

Universidad de Lima
Facultad de Ingeniería
Carrera de Ingeniería Industrial



IMPROVING DEMAND FORECASTING BY IMPLEMENTING MACHINE LEARNING IN POULTRY PRODUCTION COMPANY

Tesis para optar el Título Profesional de Ingeniero Industrial

Joaquin Antonio Garcia Arismendiz

Código 20180760

Sandra Larissa Huertas Zuñiga

Código 20173484

Asesor

Carlos Augusto Lizárraga Portugal

Lima – Perú

Mayo de 2024

Propuesta

Carrera Ingeniería Industrial

Título

IMPROVING DEMAND FORECASTING BY IMPLEMENTING MACHINE LEARNING IN POULTRY PRODUCTION COMPANY

Autores

García Arismendiz Joaquin Antonio, Huertas Zuñiga Sandra Larissa
20180760@aloe.ulima.edu.pe, 20173484@aloe.ulima.edu.pe
Universidad de Lima

Resumen: Debido al alto porcentaje de error presente en el pronóstico de la demanda de una empresa de producción avícola; lo que desencadena una pérdida monetaria mensual provocada por el método manual que utiliza la empresa para el pronóstico del plan de ventas, es por ello que se buscará reducir este porcentaje de error mediante el uso de la herramienta Machine Learning, a través de la cual se utilizará la base de datos de ventas proporcionada por la empresa que se utilizó para entrenar y utilizando el método de series de tiempo, se podrá predecir el último trimestre del 2022. Finalmente, se compararon los resultados obtenidos por el modelo de Machine Learning con la venta real de la empresa y su pronóstico. Se logro como resultado un FB de 2,44% y un FA de 97,56%, reduciendo así el error que manejaba la empresa.

Palabras Clave: *Aprendizaje automático, Pronostico de la demanda, Empresa Avícola, Sesgo de Pronóstico, Precisión de los Pronósticos.*

Abstract: The use of manual methods to forecast demand in perishable food companies is generally subject to the variability of internal and external factors in the company, causing excess inventories and significant monetary losses, so it is relevant to carry out this research with the objective of to demonstrate that by implementing Machine Learning it is possible to improve the accuracy of the demand forecast. A case study in a company in the poultry sector in Peru, forecasting the last quarter of 2022, based on a real sales database and applying the time series method, comparing the results of the Machine Learning model, and obtaining as a result in a model with high Forecast Accuracy (FA) of 97.56% and a high Forecast Bias (FB) of 2.44%. The research is an important contribution to knowledge, demonstrating that Machine Learning is an ideal tool to project the demand for perishable food products, ideal for its application in various fields, such as loss reduction control, preventive maintenance of machines and control of supplies such as water and energy, among others.

Keywords: *Machine learning, Demand forecasting, Poultry company, forecast bias, Forecast accuracy.*

Línea de investigación IDIC – ULIMA

Operations Research & Analysis

Área y Sub-áreas de Investigación:

Diseño y desarrollo de modelos para el análisis y predicción de las variables de un proceso.

Desarrollo Empresarial

Objetivo (s) de Desarrollo Sostenible (ODS)

relacionado (s) al tema de investigación.

ODS 9 – Industria, Innovación e Infraestructura

PLANTEAMIENTO DEL PROBLEMA

La predicción de la demanda de alimentos es uno de los temas críticos para las empresas y el desarrollo sostenible Lutosławski et al. 2017. Moreno et al., 2019 mencionaron que en la industria de alimentos frescos, incluyendo los refrigeradores, la vida útil corta, la necesidad de mantener la calidad en los procesos de almacenamiento y distribución hacen que la precisión del pronóstico en el plan de ventas sea un factor importante en la planificación de la producción, minimizando las pérdidas de ventas por falta de productos, reduciendo devoluciones por la proximidad de la fecha de vencimiento, y mejorando también la disponibilidad para los clientes, reflejando así mejores resultados en la empresa e incluso reduciendo el daño medioambiental. Garre et al. 2020, afirman que la producción de alimentos es un proceso complejo donde la incertidumbre es más relevante (por ejemplo, rendimiento estocástico y previsión de la demanda, variabilidad de materias primas e ingredientes...), mostrando diferencias entre la producción planificada y la producción real. Kumar et al. 2021 para hacer frente a estos problemas en las industrias alimentarias, la automatización industrial es la mejor solución posible. La automatización se basa completamente en inteligencia artificial (IA) o aprendizaje automático (ML) o algoritmos de aprendizaje profundo (DL). Mediante el uso del sistema basado en inteligencia artificial, el plan de ventas se puede administrar de manera eficiente, al tiempo que mejora la competencia operativa.

