

Modelo predictivo para el análisis de pérdidas industriales en un proceso azucarero
Predictive model for the analysis of industrial losses in a sugar process

Yasmany García López¹; Lourdes Yamen González Sáez²; Juan Alfredo Cabrera Hernández³; Janin Águila Pérez¹; Catheryn Blanco Caballero¹

¹Instituto de Investigaciones de la Caña de Azúcar, Cuba

²Departamento de Química e Ingeniería Química, Universidad de Matanzas, Cuba.

³Observatorio Ambiental COSTATENAS, Universidad de Matanzas, Cuba

e-mail: yasmanygarcia31@gmail.com

Resumen

El uso de un modelo predictivo puede aportar conocimientos importantes sobre cómo se procesan y se obtienen los bienes del proceso azucarero agroindustrial. Para este trabajo, se recopilaron 340 datos del procesamiento industrial de la caña de azúcar, en tres zafra de la agroindustria ubicada a 22° 31' 55" N y 80° 52' 8" O en el municipio Calimete, Provincia Matanzas (Cuba), con el objetivo de ajustar un modelo de regresión. La analítica predictiva se basó en el cálculo del distanciamiento entre el potencial azucarero de la caña de azúcar que es molida y la producción real obtenida. A partir de estas diferencias entre potencial y real, se determinó un índice a escala diaria (I_{p-DPRE}). Este índice fue utilizado como variable respuesta para el ajuste de un modelo predictivo, donde se alcanzó un R^2 de 0,82 y se cumplieron los test de diagnósticos y validación. De esta forma se arribó a un modelo de polinomio, capaz de realizar la predicción de perjuicio, que en el caso evaluado estuvo entre 0 y 30 USD t⁻¹ de caña molida. Este resultado evidencia la importancia de accionar sobre las causas de pérdidas industriales, además de emplear elementos de inteligencia artificial para adquirir conocimiento para la sostenibilidad de la agroindustria azucarera.

Palabras clave: Analítica predictiva, eficiencia azucarera industrial, econometría

Abstract

The use of a predictive model can provide important knowledge about how goods are processed and obtained from the agro-industrial sugar process. For this work, 340 data were collected from the industrial processing of sugar cane, in three harvests of the agroindustry located at 22° 31' 55" N and 80° 52' 8" W, in the Calimete municipality, Matanzas Province (Cuba), to adjust a regression model. Predictive analytics was based on the calculation of

the distance between the sugar potential of the sugarcane that is milled and the actual production obtained. From these differences between potential and actual, a daily scale index (I_{p-DPRE}) was determined. This index was used as a response variable for the adjustment of a predictive model, where a R^2 of 0.82 was reached and the diagnostic and validation tests were met. In this way, a polynomial model was arrived, skilled of predicting damage, which evaluated case it was between 0 to 30 USD t^{-1} of milled cane. This result shows the importance of acting on the causes of industrial losses, in addition to using elements of artificial intelligence to obtain knowledge to the sustainability of the sugar agroindustry.

Keywords: Predictive analytics; industrial sugar efficiency, econometric

Introducción

En los modelos estadísticos se establece la necesidad de crear bases de datos colectadas por métodos directos o indirectos (Bogunovic *et al.*, 2017), que permitan estudiar a los procesos y servicios ecosistémicos a diversas escalas, así como estimar los flujos de materia o energía (Willcock *et al.*, 2018). En un contexto agroindustrial azucarero, se pueden encontrar ejemplos de usos de modelos, como análisis de regresión lineal para evaluar relaciones con el rendimiento del cultivo (Rahman y Robson, 2016), modelo de polinomio para evaluar la reducción de rendimiento en azúcar (Nashiruddin *et al.*, 2020); así como combinaciones de redes neuronales y algoritmos genéticos para predecir características del jugo de la caña de azúcar (Tarafdar *et al.*, 2020). También, modelos de *random forest*, *boosting* y máquinas de soporte vectorial para predecir el rendimiento agrícola (Hammer *et al.*, 2020; Natarajan *et al.*, 2016) y métodos *stepwise regression* para la selección de variables significativas (Kumar *et al.*, 2021). En Chiang *et al.* (2018), se obtienen modelos que mostraron consistencia en comparación con los resultados de los balances de masa. Ribas *et al.* (2016), mediante un análisis de componentes principales, expone con pesos significativos, las pérdidas industriales, el aprovechamiento del RPC y los tiempos perdidos (agrícola e industrial). Igualmente Cabrera y Rodríguez (2015) proponen modelos predictivos y destacan la importancia de las variables materia extraña y frescura de la caña, así como su relación con el Pol en jugo mezclado. En Shahzad *et al.* (2017) se correlaciona el recuperado azucarero con diferentes indicadores y rasgos morfológicos de la caña de azúcar. Asimismo, García *et al.* (2022), realizaron una selección de predictores importante y ajuste de un modelo de regresión lineal para la predicción de un indicador de provisión azucarera. Estos trabajos destacan el análisis de indicadores del

proceso industrial y el uso de la estadística, para, a partir de hechos, emitir criterios que contribuyan con el incremento de los beneficios del sistema.

