

Propuesta de Proyecto: Pronóstico de Demanda de Huevos para empresa avícola mediante Machine Learning

Presentado por:

LUIS CARLOS CORREA Cod. 22501541

CARLOS ANDRES BECERRA G Cod. 22500215

JUAN DAVID DÍAZ CALERO Cod. 22502473

Curso: Aprendizaje Automático

Profesor: Francisco Jose Mercado Rivera, PhD

Fecha: 8 de Mayo de 2025

Lugar: Cali, Valle del Cauca



Introducción

Avícola productora de huevo

Empresa regional con sede en el Valle del Cauca, dedicada a la producción y comercialización de huevos frescos desde hace más de 50 años. Su actividad principal es la producción y distribución de huevos en los canales de venta como Autoservicios, TAT, Horeca y mayoristas en el territorio colombiano.

Desafío Actual

A pesar de su trayectoria, se enfrenta el desafío constante de gestionar eficientemente un producto altamente perecedero con una demanda que presenta variaciones significativas, impactando la planificación de la producción, los niveles de inventario y la logística de distribución.

Objetivo del Proyecto

Desarrollar y evaluar un modelo de Machine Learning para pronosticar la demanda semanal de huevos, utilizando el historial de ventas de 2 años (2022-2023), con el fin de mejorar la precisión de las predicciones a un error inferior al 10% (MAPE).

Descripción del Problema

Problema Central

Actualmente está avícola se basa en los métodos de pronóstico tradicionales (promedios históricos y estimaciones manuales) que resultan en una desviación promedio del 20-25% respecto a la demanda real.

Subproducción

En picos de demanda no anticipados, se calcula una pérdida de ventas del 5-7% y dificultades para cumplir con pedidos de clientes clave.

Ineficiencias Operativas

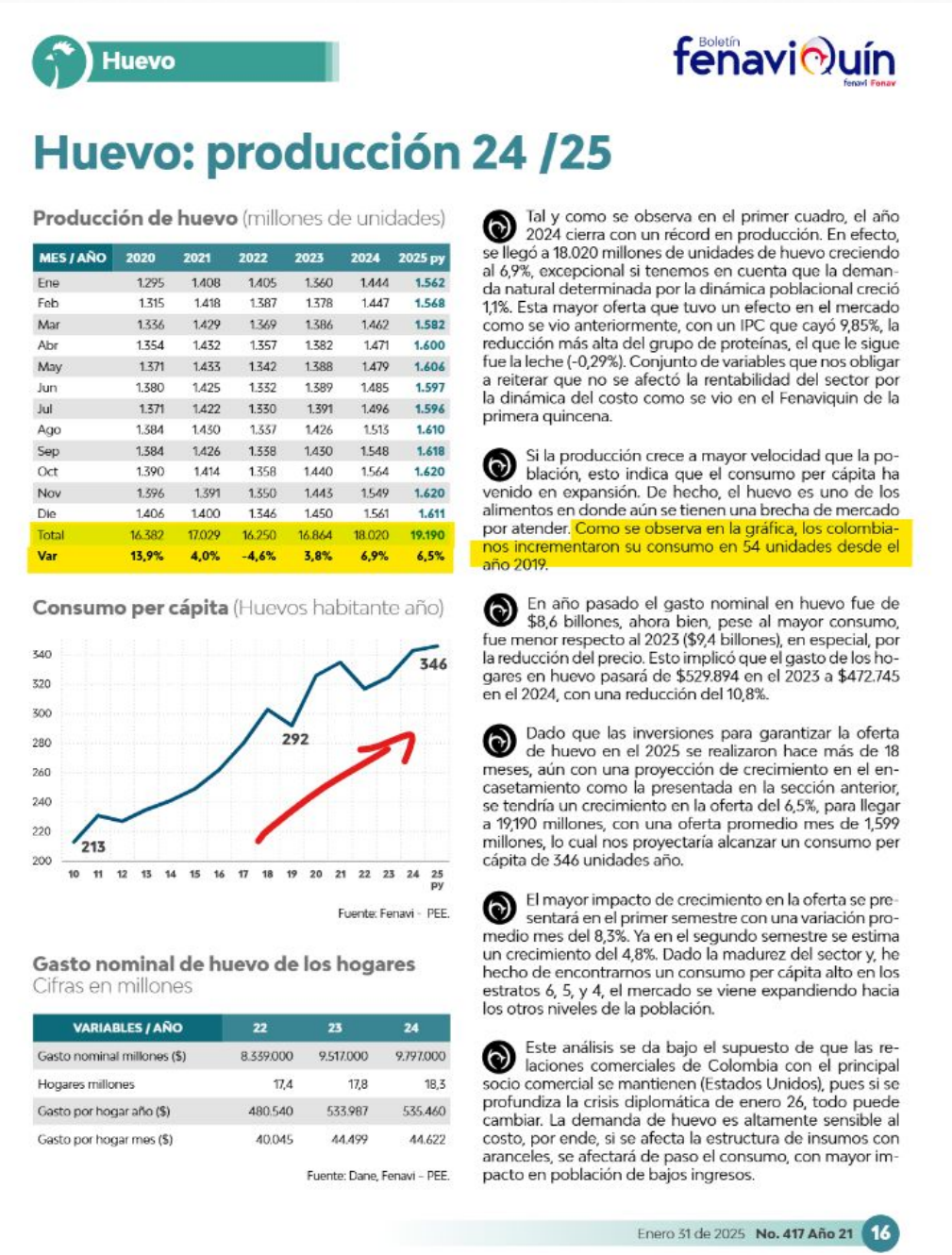
Costos elevados en transporte urgente para cubrir déficits y asignación subóptima de recursos en las granjas y centros de distribución.