Ante las exigencias de un mercado muy competitivo, resulta de vital importancia que las empresas incorporen herramientas de inteligencia de mercados que les permita tomar decisiones en forma proactiva a fin asegurar el logro de sus objetivos estratégicos (Mota, 2019). Además, tomar decisiones de estratégicas correctas depende de la precisión de la predicción de la demanda, evitando costos por sobreestimación o subestimación de la misma (Kumar et al. 2014).

Las principales razones por las que no se puede obtener un pronóstico de demanda preciso son la variabilidad y los datos atípicos, factores externos y no considerar los eventos y variables necesarios. Asimismo, existe una gran dependencia de los proveedores para entregar los productos al consumidor final.

OBJETIVOS

Objetivo general: Mejorar la proyección de la demanda implementando Machine Learning en una empresa de producción avícola.

Objetivos específicos:

- Medir el desempeño del modelo de machine Learning en el pronóstico de la demanda mediante métricas de error.
- Reducir la pérdida económica ocasionada por el desperdicio de ventas perdidas y de sobre stock de productos.

JUSTIFICACIÓN

Basándonos en los resultados del estudio, se logró demostrar que al implementar el modelo de Machine Learning, fue posible reducir en 9.7% las métricas de error en la predicción de la demanda, comparado con el modelo actual utilizado por la compañía. Además, el rendimiento del modelo de Machine Learning se midió con éxito a través del KPI de cumplimiento y el MAPE, lo que ayudó a saber que este modelo muestra una mejora significativa en la previsión del plan de ventas de la empresa de producción avícola.

Comparando los resultados del modelo de Machine Learning con las ventas reales de la empresa, se demuestra que el modelo tiene una alta Precisión de Previsión (FA) del 97,56% y un alto Sesgo de Previsión (FB) del 2,44%, consiguiendo reducir las pérdidas económicas en un 64,68% y los inventarios en un 49,77%.

Pruebas similares de otros sectores demuestran que reduciendo las métricas de error en la previsión de la demanda es posible ser más preciso, lo que en efecto reduce las pérdidas económicas y puede conducir a una mejor gestión de los inventarios en relación con el exceso de existencias o la escasez. Este estudio cumple el objetivo de aportar nuevos conocimientos y herramientas para minimizar la métrica de error en una estimación predictiva de un plan de ventas en el sector avícola. Además, cómo afecta positivamente a otras áreas, como la producción, lo que permitirá reducir los costes y mejorar la gestión de inventarios y facilitará el control de los desechos generados por la empresa cumpliendo así con la ODS de producción y consumo responsable, impactando de manera positiva en la sociedad y el medio ambiente.

