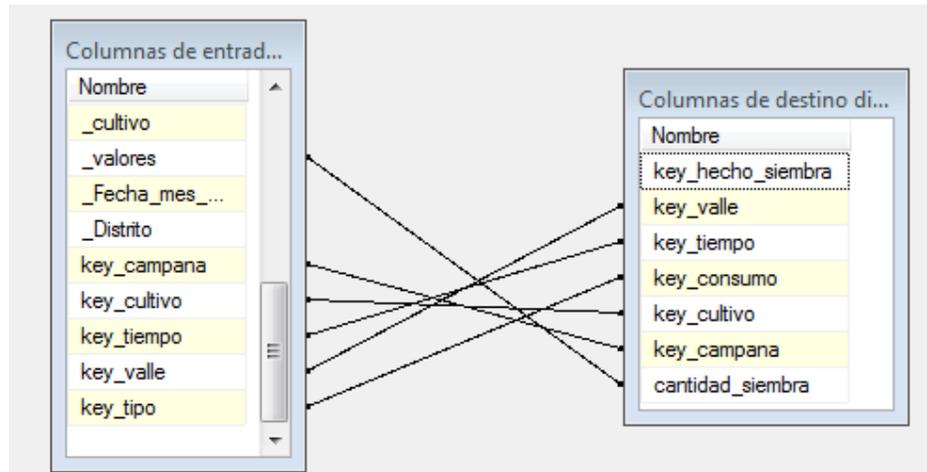


Figura 39. Relación entre campos de entrada y destino para el Hecho Siembra



Figura 40. Proceso ETL para el Hecho Cosecha

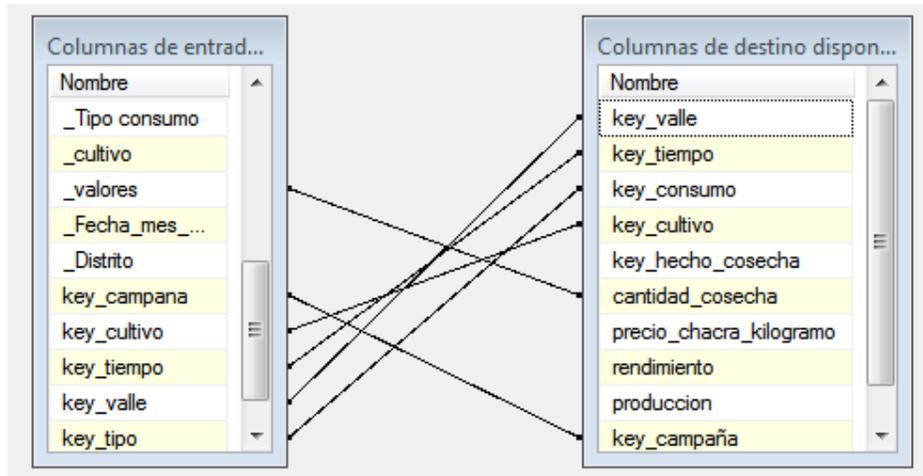


Figura 41. Relación entre campos de entrada y destino para el Hecho Cosecha

| | Nombre de columna | Tipo de datos | Permitir val... |
|-----|-------------------|---------------|-------------------------------------|
| ▶ 🔑 | key_valle | int | <input type="checkbox"/> |
| | codigo_oltp | int | <input checked="" type="checkbox"/> |
| | nombre_valle | varchar(100) | <input type="checkbox"/> |
| | distrito | varchar(100) | <input type="checkbox"/> |
| | provincia | varchar(100) | <input type="checkbox"/> |
| | | | <input type="checkbox"/> |

Figura 42. Dimensión Valle

| | Nombre de columna | Tipo de datos | Permitir val... |
|-----|-------------------|---------------|-------------------------------------|
| ▶ 🔑 | key_tiempo | int | <input type="checkbox"/> |
| | mes | int | <input type="checkbox"/> |
| | año | int | <input type="checkbox"/> |
| | fecha | smalldatetime | <input type="checkbox"/> |
| | key_campana | int | <input checked="" type="checkbox"/> |
| | | | <input type="checkbox"/> |

Figura 43. Dimensión Tiempo

| | Nombre de columna | Tipo de datos | Permitir val... |
|----|---------------------|---------------|-------------------------------------|
| ▶🔑 | key_cultivo | int | <input type="checkbox"/> |
| | codigo_oltp | int | <input checked="" type="checkbox"/> |
| | nombre_cultivo | varchar(100) | <input type="checkbox"/> |
| | nombre_tipo_cultivo | varchar(100) | <input type="checkbox"/> |
| | | | <input type="checkbox"/> |

Figura 44. Dimensión Cultivo

| | Nombre de columna | Tipo de datos | Permitir val... |
|----|-------------------|---------------|-------------------------------------|
| ▶🔑 | key_tipo | int | <input type="checkbox"/> |
| | codigo_oltp | int | <input checked="" type="checkbox"/> |
| | tipo_consumo | varchar(100) | <input type="checkbox"/> |
| | | | <input type="checkbox"/> |

Figura 45. Dimensión Tipo de Consumo

| | Nombre de columna | Tipo de datos | Permitir val... |
|----|---------------------|---------------|--------------------------|
| ▶🔑 | key_campana | int | <input type="checkbox"/> |
| | descripcion_campana | varchar(9) | <input type="checkbox"/> |
| | | | <input type="checkbox"/> |

Figura 46. Dimensión Campaña

| | Nombre de columna | Tipo de datos | Permitir val... |
|----|-------------------|----------------|--------------------------|
| ▶🔑 | key_hecho_siembra | int | <input type="checkbox"/> |
| | key_valle | int | <input type="checkbox"/> |
| | key_tiempo | int | <input type="checkbox"/> |
| | key_consumo | int | <input type="checkbox"/> |
| | key_cultivo | int | <input type="checkbox"/> |
| | key_campana | int | <input type="checkbox"/> |
| | cantidad_siembra | numeric(18, 2) | <input type="checkbox"/> |
| | | | <input type="checkbox"/> |

Figura 47. Hecho Siembra

| Nombre de columna | Tipo de datos | Permitir val... |
|---|----------------|-------------------------------------|
|  key_hecho_cosecha | int | <input type="checkbox"/> |
| key_valle | int | <input type="checkbox"/> |
| key_tiempo | int | <input type="checkbox"/> |
| key_consumo | int | <input type="checkbox"/> |
| key_cultivo | int | <input type="checkbox"/> |
| key_campaña | int | <input type="checkbox"/> |
| cantidad_cosecha | numeric(18, 2) | <input type="checkbox"/> |
| key_variable | int | <input checked="" type="checkbox"/> |
| produccion | numeric(18, 2) | <input checked="" type="checkbox"/> |
| rendimiento | numeric(18, 2) | <input checked="" type="checkbox"/> |
| sup_verde | numeric(18, 2) | <input checked="" type="checkbox"/> |
| p_chacra | numeric(18, 2) | <input checked="" type="checkbox"/> |
| sup_perdida | numeric(18, 2) | <input checked="" type="checkbox"/> |
| sup_produccion | numeric(18, 2) | <input checked="" type="checkbox"/> |
| sup_crecimiento | numeric(18, 2) | <input checked="" type="checkbox"/> |
| | | <input type="checkbox"/> |

Figura 48. Hecho Cosecha

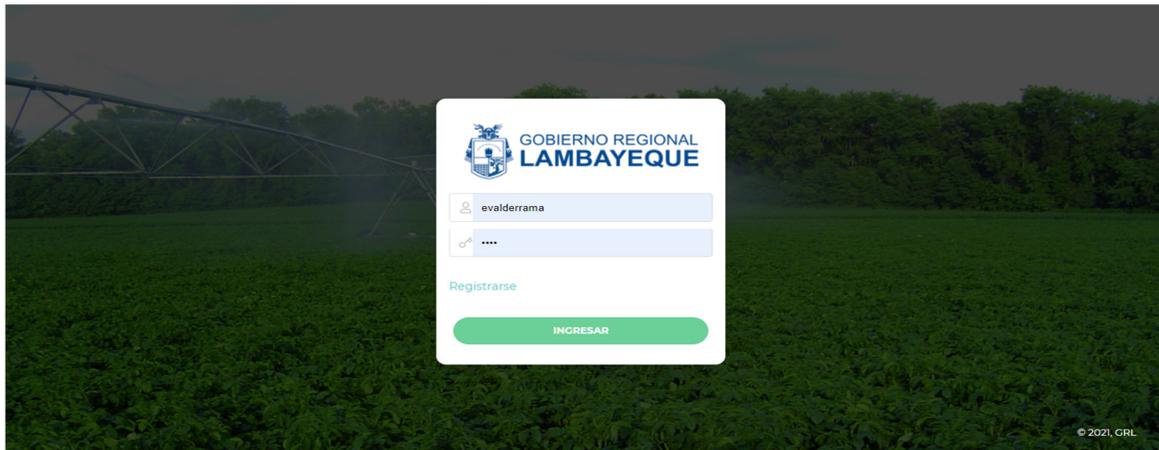


Figura 49. Inicio de Sesión de la Plataforma

Registro de usuario

 DATOS PERSONALES CUENTA DE USUARIO



CLICK AQUÍ

[SIGUIENTE](#)

Figura 50. Registro de usuarios

Registro de usuario

 DATOS PERSONALES CUENTA DE USUARIO

[ATRAS](#)[REGISTRAR](#)