Resultados esperados para la avicola

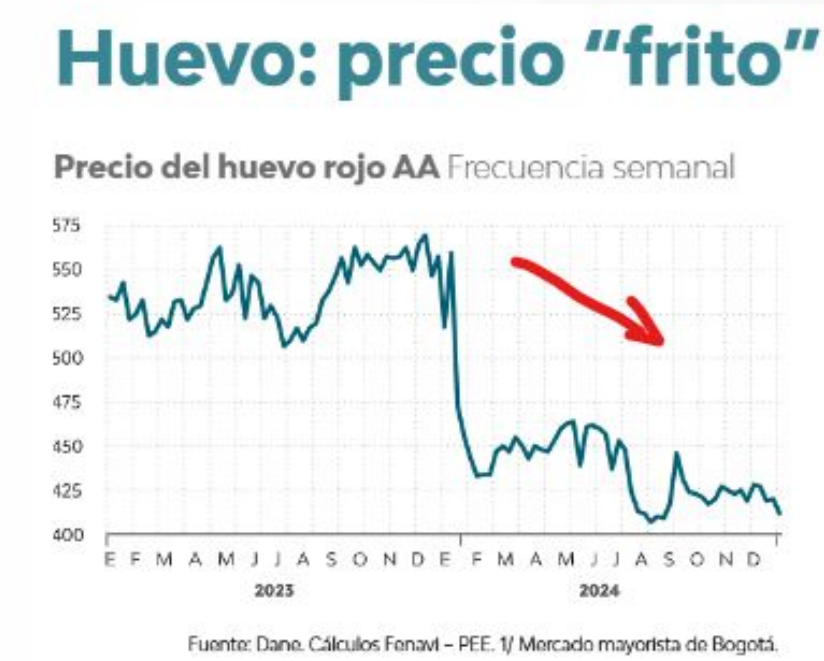


Pronóstico de Demanda en la Industria del huevo en Colombia

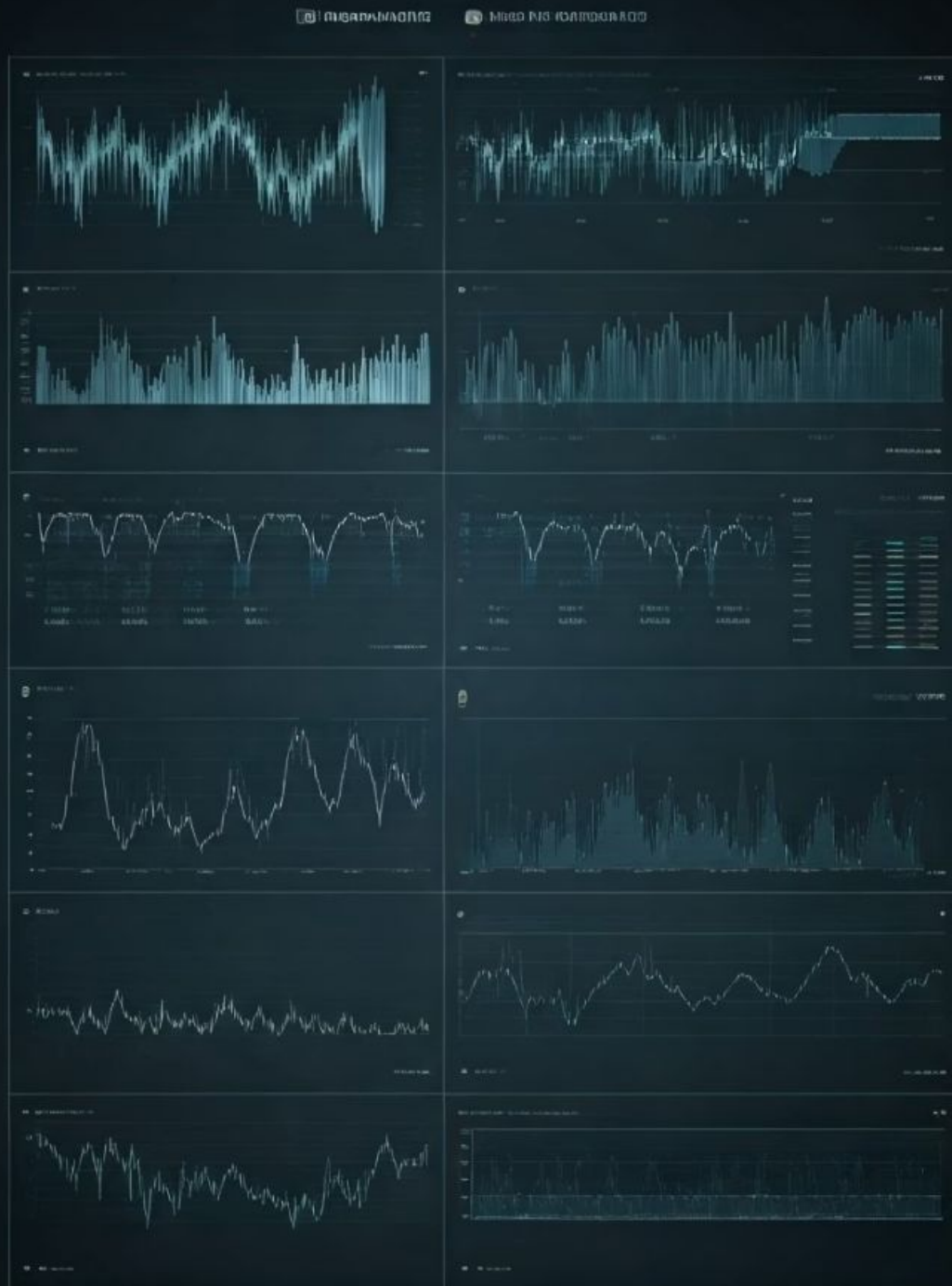


Fenavi- Federación Nacional de Avicultores de Colombia

Segun la informacion que genera el gremio de avicultores, el consumo per cápita del huevo viene en aumento en los últimos años posicionando el producto como uno de los más importantes en la canasta familiar, esto a hecho que mas avicultores incrementen su capacidades de producción pero al mismo tiempo hacen que el precio del huevo disminuya (Oferta y demanda).



