



Clasificación del Riesgo Fitosanitario asociado a productos de importación con SVM un algoritmo de Aprendizaje Automático

Ramírez-Guzmán, Martha Elva^{1*}; Villegas-Jiménez, Nancy²

¹ Colegio de Postgraduados, Campus Montecillo. Km 36.6, Carretera México-Texcoco. Montecillo, Texcoco, Estado de México. C.P. 56264.

² Organismo Internacional Regional de Sanidad Agropecuaria OIRSA, El Salvador.

* Autor para correspondencia: martharg@colpos.mx

Problema

La importación de productos agrícolas es una vía para la introducción de plagas cuarentenarias, las cuales vienen en forma de nemátodos, bacterias, virus, insectos, entre otros. Dichas plagas tienen la capacidad de producir daños económicos de consideración al transmitir enfermedades y/o daños físicos a los productos agrícolas en campo. Con el propósito de contribuir al intercambio comercial expedito y minimizar los riesgos, la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) y la Convención Internacional de Protección Fitosanitaria (CIPF), han desarrollado normas internacionales para medidas fitosanitarias (NIMF) (Figura 1). En particular, la NIMF 32: Categorización de productos según su riesgo de plagas, ha definido cuatro categorías de riesgo en función del grado de procesamiento de los artículos reglamentados, en donde la categoría 1 es la de menor riesgo y la 4 la de mayor riesgo. Sin embargo, a pesar de esta

Cómo citar: Ramírez-Guzmán, M., & Villegas-Jiménez, N. Clasificación del Riesgo Fitosanitario asociado a productos de importación con SVM un algoritmo de Aprendizaje Automático. *Agro-Divulgación*, 5(1). <https://doi.org/10.54767/ad.v5i1.413>

Editores académicos: Dra. Ma. de Lourdes C. Arévalo Galarza y Dr. Jorge Cadena Iñiguez.

Publicado en línea: Octubre, 2025.

Agro-Divulgación, 5(1). Enero-Febrero, 2025. pp: 67-72.

Esta obra está bajo una licencia de Creative Commons Attribution-Non-Commercial 4.0 International



Figura 1. Fumigación con bromuro de metilo, a fin de evitar la introducción de plagas de interés fitosanitario.

categorización y por la cantidad de productos que ingresan al país, en muchos casos no se tiene un método de identificación rápido que permita identificar el tipo de riesgo para las solicitudes de importación. Además, en muchos casos, es indispensable realizar un análisis de riesgo de plagas (ARP), que en ocasiones puede durar varios días. Esto implica dilataciones en la emisión de los requisitos de importación, lo que implica retardo y en muchas ocasiones pérdida de oportunidades en las relaciones comerciales.

Solución planteada

Con el propósito de establecer un método de pronóstico que identifique el tipo de riesgo con base en la NIMF 32, en este trabajo se propone utilizar un método de aprendizaje automático denominado máquinas de soporte vectorial (SVM, por sus siglas en inglés), que con base al país de origen ó procedencia, tipo de mercancía y nombre científico del producto, aprenda el tipo de riesgo al que pertenece una o varias mercancías en un conjunto de datos de entrenamiento, y posteriormente en función de dicho aprendizaje pronostique a que categoría de riesgo pertenece un conjunto de prueba. Cabe mencionar, que el conjunto de prueba es ajeno al modelo de aprendizaje automático. SVM es un algoritmo de aprendizaje que se utiliza para clasificación (lineal y no lineal) y regresión. Este trabaja de manera supervisada y no supervisada. Para el primer caso, se dispone de datos etiquetados con la respuesta correcta o el resultado deseado. Para el segundo caso, los datos no están etiquetados con la respuesta correcta o el resultado deseado. Puede trabajar tanto con datos numéricos como categóricos, este último es muy importante para cuando se desea clasificar objetos cuyos valores son variables de clasificación (*e.g.* país, nombre científico de mercancía agrícola de importación, etc.). Su objetivo principal es encontrar el hiperplano óptimo que separe los datos en clases diferentes. Un hiperplano es una línea que separa el espacio de características en dos regiones. En el contexto de los SVM, el hiperplano es utilizado para clasificar los datos en dos o más clases. El hiperplano óptimo es caracterizado por maximizar la distancia entre las clases y minimizar el error de clasificación. El algoritmo puede manejar datos de alta dimensionalidad, no requiere una distribución normal de los datos y es resistente a ruido y puntos aberrantes. El algoritmo consiste en seleccionar al azar, un conjunto de entrenamiento (*e.g.* 80% de los datos) y un conjunto de prueba (*e.g.* 20% de los datos restantes). Una herramienta estadística para evaluar el rendimiento del modelo, es la matriz de confusión (Cuadro1), de la cual se puede calcular la precisión global del modelo, así una medida de concordancia entre la clasificación real y la predicha, de la siguiente manera:

- Precisión global: $(VP_B + VP_M + VP_A + VP_C) / (Total\ de\ instancias)$
- Coeficiente Kappa (κ): mide la concordancia entre las clasificaciones reales y predichas.

Respecto a la clasificación de mercancías, la presente investigación propone utilizar las categorías estipuladas por la NIMF 32, las cuales son:

Cuadro 1. Matriz de Confusión Multiclase, con categorías de riesgo: Bajo (1), Moderado (2), Alto (3) y Crítico (4).

	Real 1	Real 2	Real 3	Real 4
Predicho 1	VP_1	FP_1_2	FP_1_3	FP_1_4
Predicho 2	FN_2_1	VP_2	FP_2_3	FP_2_4
Predicho 3	FN_3_1	FN_3_2	VP_3	FP_A_4
Predicho 4	FN_C_1	FN_4_2	FN_4_3	VP_4

VP_X: Verdaderos Positivos para la clase X (número de instancias correctamente clasificadas como X).

FN_X_Y: Falsos Negativos para la clase X, clasificados como Y (número de instancias de X incorrectamente clasificadas como Y).

FP_Y_X: Falsos Positivos para la clase Y, clasificados como X (número de instancias de Y incorrectamente clasificadas como X).

- Categoría 1. Productos que han sido procesados hasta el punto en que no deberían reglamentarse.
- Categoría 2: productos que han sido procesados pero podrán reglamentarse sobre la base de un ARP para las plagas cuarentenarias que puedan no ser eliminadas por el proceso aplicado.
- Categoría 3: Productos cuyo uso previsto sea consumo o procesamiento, los cuales podrán reglamentarse sobre la base de un ARP para las plagas cuarentenarias que sobrevivan al uso previsto.
- Categoría 4: Productos cuyo uso previsto sea la siembra o plantación, que supone un alto riesgo de introducción y dispersión de plagas reglamentadas, que sobre la base del ARP, generalmente estos productos se reglamentan.

El Servicio Nacional de Sanidad, Inocuidad y Calidad Agroalimentaria (SENASICA), proporcionó una base de datos de mercancías agrícolas de importación (4,301 registros) de 2022. De dicha base de datos fueron seleccionados los datos de país de origen o país de procedencia, mercancía, tipo de mercancía y especie. Posteriormente, se asignó una etiqueta para cada dato multivariado respecto al tipo de riesgo de acuerdo a la NIMF 32 (Cuadro 2). Con el propósito de tener en balance los productos categorizados, se tomaron al azar 160 observaciones de cada una de las cuatro categorías, la razón fue que esta cantidad correspondió a la categoría con la menor cantidad de datos, la cual correspondió a la categoría 1. Por lo tanto, el conjunto base para este estudio consistió de 640 observaciones. Cabe destacar, que esta clasificación no es exacta, debido a cierta ambigüedad de la clasificación de la NIMF 32, para las categorías 1 y 2 en donde la primera es un proceso industrial y el segundo semindustrial.