Los análisis predictivos persiguen obtener modelos que permitan predecir una variable objetivo para nuevas observaciones, en función de un conjunto de variables de entrada (Andrade y Flores, 2018). Lo cierto es que cada modelo tiene diferentes niveles de flexibilidad y restricciones, así como facilidad y dificultad para su interpretación (James *et al.*, 2021); las restricciones tienen que ver con el diagnóstico y cumplimiento del modelo con ciertos criterios específicos. Dentro del diagnóstico del modelo para la regresión lineal se puede mencionar la significancia de los parámetros estimados, el análisis de normalidad de los residuos, homogeneidad de varianza, puntos influyentes y la multi-co-linealidad (Ramasubramanian y Singh, 2019).

Dada la incertidumbre generalizada sobre la capacidad de la agricultura para alimentar a una población creciente, las predicciones adquieren un rol importante en los estudios de prospectiva sobre temas de seguridad alimentaria, así como de oferta y demanda (Delgadillo *et al.*, 2016). También, las predicciones son útiles para las valoraciones biofísicas y económicas que se requieren en un enfoque de servicios ecosistémicos. De acuerdo con Zeki (2020), se necesita de un grupo adecuado de indicadores para monitorear y valorar cambios en los servicios ecosistémicos e identificar los perjuicios y beneficios ocultos de la forma en que se producen y procesan los alimentos, que según Müller y Sukhdev (2018), rara vez estos se reflejan en los análisis económicos convencionales.

La relación entre la caña que es molida y el azúcar que es producido es otra medida de eficiencia para este sistema agroindustrial. Esa relación, en términos de porcentaje, es expresada como rendimiento industrial (RI) (Cala *et al.*, 2020) y requiere de diversas estrategias para su incremento, lo que ha constituido el objetivo de varias investigaciones, dentro de las cuales se pueden mencionar: la selección de variedades azucareras que produzcan la mayor cantidad de azúcar por unidad de área (Chiang *et al.*, 2018) y el estudio e implementación de métodos que permitan cuantificar con mayor precisión la influencia de variables tecnológicas (Ribas *et al.*, 2016). Además, se ha estudiado la relación entre indicadores industriales como materia extraña con los porcentajes de fibra en caña, bagazo y pérdidas en bagazo (Navarro y Rostgaard, 2014); la disminución de los tiempos de transportación entre los campos de cosecha de la caña de azúcar y su procesamiento en la industria (Rico *et al.*, 2013), así como la evaluación causa-efecto entre la calidad de la materia prima con el rendimiento industrial y producto final (Martínez y De León, 2012).

El proceso azucarero agroindustrial está incidido por múltiples variabilidades espacio-temporales que son resultado de las etapas agrícolas e industrial. Cada mes de una zafra está condicionado por lo obtenido en los diferentes campos de cultivo, como consecuencia de los manejos agrícolas, momento de cosecha y condiciones climáticas; así como, por la eficiencia del proceso de extracción de sacarosa, que ocasiona incrementos o decrecimientos en la provisión azucarera agroindustrial. La sostenibilidad de la agroindustria azucarera, entonces, estará incidida por las condiciones relacionadas con estructuras, procesos y la capacidad de una toma de decisiones que influya en manejos eficientes de recursos, mejore la calidad de la materia prima, el procesamiento en la industria, así como el impacto ambiental. En otras palabras, las condiciones del suelo, sus manejos y condiciones climáticas, influirán en la materia prima principal de esta agroindustria, mientras que ineficiencias en el procesamiento de la caña, reducirán la cantidad de azúcar que es producida (Kustiyo y Arkeman, 2019). Por tal motivo, el objetivo de este trabajo es ajustar un modelo de regresión para la predicción del perjuicio relacionado a las pérdidas industriales en un proceso azucarero.

Materiales y Métodos

La agroindustria analizada se encuentra ubicada a 22° 31' 55" N y 80° 52' 8" O en el municipio Calimete, Provincia Matanzas (Cuba). Las complejidades de este sistema, están comprendidas en diferentes escalas espaciales y temporales. Ambos sectores, agrícola e industrial, están sometidos a diversos impulsores de cambio, que fomentan beneficios o los perjudican y se relacionan con el manejo, la calidad de la caña de azúcar y la eficiencia industrial para procesarla. En donde, una analítica predictiva fue de utilidad para los fines de diagnóstico y comprensión de variabilidades temporales. En el trabajo se realizó la organización de una base de datos y con la utilización del lenguaje de programación de (R), *software* versión 3.6.1 (R Core Team, 2019). La analítica predictiva entonces, se relacionó con los métodos para la recolección, organización y tabulación, que permitieron el análisis de los datos y la predicción de comportamientos. Todo ello, fue de utilidad en una valoración de perjuicios que se basó en captar las variabilidades temporales a menores escalas de análisis, lo que denotó la utilidad de los modelos predictivos para determinar los impactos de ineficiencias (Figura 1).