Figura 51. Registro de cuenta de usuario

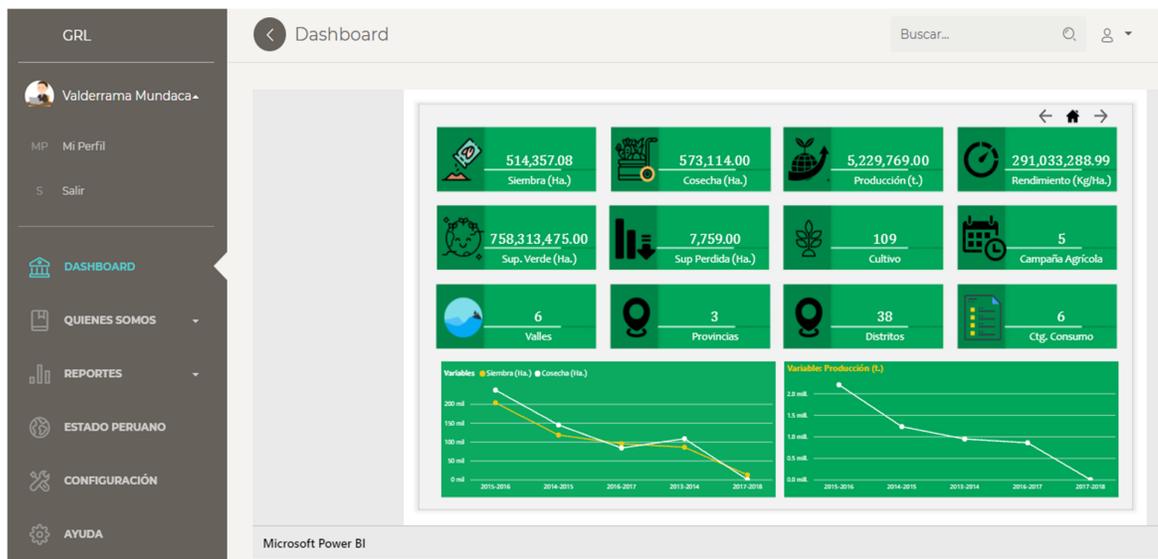


Figura 52. Navegación del Sistema

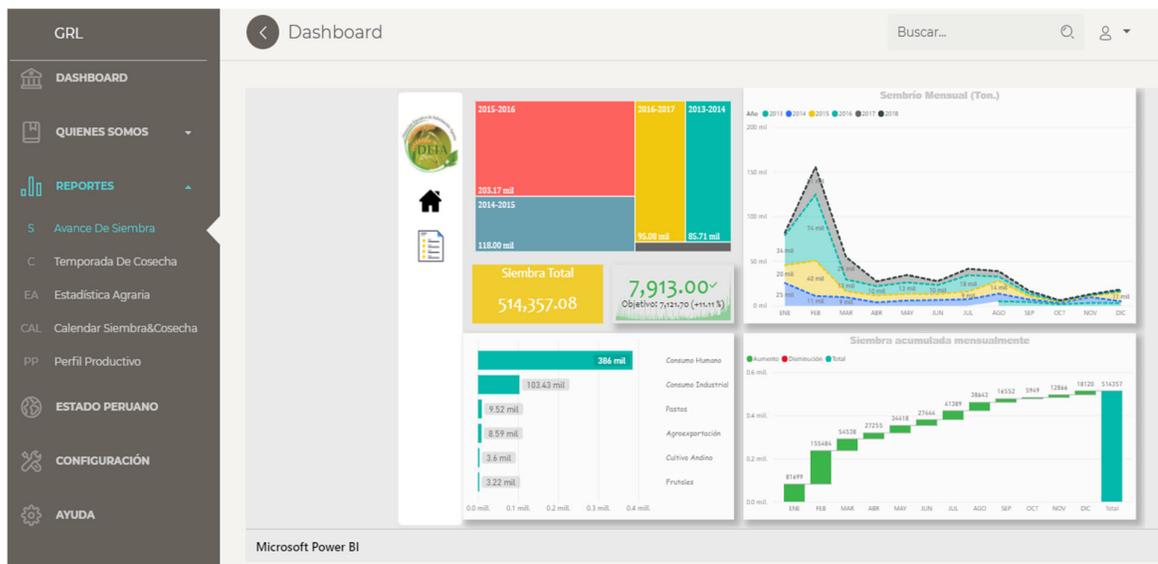


Figura 53. Reporte analítico de Siembra y Cosecha

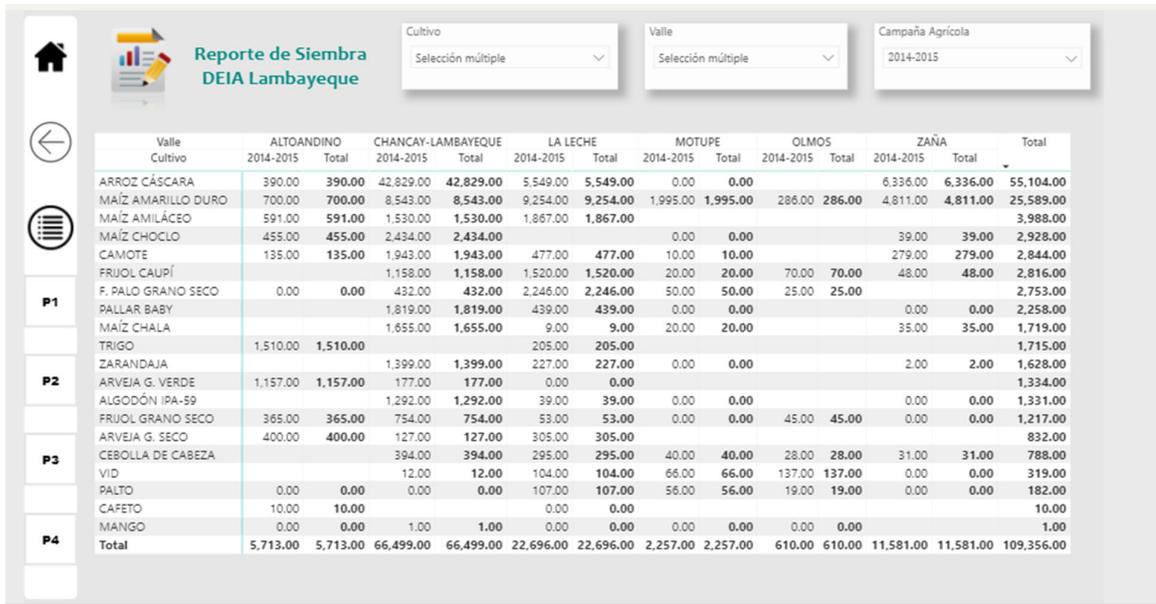


Figura 54. Reporte de Consolidado de Siembra – DEIA

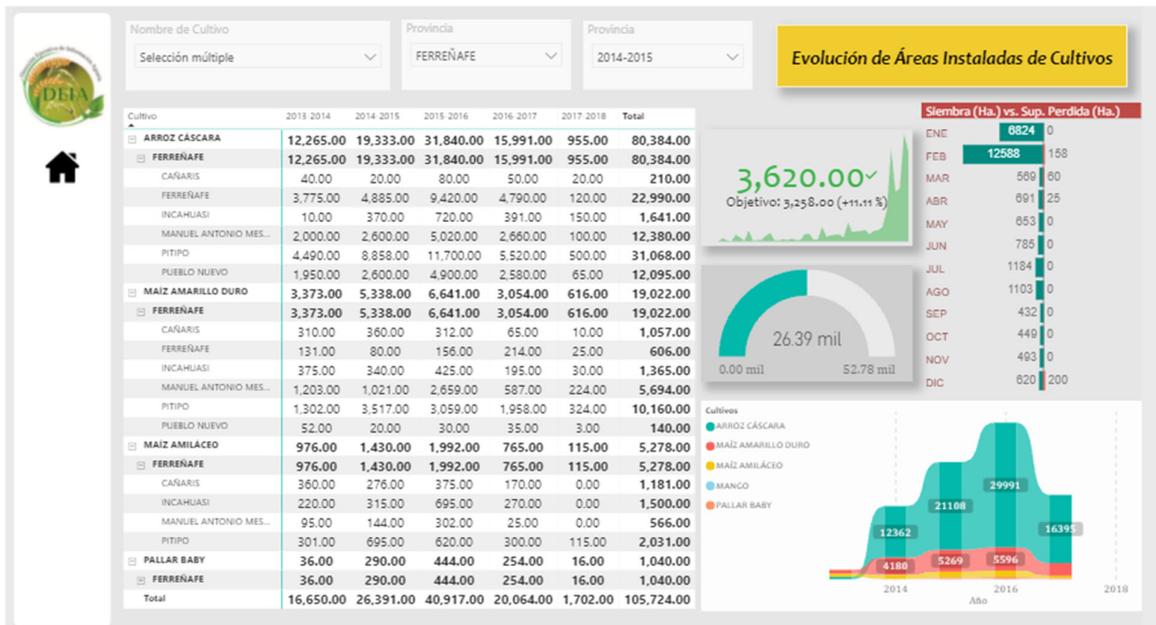


Figura 55. Tabulares, métricas y gráficas de Áreas instaladas de Cultivos

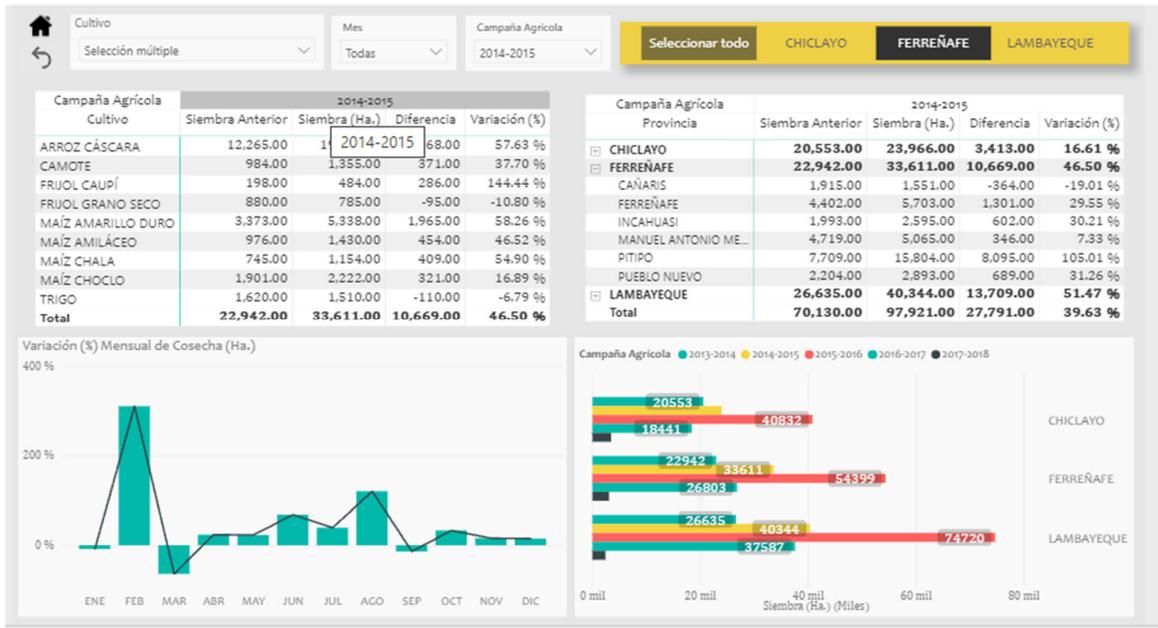


Figura 56. Reporte de variaciones y comparativos anuales

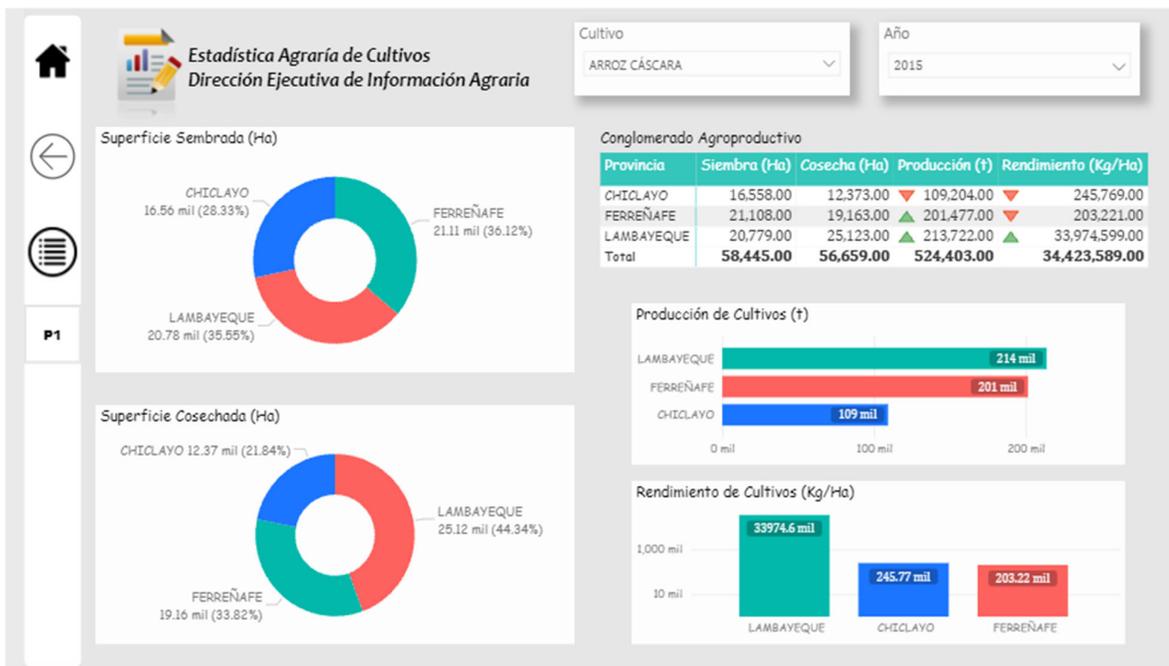


Figura 57. Estadística agraria de Cultivos

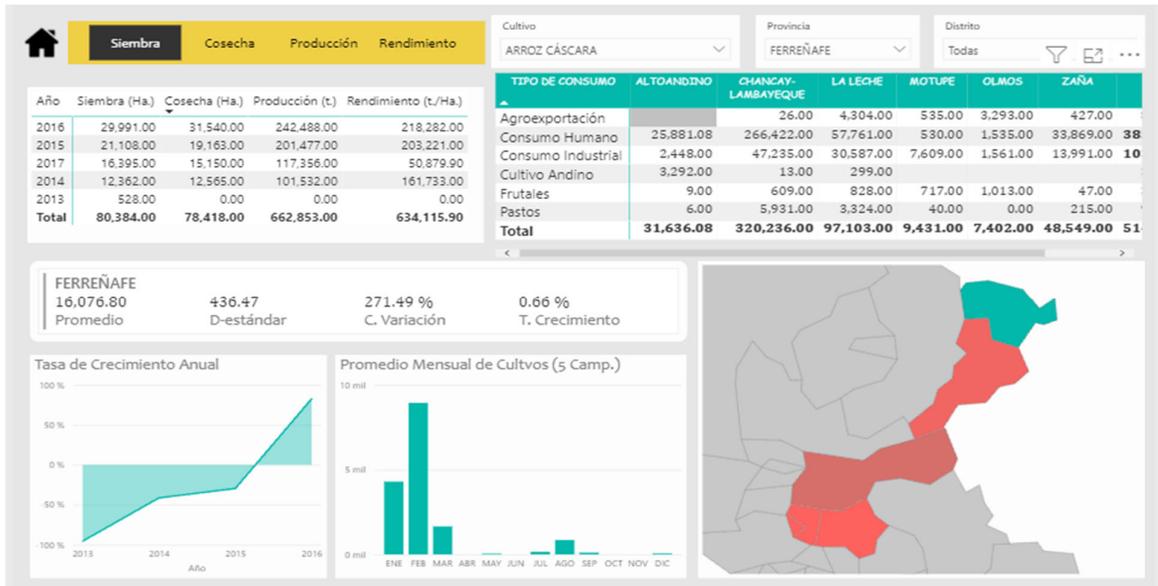


Figura 58. Reporte estadístico y visualización GIS

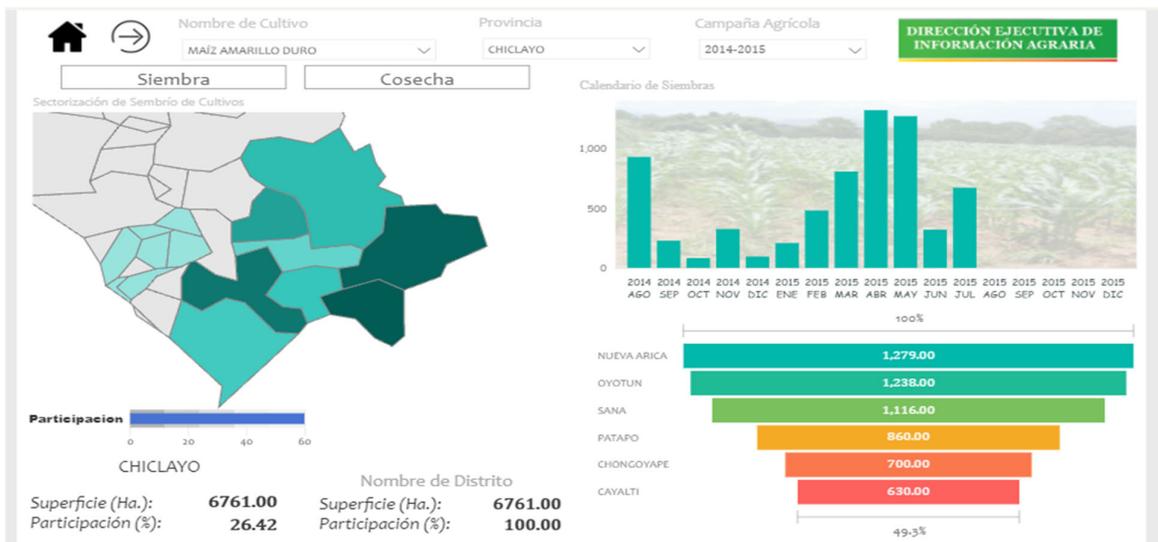


Figura 59. Calendario de siembra y cosecha

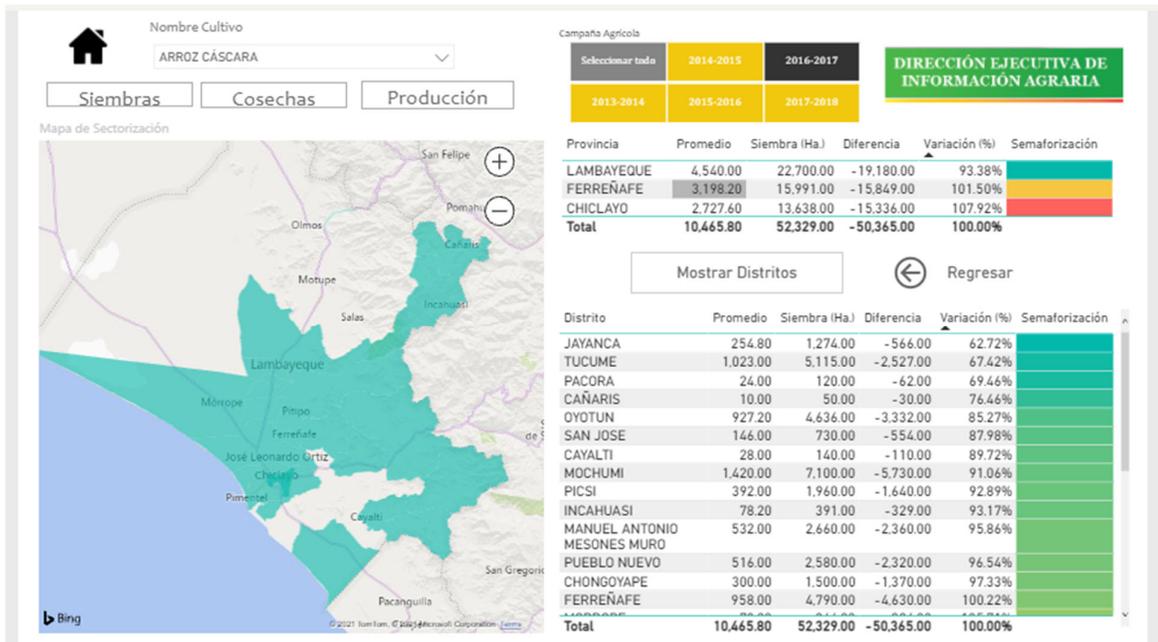


Figura 60. Semaforización de siembra, cosecha y producción