HIPÓTESIS

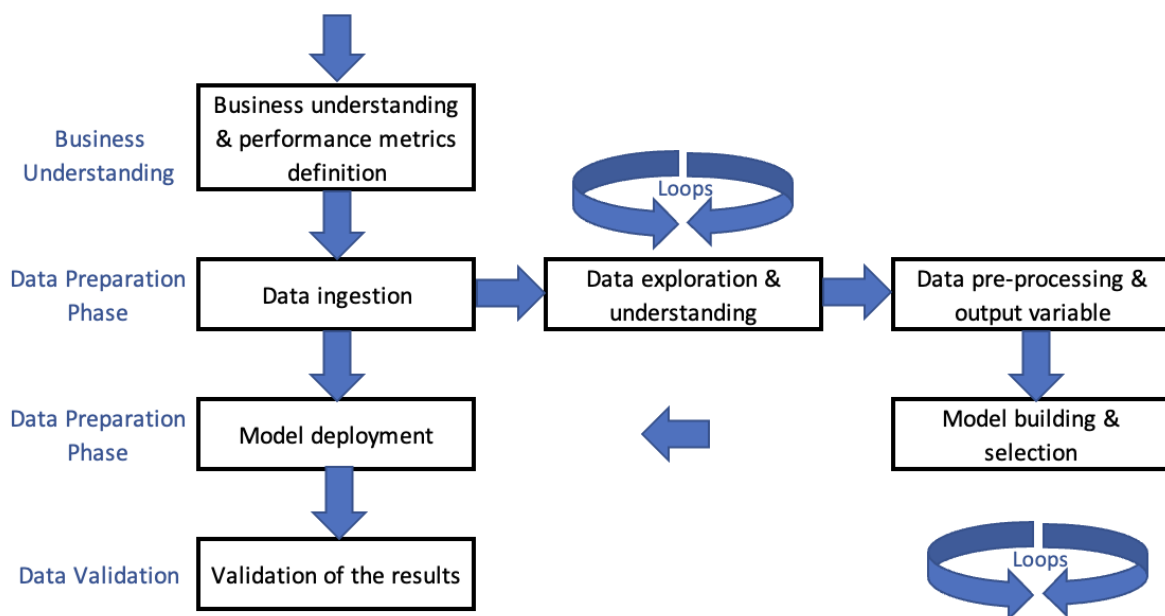
[No aplica]

DISEÑO METODOLÓGICO

Esta investigación sigue la metodología de Machine Learning, donde se realiza una comparación objetiva de los métodos de pronóstico tradicionales contra los modelos de Machine Learning en una empresa de producción avícola. Se utilizó un código en Python para crear un modelo eficiente de Machine Learning. La metodología del diagrama de bloques se muestra en la figura.

Figura 0.1

Fases para construcción de modelo de Machine Learning



Nota. De *Machine Learning or time Series Forecasting with Python* (p. 46), por Francesca Lazzeri, 2021, Wiley (<https://es.b-ok.lat/book/11081747/dc22b4>).

1.1 Entendimiento del negocio

En esta etapa se identificarán los problemas y las oportunidades de mejora del negocio mediante un diagnóstico de la empresa, el cual consiste en una visita guiada a la empresa, una entrevista con el gerente general de esta misma, donde se explicó la situación de la empresa y por último la entrega de la documentación de datos del plan de ventas. Para la identificación de problemas se empleó la herramienta de árbol de problemas, el cual ayudará a encontrar las causas raíces, para luego buscar posibles soluciones. Además, según Lutosławski et al. 2021, la mejor topología del modelo de predicción de la demanda (abaja y alta relación) para cada producto en base al valor mínimo del error absoluto medio (MAE), error porcentual absoluto medio (MAPE) y error cuadrático medio (RMSE), y luego sobre la base del mayor valor del coeficiente de determinación (R-cuadrado). mide

el porcentaje de variación en la variable de respuesta explicada por la variable explicativa. MAE, MAPE, RMSE y R-cuadrado se definen de la siguiente manera:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y} - y_i| \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y} - y_i}{y_i} \right| * 100\% \quad (2)$$

Datos pronosticados menos datos reales entre datos reales por el 100%

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y} - y_i)^2} \quad (3)$$

Datos reales menos datos pronosticados al cuadrado.

$$R^2 = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y} - \underline{y})^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \underline{y})^2}} \quad (4)$$

Donde n es el número de fuera de muestra; y_i y \hat{y}_i son los valores observados (reales) y valores ajustados de los valores dependientes de la variable Y para el i –ésimo caso, respectivamente; además \hat{Y} Es la media aritmética de Y .

Además, se utilizará un indicador de rendimiento usado por la empresa, el cual es llamado como “indicador de cumplimiento” el que se muestra a continuación:

$$\text{Indicador de cumplimiento} = \frac{\text{Venta real}}{\text{Venta planificada}} * 100\%$$

Se interpreta como:

- Si el porcentaje resulta mayor a 100%, se produjo más de lo que se vendió.
- Si el porcentaje resulta menor a 100%, se produjo menos de lo que se vendió.