Figura 1. Diagrama general de trabajo

Indicadores de extracción azucarera

En este caso, se consideró una escala de valores diarios que permitió adquirir suficiente información para un análisis de relaciones causa-efecto.

Se seleccionaron tres zafras de forma aleatoria en el período de 2004 a 2021, con un total de 340 días que abarcan la extracción azucarera en los años 2014, 2015 y 2020. La elección de una escala diaria en diferentes años responde a la necesidad de mayor cantidad de datos; pues en una sola zafra, no se dispone de información suficiente para realizar los análisis causales con múltiples predictores y criterios estadísticos de ajuste y validación. Donde se deben seleccionar entre 10 y 20 valores por cada variable (McDonald, 2014), además de dividir los datos en entrenamiento y prueba. En este caso se utilizó como predictor las pérdidas industriales y la variable respuesta se elaboró a partir de las diferencias entre el potencial azucarero y la producción real obtenida. La selección de las pérdidas industriales corresponde a investigaciones precedentes en este caso de estudio (García *et al.*, 2022)

Modelos predictivos para análisis de causa-efecto

Se analizó la relación entre los predictores y la variable respuesta. Para ello, se efectuó una prueba de hipótesis y se comprobó el valor de la probabilidad (*p-value*). Los criterios fueron: si el *p-value* > α (α es igual a 0,05), no se rechaza la hipótesis nula H_0 . Sin embargo, si *p-value* < α , se rechaza la hipótesis nula de que la pendiente es igual a cero (Carrasquilla-Batista *et al.*, 2016; James *et al.*, 2021).

Se utilizaron extensiones del modelo lineal con el uso de polinomio. En tal caso, se reemplazó el modelo lineal con una función de polinomio de grado d (1):

$$Y = B_0 + B_1x_i + B_2x_i^2 + B_3x_i^3 + \dots + B_dx_i^d + e_i \quad (1)$$

Donde los términos B_0, B_1, \dots, B_d , son los coeficientes de regresión, x los predictores y e_i es el término de error. Basado en los criterios de Ramasubramanian y Singh (2019), no se utilizó un valor (d) superior a 4 para evitar sobre ajuste del modelo

Ajuste del modelo de regresión de polinomio

En el ajuste del modelo se utilizaron los parámetros R^2 y el RSE, que según James *et al.* (2021) están entre las mediciones más comunes para tales funciones.

El error estándar de los residuos (RSE) (2).

$$RSE = \sqrt{\frac{1}{n-p-1} RSS} \quad (2)$$

Donde

RSS: Cuadrado de la suma de los residuos

Grados de libertad (G) = (n) número observaciones – (p) número predictores - 1

Mientras el coeficiente de determinación R^2 (3):

$$R^2 = \frac{SCT - SCR}{SCT} \quad (3)$$

Donde:

SCT es la suma de cuadrados totales y SCR es la suma de cuadrados residuales.

Diagnóstico del modelo predictivo

En el diagnóstico del modelo se observó cómo este se ajusta a los datos de entrenamiento (Ramasubramanian y Singh, 2019). Para ello, se consideró el *p-value* de parámetros estimados. Se recurrió al cálculo de los residuos de cada observación acorde al modelo generado. La distribución normal de los residuos se comprobó con el gráfico Q-Q y el test de *Lilliefors*. La homocedasticidad se comprobó mediante gráficos (*scatterplot*) de los residuos de cada observación. También se recurrió a contrastes de homocedasticidad, como el test de *Breusch-Pagan*. Además, se analizó, la existencia de Valores atípicos y de Alta influencia, los que se estudiaron con detenimiento, pues pueden generar una correlación que realmente no existe, u ocultar una existente. La independencia de los datos, también se tuvo en cuenta y para ello, se empleó el test de *Durbin-Watson*. Por último, se consideró factor de inflación de la varianza (*VIF*), el cual no debe exceder de cinco (James *et al.*, 2021; Ramasubramanian y Singh, 2019).

Validación del modelo de regresión

El término validación cruzada corresponde con una estrategia para estimar el error de predicción del modelo. Para ello, a la base de datos de 340 observaciones, se le extraen aleatoriamente 200 datos para el entrenamiento (60 % de los datos) del modelo y el resto (40 %) para su validación y estimación del error de predicción. Al tratarse de una variable continua se empleó como medida, los errores medios al cuadrado (MSE), que consiste en la división del RSS entre el número de observaciones (n) (4):

$$MSE = \frac{RSS}{n} \quad (4)$$

Una de las principales desventajas de la validación simple es que la estimación del error varía en dependencia de cómo se hayan repartido los datos, entre entrenamiento y validación (James *et al.*, 2021). Sin embargo, lo que no se debe obviar es la necesidad de comprobar el modelo con datos que no formen parte del entrenamiento, para verificar el comportamiento ante nuevas informaciones.