<https://fenavi.org/boletin-fenaviquin/fenaviquin-edicion-417-enero-31-de-2025/>



Modelos Comunes de Machine Learning para Series Temporales



Modelos Estadísticos Clásicos

ARIMA y SARIMA son ampliamente usados para series temporales agrícolas, capturando autocorrelacion y estacionalidad según referencias de Restack.io y Allied Academies.



Random Forest y XGBoost

Efectivos para capturar relaciones no lineales y manejar diversas características. Han mostrado buenos resultados en pronósticos de demanda en retail.



LSTMs (Redes Neuronales)

Adecuadas para secuencias temporales largas y dependencias complejas, aunque requieren más datos y cómputo. Utilizadas en pronóstico de cosechas.



Prophet (Facebook)

Diseñado para series temporales de negocios con múltiples estacionalidades y días festivos, robusto a datos faltantes y outliers. Ha sido aplicado con éxito en pronósticos de ventas minoristas.

Pronóstico de Demanda en la Industria de Alimentos Perecederos

Importancia Crítica

La literatura subraya que para bienes perecederos, como los huevos, la precisión en el pronóstico es crucial. Errores pueden llevar a pérdidas económicas directas y afectar la satisfacción del cliente.

Las referencias de MobiDev y Slimstock destacan que la volatilidad de la demanda, estacionalidad y factores externos hacen del pronóstico un desafío complejo en este sector.

Desafíos Específicos

Según estudios de ResearchGate sobre "Demand Forecasting Models for Food Industry", los productos perecederos presentan retos únicos:

- Vida útil limitada que reduce el margen de error
- Fluctuaciones estacionales pronunciadas
- Sensibilidad a eventos externos y promociones
- Variabilidad en la calidad del producto

Aplicaciones de Machine Learning en el Sector Avícola

Estudios de Producción

"Machine Learning Techniques for Egg Production Prediction" (CILAMCE) ha demostrado la viabilidad de modelos como Ridge Regression y XGBoost para predecir la producción de huevos, considerando variables como temperatura y consumo de alimento, logrando un MAPE del 3.81% en sus mejores casos.

Aplicabilidad al Proyecto

Aunque menos específico para demanda de huevos, estos trabajos validan el uso de ML en el sector y sugieren variables análogas o que podrían tener un efecto indirecto en la predicción de demanda.



Pronóstico de Oferta

Investigaciones en "Machine Learning Based Egg Supply Forecasting" (JOAAT) exploran el uso de ML para la oferta de huevos, indicando la importancia de datos históricos y características del producto.

Revisión de la literatura sobre otras soluciones al problema

Soluciones globales en productos perecederos y huevos

Support Vector Regression (SVR)

Se ha demostrado que los modelos de SVR mejoran la precisión de pronósticos en productos perecederos gracias a su capacidad de generalización y garantía de mínimos globales para los datos de entrenamiento. Por ejemplo, un estudio aplicó SVR a frutas y verduras y obtuvo mejoras significativas en el forecasting precision frente a métodos estadísticos tradicionales.

[1] S. Tefvik Altay, “Demand forecasting of perishable farm products using support vector machine,” ResearchGate, Jun. 2024.

Algoritmos de boosting: XGBoost

XGBoost se ha consolidado como uno de los más precisos para datos complejos. En particular, para predecir la tasa de producción de huevos en granjas de reproductoras, XGBoost superó a Random Forest, enfatizando su habilidad para capturar patrones sutiles en datos no lineales

[2] J. Doe et al., “Predicting egg production rate and egg weight of broiler breeders using XGBoost,” *Poult. Sci.*, vol. XX, no. X, pp. XX–XX, 2024.