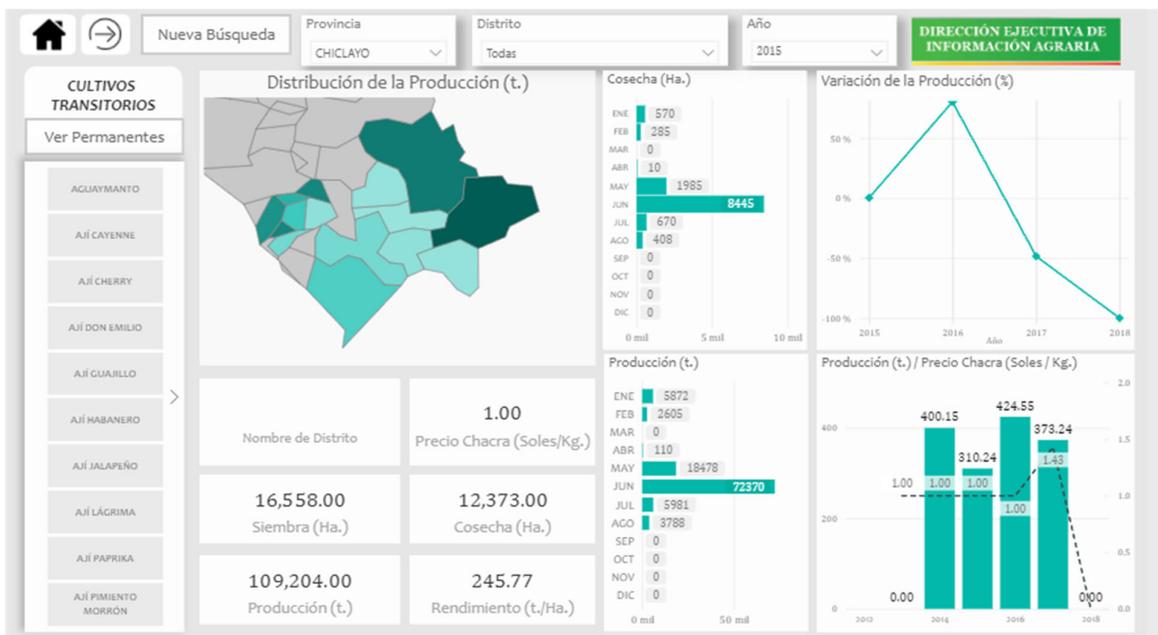


Figura 61. Perfil productivo de Cultivos transitorios

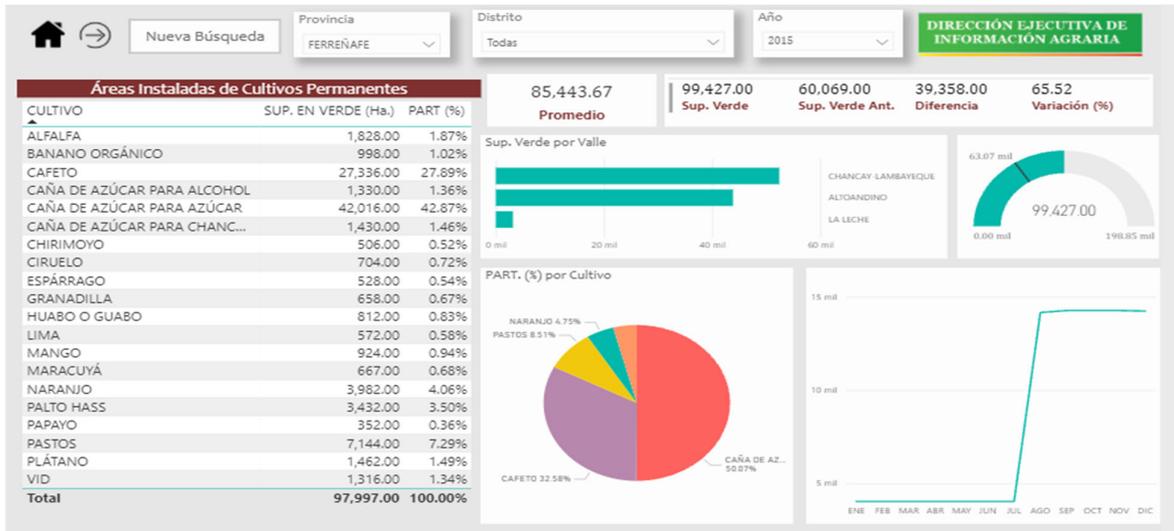


Figura 62. Reporte analítico de Sup. en Verde de Cultivos permanentes

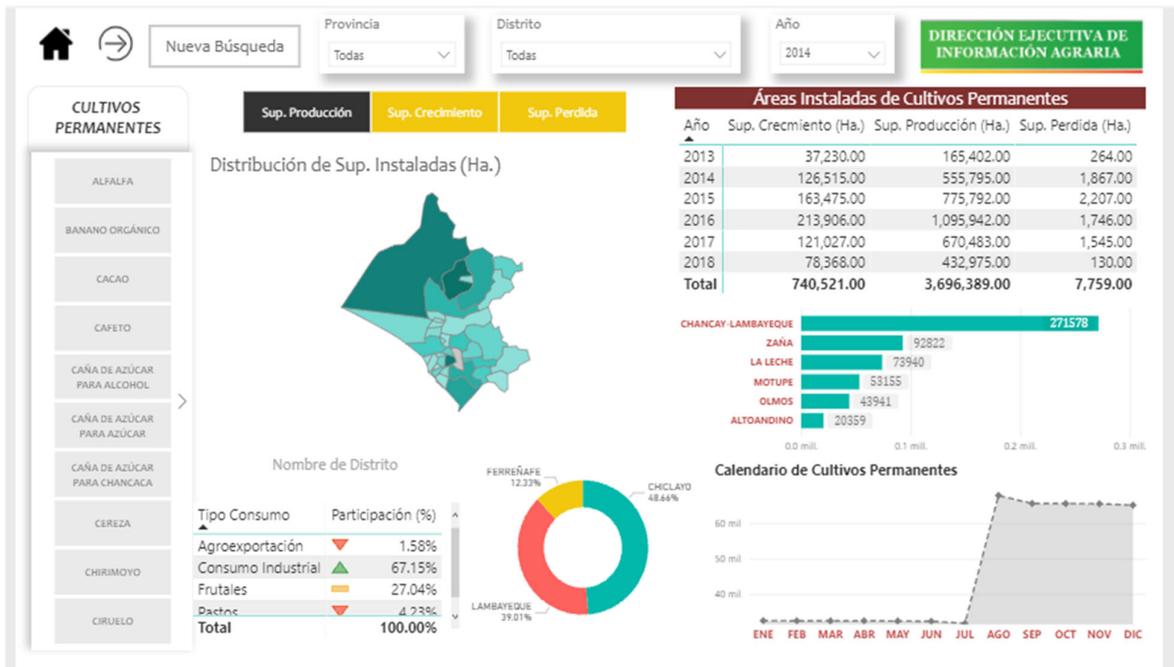


Figura 63. Reporte analítico de Sup. en Producción, Crecimiento y Perdida de Cultivos permanentes

Pronóstico de Producción de Cultivos

| ID | Fecha | Pronóstico |
|----|-----------|------------|
| 1 | Ene. 2018 | 21,695.00 |
| 2 | Feb. 2018 | 3,473.00 |
| 3 | Mar. 2018 | 0.00 |
| 4 | Abr. 2018 | 320.00 |
| 5 | May. 2018 | 88,202.00 |
| 6 | Jun. 2018 | 248,694.00 |
| 7 | Jul. 2018 | 139,247.00 |
| 8 | Ago. 2018 | 4,471.00 |
| 9 | Set. 2018 | 0.00 |
| 10 | Oct. 2018 | 0.00 |
| 11 | Nov. 2018 | 0.00 |
| 12 | Dic. 2018 | 9,293.00 |

[datos.xlsx](#)
[gráficas.png](#)
[pronostico.xlsx](#)