Perjuicio en el proceso extracción azucarera

El distanciamiento antes mencionado, respecto a la variabilidad temporal de la provisión azucarera, se relaciona con las diferencias entre el potencial azucarero del cultivo de la caña de azúcar (expresado como RPC) y la producción obtenida (expresada como azúcar B-96). Al poseer repetidas mediciones en el tiempo, junto a factores correlacionados, se puede realizar un análisis econométrico y probar el ajuste a un modelo de regresión (García *et al.*, 2022). El cálculo del perjuicio y las ecuaciones utilizadas (5 a 6) en este caso, se describe a continuación:

$$DPRE = (X_i - X_R) \quad (5)$$

Donde

DPRE: Diferencia entre potencial de provisión de azúcar y producción real ($t \text{ a}^{-1}$)

X_i : Es el Potencial azucarero y X_R : Es el azúcar B-96

$$\text{Potencial } (X_i) = Cmi * \frac{RPC}{100} \quad (6)$$

Cmi : Caña de azúcar molida (t)

RPC: Rendimiento potencial en caña (%)

$$\text{Azúcar B - 96 } (X_R) = Cmi * \frac{RB-96}{100} \quad (7)$$

RB-96: Rendimiento industrial en base 96(%)

A partir de la diferencia entre provisión azucarera agroindustrial, potencial y real se determinó el índice a escala diaria (I_{p-DPRE}) (8). Este índice fue utilizado como variable respuesta para el ajuste de un modelo predictivo.

$$I_{p-DPRE} = \frac{DPRE * P_A}{C_{mi}} \quad (8)$$

I_{p-DPRE} : Índice de perjuicio por diferencia entre provisión azucarera agroindustrial potencial y real (USD t⁻¹ de caña molida)

P_A : Precio del azúcar (USD t⁻¹ de azúcar).

Los análisis económicos se realizaron bajo la consideración de los precios del mercado internacional en dólares (USD). En el ajuste del modelo predictivo para este perjuicio, se generó una variable aleatoria de precios con media de 467,7, desviación estándar de 112,2, para un intervalo de 213,3 a 797,2 USD t⁻¹ de azúcar.

Resultados y Discusión

En la Tabla 1 se puede observar la caracterización de los indicadores que sirvieron de base para el ajuste del modelo predictivo.

Tabla 1. Descripción estadística de indicadores recopilados a escala diaria

	PERD_T	DPRE
N: tamaño de la muestra	340	340
X: media	18,5	4,0
M: mediana	18,4	3,7
DS: desviación estándar	3,1	1,8
CV: coeficiente de variación	17,0	44,7

La variable (PERD_T) corresponde a variaciones en factores dentro de la industria que originan pérdidas en el azúcar producido. Navarro y Rostgaard (2014) han encontrado que el porcentaje de extracción se ve afectado de 87,37 % a 83, 3 %, para el intervalo de 0 a 25 % de materia extraña. Para este caso se puede apreciar que el mayor valor del coeficiente de variación lo recoge la variable DPRE, que expresa que las diferencias entre potencial azucarero y el azúcar real obtenido son variables.

En los datos recolectados se observan valores de aprovechamiento del RPC variables entre 80 % y 90 %, concentrados entre el primer y tercer cuartil, respectivamente. Lo expuesto refiere a la existencia de un conflicto donde, a medida que aumentan las pérdidas, se aleja la producción de azúcar real, del potencial que posee la caña de azúcar cosechada. De

esta forma, las pérdidas en la industria implican un impacto negativo que afecta la provisión azucarera. Los coeficientes del modelo se pueden apreciar en la tabla 2.

Tabla 2. Coeficientes del modelo de regresión

	<i>Estimación</i>	<i>Error</i>	<i>valor T</i>	<i>Pr(> t)</i>	<i>Sig.</i>
Intercepto	8,9	3.42	2,60	9,91e ⁻³	**
PERD_T	-1,8	0.34	-5,21	4,68e ⁻⁰⁷	***
PERD_T ²	0,07	0.0087	8,26	2,17e ⁻¹⁴	***
PRECIO	0,01	0.0011	13,13	< 2e ⁻¹⁶	***

Significancia: 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1

Error estándar residual: 1,80 con 196 grados de libertad

R² múltiple: 0,82, R² ajustado: 0,82

Estadígrafo F: 305 en 3 y *p-value*: < 2,2e⁻¹⁶

Las determinaciones realizadas en esta investigación permiten representar en un modelo de polinomio (Figura 2), el índice de perjuicios I_{p-DPRE} , mediante el cual se aprecia cómo las pérdidas pueden ocasionar, en términos económicos, un alejamiento del potencial azucarero de hasta 30 USD t⁻¹ de caña molida a escala diaria, en función de los parámetros analizados.