Revisión de la literatura sobre otras soluciones al problema

Soluciones globales en productos perecederos y huevos

Modelos de Deep Learning (LSTM, CNN-LSTM)

Las redes neuronales recurrentes tipo LSTM han sido empleadas para series temporales de ventas frescas. Sin embargo, en comparaciones de granularidad múltiple (por categoría y ubicación), XGBoost a veces supera a LSTM en RMSE a nivel de tienda, aunque LSTM sigue siendo competitivo en pronósticos univariantes largos. También se han propuesto arquitecturas híbridas CNN-LSTM para productos perecederos, mejorando la detección de patrones espaciales y temporales simultáneamente.

[3] A. Priyadarshi and J. Huber, “Machine learning model for forecasting perishable foods in retail,” UVT, 2023.

Métodos híbridos y basados en STL

Para el mercado de huevos, se desarrolló un modelo híbrido basado en la descomposición STL (Seasonal-Trend-Loess) combinado con varios algoritmos de ML, logrando pronósticos más estables frente a cambios estacionales y ciclos irregulares propios del consumo de huevos

[4] L. Mo et al., “A Method for Predicting Egg Price Based on Multi Model Combination Optimization,” ResearchGate, 2025.

Revisión de la literatura sobre otras soluciones al problema

Soluciones globales en productos perecederos y huevos

Combinación de múltiples modelos

Recientemente, D. Kumar et al. propusieron un enfoque de combinación óptima de modelos (ensemble) para predecir el precio de huevos, donde un ensamblaje de SVR, XGBoost y redes neuronales mejoró la exactitud en un 12 % respecto al mejor modelo individual.

[5] D. Kumar et al., “A Novel Hybrid STL-Based Model for Egg Price Forecasting,” *Proc. IEIT 2023*, pp. 365–382, 2023.

Producción de huevo vs. demanda de venta

Aunque la mayoría de estudios se centran en producción y precio, la literatura de “Hen Egg Production Forecasting” revisa distintos métodos (ARIMA, SVR, Random Forest, XGBoost) comparándolos en conjuntos de datos de producción reales, sentando bases para trasladar esas arquitecturas a la predicción de ventas.

[6] C. Smith et al., “Hen Egg Production Forecasting: Capabilities of Machine Learning,” *Appl. Sci.*, vol. 13, no. 13, 2023.

Guías prácticas

Diversos artículos de industria y blogs especializados presentan pipelines completos para perecederos, desde preprocesamiento de series temporales hasta validación en ventana deslizante y despliegue en tableros de control, destacando XGBoost, Prophet y LSTM como pilares de las soluciones comerciales

[7] “How to Produce Accurate Demand Forecasts for Perishable Products,” Nexocode Blog, Oct. 2022.

Revisión de la literatura sobre otras soluciones al problema

Estudios y casos en Colombia

Dyna & Cía S.A. (sector agroalimentario)

En un estudio universitario de la UdeA, se compararán modelos de ML (SVR, Random Forest, XGBoost) frente al método estadístico de la empresa, logrando que ML redujera el error de pronóstico en más del 15 % para productos agropecuarios

[8] J. Pérez et al., “Análisis de modelos basados en Machine Learning para el pronóstico de la demanda en Dyna & Cía. S.A.,” Univ. Antioquia, 2023.

Pronóstico de perecederos en Estudios Gerenciales (Universidad Icesi)

Un artículo de la revista Estudios Gerenciales de la Universidad Icesi analiza series de tiempo de productos perecederos en Colombia, enfatizando la necesidad de modelos capaces de capturar estacionalidad y rupturas asociadas a eventos climáticos y festividades locales

[9] M. Gómez and L. Ramírez, “Análisis de series de tiempo en el pronóstico de la demanda de almacenamiento de productos perecederos,” *Estud. Gerenc.*, 2022.

