

Técnicas de clasificación para optimización de procesos de cargue de productos químicos líquidos (junio 2023)

D. Restrepo¹

¹BRINSA S.A., Medellín, Antioquia, 050021, Colombia

Corresponding author: D.Restrepo (e-mail: daniel.restrepo@brinsa.com.co).

ABSTRACT Los productos derivados de la química del cloruro de sodio, hacen parte de las materias primas que se encuentran en gran parte de los productos de nuestro día; pero como los “commodities” que son, obligan a las empresas que trabajan con ellos, a contar con operaciones seguras, pero a su vez altamente eficientes, para poder seguir siendo competitivos y conservar su participación de mercado, por lo que el uso de modelos basados en técnicas de aprendizaje automático como SVM o “Random Forest” en sus procesos de cargue, para posteriormente generar matrices de confusión, que serán evaluadas por medio de métricas como f1, puede conferirle ventajas competitivas frente a sus competidores tradicionales, ya que podrá ofrecer un mejor nivel de servicio a sus clientes, mejoras en su costo logístico y un mayor aprovechamiento de un activo valioso como es un equipo de transporte.

I. CONTEXTO

Brinsa S.A. en una compañía colombiana fundada en 1994, con presencia fabril en Colombia, Costa Rica y Republica Dominicana, dedicada a la explotación del cloruro de sodio y sus derivados por medio de un proceso de electrolisis usando tecnología de membrana, de este proceso se obtiene soda caustica y cloro gaseoso, que pueden ser empleados a su vez en reacciones subsecuentes para producir un sinnúmero de nuevos productos, que pueden ser catalogados como “commodities”, que se encuentran agrupados en lo que se conoce como el “árbol del cloro”; y que constituyen la materia prima para millares de productos imprescindibles en nuestro día a día. Ver [1]–[3].

A pesar de su importancia, no se debe dejar de lado el riesgo asociado a cada producto químico, estas etapas de prevención son necesarias no solo en producción, sino también en el transporte, y distribución de acuerdo con el cliente final, para identificar los riesgos en cada una de estas, para estructurar estrategias para prevenirlos adecuadamente, bien sea que se trate de eventos externos o internos. Ver [4].

Pero es claro que, al ser productos vitales para las cadenas productivas, y al tratarse de “commodities” es indispensable que su operación sea lo mas eficiente posible, ya que de lo contrario perdería competitividad frente a sus competidores, lo que haría perder participación de mercado a la compañía involucrada.

A. PRODUCTOS QUIMICOS LIQUIDOS

En las instalaciones principales de Brinsa S.A, ubicadas en Cajicá (Cundinamarca). se fabrican múltiples productos, pero aquellos denominados como “líquidos”, se caracterizan por tener tiempos de cargue con una alta incertidumbre, estos productos son: Ácido clorhídrico, ácido sulfónico, cloruro de calcio, hipoclorito de sodio, soda caustica y salmuera. Estos productos son cargados en un amplio abanico de presentaciones como: carotanes de diversas capacidades conocidos como venta a granel, otros productos son vendidos en presentaciones menores como IBC (1000 l de producto), tambores (200 l) o garrafas de 20l, de acuerdo con las necesidades del cliente. Ver [5].

B. PROCESO DE CARGUE

Este comienza al momento en que el vehículo llega a las instalaciones de Brinsa S.A., este es inmediatamente revisado para garantizar que esté en condiciones idóneas y el proceso de entrega hasta el cliente final se realice con los más altos estándares. Posteriormente se envía el vehículo a unos de los muelles de cargue, que de acuerdo con el volumen de vehículos podría estar o no ocupado, cuando el vehículo ya está en posición, comienza el proceso de cargue, después de terminar el vehículo debe desplazarse a bascula para la entrega de los documentos que acompañan la entrega, en el caso de que sea un producto controlado se debe hacer el registro ante el Ministerio de Justicia de la República de Colombia.