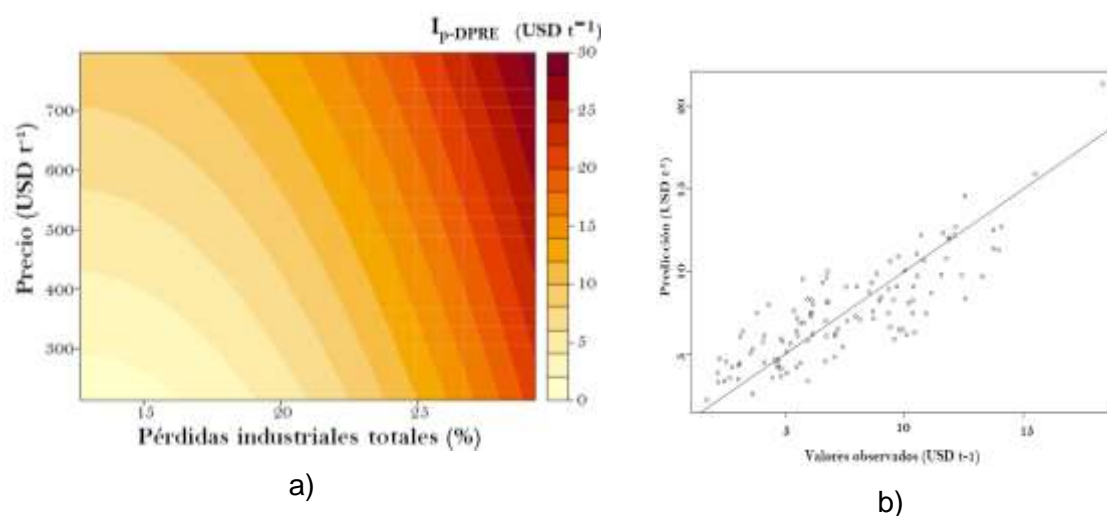


Figura 2. Modelo de efectos económicos de pérdidas totales industriales (a) y relación de valores de prueba con valores de predicción del modelo (b)

Mediante el análisis combinado de parámetros estadísticos calculados y métodos gráficos (Figura 3), se considera que el modelo de polinomio expuesto cumple con los criterios de

diagnósticos, validación y es adecuado para predecir el índice I_{p-DPRE} en función de las pérdidas industriales totales y los precios del azúcar.

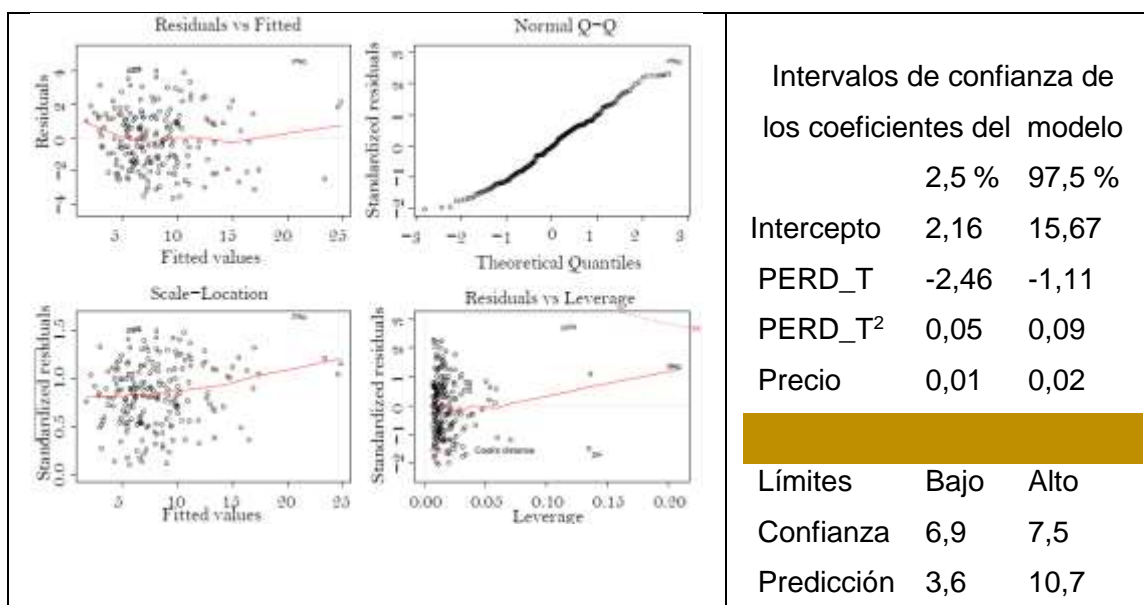


Figura 3. Validación del modelo de regresión de polinomio e intervalos de confianza