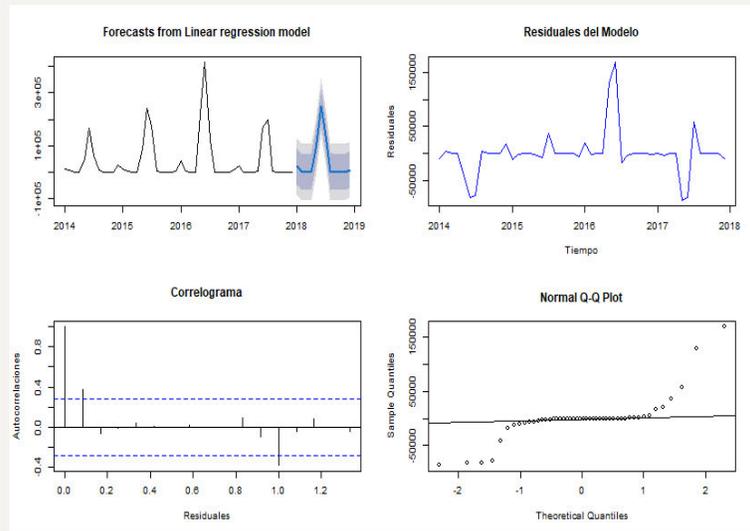


Figura 64. Módulo de pronóstico

```

-- Consulta para Pronóstico --
ALTER PROCEDURE pa_Consulta
(
    @nombre_cultivo varchar(100)
)
AS
BEGIN
    select nombre_cultivo as 'Cultivo',
           fecha as 'Fecha',
           month(fecha) as 'Mes',
           year(fecha) as 'Año',
           SUM(produccion) as 'Producción_t'
    from hecho_cosecha
    inner join dim_cultivo on hecho_cosecha.key_cultivo=dim_cultivo.key_cultivo
    inner join dim_tiempo on hecho_cosecha.key_tiempo=dim_tiempo.key_tiempo
    inner join dim_valle on hecho_cosecha.key_valle=dim_valle.key_valle
    where nombre_cultivo = @nombre_cultivo
    group by nombre_cultivo, fecha, month(fecha), year(fecha)
    order by 1, 4
END
GO

-- ILLIMO, JAYANCA
exec pa_Consulta 'ARROZ CÁSCARA'

```

Figura 65. Consulta al hecho hecho_cosecha

exec pa_Consulta 'ARROZ CÁSCARA'