Por otro lado, previamente se utilizarán los siguientes parámetros estadísticos aplicado a los datos, los cuales fueron:

1. Ljung-Box: This test has as:

H0: The residuals are distributed independently.

H1: The residuals are not distributed independently and have a correlation (Zach et al. 2020).

2. Shapiro test: In this test you have

H0: The data comes from a normal distribution.

H1: The data does not come from a normal distribution (Zach et al. 2020)

3. KPSS: The following hypothesis is proposed,

H0: The time series has a stationary tendency.

H1: The time series does not have a stationary tendency (Zah et al. 2021).

4. ADF: The following hypothesis is raised,

H0: The time series is not stationary,

H1: The time series is stationary (Zach et al. 2021).

Para el pronóstico de las ventas, los resultados se someterán a los siguientes indicadores.

Forecast accuracy (FA): mide qué tan preciso es el pronóstico [18].

$$FA = 1 - \frac{|SALES - FORECAST|}{SALES}$$

Forecast bias (FB): mide la precisión entre el margen que existe entre el pronóstico y las ventas reales (The Inventor, 2022).

$$FB = \frac{FORECAST}{ACTUAL} - 1$$

Para lograr esto, se deben incluir las fórmulas en el código de Python.

1.2 Fase de preparación de datos:

Siguiendo los siguientes pasos:

Inserción de datos: En el caso de la predicción de la demanda, los datos deberán llevar un seguimiento de forma constante y frecuente, y asegurar que los datos sin procesar fluyan mediante un proceso de ingestión de datos sólido y fiable. El proceso de inserción debe garantizar que los datos sin procesar estén disponibles para el proceso de pronóstico en el momento requerido. Eso implica que la frecuencia de ingesta de datos debe ser mayor que la frecuencia de pronóstico.

Exploración y entendimiento de datos: La exploración de datos implica resumir las características principales de un conjunto de datos, incluido su tamaño, precisión, patrones iniciales en los datos y otros atributos. La fuente de datos sin procesar, que se requiere para realizar pronósticos confiables y precisos, debe cumplir con algunos criterios básicos de calidad de datos. Si bien se pueden usar métodos estadísticos avanzados para compensar algunos posibles problemas de calidad de los datos, aún debemos asegurarnos de que estamos cruzando algún umbral de calidad de los datos básicos al ingerir nuevos datos.

Se debe tomar en cuenta las siguientes consideraciones sobre la calidad de datos sin procesar: valores faltantes, precisión de la medición, tiempo de medición, sincronización y latencia.

Preprocesamiento de datos y de la variable de salida: El preprocesamiento de datos es un paso importante para preparar, limpiar datos y crear funciones que harán que los algoritmos de aprendizaje automático funcionen mejor y reproduzcan resultados más precisos.

Construcción y selección del modelo: La fase de modelado es donde se lleva a cabo la conversión de los datos en un modelo. En el núcleo de este proceso existen algoritmos avanzados que escanean los datos históricos (datos de entrenamiento), extraen patrones y construyen un modelo. Ese modelo se puede usar más tarde para predecir nuevos datos que no se han usado para construir el modelo.

Conjunto de datos de entrenamiento: El conjunto de datos de entrenamiento representa la cantidad de datos que se deciden usar para adaptarse a los modelos de aprendizaje automático. A través del conjunto de datos de entrenamiento, se puede entrenar un algoritmo de aprendizaje automático para aprender de los datos históricos a fin de predecir puntos de datos futuros.

Conjunto de datos de validación: El conjunto de datos de validación es la cantidad de datos utilizados para ofrecer una evaluación imparcial del ajuste de un modelo en el conjunto de datos del entrenamiento, mientras se ajustan a los parámetros del modelo. Se debe aprovechar los conjuntos de datos de validación para ajustar los parámetros de un modelo de aprendizaje automático. Los parámetros son factores adicionales en el modelo, cuyos valores se utilizan para controlar y eventualmente mejorar el proceso de aprendizaje del propio modelo. Se debe observar los resultados del conjunto de validación para actualizar los niveles de los parámetros y, en consecuencia, mejorar los modelos.