El modelo con la ecuación: $I_{p-DPRE} = 8,91 - 1,79(PERD_T) + 0,07(PERD_T)^2 + 0,01\text{Precio}$, es capaz de explicar el 82,09 % de la variabilidad observada en la variable I_{p-DPRE} (R^2 de 0,82). El test F muestra que es significativo ($p\text{-value}: < 2,2 \times 10^{-16}$), con un error residual de 1,8 y 196 grados de libertad. Se cumplen los criterios estadísticos requeridos como la prueba de normalidad (prueba de *Lilliefors*: 0,05 y $p\text{-value}$: 0,2); residuos estandarizados (*Breusch-Pagan*: 7,7 y $p\text{-value}$: 0,053). La prueba de *Durbin-Watson* no encontró evidencias de autocorrelación, donde el estadístico d fue de 2,11 y $p\text{-value}$ de 0,45. Además de lo expuesto, la figura 3 mostró cuatro gráficos que corroboran los resultados de los test estadísticos.

Es importante aclarar que no se tiene la intención simplificar las complejidades del procesamiento de la caña de azúcar, sino exponer el predictor seleccionado y su relación con la existencia de variabilidades espacio-temporales en la provisión azucarera; su perjuicio en términos económicos variables y de esta forma aportar elementos de precisión para las valoraciones del proceso azucarero agroindustrial. El modelo predice el comportamiento del indicador I_{p-DPRE} en términos monetarios; lo que puede ayudar a internalizar los impactos, en la valoración económica y de acuerdo con Bastian y Grunewald (2015) ser útil para mejorar la toma de decisiones. Por ello, la sostenibilidad de la

agroindustria azucarera requiere de la utilización de algoritmos de aprendizaje estadístico y modelos predictivos, así como el enriquecimiento con nuevos datos para lograr mayores conocimientos sobre el procesamiento de la caña de azúcar. Lo que concuerda con Ribas *et al.* (2016), sobre la aplicación de la estadística en el control de procesos para la comprensión de las variables tecnológicas de mayor peso, así como, influir en la planificación y optimización de los recursos técnicos, humanos, financieros. Según, Granja y Vidal (2014), ello puede mejorar los resultados productivos, con la aplicación de diferentes herramientas gerenciales. En este sentido, se puede resumir que el diagnóstico espacio-temporal mediante la utilización de algoritmos de aprendizaje estadístico, modelos y predicciones, puede derivar en:

- Establecimiento de rangos de trabajo para las variables identificadas como importantes.
- Evaluación de procedimientos operacionales y eficiencias en el manejo de recursos, mediante las predicciones de indicadores de los procesos de la agroindustria azucarera.
- Determinación de las causas de conflictos e impactos, para aplicar las soluciones adaptadas para cada caso.
- Fomento de debates entre especialistas y aprendizaje colectivo para valorar las mejores opciones y sus resultados.

Dos importantes cuestiones para la sostenibilidad son: ¿Cómo los ecosistemas y los servicios que proveen cambiarán en el futuro? ¿Y cómo las decisiones humanas inciden en esos cambios? La respuesta a esas cuestiones requiere de la capacidad para predecir los cambios en los procesos ecológicos (Dietze *et al.*, 2018), a lo que se agregan los cambios en los procesos antrópicos y el computo de la brecha productiva entre potencial y real (García *et al.*, 2022). En otras palabras, los niveles de eficiencias con que se manejan los recursos implican diferentes impactos económico-ambientales. Un fenómeno como las pérdidas industriales, ocasiona un perjuicio por la cantidad de azúcar no extraída; e incide en un incremento de la cantidad de caña de azúcar a procesar, para cumplir los compromisos pactados. Lo que afecta la relación costo-beneficio y origina una mayor demanda de recursos naturales. Estos perjuicios relacionados con el proceso de extracción azucarera, fundamentan el uso de la inteligencia artificial, consistente con la visión de industria 4.0. Donde, se precisa de algoritmos de aprendizaje automático para analizar

diferentes tipos de datos, encontrar relaciones, perfeccionar sistemas y adelantarse a problemas futuros (Lee, 2020).

Conclusiones

El modelo predictivo evaluado constituye una herramienta para el análisis econométrico del proceso azucarero agroindustrial.

El perjuicio por distanciamiento entre el potencial que posee este cultivo y la producción de azúcar, puede llegar hasta un valor de 30 USD t⁻¹ de caña molida, en función de las condiciones analizadas en el trabajo.