II. OPORTUNIDAD

La operación de cargue y despacho, es analizada de manera constante por personas altamente capacitadas y con una amplia experiencia en el rubro, pero apoyándose en herramientas informáticas básicas. Con el propósito de buscar una mejora constante en pro de hacerlo de la manera más ágil y segura posible, para aprovechar al máximo los activos de la compañía. De allí, que estemos interesados en incorporar nuevos análisis sobre este tema, con el propósito de brindar mas herramientas al equipo de logística para la toma de decisiones.

Para empezar, se propone un modelo de clasificación para determinar si un vehículo se va a cargar en el tiempo deseado, este sería un insumo muy valioso para la compañía ya que brindaría herramientas para estimar la hora de salida a tránsito, así como una mejor estimación de llegada al cliente. Se pretende construir este modelo a partir del análisis de variables como la hora de llegada a planta, así como el día de la semana en que lo hace, el tipo de cargue, y la cantidad de producto a cargar, entre otras.

IV. REVISIÓN BIBLIOGRÁFICA

Aunque es difícil encontrar artículos, que traten los mismos temas a los planteados, es posible encontrar operaciones logísticas con retos similares que fueron afrontadas de diversas maneras y con múltiples técnicas.

A. Un modelo de árbol de decisión para la predicción del tiempo de estadía de los barcos en puertos brasileños.

Por medio del uso de árboles de decisión y “Random forest”, los autores logran la construcción de un modelo cuyo propósito es la programación de viaje, tiempos de entrega, disponibilidad en puerto y retrasos. Para la validación del mismo se recurre a métricas de desempeño como precisión y f1 con valores superiores a 70%. Ver [6].

B. Evaluación del rendimiento logístico de productos agrícolas desde una perspectiva de sostenibilidad.

En este caso, el modelo no se limitó a los tiempos de cargue de un producto específico, sino que involucro la totalidad de la cadena logística, involucrando inclusive transporte intermodal, aunque con resultados variables de acuerdo con la operación evaluada. La métrica empleada es definida por el autor como “metodología de frontera inversa”. Ver [7].

C. Integración de simulación y optimización para programar las operaciones de carga en terminales de contenedores.

Tal vez una de las operaciones logísticas mas complejas es la correspondiente al movimiento y cargue de contenedores al interior de un puerto, en este caso se emplea un algoritmo híbrido, que tiene elementos de algoritmos genéticos, así como redes neuronales de retro propagación. Para validar los resultados obtenidos se recurrió a la “Desviación relativa”. Ver [8].

D. Aprendizaje automático y realidad mixta para una aviación inteligente: aplicaciones y desafíos.

El artículo revisado, no contempla en si mismo la construcción de un modelo, pero si realiza un estado del arte de herramientas de aprendizaje automático desarrolladas para su uso en aeropuertos o por parte de las aerolíneas. Ver [9].

V. ENTRENAMIENTO Y EVALUACIÓN DE MODELOS

Para enfrentar este problema se planteó inicialmente el uso exhaustivo de las siguientes técnicas.

A. MAQUINAS DE SOPORTE VECTORIAL (SVM)

Algunas de las características de la entrada de la base de datos, se caracterizan por estar desbalanceadas, por lo que el uso de técnicas que puedan lidiar con esta característica es fundamental, una de ellas, sería el uso de SVM, con evaluación de los posibles kernel, así como de múltiples iteraciones para definir los valores óptimos de cada uno de ellos, para obtener un modelo con el mejor f1 posible, obteniéndose los resultados mostrados en la Tabla I

TABLA I
EVALUACIÓN DIFERENTES KERNEL

Kernel	Valores óptimos	f1-score
Lineal	C = 0.001	0.6755
Polinómico	C = 0.01	0.6755
	Grado = 2	
Radial	C = 0.1	0.6757
	Gamma= 0.01	
Sigmoide	C = 0.1	0.6764
	Gamma= 0.1	

En este caso, se evidencio que por muy poco el modelo “sigmoide”, fue el que obtuvo los mejores resultados, por lo cual es el kernel seleccionado para proceder.