100 %

Resultados Mensajes

| | Cultivo | Fecha | Mes | Año | Producción_t |
|----|---------------|---------------------|-----|------|--------------|
| 1 | ARROZ CÁSCARA | 2014-01-15 00:00:00 | 1 | 2014 | 12181.00 |
| 2 | ARROZ CÁSCARA | 2014-02-15 00:00:00 | 2 | 2014 | 7845.00 |
| 3 | ARROZ CÁSCARA | 2014-03-15 00:00:00 | 3 | 2014 | 0.00 |
| 4 | ARROZ CÁSCARA | 2014-04-15 00:00:00 | 4 | 2014 | 1169.00 |
| 5 | ARROZ CÁSCARA | 2014-05-15 00:00:00 | 5 | 2014 | 47711.00 |
| 6 | ARROZ CÁSCARA | 2014-06-15 00:00:00 | 6 | 2014 | 167413.00 |
| 7 | ARROZ CÁSCARA | 2014-07-15 00:00:00 | 7 | 2014 | 61830.00 |
| 8 | ARROZ CÁSCARA | 2014-08-15 00:00:00 | 8 | 2014 | 8016.00 |
| 9 | ARROZ CÁSCARA | 2014-09-15 00:00:00 | 9 | 2014 | 0.00 |
| 10 | ARROZ CÁSCARA | 2014-10-15 00:00:00 | 10 | 2014 | 0.00 |
| 11 | ARROZ CÁSCARA | 2014-11-15 00:00:00 | 11 | 2014 | 0.00 |
| 12 | ARROZ CÁSCARA | 2014-12-15 00:00:00 | 12 | 2014 | 26493.00 |
| 13 | ARROZ CÁSCARA | 2015-01-15 00:00:00 | 1 | 2015 | 9990.00 |
| 14 | ARROZ CÁSCARA | 2015-02-15 00:00:00 | 2 | 2015 | 2727.00 |
| 15 | ARROZ CÁSCARA | 2015-03-15 00:00:00 | 3 | 2015 | 0.00 |
| 16 | ARROZ CÁSCARA | 2015-04-15 00:00:00 | 4 | 2015 | 110.00 |
| 17 | ARROZ CÁSCARA | 2015-05-15 00:00:00 | 5 | 2015 | 85439.00 |
| 18 | ARROZ CÁSCARA | 2015-06-15 00:00:00 | 6 | 2015 | 241717.00 |
| 19 | ARROZ CÁSCARA | 2015-07-15 00:00:00 | 7 | 2015 | 176342.00 |
| 20 | ARROZ CÁSCARA | 2015-08-15 00:00:00 | 8 | 2015 | 4678.00 |
| 21 | ARROZ CÁSCARA | 2015-09-15 00:00:00 | 9 | 2015 | 0.00 |
| 22 | ARROZ CÁSCARA | 2015-10-15 00:00:00 | 10 | 2015 | 0.00 |

Figura 66. Consulta al hecho hecho_cosecha

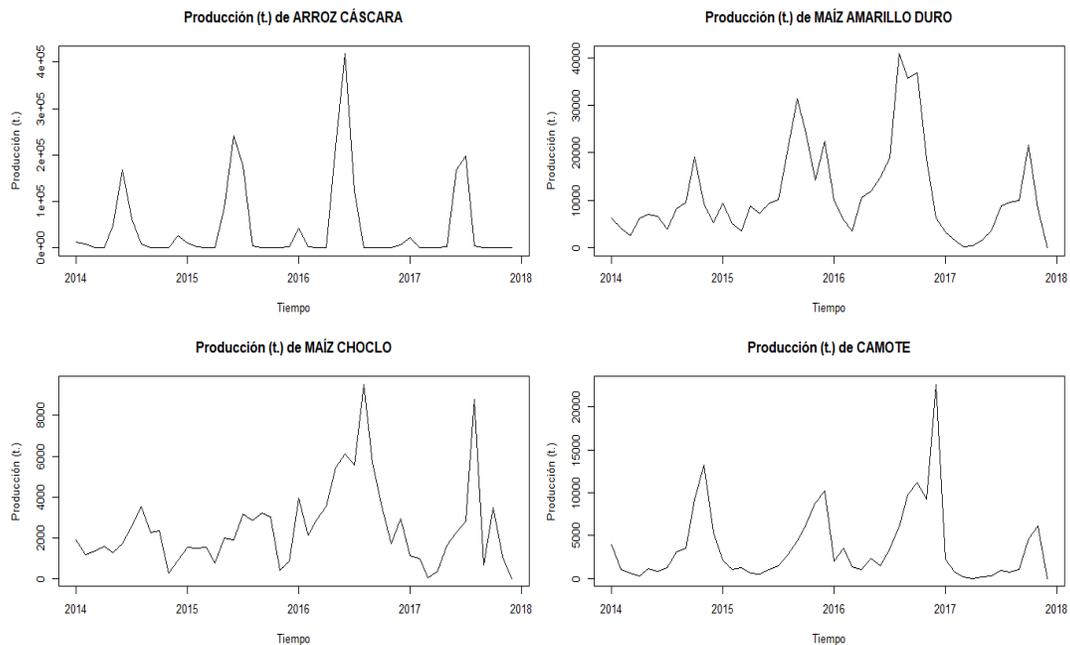


Figura 67. Producción (t.) de cultivos por años

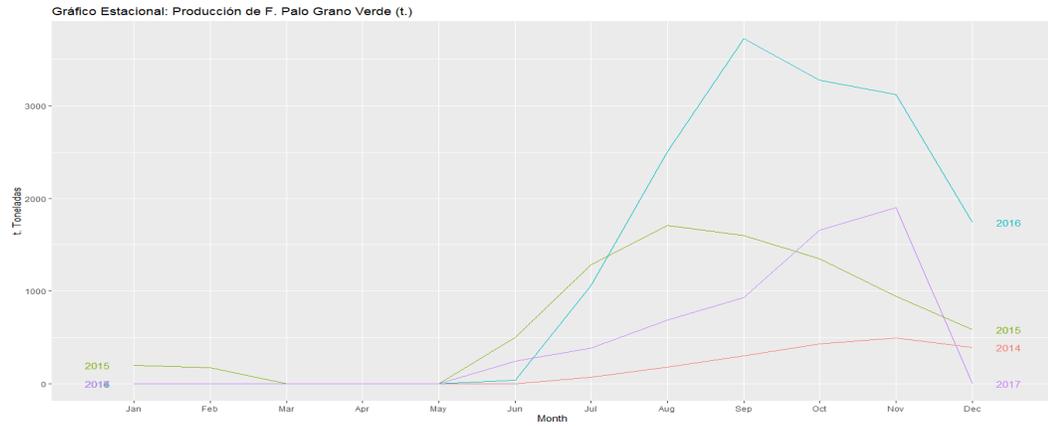


Figura 68. Gráfico Estacional: Producción de F. Palo Grano Verde (t.)

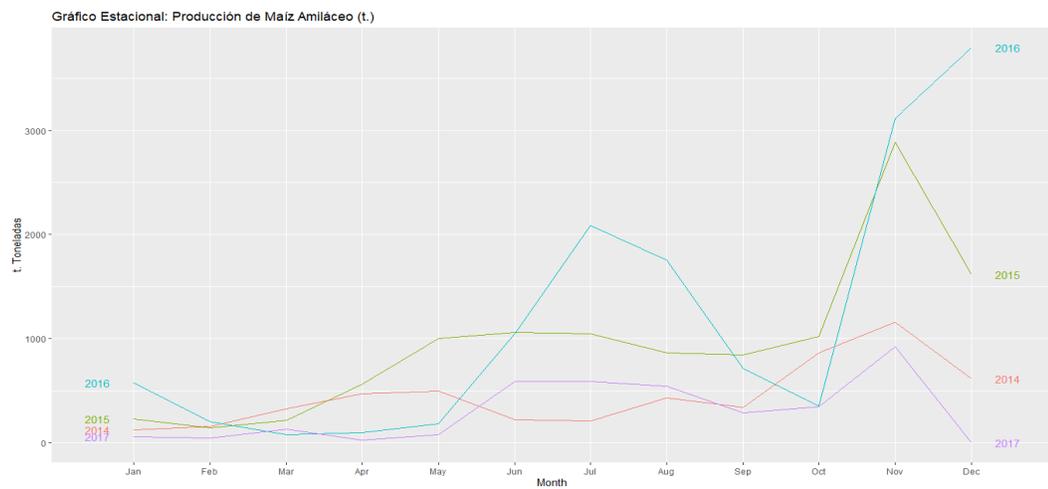


Figura 69. Gráfico Estacional: Producción de Maíz Amiláceo

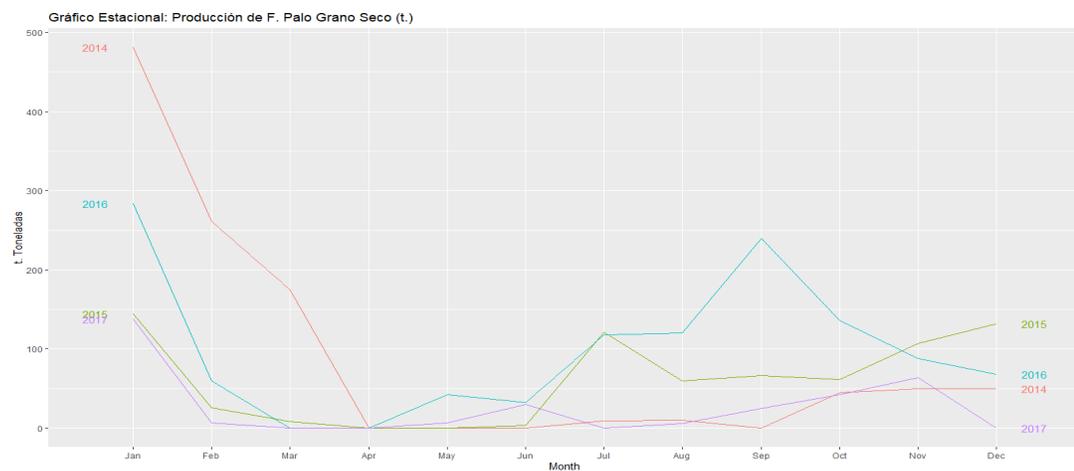


Figura 70. Gráfico Estacional: Producción de F. Palo Grano Seco (t.)