La analítica predictiva realizada, evidencia lo recomendable que es la utilización de algoritmos de aprendizaje estadístico, consistente con una visión de industria 4.0, para mejorar la toma de decisiones, en un contexto azucarero agroindustrial.

Bibliografía

- Andrade V. y Flores, P. (2018). Comparativa entre classification trees, random forest y gradient boosting, en la predicción de la satisfacción laboral en Ecuador. *Ciencia Digital*, 2, (4.1), 42-54. <https://doi.org/10.33262/cienciadigital.v2i4.1..189>
- Bastian, O. y Grunewald, K. (2015). Properties, Potentials and Services of Ecosystems". In: Grunewald, K., & Bastian, O. (Eds), *Ecosystem Services – Concept, Methods and Case Studies* (pp. 36-45). Berlin (Germany): Springer. doi:10.1007/978-3-662-44143-5_3
- Bogunovic, I., Trevisani, S., Seput, M., Juzbasic, D., Durdevic, B. (2017). Short-range and regional spatial variability of soil chemical properties in an agro-ecosystem in eastern Croatia. *Catena*, 154, 50–62. DOI: <http://dx.doi.org/10.1016/j.catena.2017.02.018>
- Cabrera, A. y Rodríguez, R. (2015). Perfeccionamiento de estimados fabriles en la Unidad Empresarial de Base central azucarero "Paquito Rosales", de la provincia Santiago de Cuba. *Anuario Facultad de Ciencias Económicas y Empresariales*, Número Especial. 73-84. <https://revistas.uo.edu.cu/index.php/aeco/article/viewFile/545/521>
- Cala, Y., Pacheco, U. y Sánchez, M. (2020). Análisis de indicadores de eficiencia productiva y perspectivas de la industria azucarera en Santiago de Cuba. *AFCEE*, Número especial. <https://www.google.com/url?esrc=s&q=&rct=j&sa=U&url=https://anuarioeco.uo.edu.cu/index.php/aeco/article/view/5144/4617&ved=2ahUKEwiMI93V3fvxAhXqRTABHR9qBhYQFjADegQICBAB&usq=AOvVaw0TQqlq602lyZw1ZZAYgaFZ>

- Carrasquilla, A., Chacón, A., Núñez, K., Gómez, O., Valverde, J., Guerrero, M. (2016). Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal. *Tecnología en Marcha*, 29, 33-45. doi:10.18845/tm.v29i8.2983
- Chiang, J., González, V., Reyes, Y., Miño, J. (2018). Influencia de las variedades de caña sobre la eficiencia industrial en la fábrica "14 de julio de Cienfuegos. *Centro Azúcar*, 45, 41-49.
http://centroazucar.uclv.edu.cu/index.php/centro_azucar/article/download/52/45/73
- Delgadillo, O., Ramírez, P., Leos, J., Salas, J., Valdez, R. (2016). Pronósticos y series de tiempo de rendimientos de granos básicos en México. *Acta Universitaria*, 26 (3). doi:10.15174/au.2016.882
- Dietze, M., Fox, A., Beck, L., Betancourt, B., Jarnevich, S., Keitt, H., Kenney, A., Laney, M., Larsen, G., Loescher, W., Lunch, K., Pijanowski, C., Randerson, T., Read, K., Tredennick, T., Vargas, R., Weathers, C., White, E. (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences*. 115 (7), 1424-1432. doi: 10.1073/pnas.1710231115
- García, Y., González, L. y Cabrera, J. (2022). Aplicaciones de aprendizaje automático para el análisis industrial de la provisión azucarera en Matanzas, Cuba. *Rev. U.D.C.A Act. & Div. Cient.* 25(2):e2334. <http://doi.org/10.31910/rudca.v25.n2.2022.2334>
- Granja, M. y Vidal, H. (2014). Pérdidas indeterminadas en la producción de azúcar. Imbabura (Ecuador): Derechos Reservados de Autor.
https://www.academia.edu/29924196/P%C3%89RDIDAS_INDETERMINADAS_EN_LA_PRODUCI%C3%93N_DE_AZ%C3%9ACAR
- Hammer, R., Sentelhas, P., y Mariano, J. (2020). Sugarcane yield prediction through data mining and crop simulation models. *Sugar Tech*, 22, 216-225.
<https://doi.org/10.1007/s12355-019-00776-z>
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). *An Introduction to Statistical with Applications in R*. Second Edition. New York (USA): Springer.
<https://doi.org/10.1007/978-1-0716-1418-1>
- Kumar, A.V., Kumar, P.G., Hari, K.S.P., Kumar, V.D., Kumar, S.D., Kumar, A. (2021). Sugarcane Yield Forecasting Model Based on Weather Parameters. *Sugar Tech*, 23, 158-166. <https://doi.org/10.1007/s12355-020-00900-4>

