



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS AGROPECUARIAS Y

AMBIENTALES

CARRERA DE AGROINDUSTRIA

INFORME FINAL DEL TRABAJO DE INTEGRACIÓN CURRICULAR,

MODALIDAD PRESENCIAL

TEMA:

“DESARROLLO DE UN MÉTODO COMPUTACIONAL PARA LA

CLASIFICACIÓN DE NARANJAS (*Citrus x sinensis* O.),

VARIEDADES VALENCIA Y WASHINGTON”

Trabajo de titulación previo a la obtención del título de ingeniero agroindustrial

**Línea de investigación: Gestión, producción, productividad, innovación y desarrollo socio
económico**

Autor: Moreno Mendoza Angelo Emanuel

Director: Ing. Juan Carlos De La Vega MSc

Ibarra-2023



UNIVERSIDAD TÉCNICA DEL NORTE BIBLIOTECA UNIVERSITARIA

AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

1. IDENTIFICACIÓN DE LA OBRA

En cumplimiento del Art. 144 de la Ley de Educación Superior, hago la entrega del presente trabajo a la Universidad Técnica del Norte para que sea publicado en el Repositorio Digital Institucional, para lo cual pongo a disposición la siguiente información:

DATOS DE CONTACTO			
CÉDULA DE IDENTIDAD:	1312215666		
APELLIDOS Y NOMBRES:	Moreno Mendoza Angelo Emanuel		
DIRECCIÓN:	Esmeraldas, Rioverde, recinto palestina calle principal barrio la bahía		
EMAIL:	aemorenom1@utn.edu.ec		
TELÉFONO FIJO:	0622744085	TELÉFONO MÓVIL:	0939961771

DATOS DE LA OBRA	
TÍTULO:	DESARROLLO DE UN MÉTODO COMPUTACIONAL PARA LA CLASIFICACIÓN DE NARANJAS (<i>Citrus x sinensis</i> O.), VARIEDADES VALENCIA Y WASHINGTON.
AUTOR (ES):	Moreno Mendoza Angelo Emanuel
FECHA: DD/MM/AAAA	29/05/2023
SOLO PARA TRABAJOS DE GRADO	
PROGRAMA:	<input checked="" type="checkbox"/> PREGRADO <input type="checkbox"/> POSGRADO
TÍTULO POR EL QUE OPTA:	Ingeniero Agroindustrial
ASESOR /DIRECTOR:	Ing. Juan Carlos De La Vega, MSc.

2. CONSTANCIAS

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto la obra es original y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

Ibarra, a los 29 días del mes de mayo de 2023

EL AUTOR:

(Firma).....


Nombre: Moreno Mendoza Angelo Emanuel

CERTIFICACIÓN DIRECTORIO DEL TRABAJO DE INTEGRACIÓN CURRICULAR

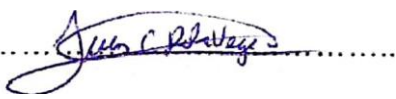
Ibarra, 25 de mayo de 2023

Ing. Juan Carlos De La Vega MSc

DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR

CERTIFICA

Haber revisado el presente informe final del trabajo de Integración Curricular, el mismo que se ajusta a las normas vigentes de la Universidad Técnica del Norte; en consecuencia, autorizo su presentación para los fines legales pertinentes


(f) 

Ing. Juan Carlos De La Vega MSc

C.C: 1002958856

APROBACIÓN DEL COMITÉ CALIFICADOR

El Comité Calificador del trabajo de Integración Curricular "DESARROLLO DE UN MÉTODO COMPUTACIONAL PARA LA CLASIFICACIÓN DE NARANJAS (*Citrus x sinensis* 0), VARIEDADES VALENCIA Y WASHINGTON" elaborado por Moreno Mendoza Angelo Emanuel previo a la obtención del título del Ingeniero Agroindustrial, aprueba el presente informe de investigación en nombre de la Universidad Técnica del Norte,

(f) 

Ing. Juan Carlos De La Vega MSc

C.C: 1002958856

(f) 

Ing. Nicolás Pinto MSc

C.C:1712640935

DEDICATORIA

El esfuerzo realizado a través de estos años y del trabajo de investigación va dedicado especialmente a mis padres Shirley Mendoza y Stalin Moreno, a mis dos grandes tesoros de la familia Keyla Saleth y Jesús Abel, asimismo a mis amigos que buscan y persiguen sus metas.

AGRADECIMIENTO

Agradezco fundamentalmente a Dios todo poderoso, sin bendición no hubiese sido posible esta nueva meta alcanzada, a mis padres esencialmente quienes fueron ente y pilares de mi crecimiento, les agradezco de todo corazón por el apoyo incondicional demostrado en esta etapa de mi vida.

A mis hermanos quienes supieron apoyarme y llenarme de alegría, estoy profundamente gratificado con todos los amigos de las diferentes carreras de la universidad, en especial las carreras de Medicina, Fisioterapia y Gastronomía, pero sobre todo con los compañeros de carrera con quienes compartimos conocimiento y momentos especiales. A ustedes Belén, Alexis, Boris, Lenin, Andrés y Priscila que sin importar cada una de nuestras caídas nunca faltó ese apoyo total.

A la universidad por permitirme desarrollar intelectualmente con excelentes docentes quienes ayudaron a mi formación profesional, entre ellos al Ing. Juan Carlos De La Vega que sin su guía y conocimiento no hubiera sido posible el alcance de esta meta.

RESUMEN EJECUTIVO

La naranja dulce (*Citrus sinensis*) se caracteriza por tener una gama de colores que no corresponde con su estado de madurez, ya que su corteza puede estar verde, pero en el interior madura. El objetivo de este estudio es desarrollar un método computacional que permita clasificar las naranjas (*Citrus x sinensis Osbek*), variedades Valencia y Washington mediante un sistema sencillo que no dependa del color de la fruta. A través del uso de técnicas de Machine Learning (Neural Net Fitting y Regression Learner) con el manejo de MATLAB, se logró relacionar el peso, diámetro ecuatorial, diámetro polar, la variedad y los sólidos solubles de la fruta. Como resultado final se entrenó cada uno de los métodos disponibles en el campo de inteligencia artificial. La validación de los métodos se cumplió con el 10% de datos como prueba real en campo. Finalmente se concluyó que el mayor coeficiente de determinación alcanzado fue 0.92 por el método de Regression Learner, representando así con fidelidad la clasificación de las naranjas (*Citrus x sinensis*) variedades Valencia y Washington, entre todos los modelos evaluados en este estudio.

Palabras clave: índice de madurez, MATLAB, Regression Learner, Neural Net Fitting

ABSTRACT

The sweet orange (*Citrus sinensis*) is characterized by its color range that does not correspond to its state of maturity, since its rind could be green, but it is ripe inside. The aim of the study is to develop a computational method that allows classifying of the oranges (*Citrus sinensis Osbek*), Valencia and Washington varieties by means of a simple system that does not depend on the color of the fruit. Through the use of Machine Learning (Neural Net Fitting and Regression Learner) with the management of MATLAB, it was achieved to relate the weight, equatorial diameter, polar diameter, variety and soluble solids of the fruit. As a result, each of the methods available in the field of artificial intelligence was trained. The validation of the methods was fulfilled with 10% of the data as a real test in the field. Finally, it was concluded that the highest correlation coefficient reached was 0.92 by the Regression Learner method, thus faithfully representing the classification of oranges (*Citrus x sinensis*) Valencia and Washington varieties, among all the models evaluated in this study.

Keywords: maturity index, MATLAB, Regression Learner, Neural Net Fitting

ÍNDICE DE CONTENIDOS

INTRODUCCIÓN	14
Problema	14
Justificación.	16
Objetivos	18
Objetivo general.....	18
Objetivos específicos	18
Preguntas directrices	18
Hipótesis	18
Alternativa	18
Nula.....	19
CAPÍTULO I	20
MARCO TEÓRICO	20
1.1. La naranja	20
1.1.1. Característica de la naranja	20
1.1.2. Taxonomía	21
1.1.3. Variedades	22
1.2. Producción.....	24
1.3. Situación Actual	25
1.4. Red neuronal Artificial (RNA).....	26
1.4.1. Técnicas de redes neuronales.....	27
1.4.2. Estudios con redes neuronales de clasificación en la industria alimentaria	28

	10
1.5. Técnica de Neural Net Fitting	32
1.5.1. Métodos de entrenamiento de Neural Net Fitting de MATLAB.	32
1.6. Técnica de Regression Learner.....	33
1.6.1. Validación cruzada	34
CAPÍTULO II.....	35
METODOLOGÍA.....	35
2.1. Materiales y equipos.....	35
3.1.1. Materiales y equipo.....	35
3.2. Métodos.....	36
2.2.1. Desarrollo del modelo Neural Net Fitting	36
3.2.2. Validación del método	40
3.3. Descripción del proceso.....	41
2.4.1. Ubicación de materia prima.....	41
3.4.2. Selección de la naranja	42
3.4.3. Toma de muestra.....	42
3.4.4. Envasado y transporte de muestra	43
3.4.5. Recepción de la materia prima.....	44
3.4.6. Conservación de la naranja en fase experimental.....	45
3.4.7. Determinación del peso	46
3.4.8. Determinación del diámetro ecuatorial.....	46
3.4.9. Determinación del índice de madurez	46
CAPÍTULO III.....	49
RESULTADOS Y DISCUSIÓN	49

