

UNIVERSIDAD PERUANA UNIÓN

ESCUELA DE POSGRADO

Unidad de Posgrado de Ingeniería y Arquitectura



**Aplicación de aprendizaje automático en espectros de infrarrojo
cercano para la predicción de humedad en comino molido**

Tesis para obtener el Título de Segunda Especialidad Profesional de
Ingeniería: Estadística Aplicada para Investigación

Autor:

Segundo Ramos Villalta Arellano

Asesor:

PhD. Javier Linkolk López Gonzales

Lima , Febrero 2025

DECLARACIÓN JURADA DE ORIGINALIDAD DE TESIS

Yo Javier Linkolk López Gonzales, docente de la Unidad de Posgrado de Ingeniería y Arquitectura , Escuela de Posgrado de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **“Aplicación de aprendizaje automático en espectros de infrarrojo cercano para la predicción de humedad en comino molido”** del autor Segundo Ramos Villalta Arellano tiene un índice de similitud de 15% verificable en el informe del programa Turnitin, y fue realizada en la Universidad Peruana Unión bajo mi dirección.

En tal sentido asumo la responsabilidad que corresponde ante cualquier falsedad u omisión de los documentos como de la información aportada, firmo la presente declaración en la ciudad de Lima , a los 27 días del mes de Febrero del año 2025



PhD. Javier Linkolk López Gonzales

Asesor

ACTA DE SUSTENTACIÓN DE TESIS

En Lima, Ñaña, Villa unión a 27 días del mes de febrero del año 2025, siendo las 09:00 horas, se reunieron de forma online sincrónica, bajo la dirección del presidente del jurado Mg. Junior Israel Pacheco Espinoza, secretario Mtro. Carlos Daniel Abanto Ramirez; los demás miembros: Mg. Lizeth Geanina Huanca López, Mg. Esteban Tocto Cano y el asesor PhD. Javier Linkolk López Gonzales, con el propósito de administrar el acto académico de sustentación de tesis de la segunda especialidad titulada "Aplicación de aprendizaje automático en espectros de infrarrojo cercano para la predicción de humedad en comino molido", conducente a la obtención del título de segunda especialidad en estadística aplicada para investigación.

El presidente inició el acto académico de sustentación invitando al candidato a hacer uso del tiempo determinado para su exposición. Concluida la exposición, el presidente invitó a los demás miembros del jurado a efectuar las preguntas, cuestionamientos y aclaraciones pertinentes, los cuales fueron absueltos por el candidato. Luego se produjo un receso para las deliberaciones y la emisión del dictamen del jurado.

Posteriormente, el jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:


Candidato: Segundo Ramos Villalta Arellano

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
Aprobado	18	A-	Muy bueno	Sobresaliente

Finalmente, el presidente del jurado invitó al candidato a ponerse de pie, para recibir la evaluación final. Además, el presidente del jurado concluyó el acto académico de sustentación, procediéndose a registrar a registrar las firmas respectivas.



Presidente



Secretario



Asesor



Miembro



Miembro



Candidato

Índice

Resumen	1
Abstract	1
Palabras clave	2
Keyword:	2
Introducción	2
Materiales y métodos	3
<i>Materia prima</i>	3
<i>Adquisición de espectros NIR</i>	4
<i>Preprocesamiento de espectros</i>	4
<i>a)</i> Suavizado de Savitzky-Golay (SG):	4
<i>b)</i> Normalización:	4
<i>Desarrollo de modelos</i>	5
<i>a)</i> Regresión lineal múltiple (MLR): MLR	5
<i>b)</i> Máquina de vectores de soporte cuadrático (QSVM)	5
<i>Optimización de modelos</i>	5
<i>Cálculo del rendimiento de los modelos</i>	6
<i>Análisis referencial</i>	6
<i>Análisis de datos</i>	6
Resultados y discusión	7
<i>Contenido de humedad</i>	7
<i>Espectros NIR originales y pretratados</i>	7
<i>Modelos MLR y QSVM completos</i>	9
<i>PCA</i>	10
<i>Modelos MLR y QSVM optimizados</i>	12
Conclusiones	14
Referencias	15

Aplicación de aprendizaje automático en espectros de infrarrojo cercano para la predicción de humedad en comino molido.

Segundo Ramos Villalta Arellano, Javier Linkolk López Gonzales

Resumen

El comino (*Cuminum cyminum* L.) es una de las especias más valoradas a nivel mundial, y el contenido de humedad constituye un parámetro crítico para evaluar su calidad y seguridad. Este estudio tuvo como objetivo evaluar el rendimiento de algoritmos de aprendizaje automático en el análisis de espectros de infrarrojo cercano (NIR) para la predicción de humedad en comino molido. Se recolectaron seis muestras de comino y se obtuvieron espectros NIR en el rango de 1100-2100 nm. Los espectros, expresados en absorbancia, fueron pretratados utilizando Savitzky-Golay (SG) y normalización de manera independiente, seguidos por una optimización mediante análisis de componentes principales (PCA). Los datos completos y los optimizados se modelaron empleando regresión lineal múltiple (MLR) y máquinas de vectores de soporte cuadrático (QSVM). Los modelos se entrenaron mediante validación cruzada de 5 particiones con 30 repeticiones, evaluando su desempeño mediante el coeficiente de determinación (R^2) y la raíz del error cuadrático medio (RMSE). En el conjunto de datos completo, QSVM con pretratamiento SG obtuvo los mejores resultados ($R^2 = 0.99$; RMSE = 0.0582). Para los datos optimizados, QSVM combinado con normalización presentó un rendimiento destacado ($R^2 = 0.97$; RMSE = 0.0919). Estos hallazgos resaltan que QSVM, gracias a su capacidad para capturar relaciones no lineales, ofrece predicciones más precisas y robustas en comparación con MLR. Este estudio sugiere explorar técnicas avanzadas basadas en aprendizaje profundo y/o algoritmos complementarios que puedan mejorar el rendimiento de los modelos predictivos