- Kustiyo, A., y Arkeman, Y. (2019). Design for improvement of sugar factory performance based on statistical thinking. *Earth and Environmental Science*, 335, 012033, 1-10. doi:10.1088/1755-1315/335/1/012033.
- Lee, J. (2020). *Industrial AI: Applications with Sustainable Performance*. Singapore: Springer. <https://doi.org/10.1007/978-981-15-2144-7>
- Martínez, C. y De Leon, J. (2012). Influencia de la calidad de la materia prima en el proceso tecnológico, calidad del producto final, y el rendimiento industrial en una fábrica de azúcar. *Centro Azúcar*, 39(3), 28-34. http://centroazucar.uclv.edu.cu/index.php/centro_azucar/article/download/344/336/373
- McDonald, J. (2014). *Handbook of biological statistics*. Third edition. Maryland (USA): Sparky House Publishing. https://www.researchgate.net/publication/267210763_The_Provision_of_Forest_Ecosystem_Services_-_What_Science_Can_Tell_Us_5_2014_Volume_I_Quantifying_and_valuing_non-marketed_ecosystem_services_and_Volume_II_Assessing_cost_of_provision_and_designing_eco/download
- Müller, A. y Sukhdev, P. (2018). Midiendo lo que importa en la agricultura y los sistemas alimentarios: síntesis de los resultados y recomendaciones del Informe sobre los Fundamentos Científicos y Económicos de la iniciativa TEEB para la Agricultura y la Alimentación. Ginebra (Suiza): TEEBAgriFood. http://digital.csic.es/bitstream/10261/156293/1/Maranon_Mat_org_octubre2017.pdf
- Nashiruddin, I., Fadziyana, A., Rahman, A., Ilias, M., Wan, H. (2020). Process parameter optimization of pretreated pineapple leaves fiber for enhancement of sugar recovery. *Industrial Crops and Products*, 152, 112514. <https://doi.org/10.1016/j.indcrop.2020.112514>
- Natarajan, R., Subramanian, J., y Papageorgiou, E. (2016). Hybrid learning of fuzzy cognitive maps for sugarcane yield classification. *Computers and Electronics in Agriculture*, 127, 147–157. <http://dx.doi.org/10.1016/j.compag.2016.05.016>
- Navarro, H. y Rostgaard, L. (2014). Impacto de la materia extraña en la calidad de los jugos de caña y en los indicadores de eficiencia de un central azucarero. *Centro Azúcar*, 41. http://centroazucar.uclv.edu.cu/index.php/centro_azucar/article/download/245/236/272
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, (Austria). <https://www.R-project.org/>

- Rahman, M., y Robson, A. (2016). A Novel Approach for Sugarcane Yield Prediction Using Landsat Time Series Imagery: A Case Study on Bundaberg Region. *Advances in Remote Sensing*, 5, 93-102. <http://dx.doi.org/10.4236/ars.2016.52008>
- Ramasubramanian, K., y Singh, A. (2019). *Machine Learning Using R: With Time Series and Industry-Based Use Cases in R*. Second Edition. New York (USA): Springer. <https://doi.org/10.1007/978-1-4842-4215-5>
- Ribas, M., Consuegra, R. y Alfonso, M. (2016). Análisis de los factores que más inciden sobre el rendimiento industrial azucarero. *Centro Azúcar*, 43, 51-60. <http://scielo.sld.cu/pdf/caz/v43n1/caz06116.pdf>
- Rico, O., Peñalver, Y., Leiva, J., González, V. (2013). Efecto del tiro directo de la caña en el rendimiento y la eficiencia energética. *Centro Azúcar*, 40, 77-82. http://centroazucar.uclv.edu.cu/index.php/centro_azucar/article/download/320/312/349
- Shahzad, S., Shokat, S., Fiaz, N., Hameed, A. (2017). Impact of Yield and Quality-Related Traits of Sugarcane on Sugar Recovery. *J. Crop Sci. Biotech*, 20 (1), 1 – 7. doi: No. 10.1007/s12892-016-0048-2
- Tarafdar, A., Kaur, B., Nema, P., Babar, O., Kumar, D. (2020). Using a combined neural network – genetic algorithm approach for predicting the complex rheological characteristics of microfluidized sugarcane juice. *Food Science and Technology*, 123, 109058. <https://doi.org/10.1016/j.lwt.2020.109058>
- Willcock, S., Martínez, J., Hooftman, D., Bagstad, K., Balbi, S., Marzo, A., Prato, C., Sciandrello, S., Signorello, G., Voigt, B., Villa, F., Bullock, J., Athanasiadis, I. (2018). Machine learning for ecosystem services. *Ecosystem Services*, 33, 165-174. <https://doi.org/10.1016/j.ecoser.2018.04.004>
- Zeki, E. (2020). A Framework for Characterizing and Regulating Ecosystem Services in a Management Planning Context. *Forests*, 11, 102. doi:10.3390/f11010102.