3.1. Evaluar diferentes técnicas de Machine-Learning para la modelación de los datos experimentales.....	49
3.1.1. Técnica Neural Net Fitting	49
3.1.2. Regression Learner	51
3.2. Validar los modelos para la selección del que se ajuste en mayor medida a los datos. 52	
CONCLUSIONES Y RECOMENDACIONES	57
Referencias Bibliográficas	59
Anexo.....	65

ÍNDICE DE TABLAS

Tabla 1 <i>Taxonomía</i>	222
Tabla 2 <i>Antecedentes de investigaciones con redes neuronales</i>	28
Tabla 3 <i>Equipo, materiales e insumos</i>	355
Tabla 4 <i>Locación de la muestra</i>	422
Tabla 5 <i>Ubicación y datos meteorológicos del lugar de experimentación</i>	455
Tabla 6 <i>Resultados de la técnica Neural Net Fitting</i>	500
Tabla 7 <i>Validación de modelos Neuronales</i>	544
Tabla 8 <i>Comparativa de los objetivos alcanzados</i>	555
Tabla 9 <i>Cálculo del MSE del modelamiento por variedad</i>	555

ÍNDICE DE FIGURAS

Figura 1 <i>Naranja Washington</i>	233
Figura 2 <i>Naranja Valencia</i>	244
Figura 3 <i>Arquitectura típica de una red neuronal</i>	277
Figura 4 <i>Diagrama secuencial del diseño de una red neuronal</i>	311
Figura 5 <i>Modelado de una red neuronal</i>	322
Figura 6 <i>Flujo de trabajo en la aplicación Regression Learner</i>	333
Figura 7 <i>Varios tipos de modelos de regresión de MATLAB entrenándose al mismo tiempo</i>	400
Figura 8 <i>Relación del índice de madurez de la fruta con la maduración de esta y la pérdida de calidad organoléptica de su zumo.</i>	477
Figura 9 <i>Respuesta de Regression Learner</i>	511
Figura 10 <i>Comparativa de valores predichos y valores verdaderos</i>	533

INTRODUCCIÓN

Problema

Las naranjas son de vital importancia para el mercado de frutas frescas y la industria de concentrados. El ser cosechado en un momento inadecuado permite llegar al procesador con características que reducen la eficiencia industrial (Niño et al., 2020), incidiendo negativamente en la deficiencia de la materia prima que cumpla con requerimientos de calidad, destacando la falta de integración del productor en el sector primario. (Durán Barón & Villa, n.d.a)

Existen una serie de razones que no contribuyen al desarrollo de alternativas al manejo precosecha de la naranja, en general, es difícil contar con tecnología relevante que permita la estandarización (Durán Barón & Villa, n.d.b). Es fundamental determinar la madurez óptima de la fruta dada su naturaleza no climatérica para lograr una mejor cosecha.(Georgi Alexander Campelo Valle, 2020a)

Estudios demuestran (Estefanía Romero Rodríguez, 2020) que la medición de parámetros físicos químicos constituye alternativas que pueden ayudar con el control de calidad, entre estos se encuentran: pH, acidez, sólidos solubles, etc. Debido a la heterogeneidad de los frutos cosechados resulta difícil establecer un modelo representativo.(Torres et al., 2013a)

Investigaciones recientes exponen que una de las mediciones que accede a solventar esta problemática y que no implica procedimientos especializados, es a través de la coloración del producto mediante un sistema de visión computacional y el reconocimiento de imágenes (Torres et al., 2013b) basado en redes neuronales se logra alcanzar rendimientos del 90% en la clasificación de frutos (Victor Aguilar-Alvarado & Alfredo Campoverde-Molina, 2019a), pero

por su tecnología, precio, y la complejidad que requieren hacen que sea inaccesible al productor.

En circunstancias normales serían métodos eficaces, sin embargo, en la actualidad constituyen alternativas que no pueden aplicarse con ciencia cierta en el medio, principalmente por dos razones; la primera observa que estos modelos poseen limitaciones al momento de entrenar con coloraciones muy cercanas, estos dependen mucho de la calidad de la imagen, la iluminación y la complejidad del algoritmo obteniendo como resultados finales de efectividad menores del 87%. (Herrera, 2015; Victor Aguilar-Alvarado & Alfredo Campoverde-Molina, 2019b)

En segunda instancia la coloración del fruto depende de factores externos a la madurez, el mismo que manifiesta una correlación positiva con la intensidad de la luz solar y los cambios de temperaturas (Martínez-González et al., 2017). Por tal motivo no se puede optar por una normativa extranjera como guía para establecer que el estándar de calidad en frutas como la naranja en función del índice de color, más aún solo puede apreciarse de forma subjetiva en especial en una industria que puede tener un alto valor a futuro.

El desarrollo de la producción de naranja básicamente es destino de comercio local, pero en temporada de pico de cosecha sobrepasa la demanda interna del producto provocando que el precio en finca se desplome (El Telégrafo, 2017) , una realidad que vive Esmeralda cantón Rioverde. La información para procesamiento de esta materia prima es casi nula, más aún no hay registro de una normativa que amerite información concreta de factores intrínsecos que permitan parametrizar cuantitativamente la madurez de la fruta; por lo que es necesario implementar un nuevo método programable de clasificación que brinde una valoración justa a la fruta.(Georgi Alexander Campelo Valle, 2020a)

Justificación.

La presente investigación está orientada en proponer una alternativa a los parámetros convencionales que se toman en cuenta al momento de determinar el índice de madurez de la naranja. Teniendo en cuenta que el estado de madurez de la fruta permitirá conocer la trazabilidad adecuada del producto. Una estandarización de la materia prima permite obtener un abanico de posibilidades para la elaboración de productos agroindustriales.(ISOTools, 2015)

Entre los principales parámetros de calidad de las frutas se encuentran el peso, diámetro ecuatorial, sólidos solubles, acidez, pH, color e índice de madurez(Estefanía Romero Rodríguez, 2020). Normalmente estos procesos llevan a la destrucción de la fruta, ya que se los realiza con equipos y procedimientos especializados, algunos complejos que son ejecutados con personal entrenado en laboratorios lejos del productor(Torres et al., 2013b).

A diferencia de métodos tradicionales que no logran determinar una representatividad fuerte de frutas heterogéneas, la implementación de un modelamiento de Redes neuronales con las técnicas de Neural Net Fitting o Regression Learner para la predicción del índice de madurez de la naranja se ve reflejada en la eficiencia y la optimización de recursos humanos para los procesos de la inspección de alimentos, facilitando así la toma de decisiones post cosecha.

Para (Figueredo-Ávila & Ballesteros-Ricaurte, 2016)las RNA han demostrado poder efectivamente reemplazar sistemas automáticos que se ocupaban de la clasificación e inspección de alimentos. Debido a la importancia de obtener frutos con homogeneidad y madurez óptimas para determinar el momento recomendable de recolección.

Por lo que es necesario implementar un sistema computacional que favorezca esta operación respecto a la heterogeneidad de las naranjas, lo que contribuye a mejorar la economía de la Zona 1 que presenta una productividad de 1637 Tm.(INEC, 2020)

Al considerar el presente tema de investigación se podrá mejorar la correlación de propiedades físicas con el índice de madurez, a través de la generación de base de datos de propiedades físicas cuantificables como el peso, diámetro ecuatorial y la altura (Durán Barón & Villa, n.d.); con esto el productor podrá estar al tanto del estado de madurez de su producto, teniendo en cuenta que el requerimiento del índice de madurez es de 6.5 mínimo para su comercialización (Comisión Europea, 2011)

A través de técnicas de Machine-Learning ejecutadas con el programa de MATLAB con diferentes modelamientos, ya que, se pretende desarrollar y entrenar un algoritmo que utilice la información de datos obtenidos a través de la practica experimental de la naranja en campo al momento de ser cosechada, determinando la función de variables cuantificables que identificarán el índice de madurez.

Objetivos

Objetivo general

- Desarrollar un método computacional que permita clasificar las naranjas (*Citrus x sinensis Osbek*), variedades Valencia y Washington según su índice de madurez.

Objetivos específicos

- Evaluar diferentes técnicas de Machine-Learning para la modelación de los datos experimentales.
- Validar los modelos para la selección del que se ajuste en mayor medida a los datos.

Preguntas directrices

¿El modelo propuesto puede ser replicable para dos variedades de naranja *Citrus x sinensis Osbek*?

¿Cuál es la muestra representativa para validar el modelo computacional?

¿El modelo computacional es susceptible para ser aplicado en campo?

Hipótesis

Alternativa

La correlación de las variables medidas en campo y las características fisicoquímicas permiten al método computacional determinar el índice de madurez de la naranja.

Nula

La correlación de las variables medidas en campo y las características fisicoquímicas no permiten al método computacional determinar el índice de madurez de la naranja.