El conjunto de entrenamiento fue constituido por el 80% (512) de las observaciones, las cuales fueron tomadas al azar, mientras que el resto constituyó el conjunto de prueba (128). La matriz de confusión derivada del modelo SVM (Cuadro 3) para los datos de entrenamiento, proporcionó una visión detallada de cómo el modelo clasificó los datos multivariados. La precisión del modelo (proporción de predicciones correctas) fue de 88.67%. El coeficiente de Kappa (medida de concordancia entre predicciones y valores reales) fue de 0.85. La probabilidad de que la precisión sea mayor que la tasa de no información por

Cuadro 2. Muestra de las variables categóricas que entrenaron al algoritmo SVM.

	Origen	Procedencia	Mercancía	Tipo	Nombre científico	Riesgo
1	BOL	BOL	Chía	Grano	<i>Salvia hispanica</i>	2
2	NZL	USA	Kiwi	Fruto fresco	<i>Actinidia chinensis</i>	3
3	USA	USA	Estragón	Estacas	<i>Artemisia</i> spp.	4
4	NED	NED	Apio	Semillas	<i>Apium graveolens</i>	4
5	FRA	FRA	Colza/Canola	Semilla excepto para siembra	<i>Brassica napus</i>	2
6	NED	NED	Tulipán	Bulbos	<i>Tulipa</i> spp.	4
7	USA	AUS	Maíz	Semillas	<i>Zea mays</i>	4
8	USA	USA	Colorín	Estacas	<i>Erythrina abyssinica</i>	4
9	PRY	USA	Ajonjolí	Semilla excepto para siembra	<i>Sesamum indicum</i>	2
10	NED	NED	Fresia	Flor cortada y_o follaje fresco	<i>Freesia alba</i>	3
11	ITA	USA	Castaña	Fruto fresco	<i>Castanea sativa</i>	3
12	IND	IND	Cacahuete	Con o sin cáscara, entero o triturado	<i>Arachis hypogaea</i>	2
13	TZA	USA	Cacao	Tostado	<i>Theobroma cacao</i>	1
14	NED	USA	Cacao	Tostado	<i>Theobroma cacao</i>	1
15	GHA	BEL	Cacao	Tostado	<i>Theobroma cacao</i>	1
16	MYS	USA	Cacao	Tostado	<i>Theobroma cacao</i>	1

Cuadro 3. Matriz de Confusión Multiclase para los datos de entrenamiento, con categorías de riesgo: Categoría 1 (Industrial), Categoría 2 (Semiindustrial), Categoría 3 (Consumo) y Categoría 4 (Simiente).

	Real 1	Real 2	Real 3	Real 4
Predicho 1	93	0	0	0
Predicho 2	4	116	3	1
Predicho 3	3	3	122	4
Predicho 4	28	9	3	123

Cuadro 4. Matriz de Confusión Multiclase para los datos de prueba, con categorías de riesgo: Categoría 1 (Industrial), Categoría 2 (Semi industrial), Categoría 3 (Consumo) y Categoría 4 (Simiente).

	Real 1	Real 2	Real 3	Real 4
Predicho 1	18	0	0	0
Predicho 2	1	29	1	0
Predicho 3	1	2	30	2
Predicho 4	12	1	1	30

casualidad fue menor a $2.2e-16$, lo cual implica que la precisión del modelo fue estadísticamente significativa.

Respecto al conjunto de prueba, las estadísticas de precisión e índice de Kappa fueron 83.59% y 0.78, respectivamente. La probabilidad de que la precisión sea mayor que la tasa de no información por casualidad fue menor a $2.2e-16$, lo cual implica que la precisión del modelo fue estadísticamente significativa.

La Figura 2 muestra que la Categoría 4 (e.g. material vegetal propagativo como plantas y semillas), que es la más riesgosa, es la mejor predicha por el modelo, en tanto que las categorías 1 y 2 son menos afortunadas en su clasificación. Esto implica, que se debe mejorar la clasificación de origen.

Los resultados indican que el modelo tiene un buen rendimiento en la clasificación, con una precisión del 88.67% (para entrenamiento) y 83.59%, para prueba, con una concordancia (Kappa) de 0.85 y 0.78, para entrenamiento y prueba, respectivamente. El uso del modelo de aprendizaje automático SVM, representa una alternativa de apoyo, para identificar muy rápidamente, el tipo de riesgo que puede tener una mercancía, considerando variables categóricas como país de origen, país de procedencia, tipo de mercancía y nombre científico. Cabe destacar, que el modelo puede ser actualizado continuamente, incorporando otras características del producto importado ya sean numéricas y/o no numéricas. Se concluye que esta metodología puede ser de gran utilidad, para disminuir costos y tiempo durante el proceso de inspección de productos de importación que llegan por puertos, aeropuertos y fronteras.