Tabla 19.
Estadísticos descriptivos de la variable Producción (t.) de Arroz cáscara

| PRODUCCIÓN T | |
|---------------------------|------------------|
| Media | 38,814.98 |
| Error típico | 9,777.64 |
| Mediana | 4,432.50 |
| Moda | 0.00 |
| Desviación estándar | 67,741.49 |
| Varianza de la muestra | 4,588,909,144.83 |
| Curtosis | 2.08 |
| Coefficiente de asimetría | 1.85 |
| Rango | 241,717.00 |
| Mínimo | 0.00 |
| Máximo | 241,717.00 |
| Suma | 1,863,119.00 |
| Cuenta | 48.00 |

```
> summary(ur.df(data, type = "trend", selectlags = "AIC")) #Intercepto y Tendencia
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min     1Q  Median     3Q     Max
-81070 -30662 -24977  23297 160650

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 34226.1368  17193.5970   1.991 0.053055 .
z.lag.1      -0.7848     0.1325  -5.923 5.13e-07 ***
tt           -99.7471    588.1990  -0.170 0.866154
z.diff.lag    0.5265     0.1312   4.012 0.000242 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 52930 on 42 degrees of freedom
Multiple R-squared:  0.4651,    Adjusted R-squared:  0.4268
F-statistic: 12.17 on 3 and 42 DF,  p-value: 7.251e-06

Value of test-statistic is: -5.923 11.7088 17.5605

Critical values for test statistics:
      1pct   5pct  10pct
tau3  -4.15  -3.50  -3.18
phi2   7.02   5.13   4.31
phi3   9.31   6.73   5.61
```

Figura 71. Resultados de la Prueba de Dickey-Fuller Aumentada en R

H_0 : Raíz unitaria. (No es estacionaria) p-valor > .05
 H_1 : No hay raíz unitaria. (Es estacionaria) p-valor < .05

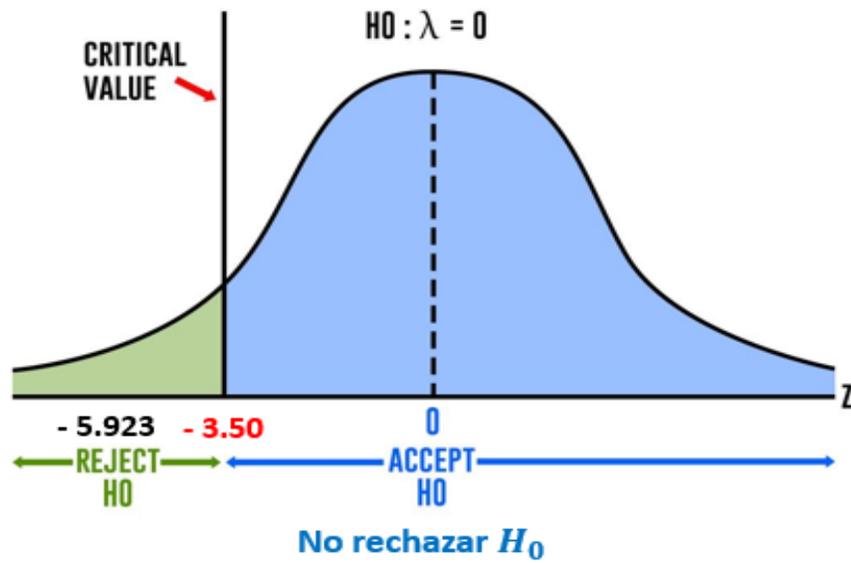


Figura 72. Prueba de Dickey-Fuller Aumentada

Tabla 20.

Prueba de Levene para la Homogeneidad de la varianza (Centro = Mediana)

| | Df | F-value | Pr (>F) |
|-------|----|---------|---------|
| Group | 3 | 0.0459 | 09868 |
| | 44 | | |

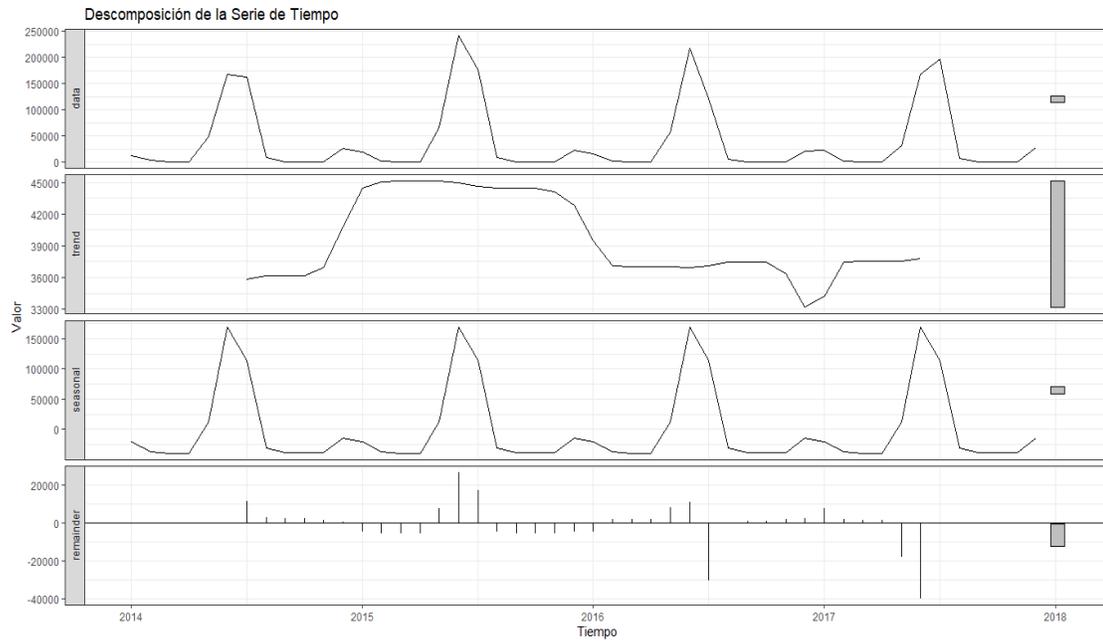


Figura 73. Gráfica de Descomposición de la Serie de Tiempo

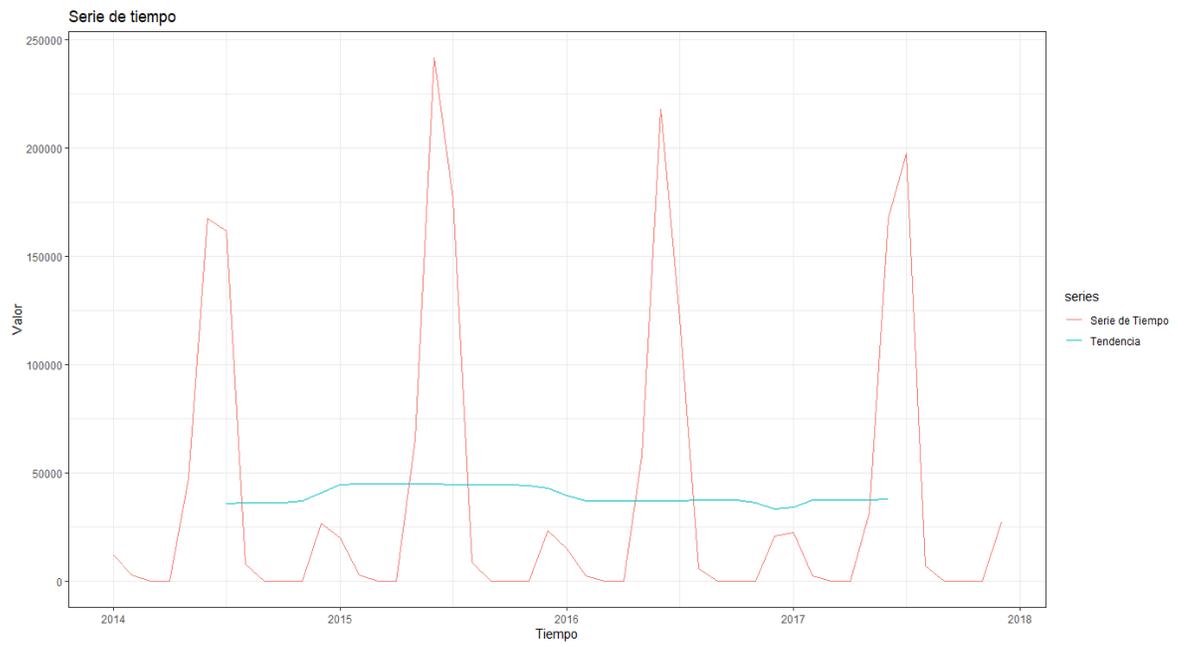


Figura 74. Gráfica de la Serie de Tiempo con su Tendencia

```

### MODELOS PREDUCTIVOS ###
# 1. Modelo Naive Simple
mNS <- naive(data2, h=plazo)

# 2. Modelo Naive + Estacionalidad
mSN <- snaive(data2, h=plazo)

# 3. Modelo de Regresión
mTen <- forecast(tslm(data2 ~ trend), h=plazo) #Tendencia
mStac <- forecast(tslm(data2 ~ season), h=plazo) #Estacionalidad
mStdTen <- forecast(tslm(data2 ~ trend + season), h=plazo) #Tendencia + Estacionalidad

# 4. Modelo Suavizado exponencial
## Suavizado exponencial simple
modSes <- ses(data2, h = plazo)

## Suavizado doble o Método Holt
mHoltAdd <- holt(data2, type=c("additive"))
mHoltMul <- holt(data2, type=c("multiplicative"))
mDobAdd=forecast(mHoltAdd, h=12)
mDobMul=forecast(mHoltMul, h=12)

## Método de Holt-Winters
mHW=forecast(HoltWinters(data2), h=plazo)

# 5. Modelo Medias móviles
## Método MA
# Elaborando el Modelo
modMA <- ma(data2, order = 3, centre = TRUE)
#forecast(modMA, h = 13)

## Método de Arima
mod <- auto.arima(data2, stepwise = FALSE, approximation = FALSE, trace = FALSE)
mod

mARIMA <- forecast(auto.arima(data2, stepwise = FALSE, approximation = FALSE, trace = FALSE),
                  level = c(95), h = plazo)