CAPÍTULO I

MARCO TEÓRICO

1.1. La naranja

Es un fruto proveniente del árbol de naranjo perteneciente a la familia Rutaceae. La naranja dulce es conocida como (*Citrus sinensis*) mientras que la variedad amarga (*Citrus aurantium*), estas eran cultivadas como plantas ornamentales desde tiempos inmemoriales. Estas frutas, tienen su particularidad en su pulpa, está formada por muchas burbujas llenas de jugo. Son de color naranja, de ahí su nombre, aunque algunas especies cuando está maduro es casi verde. Su sabor va de amargo a dulce.(EFSA, 2013)

Es nativo del sureste de China y el norte de Birmania, aunque es de región mediterránea durante unos tres mil años desde su lugar origen, el naranjo se extendió a Japón y toda la India al oeste, aunque el naranjo dulce no se conoció hasta 1450.(Almache et al., 2021) Desde este momento se extendió por toda Europa y se convirtió en un gran éxito al año siguiente a primera mitad del siglo XV.(EFSA, 2013)

1.1.1. Característica de la naranja

Según (Spanish Fruits & Delicacies, 2019a) describe a continuación a algunas particularidades a ser tomada en cuenta de las naranjas son las siguientes.

- **Tamaño y peso:**

Mantiene un diámetro de 6 a 10 cm. El peso recomendado de cítrico es de 150 gramos y 200 gramos, sin pelar. La escala de clasificación se escala de 0 a 13 y se representa de la

siguiente manera: 13 corresponde a la fruta de menor diámetro y 0 corresponde a la fruta de mayor diámetro.(CODEX STAN 245-2004, n.d.)

- **Forma del fruto:**

Las frutas normalmente son de una forma esférica, redonda y aplastadas a los extremos. (Spanish Fruits & Delicacies, 2019b)

- **Color:**

Normalmente la cáscara de la naranja posee tonalidad verde con pigmentos amarillentos o anaranjados, además está compuesta de corteza liza o rugosa que posee una segunda piel blanca la cual cubre al fruto y protege la pulpa. (Spanish Fruits & Delicacies, 2019c)

- **Sabor:**

La pulpa del fruto normalmente contiene entre 8 y 12 gajos que proveen de abundante jugo, de sabor dulce, y matices agrios dependiendo de la variedad y el tamaño del fruto.

- **Raíz:**

Está compuesta por raíces primarias y secundarias que se hunden hasta alcanzar un metro de profundidad, mismas que son el soporte alimenticio del árbol fruta(Almache Almache Mayra Alexandra & Villacres Escudero Andrea Estefania, 2021a)

1.1.2. Taxonomía

De acuerdo con (Almache Almache Mayra Alexandra & Villacres Escudero Andrea Estefania, 2021b) la información taxonómica de la naranja (*Citrus sinensis*) es la que se muestra en la tabla 1.

Tabla 1*Taxonomía*

Familia	Rutaceae
Género	Citrus
Especie	<i>Citrus sinensis</i>

Nota. (Almache Almache Mayra Alexandra & Villacres Escudero Andrea Estefania, 2021)

1.1.3. Variedades

Se pueden hallar tres tipos primordiales de naranjas, las cuales se clasifican en: navel, blancas y sanguíneas, las cuales de la misma manera se subdividen en 21 variedades que se las puede diferenciar a partir del sabor y dulzor que va descendiendo o incrementado dependiendo de la variedad, por la cantidad de semillas, el tamaño y corteza de la fruta (ASTUDILLO, 2018).

Las variedades de naranja que se cultivan con frecuencia en el Ecuador son: Valencia tardía, Valencia común, Valencia delta, Thompson, Washington, Naranja lima, Naranja pomelo y Naranja Agria. (De Investigación et al., n.d.b) La más consumida es la variedad Valencia común que sobresale por su aceptabilidad entre los consumidores y su alto contenido de jugo y azúcares (Georgi Alexander Campelo Valle, 2020a). En este trabajo se efectuará con la taxonomía de la tabla 1 para ambas variedades.

- ***Variedad Washington***

Esta variedad de naranja tiene una formación en España, se caracteriza por presentar una fuerte coloración y una excelente calidad como se muestra en la figura 1. Suele ser consumida desde principios de febrero hasta mediados de abril.(Spanish Fruits & Delicacies, 2019d)

Figura 1

Naranja Washington



Nota. Adaptado de (Spanish Fruits & Delicacies, 2019)

- ***Variedad Valencia***

Se destaca por ser una naranja dulce de producción tardía que pertenece al grupo de naranjas blancas, esta fruta se consume directamente en fresco y también se la utiliza en el campo agroindustrial en los pueblos citrícolas del mundo(Madeleine Hanco Condori, 2017). Por otro lado, esta variedad es muy empleada por su alto contenido de zumo con aroma excelente y sabor levemente ácido como se muestra en la figura 2, es apta para industrialización de zumos ácidos, bebidas, dulces entre otros subproductos.(De Investigación et al., n.d.c).

Figura 2

Naranja Valencia



Nota. Adaptado de (Spanish Fruits & Delicacies, 2019)

1.2. Producción.

El cultivo de naranja es de gran relevancia para los productores que se dedican a este tipo de cultivo (Georgi Alexander Campelo Valle, 2020b). En la actualidad podemos evidenciar que en el Ecuador existe una producción aproximada de 146.159 Tm referente a la naranja para consumo en fresco, pero esto no es aprovechado de manera eficiente. (INEC, 2020)

Según (Robalino, 2010) dentro de la cadena agroindustrial, la naranja representa el 14% del comercio mundial de frutas, como fruta fresca representa el 42% del total de exportaciones mientras que el resto se destina para jugo. Entre los mayores productores se encuentran: Brasil, Bélgica, España, Estados Unidos.

Tomando sus consideraciones de volúmenes de producción y una industrialización fuertemente desarrollada hacen que la participación del país concerniente a esta fruta sea

prácticamente nula en el mercado internacional por tal motivo es necesario la implementación de nuevas tecnologías que permitan el crecimiento y la optimización de este.(Georgi Alexander Campelo Valle, 2020a)

1.3. Situación Actual

El desconocimiento de los productores y la poca información amerita a que el producto sea despreciado y no valorado como corresponde, provocando que los productos sean devaluados por intermediarios, quienes se aprovechan del sacrificio y los gastos generados para producir la fruta o en peores casos la cosecha termina desperdiciándose. (Georgi Alexander Campelo Valle, 2020b)

El (el Telégrafo, 2017a) indica que en la temporada optima de producción aumenta 30% de la cosecha normal convirtiéndose en una situación crítica, esta realidad afecta a las dos partes involucradas directamente en los cultivos tanto productores como trabajadores ya que involucra más trabajo, pero el precio de venta hace que no puedan aumentar el sueldo. (El Telégrafo, 2017) informa que el que la sobreproducción que es la temporada alta de cosecha provoca que el valor en finca decaiga, motivo por la cual en algunos sectores el precio sea de 3\$ USD; mientras que, en temporadas bajas el precio se encuentra normalmente en 6 \$ USD.

La investigación de (Solano Robinzon Hugo Roberto, 2018) determinó que en el campo los productores de naranja tienen muy poca asistencia técnica apoyadas por alguna institución encargadas de la producción agrícola, este factor provoca que tanto la producción, comercialización y rentabilidad se vean afectadas.

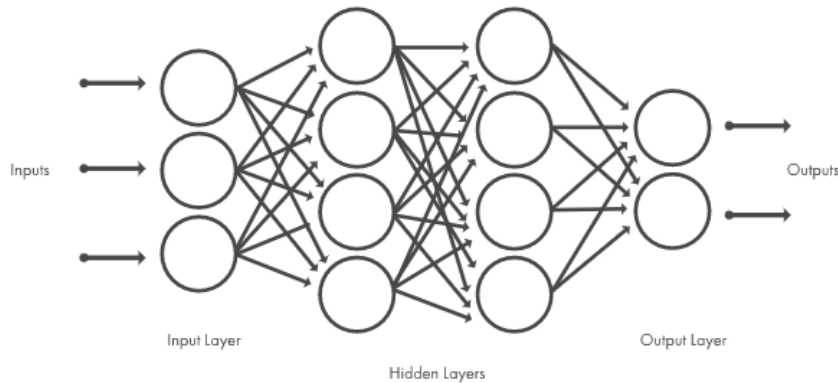
1.4. Red neuronal Artificial (RNA)

La RNA es un modelo computacional de estructura semejante a las neuronas en el cerebro, capaz de aprender de los datos, de modo que estos se entrenen consecutivamente para que a través de las reiteraciones aprendan a reconocer ejemplos, clasifique los datos y prediga casos futuros.(¿Qué Es Una Red Neuronal? - MATLAB & Simulink, n.d.a). Para (Gustavo Andrés Figueredo-Ávila, 2015) es un sistema que se adapta a las exigencias del entorno, aunque de manera artificial se dirigen a imitar el comportamiento de los entes biológicos; para ello, modelizan matemáticamente dicho comportamiento.

El funcionamiento de una red neuronal se basa en la combinación de varias capas de procesamiento, consta de una capa de entrada o input, una capa oculta cuando es Machine Learner y más de una capa oculta cuando es Deep Learner y finalmente una capa para la salida. Las capas están interconectadas mediante nodos, o neuronas; cada capa utiliza la salida de la capa anterior como entrada como se indica en la figura 3.(¿Qué Es Una Red Neuronal? - MATLAB & Simulink, n.d.b)

Figura 3

Arquitectura típica de una red neuronal



Nota. Adaptado de (Bazalar Pereda et al., n.d.)