Al construir la matriz de confusión, mostrada en la figura 1, para el modelo SVM, con las condiciones kernel (sigmoide), c (0.1) y gamma (0.1).

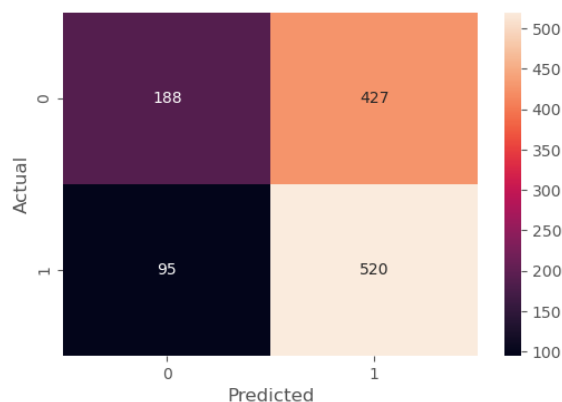


FIGURA 1. Matriz de confusión para un modelo SVM, con un kernel Sigmoide.

0 = Cargue con retraso
1 = Cargue en el tiempo deseado

A pesar de ofrecer una métrica f1-score promedio de 0.58, como se muestra en la Figura 2 es un modelo que aún es impreciso, y con una fuerte tendencia a pronosticar, que gran parte de los cargues se realizan dentro de los tiempos estimados, generando una cantidad numerosa de falsos positivos (en esta matriz de confusión, estos valores se muestran en la esquina superior derecha), a pesar de que la variable respuesta se encuentra correctamente balanceada. Ver [10].

	precision	recall	f1-score	support
0	0.66	0.31	0.42	615
1	0.55	0.84	0.66	615
accuracy			0.58	1230
macro avg	0.61	0.58	0.54	1230
weighted avg	0.61	0.58	0.54	1230

FIGURA 2. Métricas del modelo para un modelo SVM, con un kernel Sigmoide.

0 = Cargue con retraso
1 = Cargue en el tiempo deseado

En el caso de la curva ROC (Receiver Operating Characteristic), en donde el área bajo la curva, indica un valor de 0,58.

B. RANDOM FOREST

Otra técnica prometedora para enfrentar este problema son los Random Forest, apoyándonos en técnicas de “grid search” basado en “out-bag score” y en validación cruzada. En donde se hacen iteraciones de los hiperparámetros con el máximo de características, la profundidad máxima y los criterios disponibles como “Gini” o “entropía”.

En este punto, se observan comportamientos similares en “out-bag score”, tanto para entropía (exactitud de 0.6216, con 7 características y una profundidad máxima de 20), mientras que con Gini (exactitud de 0.6214, con 7 características y una profundidad máxima de 10).

En el caso de validación cruzada, la exactitud precisión obtenida es muy similar con un valor de 0.6204, por medio del criterio de entropía, 5 características y una profundidad máxima de 10, la matriz de confusión para este modelo se muestra en la Figura 3.

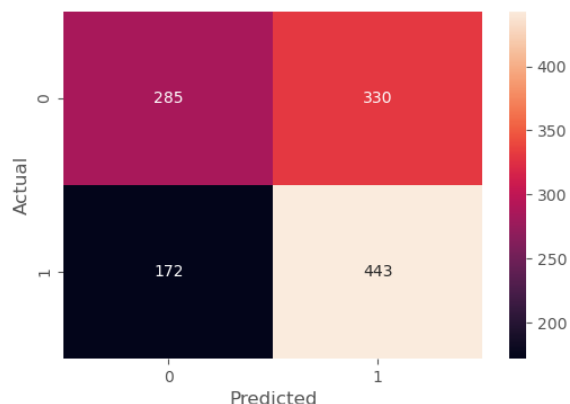


FIGURA 3. Matriz de confusión para un modelo Random Forest, criterio de entropía, 5 características y una profundidad máxima de 10