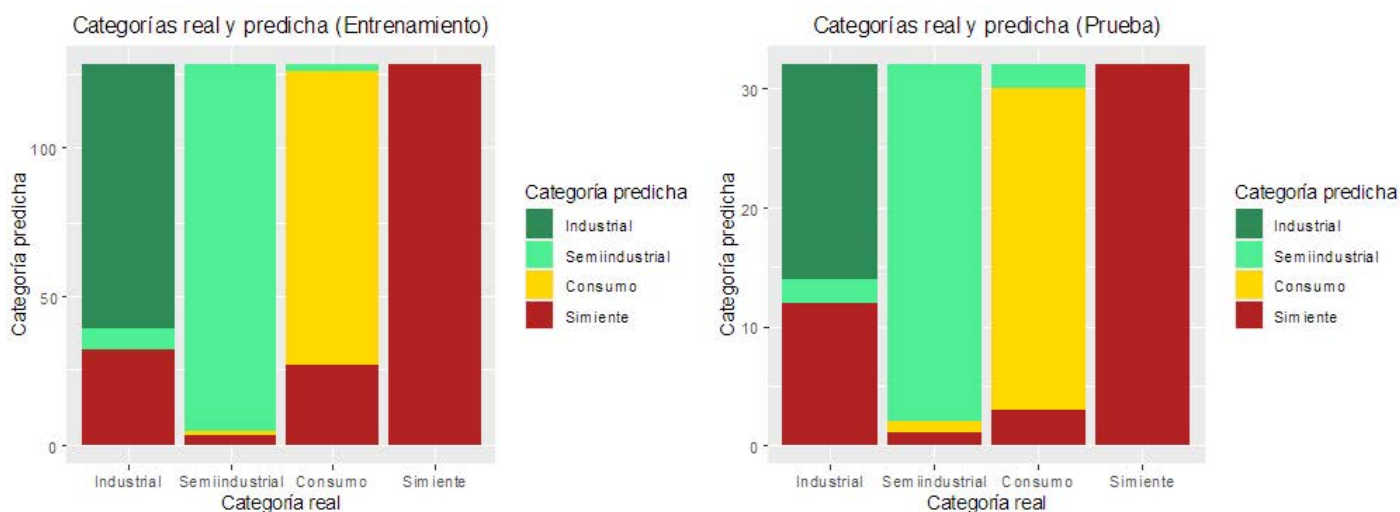


Figura 2. Frecuencias de las Categorías de los productos de importación.

Innovación, Impactos e Indicadores

Nivel de Innovación	Descripción	Transferido	Impacto		Indicador General de Políticas Públicas	Indicadores Específicos	Subindicador
			Sector	Ámbito			
Incremental	Busca mejorar los sistemas que ya existen haciéndolos mejores, más rápidos, más baratos, etc.	Asociaciones de Productores Gobierno de los Estados Productores independientes	Primario: Agricultura, Ganadería, Pesca, Explotación forestal, Minería Cuaternario: Servicios basados en el conocimiento que prestan industrias de las Tecnologías de Información y comunicación, de consultoría empresarial, de planificación financiera, de informática y de investigación científica.	Social Económico Ambiental Conocimiento	Ciencia y Tecnología Económico Responsabilidad Ambiental Salud Pública	Competitividad Comercio Capacitación Finanzas Públicas	Número de publicaciones Transferencias tecnológicas Aplicación de técnicas y conocimientos tecnológicos para el desarrollo social y económico
Procesos	Implementación de una nueva o significativa mejora de un método de producción o de suministro.	Comunidades Agrarias					
Innovación sostenible	Desarrollo de productos y procesos que contribuyen al desarrollo sostenible						