1.4.1. Técnicas de redes neuronales

Algunas técnicas habituales de Machine Learning que se emplean para diseñar aplicaciones de redes neuronales son el aprendizaje supervisado y no supervisado (Entrenar Modelos de Regresión Para Predecir Datos Mediante Aprendizaje Automático Supervisado - MATLAB - MathWorks América Latina, n.d.a), el reconocimiento de patrones, y el clustering en este proyecto se implementará dos técnicas como es la clasificación y la regresión.

- ***Clasificación***

La clasificación es un tipo de Machine Learning supervisado en el que un algoritmo aprende, modela y entrena para clasificar nuevas observaciones a partir de ejemplos de datos etiquetados. (¿Qué Es Una Red Neuronal? - MATLAB & Simulink, n.d.c)

- **Regresión**

Los modelos de regresión representan la relación entre una variable de respuesta en targets (salida) y una o varias variables de predicción inputs (entrada). (¿Qué Es Una Red Neuronal? - MATLAB & Simulink, n.d.d)

1.4.2. Estudios con redes neuronales de clasificación en la industria alimentaria

Estudios realizados para la clasificación por color de frutas climatéricas como las manzanas que se encuentran citados (Gustavo Andrés Figueredo-Ávila, 2015) demuestran que, en estos tipos de clasificación, se representa una selección de datos difusos por lo que los resultados llegan a ser aproximados, a pesar de esto, la precisión fue 96,6% por lo que se puede determinar que es un método muy fiable.

La siguiente tabla 2 proporcionará información de algunas investigaciones de modelamiento de redes neuronales para el método de clasificación.

Tabla 2

Antecedentes de investigaciones con redes neuronales

ORD	Investigación	Parámetros	Observaciones
1	“Vista de diseño de un sistema automático de selección de frutos de café mediante técnicas de visión artificial” ^a	Análisis de color de imágenes, Binarización, Clasificación de imágenes, Extracción de características, Fruto de café, Proceso de selección, Segmentación de	Muestra el desarrollo de dos algoritmos con sva capaz de clasificar el fruto de café maduros o no y algoritmo otro que detecta plagas.

		imágenes, Visión artificial.	
2	“Classification of fruit based on convolutional neural networks” ^b	Forma, tamaño, color	Una RNA para considerarse un identificador autónomo de frutas competitivo y de valor en el mercado debe ser capaz de identificar 3 parámetros esenciales de esta, entre las cuales destaca la forma, el tamaño y el color. En la actualidad se busca identificadores capaces de reconocer estos parámetros básicos con efectividad y rapidez.
3	“Application of Non-Linear mixed models in the Growth of pear fruits” ^c	Tamaño (longitudinal y transversal)	Se propone lograr un modelo mixto no lineal con alta capacidad predictiva de crecimiento a través de la medición del diámetro ecuatorial para frutos de peras
4	“Implementación de un algoritmo de control de calidad para la selección de productos agrícolas utilizando visión artificial” ^d	MATLAB, Algoritmos de clasificación, procesamiento de imágenes.	El siguiente trabajo propone un algoritmo de detección de fruta en mal estado como control de calidad; para ello se utilizan técnicas de procesamiento digital de imágenes como histogramas, uso de operadores morfológicos, cambios a otros espacios de color entre otras.
5	“Modelos matemáticos para estimar el crecimiento del fruto de chile manzano” ^e	Peso, longitud, ancho, volumen de agua desplazada y volumen del fruto	El estudio propone modelos matemáticos utilizando variables no destructivas permite para predecir el crecimiento del fruto chile manzano

6	<p>“Selección de una Arquitectura de Red Neuronal Artificial eficiente para predecir el coeficiente de difusividad aguaymanto peruviana deshidratado osmoconvectivamente”^f.</p>	<p>RNA con algoritmo de entrenamiento Backpropagation y ajustes de pesos Levenberg-Marquart.</p>	<p>Se determinó la arquitectura de la RNA al comparar el error de los valores reales con los predichos por la red en el pronóstico del coeficiente de difusividad.</p>
7	<p>“Implementación de un Sistema de Visión Artificial para la clasificación de naranja producida en el departamento del Quindío”^g</p>	<p>Medidas de color a muestras de naranja con diferente grado de madurez en el SVA y en un espectrofotómetro HunterLab, los valores de L*, a*, b*</p>	<p>Mediante un sistema de visión artificial con medidas de color puede indicar el estado de madurez de las naranjas tipo valencia, de una forma sencilla, rápida, no invasiva, en tiempo real y bajo costo.</p>
8	<p>“Predicción de procesos basados en masa diseñando parámetros de fruta utilizando diferentes técnicas de modelado”^h</p>	<p>Altura, diámetro, área proyectada y volumen modelo matemático (MM), red neuronal artificial (ANN) y regresión vectorial de soporte (SVR)</p>	<p>Se observa que las propiedades físicas de la fruta se pueden predecir mediante modelos basados en ANN que luego podrían en conjunto ser útiles en la automatización y el diseño eficiente de equipos posteriores a la cosecha.</p>

Nota. ^a (Herrera, 2015). ^b (Victor Aguilar-Alvarado & Alfredo Campoverde-Molina, 2019).

^c (VALENTÍN TASSILE, 2020). ^d (La Madrid Távara et al., 2019). ^e (Rojas-Lara et al., 2008).

^f (Bazalar Pereda et al., n.d.). ^g (Lucía et al., 2018). ^h (Tomar & Pradhan, 2022).

Otros estudios mencionados por (Gustavo Andrés Figueredo-Ávila, 2015) muestran la necesidad de automatizar procesos manuales para sustituir la inspección visual del trabajador,

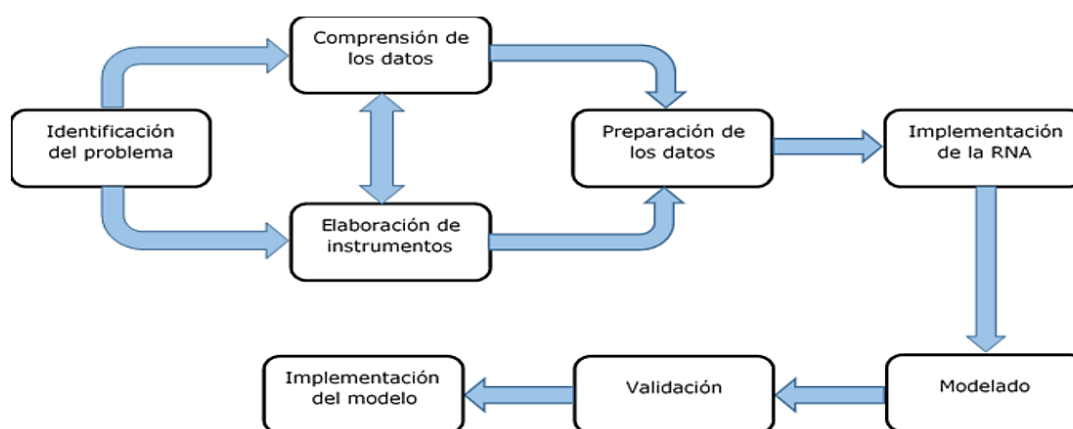
para el desarrollo de una red neuronal de clasificación no solo el conjunto de datos de entrada proporciona una predicción correcta de datos, sino que también la topología proporcionará la garantía de los resultados.(Gustavo Andrés Figueredo-Ávila, 2015)

Para la evaluación de calidad de las frutas se le propone evaluar con diferentes características, evaluar cuantitativamente el cambio de forma de la fruta haciendo uso de una red neuronal probabilística para obtener una clasificación más sofisticada(Gustavo Andrés Figueredo-Ávila, 2015).

Por ejemplo (Ashok & Vinod, 2014) propuso evaluar los atributos externos de la manzana con parámetros como el color, el tamaño, la forma, la textura y la presencia de daños, su sistema secuencial como se muestra en la figura 4 fue capaz de distinguir las características de las frutas con una exactitud 86,52% y 88,33%.

Figura 4

Diagrama secuencial del diseño de una red neuronal



Nota. Adaptado de (Flores, 2021)

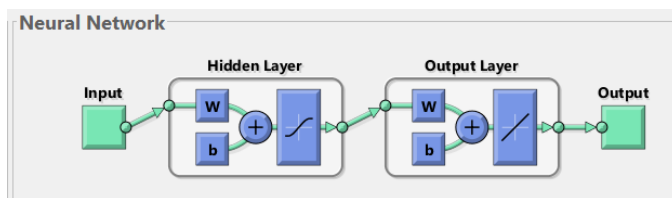
1.5. Técnica de Neural Net Fitting

La red neuronal artificial BP que se basa en la plataforma de MATLAB, utiliza un algoritmo de propagación de errores para verificar el peso de la red tiene las características de un marco dinámico, lo cual es conveniente para la construcción y programación de redes.