Revisión de la literatura sobre otras soluciones al problema

Estudios y casos en Colombia

Tendencias de adopción en la industria colombiana

Aunque no hay casos públicos específicos de ML para predicción de ventas de huevos en Colombia, reportes de mercado (IndexBox) muestran que la demanda de huevos crece a ritmos del 4–6 % anual, incentivando a las empresas a explorar ML para optimizar inventarios y reducir mermas.

[10] “Colombia’s Egg Product Market Report 2025,” IndexBox, May 2025.

Enfoque de la Solución y Proceso



Fase de comprensión del negocio y de los datos

La primera etapa consiste en entender los objetivos de la empresa avícola y las características de los clientes. Historial de ventas de Avicola de Enero 2023 a Diciembre 2024 (24 meses), extraídos de SAP y planillas Excel complementarias. Variables clave incluyen fecha de venta, cantidad vendida semanal, tipo de huevo, precio promedio e indicadores de festivos y quincenas.



Preparación y limpieza de datos (ETL)

Se importan los datos históricos (2023–2024) y se unifican formatos de fecha y texto, se detectan y corrigen valores faltantes o atípicos con imputación, y se agregan niveles mensuales por cliente y por tipo de empaque. Además, se generan variables de calendario (mes, año, festivos), transformación mediante agregación de ventas diarias a semanales y análisis de estacionariedad con pruebas como Dickey-Fuller Aumentada.



Ingeniería de características y selección

Creación de características basadas en tiempo (semana, mes, festivos), variables de retraso (lags) y ventanas móviles. Exploración de características externas como precios de competencia y eventos locales. Para evitar overfitting, se aplica selección de variables con técnicas como Boruta o importancia de características en Random Forest, reduciendo la dimensionalidad y mejorando interpretabilidad.



Selección y Entrenamiento de Modelos

Consideración de modelos baseline (SNaive), estadísticos (SARIMA) y de Machine Learning (Random Forest, XGBoost, Prophet). División de datos en entrenamiento (18 meses) y prueba (6 meses) con validación cruzada para series temporales.



Evaluación de Modelos

Uso de métricas como MAPE (objetivo $< 10\%$), RMSE y MAE, junto con evaluación cualitativa mediante análisis visual de predicciones vs. valores reales.

EXPLORACIÓN INICIAL DE DATOS

Se realizó una exploración inicial de los datos disponibles relacionados con la fuerza de ventas y su desempeño a medir, encontrado diferentes archivos de Excel no estandarizados donde se llevan los presupuestos de los vendedores y se revisa en SAP las diferentes tablas donde posiblemente se pueda extraer la información:

Fuentes de Datos Identificadas:

1. Archivos Excel de la Fuerza de Ventas: Archivos donde se encuentran sus presupuestos de ventas, por cliente, producto, canal de distribución y centro logístico.
2. ERP SAP R/3 Se identifican las siguientes tablas:

Tablas con información de clientes.

KNA1: Datos generales del maestro de clientes (nombre, Nit o cedula del cliente, dirección, teléfono).

KNVV: Información específica de ventas del cliente (Tipo de cliente, a que canal de distribución pertenece, zona de ventas, vendedor condiciones de pago).

Tabla con información de los productos.

MAKT: Descripciones de los productos.

Tablas con información de la facturación.

VBRK: Datos de cabecera de las facturas (fechas, cliente, destinatario, canal de distribución, totales).

VBRP: Datos de posición de las facturas (cantidad, producto, descuentos precio).

LIKP: Datos de cabecera de los documentos de entrega (fechas y números de entrega).