0 = Cargue con retraso
1 = Cargue en el tiempo deseado

Con respecto a las métricas de desempeño de este modelo, estas se muestran en la Figura 4.

	precision	recall	f1-score	support
0	0.62	0.46	0.53	615
1	0.57	0.72	0.64	615
accuracy			0.59	1230
macro avg	0.60	0.59	0.59	1230
weighted avg	0.60	0.59	0.59	1230

FIGURA 4. Métricas del modelo para un modelo Random Forest, criterio de entropía, 5 características y una profundidad máxima de 10.

0 = Cargue con retraso
1 = Cargue en el tiempo deseado

En el modelo Random Forest optimizado, se vislumbra unas métricas, solo un poco mejores con respecto al modelo SVM optimizado, ya que este primero indica un f1-score de 0.59, aunque con un f1-score para cada una de las clases más cercano.

Para el caso de la curva ROC, los valores obtenidos para la métrica AUC es de 0.59

VI. RESULTADOS Y DISCUSIÓN

Aunque es claro el potencial de las técnicas de aprendizaje supervisado, sería imposible quedar satisfecho con las métricas alcanzadas para cada uno de los modelos evaluados, para esto es necesario, volver a evaluar la preparación de datos, buscando incorporar mayor información que pueda servir de insumo a los modelos evaluados, para aumentar su capacidad de precisión, adicionalmente es valido evaluar técnicas como “stepwise feature selection” o en su defecto “stepwise feature elimination”

REFERENCES

- [1] Historia Brinsa. Brinsa S.A. [Online]. Available: <https://www.brinsa.com.co/historia/>. Accessed on: Junio 16, 2023.
- [2] J. Fauvarque, “The chlorine industry,” *Pure & Appl. Chem*, vol. 68, no. 9, pp. 1713–1720, 1996. Accessed on: Junio 16, 2023., DOI: 10.1351, [Online].
- [3] Árbol de Cloro y Soda. Clorosur [Online]. Available: <https://www.clorosur.org/es/arbol-de-cloro>. Accessed on: Junio 16, 2023.
- [4] C. Straut. Chemical Transportation Security. Sandia National Laboratories - U.S. Department of Energy [Online]. Available: <https://www.osti.gov/biblio/1644830> Accessed on: Junio 17, 2023
- [5] Productos Químicos Brinsa S.A. Brinsa S.A. [Online]. Available: <https://brinsaquimica.com.co/productos/>. Accessed on: Junio 16, 2023.
- [6] L.R. Abreu, I.S.F. Maciel, J.S. Alves, L.C. Braga, H.L.J. Pontes, “A decision tree model for the prediction of the stay time of ships in Brazilian ports” *Engineering Applications of Artificial Intelligence*, vol. 117, part B, 2023. Accessed on: Junio 19, 2023., [Online].
- [7] A.L. Ramos de Oliverira, K. Braga, A.P. Milanez, S L. Renner, “Evaluación del desempeño de la logística de productos agrícolas desde una perspectiva de sostenibilidad.” *Case Studies on Transport Policy* vol 10, p 674-685, 2022. Accessed on: Junio 19, 2023., [Online].
- [8] Q. Zeng, Z. Yang “Integración de simulación y optimización para programar las operaciones de carga en terminales de contenedores” *Computers & Operations Research* vol 36, 2009. Accessed on: Junio 19, 2023., [Online].
- [9] Y. Jiang, T. H. Tran, L. Williams “Aprendizaje automático y realidad mixta para una aviación inteligente: aplicaciones y desafíos” *Journal of Air Transport Management* vol 111, 2023. Accessed on: Junio 19, 2023., [Online].
- [10] S. Raschka, V. Mirjalili, “Aprender las buenas prácticas para la evaluación de modelos y el ajuste de hiperparámetros” en *Python Machine Learning*, 2th ed., Marcombo, 2019, pp. 207–239.