```

Figura 75. Modelos predictivos en R

```

> summary(mSN)

Forecast method: Seasonal naive method

Model Information:
Call: snaive(y = data2, h = plazo)

Residual sd: 20221.9547

Error measures:

```

| | ME | RMSE | MAE | MPE | MAPE | MASE | ACF1 |
|--------------|---------|----------|----------|-----------|---------|------|-----------|
| Training set | 632.625 | 20221.95 | 8951.958 | -4.737964 | 20.8456 | 1 | 0.4126462 |

```

Forecasts:

```

| | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|----------|----------------|------------|-----------|----------|----------|
| Jan 2017 | 15128 | -10787.478 | 41043.48 | -24506.3 | 54762.3 |
| Feb 2017 | 2640 | -23275.478 | 28555.48 | -36994.3 | 42274.3 |
| Mar 2017 | 0 | -25915.478 | 25915.48 | -39634.3 | 39634.3 |
| Apr 2017 | 0 | -25915.478 | 25915.48 | -39634.3 | 39634.3 |
| May 2017 | 57868 | 31952.522 | 83783.48 | 18233.7 | 97502.3 |
| Jun 2017 | 218018 | 192102.522 | 243933.48 | 178383.7 | 257652.3 |
| Jul 2017 | 121198 | 95282.522 | 147113.48 | 81563.7 | 160832.3 |
| Aug 2017 | 6020 | -19895.478 | 31935.48 | -33614.3 | 45654.3 |
| Sep 2017 | 0 | -25915.478 | 25915.48 | -39634.3 | 39634.3 |
| Oct 2017 | 0 | -25915.478 | 25915.48 | -39634.3 | 39634.3 |
| Nov 2017 | 0 | -25915.478 | 25915.48 | -39634.3 | 39634.3 |
| Dec 2017 | 20800 | -5115.478 | 46715.48 | -18834.3 | 60434.3 |

```

> |

```

Figura 76. Resultados del método Naive estacional en R

H_0 : Hay ruido blanco. p-valor > .05

H_1 : No hay ruido blanco. p-valor < .05

Tabla 21.

Resultados de Box-Ljung para Naive estacional

| Prueba | Chi-cuadrado | p-valor |
|----------------|--------------|---------|
| Box-Ljung test | 0.26925 | 0.6038 |

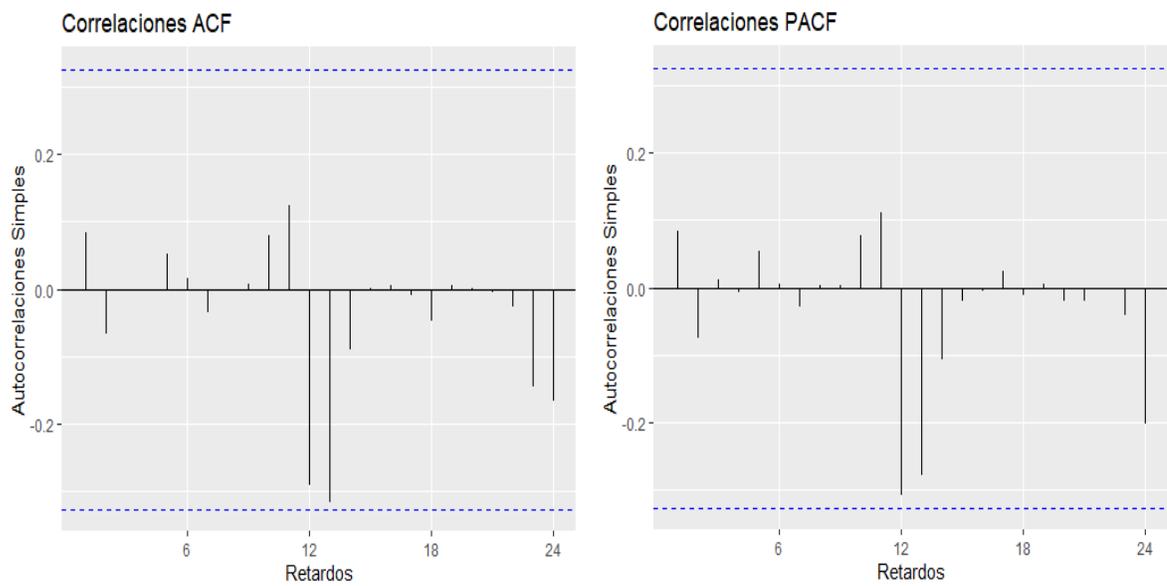


Figura 77. Gráfica de Auto correlación de residuales

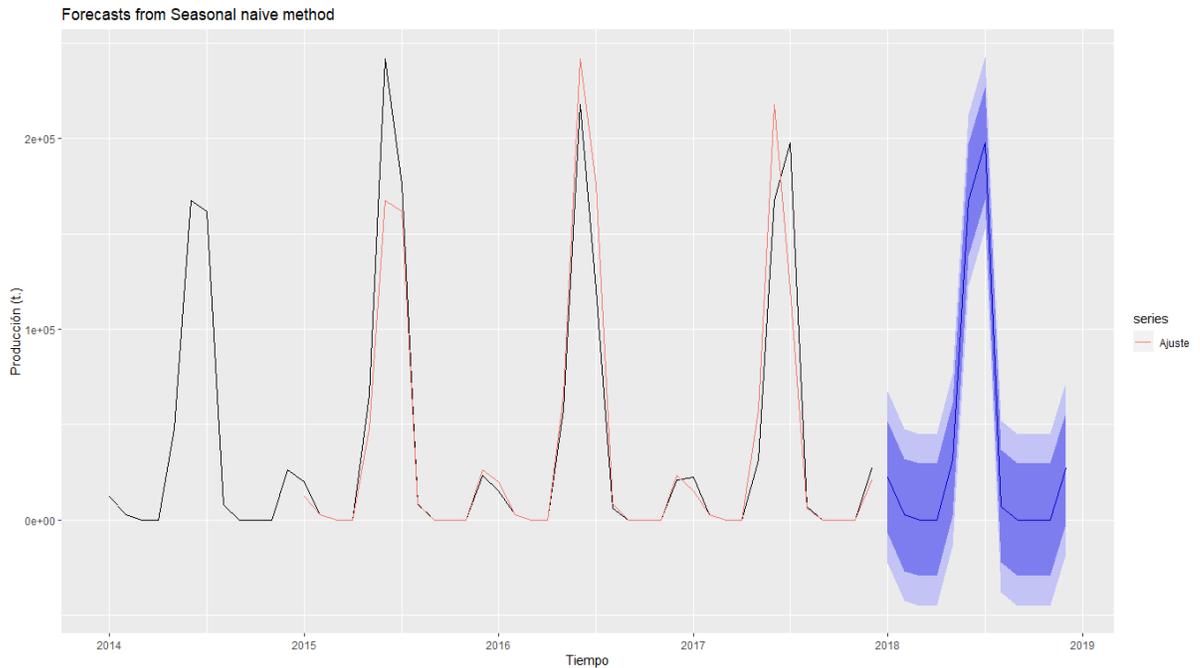


Figura 78. Valores observados y ajuste con Naive estacional

Tabla 22.
Resultados de la Prueba Box-Ljung en R

| Prueba Box-Ljung test | Chi-cuadrado | p-valor |
|---|--------------|----------|
| Naïve simple | 0.91391 | 0.3391 |
| Naïve con ajuste estacional | 0.26925 | 0.6038 |
| Regresión lineal con ajuste de Tendencia | 9.8088 | 0.001737 |
| Regresión lineal con ajuste estacional | 1.3285 | 0.2491 |
| Regresión lineal con ajuste de Tendencia y estacionalidad | 1.3054 | 0.2532 |
| Suavizado exponencial Simple | 1.3795 | 0.2402 |
| Suavizado exponencial doble Holt Aditivo | 9.944 | 0.001614 |
| Suavizado exponencial doble Holt Multiplicativo | 9.944 | 0.001614 |
| Suavizado exponencial Holt-Winters | 0.03634 | 0.8488 |
| Modelo ARIMA | 0.093694 | 0.7595 |
| Red neuronal auto regresiva | 3.1009 | 0.07825 |

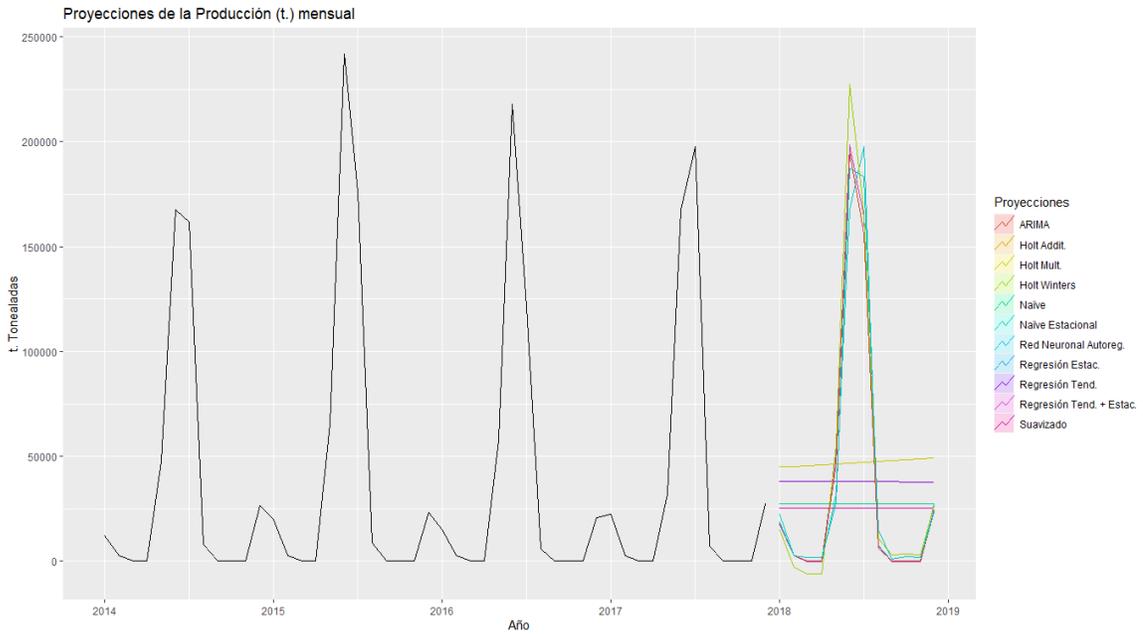


Figura 79. Valores observados y ajuste con Naive estacional

Tabla 23.
Fórmulas de Errores de Pronóstico

$$RMSE = \frac{\sum \text{Error de pronóstico}^2}{n}$$

$$MAD = \frac{\sum | \text{Real} - \text{Pronóstico} |}{n}$$

MAPE

$$= \frac{\sum_{i=1}^n 100 \frac{| \text{Real}_i - \text{Pronóstico}_i |}{\text{Real}_i}}{n}$$