Conjunto de datos de prueba: El conjunto de datos de prueba es la cantidad de datos utilizados para determinar si un modelo no se ajusta bien o se está sobre ajustando. El conjunto de datos de prueba solo se usa una vez que el modelo está completamente entrenado y validado a través

de los conjuntos de datos de entrenamiento y validación. Se debe aprovechar el conjunto de datos de prueba para evaluar diferentes modelos de aprendizaje automático.

Implementación del modelo: La implantación del modelo se debe ejecutar mediante un editor de código y un lenguaje de programación, en este caso Python, ajustando los datos en estudio y luego ejecutando el algoritmo de Machine Learning, para finalmente predecir el periodo deseado.

1.3 Fase de validación de datos

La validación de los resultados es la última etapa de la metodología, donde se realizará la comparación de las métricas de error obtenidas por el modelo propuesto con las métricas otorgadas por la empresa. Para luego realizar una prueba de hipótesis con la finalidad de averiguar si hay diferencia significativa entre las dos muestras, además se realizará predicciones de ciertos periodos, los cuales no se tengan datos, para luego compararlo con datos de ventas reales otorgados por la empresa, con el fin de validar si es que el modelo está funcionando bien o no.

NOTAS (AGRADECIMIENTOS)

Nuestro más profundo y sincero agradecimiento a nuestro asesor de tesis el Doctor Carlos Lizárraga, cuya paciencia y dedicación nos guio para poder culminar este trabajo de investigación y a nuestras familias, profesores y amigos por haber sido un apoyo incondicional y que estuvieron involucrados de una u otra manera como un soporte para poder culminar este trabajo.

REFERENCIAS

- Amat, J., Escobar, J. (2023). Forecasting time series with gradient boosting: Skforecast, XGBoost, LightGBM, Scikit-learn and CatBoost. *Skforecast*. [Online]. Available: <https://www.cienciadedatos.net/documentos/py39-forecasting-time-series-with-skforecast-xgboost-lightgbm-catboost.html>
- AWS, (2022). Amazon Forecast. [Online]. <https://docs.aws.amazon.com/forecast/latest/dg/aws-forecast-recipe-ets.html>
- Garre, A., Ruiz, M., and Hontoria, E. (2020). Application of machine learning to support production planning of a food industry in the context of waste generation under uncertainty. *Operations Research Perspectives*, 7, pp. 100147.
- Institute of Business Forecasting & Planning. (2022). <https://ibf.org/knowledge/glossary/forecast-accuracy-2-128>
- Ifft, J., Kuhns, R., and Patrick, K. (2018). Can machine learning improve prediction - an application with farm survey data. *Wageningen Academic IFAMA*, 21(8), pp. 1-16.
- Kumar, I., Rawat, J., Mohd, N., and Husain, S. (2021). Opportunities of Artificial Intelligence and Machine Learning in the Food Industry. *Hindawi Journal of Food Quality*, pp. 1-10.