Después de aprender con patrones de predicción de entrada, el modelo de predicción de red de tres neuronas como indica la figura 5 tiene una buena capacidad de generalización. El modelo de regresión prueba la velocidad de ajuste y la precisión del modelo de red, y el modelo de red tiene una clara ventaja sobre el modelo de regresión. (Gan et al., 2011)

Figura 5

Modelado de una red neuronal



Nota. Adaptado de (Flores, 2021)

1.5.1. Métodos de entrenamiento de Neural Net Fitting de MATLAB.

El método de entrenamiento de Neural Net Fitting se realizará con:

- **Levenberg-Marquardt, Bayesian Regularization y Scaled Conjugate Gradient**

Estos algoritmos habitualmente solicita más memoria, pero menos utiliza tiempo. El entrenamiento se detiene automáticamente cuando la generalización deja de mejorar repercutiendo en un aumento en el error cuadrático medio en las muestras de

validación. (*Retropropagación Levenberg-Marquardt - MATLAB Trainlm - MathWorks América Latina*, n.d.)

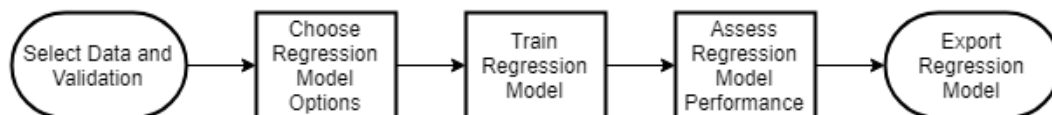
1.6. Técnica de Regression Learner.

La aplicación Regression Learner entrena modelos de regresión para predecir datos como se muestra en la figura 6. Con ella puede examinar datos, seleccionar los parámetros a evaluar, especificar esquemas de validación, entrenar todos los modelos y evaluar consecutivamente los resultados. (*Entrenar Modelos de Regresión Para Predecir Datos Mediante Aprendizaje Automático Supervisado - MATLAB - MathWorks América Latina*, n.d.b)

Puede realizar un entrenamiento automatizado para buscar el mejor tipo de modelo de regresión, incluidos modelos de regresión lineal, árboles de regresión, modelos de regresión de procesos gaussianos, máquinas vectoriales de soporte, modelos de aproximación de kernel, conjuntos de árboles de regresión y modelos de regresión de redes neuronales.

Figura 6

Flujo de trabajo en la aplicación Regression Learner



Nota. Adaptado de (*Entrenar Modelos de Regresión Para Predecir Datos Mediante Aprendizaje Automático Supervisado - MATLAB - MathWorks América Latina*, n.d.c)

1.6.1. Validación cruzada

Al entrenar un algoritmo, debe haber una forma de verificar que sea correcto. La validación cruzada relaciona los datos correctamente. Idealmente es que logre la generalización de datos porque el ensamblaje sobre ajustado hace que sea imposible predecir un valor que no esté en el entrenamiento. Por otro lado, si hay un desajuste, esto significa que el algoritmo no aprende. El conjunto de datos recopilados se divide en k grupos. Uno de los grupos servirá para autenticar, mientras que los demás permiten el entrenamiento. (Kevin Javier & Vergara Valenzuela, 2021a)

El proceso de entrenamiento se puede repetir de nuevo hasta que cada grupo sea un conjunto de validaciones. Esto le permite ajustar los hiperparámetros del modelo para elegir el modelo con el mejor rendimiento. (Kevin Javier & Vergara Valenzuela, 2021b)

CAPÍTULO II

METODOLOGÍA

2.1. Materiales y equipos.

El trabajo se realizó con diferentes materiales, instrumentos e insumos necesarios en los diferentes procesos cumpliendo los objetivos establecidos.

3.1.1. Materiales y equipo

La tabla 3 expone los diferentes equipos, insumos y materiales requeridos en el estudio planteado.

Tabla 3 *Equipo, materiales e insumos*

Equipos	Materiales	Insumos	Reactivos
Refractómetro digital	Soporte universal	Naranja valencia (<i>Citrus x sinensis</i>)	Solución 0,1 N de hidróxido de sodio
Balanza	Bureta de vidrio (25 o 50 mL)		Fenolftaleína
Prensa exprimidor	o Probeta de 100 mL		
calibre	Frasco de Erlenmeyer 250 mL		
	Pipetas volumétricas de 5 o 10 mL		

Papel industrial.

Vaso de
precipitado 300
mL

Decantador

Cuchillo

3.2. Métodos.

Se detalla a continuación diferentes métodos para cumplir con los objetivos planteados en este documento.

2.2.1. Desarrollo del modelo Neural Net Fitting

Se registró los valores pertinentes de cada una de las naranjas en base al estudio que se realizó, para ello se diseñó una tabla en una hoja electrónica de Excel versión 2020 donde se colocó las variables (peso, diámetro ecuatorial, altura, variedad, sólidos solubles, acidez titulable, índice de madurez) para alimentar el programa MATLAB.

La base de datos se almacenó acorde a los requerimientos de los dos modelos computacionales, el cual se explica en los siguientes párrafos; para evitar los errores y los sesgos de datos se ejecutó una aleatorización de los datos antes de iniciar el modelado de los métodos de Machine Learning. Para el estudio de la Neural Net Fitting se creó un archivo de Excel con cuatro hojas, la primera y tercera hoja las variables medidas o inputs (peso, variedad, diámetro ecuatorial y altura), mientras que las hojas dos y cuatro las variables de respuesta o targets (el índice de madurez).

Las dos primeras hojas se utilizaron para entrenar, validar y comprobar los modelos, mientras que la tercera y cuarta hoja se usó como prueba de la validación real medida en campo para comprobar el modelo que en mayor medida se ajustó a los datos. Por tanto, del universo de muestras se utilizó el 70% para el entrenamiento de la red neuronal, 10% para la validación y 10% para la prueba que realizó el programa. El último 10% se reservó para la verificación del método.(Atecnea, 2021)

Para el método de Regression Learner se diseñó un segundo archivo de Excel, donde se almacenaron los mismos datos del primer documento pero en diferente estructura, fue creado con tres hojas. La primera poseyó el 90% de los datos para realizar el entrenamiento, incluyendo entradas y salidas. Mientras que en la segunda hoja se almacenó el 10% restante con las mismas variables para desarrollar la validación. Por último, la variable respuesta de los datos reservados para la validación, se guardaron en la hoja tres; para luego evaluar el MSE y el valor de la regresión.

Para evitar errores al momento de hacer predicciones, se debió proporcionar los datos con el mismo formato de los datos de entrenamiento utilizados en la aplicación. Recordando que en las tablas no se tomó en cuenta los sólidos solubles totales ni la acidez titulable, porque éste es un factor para determinar el índice de madurez (variable de respuesta) para ambos métodos.

- ***Desarrollo de la técnica Neural Net Fitting***

La red Neural Net Fitting está diseñada con el Software de MATLAB versión estudiantil 2020, el cual cuenta con el paquete Statistics and Machine Learning Toolbox. Mediante esta se pudo evaluar su rendimiento utilizando el error cuadrático medio (MSE) y el análisis de

regresión (R^2). (Function Fitting Neural Network - MATLAB Fitnet - MathWorks América Latina, n.d.b)

Una vez seleccionado los datos se procedió a definir la estructura de la red neuronal. No hay ninguna regla que indique el número óptimo de entrenamiento, por lo tanto, la evaluación de los datos se realizó con las tres técnicas que proporciona el programa (Levenberg-Marquardt, Bayesian Regularization y Scaled Conjugate Gradient).

En cada técnica se ensayó con distintos números de neuronas para organizar la representación interna y escoger el mejor valor de regresión (R^2). (Bazalar Pereda et al., n.d.) Por consecuencia el número de neurona se varió de 1 en 1 a partir de la décima neurona hasta que alcanzó una diferencia significativa representada por un valor superior al mínimo de $R^2=0.75$. (Function Fitting Neural Network - MATLAB Fitnet - MathWorks América Latina, n.d.a)

Esta es una técnica de entrenamiento supervisado, donde los pesos y los sesgos inician al azar al empezar el entrenamiento. (Function Fitting Neural Network - MATLAB Fitnet - MathWorks América Latina, n.d.b) Para un determinado conjunto de datos de entrada a la red se calcula una respuesta para cada neurona de salida y se compara con la respuesta deseada correspondiente. (Bazalar Pereda et al., n.d.).

La configuración inicial predefinida por la red neuronal de MATLAB fue la siguiente para encontrar el error cuadrático medio de menor valor según (*Retropropagación Levenberg-Marquardt - MATLAB Trainlm - MathWorks América Latina, n.d.*):

- net.trainParam.epochs — Número máximo de épocas a entrenar. El valor predeterminado es 1000.

- `net.trainParam.goal` — Objetivo de rendimiento. El valor predeterminado es 0.
- `net.trainParam.max_fail` — Máximo de fallos de validación. El valor predeterminado es .6
- `net.trainParam.time` — Tiempo máximo para entrenar en segundos.
- El valor predeterminado es infinito.

Una vez seleccionada la técnica de entrenamiento el programa quedó listo para realizar el entrenamiento. Se escogió el valor de la regresión promedio más cercano a uno para escoger el mejor algoritmo. El número de neuronas ocultas intervienen en la eficacia del aprendizaje y generalización de la RNA.