$$MASE = \frac{| \text{Real} - \text{Pronóstico} |}{MAD}$$

$$sAPE = 2 * \frac{|X_i - f_i|}{|X_i + f_i|}$$

$$SMAPE = \frac{1}{m} * \sum_{i=1}^m sAPE$$

Tabla 24.
Resultados obtenidos de la función accuracy de R [Train]

| Training set | RMSE | MAE | MAPE | MASE |
|------------------------------|-----------|-----------|---------|--------|
| Método naif simple | 68,076.67 | 39,630.37 | Inf | 4.4270 |
| Método naif estacional | 20,221.95 | 8,951.96 | 20.8456 | 1.0000 |
| Regresión + Trend | 67,407.79 | 50,326.94 | Inf | 5.6219 |
| Regresión + Seasonal | 11,456.18 | 5,092.67 | Inf | 0.5689 |
| Regresión + Trend + Seasonal | 11,444.53 | 5,338.69 | Inf | 0.5964 |
| Suavizado exponencial | 67,996.31 | 40,171.31 | Inf | 4.4874 |
| Método Holt Winters | 10,964.59 | 6,451.41 | Inf | 0.7207 |
| Método ARIMA | 13,726.78 | 4,804.264 | 10.8533 | 0.5367 |
| Red Neuronal Autoregres. | 15,424.69 | 8,147.95 | Inf | 0.9102 |

Tabla 25.
Resultados obtenidos de los errores de c/Modelo predictivo

| MODELO | RMSE | MAE | MAPE | SMAPE | MASE |
|-----------------------------------|-----------|--------|--------|--------|--------|
| SNAIVE | 27,623.19 | 13,994 | 18.75% | 18.13% | 1.0000 |
| TSLM (~ season) | 19,127.42 | 10,167 | 14.79% | 13.43% | 1.0000 |
| TSLM (~ trend + season) | 19,235.88 | 10,799 | 19.68% | 16.92% | 1.0000 |
| HoltWinters ($\alpha = 0.7631$) | 24,957.48 | 17,125 | 51.25% | 92.22% | 1.0000 |
| ARIMA (0,0,1) (0,1,0) [12] | 24,323.14 | 12,468 | 16.67% | 14.21% | 1.0000 |
| NNAR (1,1,2) [12] | 25,933.32 | 14,575 | 23.74% | 22.37% | 1.0000 |

*Tabla 26.
Proceso ETL para la Dimensión Valle*

| | Datos de Origen | | | Proceso de ETL | | | Datos Destino | | |
|---|-----------------|--------------|----------|----------------|----------------|------------|---------------|--------------|----------|
| | | | | Extracción | Transformación | Carga | | | |
| | Nombre | Tipo de Dato | Longitud | | | | Nombre | Tipo de Dato | Longitud |
| 1 | Valle | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | nombre_valle | DT_STR | 100 |
| 2 | Distrito | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | distrito | DT_STR | 100 |
| 3 | Provincia | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | provincia | DT_STR | 100 |

*Tabla 27.
Proceso ETL para la Dimensión Campaña*

| | Datos de Origen | | | Proceso de ETL | | | Datos Destino | | |
|---|-----------------|--------------|----------|----------------|----------------|------------|---------------------|--------------|----------|
| | | | | Extracción | Transformación | Carga | | | |
| | Nombre | Tipo de Dato | Longitud | | | | Nombre | Tipo de Dato | Longitud |
| 1 | Campaña | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | descripcion campaña | DT_STR | 9 |

Tabla 28.
Proceso ETL para la Dimensión Cultivo

| | Datos de Origen | | | Proceso de ETL | | | Datos Destino | | |
|----|-----------------|--------------|----------|----------------|----------------|------------|---------------------|--------------|----------|
| | | | | Extracción | Transformación | Carga | | | |
| | Nombre | Tipo de Dato | Longitud | | | | Nombre | Tipo de Dato | Longitud |
| 1 | Campaña | DT_WSTR | 255 | Archivos XLS | - | - | - | | |
| 2 | Tipo Consumo | DT_WSTR | 255 | Archivos XLS | - | - | - | | |
| 3 | Tipo | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | nombre_tipo_cultivo | DT_STR | 100 |
| 4 | Provincia | DT_WSTR | 255 | Archivos XLS | - | - | - | | |
| 5 | Distrito | DT_WSTR | 255 | Archivos XLS | - | - | - | | |
| 6 | Valle | DT_WSTR | 255 | Archivos XLS | - | - | - | | |
| 7 | Cultivo | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | nombre_cultivo | DT_STR | 100 |
| 8 | Variables | DT_WSTR | 255 | Archivos XLS | - | - | - | | |
| 9 | Int#Siemb# | DT_WSTR | 255 | Archivos XLS | - | - | - | | |
| 10 | Año | DT_R8 | 0 | Archivos XLS | - | - | - | | |
| 11 | Mes | DT_R8 | 0 | Archivos XLS | - | - | - | | |
| 12 | Valores | DT_R8 | 0 | Archivos XLS | - | - | - | | |
| 13 | Fecha_mes_año | DT_WSTR | 255 | Archivos XLS | - | - | - | | |

Tabla 29.
Proceso ETL para la Dimensión Tipo de Consumo

| | Datos de Origen | | | Proceso de ETL | | | Datos Destino | | |
|----|-----------------|--------------|----------|----------------|----------------|------------|---------------|--------------|----------|
| | | | | Extracción | Transformación | Carga | | | |
| | Nombre | Tipo de Dato | Longitud | | | | Nombre | Tipo de Dato | Longitud |
| 1 | Campaña | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 2 | Tipo Consumo | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | tipo_consumo | DT_STR | 100 |
| 3 | Tipo | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 4 | Provincia | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 5 | Distrito | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 6 | Valle | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 7 | Cultivo | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 8 | Variables | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 9 | Int#Siemb# | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |
| 10 | Año | DT_R8 | 0 | Archivos XLS | - | - | - | - | - |
| 11 | Mes | DT_R8 | 0 | Archivos XLS | - | - | - | - | - |
| 12 | Valores | DT_R8 | 0 | Archivos XLS | - | - | - | - | - |
| 13 | Fecha_mes_año | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - |

Tabla 30.
Proceso ETL para el Hecho Siembra y Cosecha

| | Datos de Origen | | | Proceso de ETL | | | Datos Destino | | | | |
|---|-----------------|--------------|----------|----------------|----------------|------------|---------------------|----------------|----------|-----------|--------|
| | | | | Extracción | Transformación | Carga | | | | | |
| | Nombre | Tipo de Dato | Longitud | | | | Nombre | Tipo de Dato | Longitud | Precisión | Escala |
| 1 | Campaña | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | descripcion_campana | DT_STR | 100 | - | - |
| 2 | Valle | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | nombre_valle | DT_STR | 100 | - | - |
| 3 | Tipo consumo | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | tipo_consumo | DT_STR | 100 | - | - |
| 4 | Cultivo | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | nombre_cultivo | DT_STR | 100 | - | - |
| 5 | Fecha_mes_año | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | - | DT_DBTIMESTAMP | - | - | - |
| 6 | Valores | DT_R8 | 0 | Archivos XLS | Conversión | SQL Server | - | DT_NUMERIC | - | 18 | 2 |
| 7 | Variables | DT_WSTR | 255 | Archivos XLS | - | - | - | - | - | - | - |
| 8 | Distrito | DT_WSTR | 255 | Archivos XLS | Conversión | SQL Server | distrito | DT_STR | 100 | - | - |

Tabla 31.
Encuesta de Validación

| Indicador | Ítem | | | | | Medida de Operacionalización (¿Qué se mide?) |
|--|--|---------------|--------------------------------|------------|-----------------------|---|
| Nivel de calidad de la información reportada por la Solución de BI. | 1.- El sistema provee reportes de rápida comprensión. 2.- El sistema provee reportes con información actualizada. 3.- A través de los reportes obtengo información suficiente y necesaria para abarcar todos los factores críticos que influyen en la correcta ejecución de mis funciones. 4.- Los reportes son consistentes, es decir, la información que brinda es congruente con todos los objetivos de negocio. 5.- Los reportes son concisos, puntuales, selectivos y resumidos adecuadamente. | | | | | <ul style="list-style-type: none"> ▪ Comprensibilidad ▪ Actual ▪ Completa ▪ Relevante ▪ Síntesis |
| | 1 | 2 | 3 | 4 | 5 | |
| | Totalmente en desacuerdo | En desacuerdo | Ni de acuerdo ni en desacuerdo | De acuerdo | Totalmente de acuerdo | |
| Nivel de satisfacción del usuario con respecto a la información requerida. | 6.- Confío plenamente en los reportes que me brinda el sistema. 7.- Si me fuera a trabajar a una empresa de la competencia, recomendaría que implanten un sistema igual por la eficiencia de sus reportes. 8.- Los reportes me permiten adquirir un conocimiento valioso que de otra forma me resultaría muy difícil o imposible de alcanzarla. 9.- Cuando estoy usando las funcionalidades de los reportes del sistema mi estado es de disfrute. 10.- En términos generales estoy satisfecho con el sistema de información. | | | | | <ul style="list-style-type: none"> ▪ Confianza ▪ Eficiencia ▪ Disfrute ▪ Satisfacción general |
| | 1 | 2 | 3 | 4 | 5 | |
| | Totalmente en desacuerdo | En desacuerdo | Ni de acuerdo ni en desacuerdo | De acuerdo | Totalmente de acuerdo | |