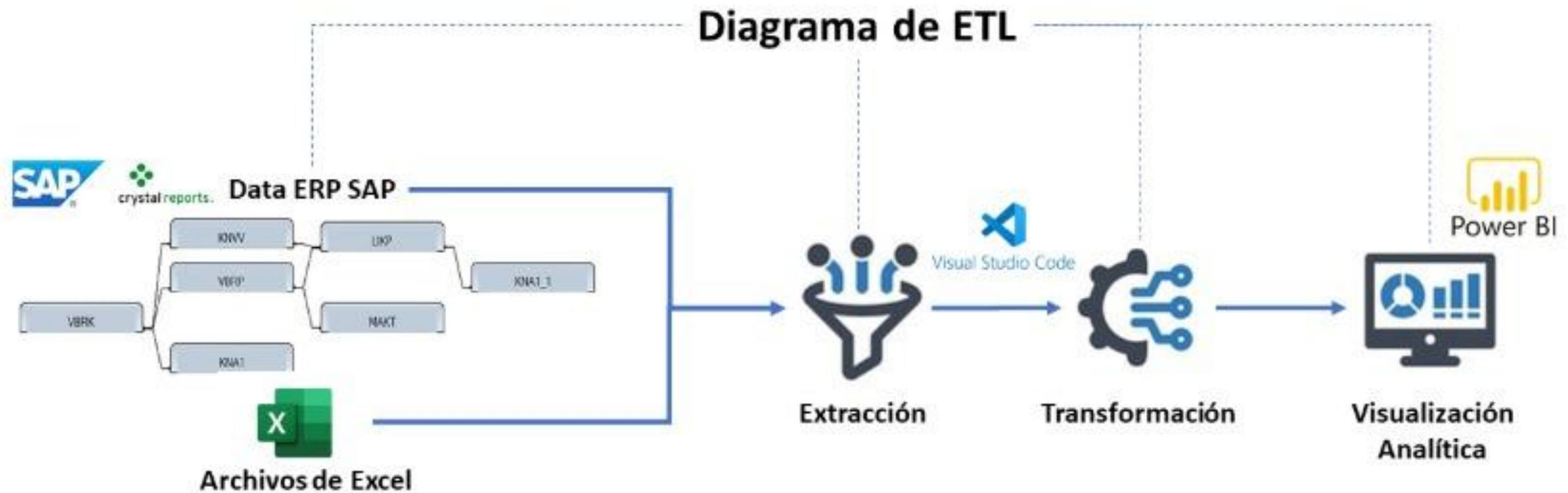
EXTRACCIÓN, TRANSFORMACIÓN Y VALIDACIÓN DE DATOS

Extracción de Datos

- Para la extracción directa desde SAP utilizaremos la herramienta con la que cuenta la empresa SAP Crystal Reports, permite consultar varias tablas al mismo tiempo de SAP y crear modelos relacionales entre ellas.
- La información generada puede descargarse en formato en XLS, XLSX o CSV.
- La extracción de datos mediante Crystal Reports se realizará mediante una carga incremental diaria, permitiendo la actualización continua sin duplicar registros innecesarios.
- Archivos Excel: Consolidación de los presupuestos para ser comparables con la venta real realizada en el mes.