- ***Desarrollo de la técnica de Regression Learner.***

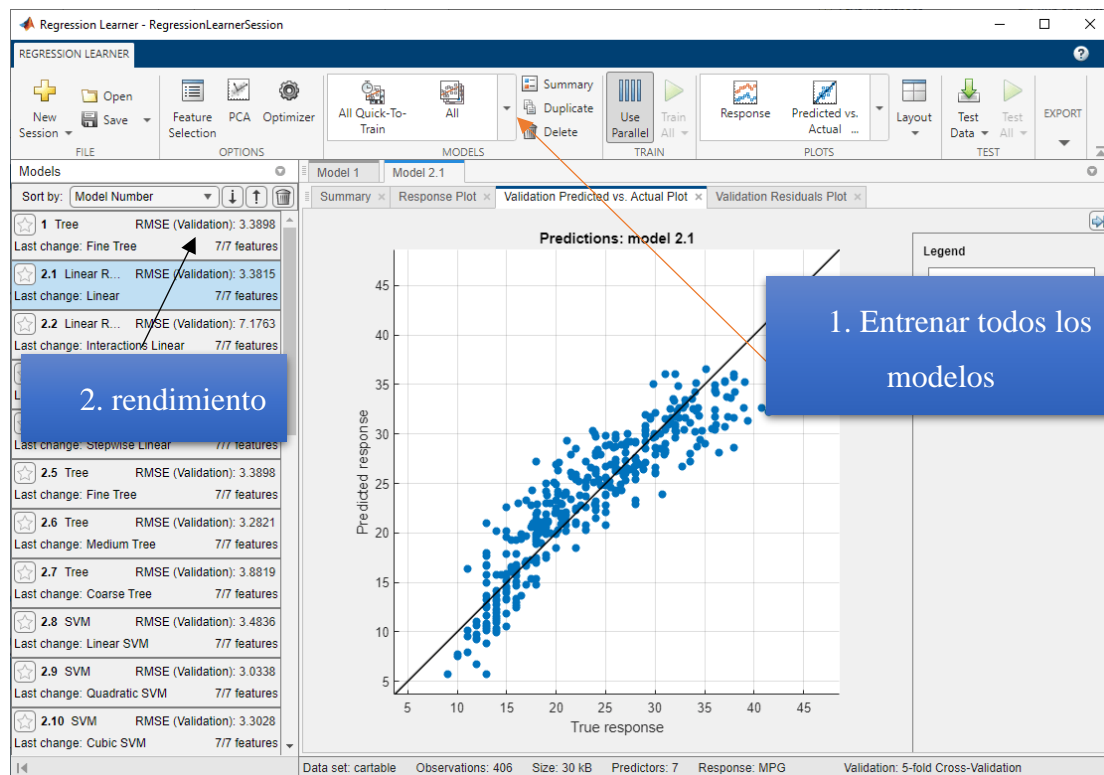
La aplicación está diseñada con el mismo Software utilizando la aplicación Regression Learner, esta entrena modelos de regresión para predecir datos, ofrece la facilidad de realizar un entrenamiento automatizado para descubrir el mejor tipo de modelo de regresión como indica la figura 7. (*Red de Trenes Con Funciones Numéricas - MATLAB & Simulink - MathWorks América Latina*, n.d.)

Para evitar el sobre ajuste, se implementará el método de validación cruzada (Kevin Javier & Vergara Valenzuela, 2021), luego de lo cual se entrenó los modelos y se comparó los resultados de ajuste de cada uno, se eligió el método que en mayor medida representó los

datos.(Entrenar Modelos de Regresión Para Predecir Datos Mediante Aprendizaje Automático Supervisado - MATLAB - MathWorks América Latina, n.d.d)

Figura 7

Varios tipos de modelos de regresión de MATLAB entrenándose al mismo tiempo



Nota. Adaptado de (Red de Trenes Con Funciones Numéricas - MATLAB & Simulink - MathWorks América Latina, n.d.)

3.2.2. Validación del método

Para el caso de la RNA una vez evaluado la eficiencia del método se procedió a exportar el algoritmo para evaluar el error de este programa. Se ingresó el 10% de los datos originales que no fueron utilizados para el entramiento en el algoritmo generado, luego se verificó el margen

de error entre la respuesta predictiva del programa y el valor verdadero determinado en laboratorio.

En el caso de Regression Learner se eligió el mejor modelo de todos los entrenados, procedente esto el algoritmo fue exportado y ejecutado desde la aplicación para la predicción con los datos previamente almacenados en la hoja Excel. De la misma manera que en RNA, se procedió a determinar el MSE Ec. (1) y el valor de la regresión en el Excel.

$$MSE = \frac{\sum \text{Error de pronóstico}^2}{n} \quad (1)$$

Fuente: (Betancourt, 2016)

Donde:

Error de pronóstico = diferencia entre el valor real y el valor pronosticado

n= número de datos

Como paso final, el mejor método se determinó por el menor valor de MSE y el valor de la regresión más cercano a uno; el modelo que más se aproximó a esta relación es el método computacional validado para la clasificación de la naranja (*Citrus x síntesis*) en las variedades Valencia y Washington.

3.3. Descripción del proceso

2.4.1. Ubicación de materia prima

La ubicación de la materia prima se define en la tabla 4

Tabla 4*Locación de la muestra*

Provincia	Esmeraldas
Cantón	Rioverde
Altitud	6 m.s.n.m
Humedad Relativa Promedio	87%
Temperatura media	28 °C
Pluviosidad	1000-1100 mm. Año
Sitio	San Vicente, finca La Patrona, localizada a 6 km del Cantón Rioverde

Nota. Adaptado de (*Cantón Rioverde - Wikipedia, La Enciclopedia Libre*, n.d.)

3.4.2. Selección de la naranja

Para la recolección de la naranja se tomó en cuenta ciertos indicadores, al menos 20% de coloración amarillenta para la cosecha, a partir de esta se recolectó de diámetros semejantes de la misma forma con el peso, esto aplica para las dos variedades de naranja (Ing. Vicente Gutiérrez Rico, 2017)

3.4.3. Toma de muestra

La muestra estuvo compuesta por un número de frutos representativos para una plantación de estándar de árboles de naranja de las dos variedades, la toma de muestra fue procedente de

las cuatro orientaciones de las copas de los árboles (norte, sur, este y oeste), se recolectó de forma manual para evitar posibles daños y contaminación, la cosecha se hizo en días secos.(Estefanía Romero Rodríguez, 2020). Para asegurar que la muestra represente confiablemente la población se aplicó la Ec. 2 para definir el tamaño de muestreo.

$$n = \frac{Z^2 * p * q * N}{e^2 * (N - 1) + Z^2 * p * q} \quad (2)$$

Fuente:(Normas APA, 2016)

Donde:

n= tamaño de la muestra

N= población

Z=nivel de confianza (95%=1,96)

p=probabilidad a favor (0,5)

q=probabilidad en contra (0,5)

e=error muestral (5%= 0,05)

3.4.4. Envasado y transporte de muestra

Los envases cumplieron las características de calidad e inocuidad necesarias para aseverar la manipulación, el transporte y la conservación apropiados de las naranjas; los envases estuvieron lejos de focos de contaminación, sobre todo no permaneció en contacto con el suelo.(NTE INEN 2844 NORMA PARA LA NARANJA (CODEX STAN 245-2004, MOD), 2004).

Se almacenó en un recipiente tipo costal y el transporte de la materia prima se realizó en transporte público en horario nocturno de 11 pm, estableciendo su arribo en menos de doce horas desde su cosecha, el análisis fisicoquímico correspondiente se ejecutó en las unidades Edu-productivas de la FICAYA, a fin de evitar errores de muestreo o confusión se efectuó un viaje por cada variedad de naranja.

3.4.5. Recepción de la materia prima

La fase de experimentación se ofició en la ciudad de Ibarra, en los diferentes laboratorios que se encuentran a disposición de la Facultad de Ingeniería en Ciencias Agropecuarias y Ambientales (FICAYA) de la Universidad Técnica del Norte, en la Tabla 5 se detalla los laboratorios y la ubicación del experimento.

Tabla 5*Ubicación y datos meteorológicos del lugar de experimentación*

Provincia	Imbabura
Cantón	Ibarra
Altitud	2225 m.s.n.m
Humedad Relativa Promedio	62%
Temperatura media	17 °C
Pluviosidad	503-1000 mm. Año
Sitio 1	Campus Colegio Universitario en las unidades eduproductivas de la Universidad Técnica del Norte
Sitio 2	Se repitió los análisis en el Laboratorio de Análisis de la finca la Patrona Cantón Rioverde

Nota. Adaptado de (Meteored, n.d.)

3.4.6. Conservación de la naranja en fase experimental

Esta actividad estuvo dirigida en dos momentos, en primera instancia se realizó un preenfriamiento luego de la cosecha en un rango de 4 a 5 ° C hasta el momento de transporte. Luego de ser transportada las naranjas permanecieron almacenadas en refrigeración para retardar el tiempo de maduración, con el fin de proporcionar una muestra homogénea, se

colocó a temperatura de 5° C, lo que le ayudó a reducir la actividad fisiológica para evitar el sesgo y el error del análisis.