- Kumar, P., and Kumar, R. (2014). The Evaluation of Forecasting Methods for Sales of Sterilized Flavoured Milk in Chhattisgarh, *International Journal of Engineering Trends and Technology (IJETT)*, 8(2), pp. 98-104., 2014.
- Latha, P.G. (2018). A Machine Learning Approach for Generation Scheduling in Electricity Markets. *International Journal of Electrical Engineering & Technology*, 9(3), 2018, pp. 69-79.
- Lazzeri, F. (2020). *Machine Learning for Time Series Forecasting with Python (1st Ed)*. Willey. <https://www.perlego.com/book/2050082/machine-learning-for-time-series-forecasting-with-python-pdf>
- Lutosławski, K., Hernes, M., Radomska, J., Hajdas, M., Walaszczyk, E., and Kozina, A.(2021). Food demand prediction using the Nonlinear Autoregressive Exogenous Neural Network. *IEEE Access*, 9, pp.146123-146136.
- Machine Learning Mastery. (2022). [Online]. <https://machinelearningmastery.com/random-forest-for-time-series-forecasting/>
- Moreno, C., Montoya, J., Jaegler, A., and Gondran, N. (2019). Sustainability metrics for real case applications of the supply chain network design problem: A systematic literature review, *Journal of Cleaner Production*, 231, pp. 600-618.
- Mota, R., Bittencourt, T., Fialho, M. (2019). Market Intelligence Tools For Improving Decision-Making In A Textile Company. *International Journal of Engineering Trends and Technology (IJETT)*, 67(10), pp. 141-146.
- Shabani, S., Candelieri, A., Archetti, F., and Gholamrezaa, N. (2018). Gene expression programming coupled with unsupervised learning: A two-stage learning process in multi-scale, short-term water demand forecasts. *MPDI Water*, 10(2), pp. 1-13, 2018.
- Srivatsan, R., Cherian, S., Venugopal P., Kumar, R. (2020). Early Maintenance and Diagnosis of Connected Machines Using Machine Learning. *International Journal of Electrical Engineering and Technology*, 11(4), pp. 412-421.
- Tarallo, E., Akabane, G., Shimabukuro, C., Mello, J., and Amancio, D. (2019). Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research. *IFAC-PapersOnLine*, 52(13), pp. 737-742.
- The Inventoro website. (2022). [Online]. <https://inventoro.com/biased-forecasts-can-cause-higher-inventory-and-logistics-costs/>
- Wu, F., Hong, S., Zhao, W., Wang, X., Shao, X., Wang, X., and Zheng, X. (2021). Neural Networks with Improved Extreme Learning Machine for Demand Prediction of Bike-sharing. *Springer Nature Mobile Networks and Applications*, 26(5), pp. 2035-2045.
- Zach. (2020, septiembre 3). How to Perform a Shapiro-Wilk Test in Python, Statology. <https://www.statology.org/shapiro-wilk-test-python/>.
- Zach. (2020, octubre 15). How to Perform a Ljung-Box Test in Python. Statology. <https://www.statology.org/ljung-box-test-python/>.

Zach. (2021, mayo 25). Augmented Dickey-Fuller Test in Python (With Example). Statology. <https://www.statology.org/dickey-fuller-test-python/>.

Zach. (2022, enero 20). How to Perform a KPSS Test in Python, Statology. <https://www.statology.org/kpss-test-in-python/>.

ANEXO. Datos del artículo publicado

- **Nombre del artículo:** Improving Demand Forecasting by Implementing Machine Learning in Poultry Production Company
- **Co autor(es):** Carlos-Augusto Lizárraga-Portugal, Juan Carlos Quiroz-Flores e Yvan Jesús García-Lopez

Publicación en revista

- **Nombre de la revista:** International Journal of Engineering Trends and Technology
- **Volumen:** 71
- **Número:** 2
- **Año:** 2023
- **Pp:** 39 - 45
- **Enlace web donde se encuentra publicado el artículo (identificador DOI, ISBN, ISSN o equivalentes):** <https://doi.org/10.14445/22315381/IJETT-V71I2P205>

MEJORA EN LA PROYECCIÓN DE LA DEMANDA IMPLEMENTANDO MACHINE LEARNING EN UNA EMPRESA DE PRODUCCIÓN AVÍCOLA

INFORME DE ORIGINALIDAD

| | | | |
|---------------------|---------------------|---------------|-------------------------|
| 9% | 7% | 2% | 6% |
| INDICE DE SIMILITUD | FUENTES DE INTERNET | PUBLICACIONES | TRABAJOS DEL ESTUDIANTE |

FUENTES PRIMARIAS

| | | |
|---|--|----|
| 1 | repositorio.unap.edu.pe | 1% |
| | Fuente de Internet | |
| 2 | www.repo.uni-hannover.de | 1% |
| | Fuente de Internet | |
| 3 | Submitted to University of Nottingham | 1% |
| | Trabajo del estudiante | |
| 4 | Submitted to University of the Andes | 1% |
| | Trabajo del estudiante | |
| 5 | Submitted to Universidad Internacional de la Rioja | 1% |
| | Trabajo del estudiante | |
| 6 | Submitted to Universitat Politècnica de València | 1% |
| | Trabajo del estudiante | |
| 7 | academic.oup.com | 1% |
| | Fuente de Internet | |