EXTRACCIÓN, TRANSFORMACIÓN Y VALIDACIÓN DE DATOS

Diagrama del Pipeline ETL



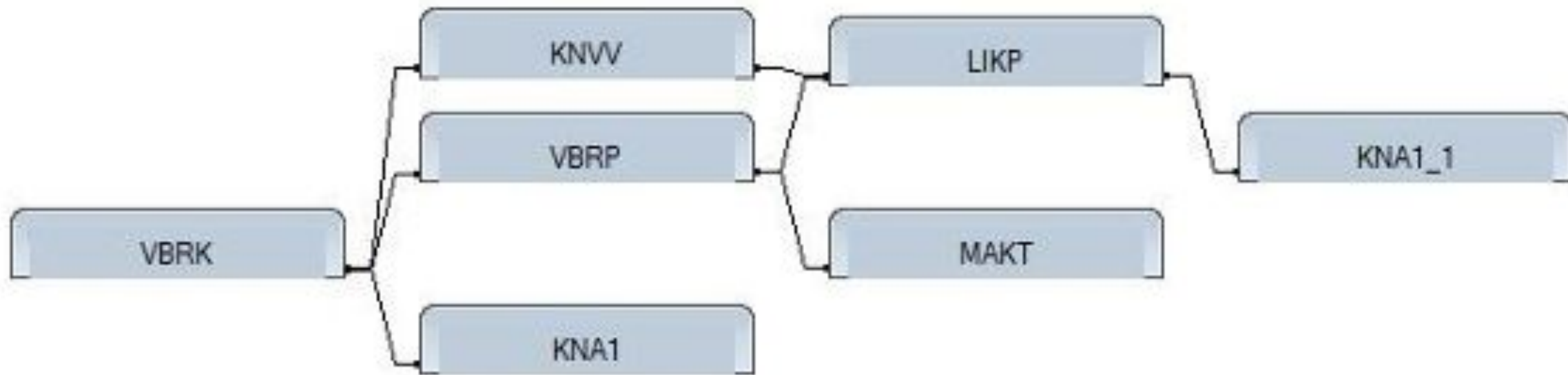
EXTRACCIÓN, TRANSFORMACIÓN Y VALIDACIÓN DE DATOS

Transformación de Datos

La transformación de datos se realizará en dos etapas. Primero, en SAP Crystal Reports, se organizarán y relacionarán las tablas extraídas del ERP para estructurar la información de manera coherente. Luego, en Python, se aplicarán procesos de limpieza, estandarización y consolidación para garantizar la calidad de los datos antes de pasarlos a Power BI.

Estos son los procesos que se realizarán:

- **Organización en Crystal Reports:** Se estructurará la relación de las tablas de SAP y generando el siguiente modelo relacional.



EXTRACCIÓN, TRANSFORMACIÓN Y VALIDACIÓN DE DATOS

- **Limpieza de datos en Python:** Se eliminarán valores nulos, registros duplicados y datos inconsistentes para mejorar la calidad del análisis.
- **Estandarización de formatos:** Se unificarán las estructuras de datos provenientes de Excel y Crystal Reports, asegurando consistencia en nombres de columnas, formatos de fechas, montos y códigos de identificación.
- **Automatización del proceso:** Se desarrollarán scripts en Python que permitirán programar la transformación de datos, asegurando que la información se procese de manera eficiente y replicable en cada actualización.
- **Protección de la información confidencial:** Para proteger la información confidencial, se desarrollará un script en Python que reemplazará los nombres de clientes con identificadores genéricos (Cliente 1, Cliente 2...), además de renombrar zonas de ventas y canales de distribución para garantizar la privacidad de la información.

EXTRACCIÓN, TRANSFORMACIÓN Y VALIDACIÓN DE DATOS

Validación de Datos

Para asegurarnos de que los datos sean confiables antes de integrarlos en Power BI, realizaremos varias validaciones clave:

- **Revisión de consistencia:** Compararemos los datos transformados con el reporte de ventas de ese mes para verificar que la información extraída desde SAP y los archivos Excel coincida con indicadores como ventas totales, cantidad de unidades vendidas y cantidad de clientes facturados.
- **Verificación de claves únicas:** Nos aseguraremos de que los identificadores de clientes, productos y vendedores sean correctos y estén correctamente relacionados en todas las tablas.
- **Detección de valores inusuales:** Usaremos Python para identificar posibles errores en los datos, como precios de venta fuera del rango esperado o nombres de productos que no correspondan a la misma categoría.
- **Protección de datos confidenciales:** Usaremos Python para revisar que los nombres de los clientes, zonas de ventas y canales de distribución hayan sido reemplazados correctamente por identificadores genéricos (Cliente 1, Cliente 2...).
- **Automatización de la validación:** Implementaremos scripts en Python para que estos controles se realicen automáticamente en cada actualización de datos.
- **Notificación de errores detectados:** Se generará un reporte automático que documentará cualquier anomalía encontrada en los datos, permitiendo su corrección antes de la carga en Power BI.