3.4.7. Determinación del peso

Se escogió cada naranja y se procedió a ubicarla en la plataforma de peso de una balanza digital para determinar el peso total en gramos, se tomó registro con dos decimales que el instrumento arrojó. (Estefanía Romero Rodríguez, 2020a)

3.4.8. Determinación del diámetro ecuatorial.

Primero se midió el valor del mayor diámetro ecuatorial, posteriormente se obtuvo al diámetro polar, para este procedimiento el instrumental necesario fue un calibre (Pie de Rey) digital, cuya unidad de medida constó en mm (Estefanía Romero Rodríguez, 2020b).

3.4.9. Determinación del índice de madurez

Se consiguió de la relación entre el valor mínimo de los sólidos solubles (%Brix) y el valor de la acidez titulable, mediante la expresión de la Ec. (3). (INEN 380, n.d.)

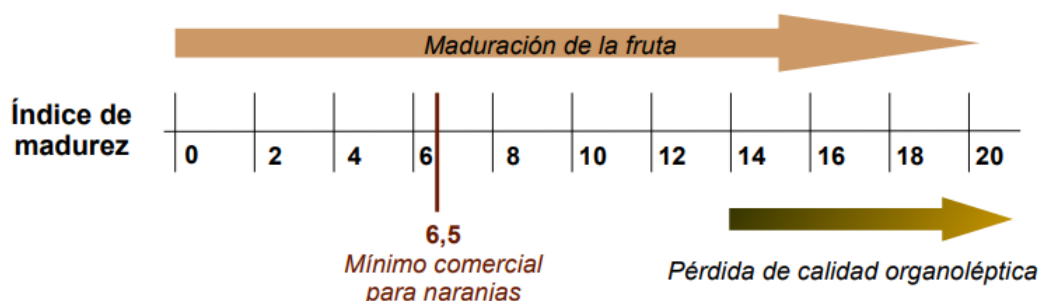
$$\text{índice de madurez} = \frac{\text{sólidos solubles totales}(\%Brix)}{\text{acidez titulable}(\%)} \quad (3)$$

Fuente: (INEN 380, n.d.)

El índice de madurez es un indicador del sabor del zumo y el momento óptimo de la cosecha, la normativa establece un índice mínimo 6,5 ° brix para su comercialización para los cítricos en función de su grupo varietal como indica la figura 8. (Comisión Europea, 2011)

Figura 8

Relación del índice de madurez de la fruta con la maduración de esta y la pérdida de calidad organoléptica de su zumo.



Nota. Adaptado de(Comisión Europea, 2011)

- ***Determinación de la acidez titulable***

La determinación de la acidez titulable acorde a la normativa (INEN 381, n.d.) se ejecutó con la toma de 5 mL de zumo en un vaso de precipitación, al cual se le añadió tres gotas de fenolftaleína. A continuación, se agregó el hidróxido sódico 0,1 N lentamente, hasta que el zumo cambió ligeramente de color, al finalizar se registró los mililitros de hidróxido sódico que se gastaron hasta el viraje de color; con este dato se procedió a la Ec. (4) para estimar la valoración de acidez.

$$Acidez\ total = \frac{(V_1 * N_1 * M)10}{V_2} \quad (4)$$

Fuente: (INEN 381, n.d.)

Donde:

A= g de ácido en 1000 mL de producto.

V_1 = mL de NaOH usados para la titulación de la alícuota.

V_2 = volumen de la alícuota tomada.

N_1 = normalidad de la solución de NaOH

M= peso molecular del ácido considerado como referencia (0,0064)

- ***Determinación de sólidos solubles totales***

Este procedimiento se trabajó conforme a la norma (INEN 380, n.d.), se realizó con un refractómetro digital, equipo con el cual se determina el contenido de sólidos solubles totales, en este instrumento se tomó tres medidas independientes de una misma muestra, empleando para cada una de ellas unas gotas de zumo, tomando como valor definitivo la media de los tres valores obtenidos. Ec. (5). (Estefanía Romero Rodríguez, 2020)

$$\text{sólidos solubles totales} = \sum 3 \text{ muestras medidas } (^{\circ}\text{Brix}) / 3 \quad (5)$$

CAPÍTULO III

RESULTADOS Y DISCUSIÓN

En la presente investigación “Desarrollo de un método computacional para la clasificación de naranjas (*Citrus x sinensis*) variedades Valencia y Washington” realizada en la Provincia de Esmeralda, cantón Rio verde se obtuvieron los subsiguientes resultados.

3.1. Evaluar diferentes técnicas de Machine-Learning para la modelación de los datos experimentales.

Para evaluar las diferentes técnicas de Machine-Learning se elaboró una base de datos con 388 datos calculados previamente en la Ec (2). Los análisis ejecutados fueron: peso, diámetro, altura e índice de madurez; obtenidas en las dos variedades de naranja; para evitar los sesgos y errores se realizó una aleatorización ejecutada con el programa de MATLAB procedente a esto, se almacenó acorde a los requerimientos computacionales del software con un 90% del conjunto de datos total.

3.1.1. Técnica Neural Net Fitting

El entrenamiento de la red para evaluar la técnica Neural Net Fitting se inició con la configuración estándar del programa aplicando los tres métodos de modelamiento ofrecidos por este (Levenberg-Marquardt, Bayesian Regularization y Scaled Conjugate Gradient). En cuanto al número de neuronas, se varió de 1 a 1 en la capa oculta hasta llegar el máximo de 100 neuronas de acuerdo con investigaciones anteriores(Bazalar Pereda et al., n.d.) para la predicción del índice de madurez.

Con el resultado de las 100 neuronas, la mejor representación de la estructura del programa se logró con el modelamiento de Leven-Marquardt al cabo de 70 neuronas en la capa oculta como muestra la tabla 6, por ende, fue seleccionado como la mejor técnica Neural Net Fitting a exportar para ser validado.

Tabla 6

Resultados de la técnica Neural Net Fitting

N° de neurona	Levenberg-Marquardt		Bayesian Regularization		Scaled Conjugate Gradient	
	MSE	R ²	MSE	R ²	MSE	R ²
10	2.591	0.8998	3.088	0.878	3.444	0.868
11	2.959	0.8891	3.171	0.884	3.791	0.853
12	3.113	0.891	3.001	0.887	2.915	0.888
13	2.566	0.904	3.074	0.886	2.903	0.89
14	2.984	0.894	3.297	0.878	4.184	0.838
15	3.033	0.882	2.986	0.89	3	0.883
20	3.092	0.892	3.131	0.883	2.603	0.902
23	2.541	0.906	3.128	0.878	3.609	0.869
25	2.503	0.905	3.36	0.874	3.086	0.888
26	2.532	0.906	2.974	0.886	3.152	0.887
28	2.599	0.907	2.911	0.889	2.815	0.894
30	2.354	0.918	3.088	0.881	3.088	0.886
36	2.114	0.924	3.222	0.878	3.319	0.863
60	1.973	0.927	3.043	0.884	3.323	0.873
67	1.751	0.934	3.099	0.883	2.858	0.89
70	1.6707	0.937	3.058	0.885	3.359	0.88
100	1.811	0.936	3.085	0.885	3.8	0.854

Se escogió esta técnica dado el valor representativo 0.937 con respecto al coeficiente de correlación superando al mínimo 0.5 mencionado por (Lalinde et al., 2018a) demostrando así la correlación fuerte, de la misma forma, el entrenamiento denota un bajo error cuadrático 1.6707 medio para la predicción del índice de madurez.

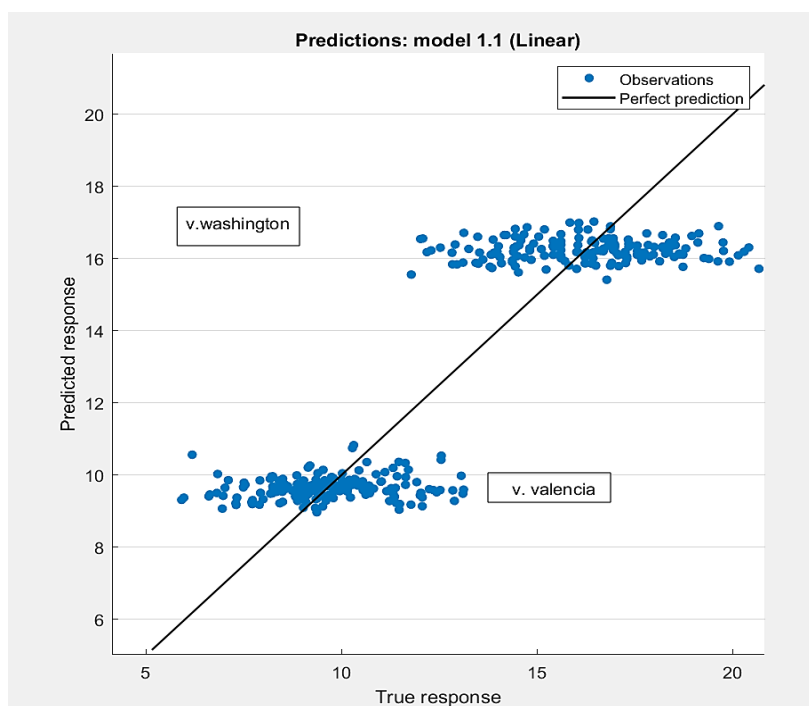
3.1.2. Regression Learner

Para el método de regresión se utilizó los mismos datos que sirvieron para entrenar el método anterior, aplicando los 20 modelados matemáticos que posee la aplicación para el modelamiento.

Como resultado final se evaluó cada uno de los métodos y se optó por el método de regresión lineal teniendo el mejor resultado evaluado, representado por R^2 0.77 y un MSE 3.19. La figura 9 evidencia la comparación entre los valores predichos por el modelo y los valores experimentales evaluados.

Figura 9

Respuesta de Regression Learner



Nota. El grafico representa la dispersión con respecto a las dos variedades de naranjas y las clasifica según el índice de madurez

El gráfico de dispersión de la figura 9, permite evidenciar el porqué del valor obtenido para el coeficiente de regresión. Ya que, una representación cercana a 1, implicaría que los puntos se adapten completamente a la línea central negra (Lalinde et al., 2018b). La dispersión evidenciada en la figura 9, demuestra el valor 0.77 alcanzado con este método computacional, que a pesar de ser menor que el alcanzado por Neural Net Fitting, sigue estando en el rango de representatividad alta de sistemas según (Lalinde et al., 2018c).

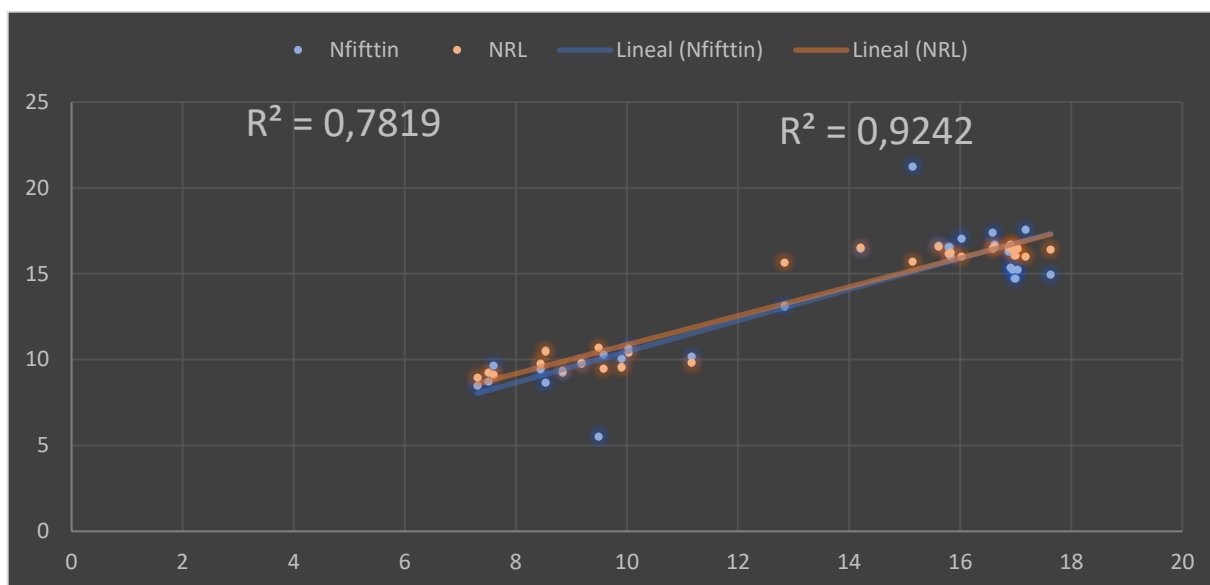
3.2. Validar los modelos para la selección del que se ajuste en mayor medida a los datos.

Se entrenó la RNA con la topología y configuración final del objetivo anterior, utilizando los datos separados en un principio por la programación de MATLAB con el fin de evitar las generalidades y sobreajuste de los mejores modelos seleccionados con respecto a la validación.

El cumplimiento de este objetivo se alcanzó con 10% de datos restantes que no fueron considerados en el entrenamiento como datos verdaderos; realizando la correspondiente comparación entre los valores reales y los valores predichos por ambas técnicas como indica la figura 10, obteniéndose como resultados un coeficiente de correlación 0.9242 del método de Regression Learner Tabla 7.

Figura 10

Comparativa de valores predichos y valores verdaderos



Nota. En el grafico muestra el ajuste de los modelos validados con respecto a los datos medido en campo, la pendiente de color anaranjado responde a Regression Learner, mientras que, la coloración azul pertenece a Neural Net Fitting.

La figura 10 evidencia la importancia de determinar el coeficiente de regresión por el hecho de que este valor indica la confiabilidad del modelo evaluado, de esta misma forma, se corrobora que la técnica Neural Net Fitting a pesar de ser menor que la técnica de Regression Lerner no deja de estar en un rango alto de representatividad de sistemas de acuerdo con la afirmación de (Lalinde et al., 2018d).

Por lo que se concluye que el método de Regression Learner representa con mayor fidelidad la clasificación de naranjas (*Citrus x sinensis*) variedades Valencia y Washington, entre todos los modelos evaluados en este estudio.

Tabla 7*Validación de modelos Neuronales*

	Neural Net Fitting	Regression Learner
R ²	0.78	0.9242
MSE	3.33	1.31

Sin embargo, a pesar de que en la validación los mejores indicadores se obtuvieron para el método de Regression Learner, no fue así para la etapa de entrenamiento. Esto puede deberse a la atipicidad de los datos en la prueba real de la naranja, de lo que se puede inferir que existe otro factor que debe tomarse en cuenta para el modelamiento del sistema de estudio.

La variable que probablemente puede estar faltando para mejorar la representación es la dureza del fruto (Estefanía Romero Rodríguez, 2020), dado que en la experimentación se obtuvo la atipicidad con frutos de igual color, pero de textura dura a diferencia de los frutos blandos. La tabla 8 muestra los indicadores del entrenamiento y la validación de las técnicas más representativas.

Tabla 8*Comparativa de los objetivos alcanzados*

	Neural Net Fitting		Regression Learner	
	R ²	MSE	R ²	MSE
Entrenamiento	0.94	1.671	0.77	3.19
Validación	0.78	3.33	0.92	1.31

En definitiva, se puede concluir que la técnica desarrollada por Regression Learner representa el método computacional que permite la clasificación de naranjas (*Citrus x sinensis*) variedades Valencia y Washington.

Se realizaron otras representaciones como el modelamiento individual de las variedades de naranja como alternativa para mejorar la representación del algoritmo, pero debido a la ineficacia del método por errores altos como observa la tabla 9 no se ofreció un modelado, los resultados demostraron que una representación global es la mejor como indica la tabla 7.

Tabla 9*Cálculo del MSE del modelamiento por variedad*

	Naranja Valencia	Naranja Washington
MSE	8	13

Las investigaciones de redes neuronales alcanzan rendimientos mayores al 80%, demostrando que tienen un amplio alcance, por lo cual es factible que sea implementada en la industria alimentaria; entre ellos destaca el trabajo presentado por Gatica citado por (Figueredo-Ávila & Ballesteros-Ricaurte, 2016) con la finalidad de lograr una óptima recolección del olivo; modelado que alcanzó un rendimiento del 90% , a pesar de ser este inferior al estudio realizado no deja de estar en el rango de representatividad alta de un sistema.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

El análisis y el estudio del índice de madurez a partir de los parámetros convencionales tomados en cuenta en este estudio se representaron con mayor fidelidad a través de la técnica de Neural Net Fitting con Levenberg-Marquardt y 70 neuronas en la capa oculta, bajo la consideración del error medio cuadrado y el coeficiente de determinación.

Tras la validación, se identificó al método de Regression Learner como el más representativo de la clasificación de naranja, por su mayor coeficiente de determinación y menor error mínimo cuadrado.

La diferencia de resultados entre la validación y el entrenamiento establece que se requiere una variable adicional para representar el sistema. A pesar de ello, los indicadores de regresión muestran que la regresión alcanzada es confiable.

Se acepta la hipótesis alternativa ya que las variables medidas en campo y las características fisicoquímicas permitieron al método computacional determinar el índice de madurez de la naranja

Recomendaciones

Recomendaciones para futuros estudios

- Realizar una base de datos con los mismos parámetros de evaluación, modelar en diferentes épocas de producción de naranja y en distintas zonas para ver la reproducibilidad y replicabilidad del método seleccionado.

- Se recomienda en el momento de seleccionar la muestra, aparte de la coloración verdosa o manchas amarillentas, tomar como indicador la dureza de la fruta
- En futuras investigaciones de clasificación de frutas relacionar las características fisicoquímicas con valoraciones de color realizados con espectrofotometría para la determinación del índice de madurez.

Referencias Bibliográficas

- Almache, A., Alexandra, M., Escudero, V., & Estefania, A. (2021). *UNIVERSIDAD TÉCNICA DE COTOPAXI FACULTAD DE CIENCIAS AGROPECUARIAS Y RECURSOS NATURALES CARRERA DE INGENIERÍA AGROINDUSTRIAL NARANJA, VALENCIA (Citrus x sinensis) Y NARANJA AGRIA (Citrus x aurantium)*.
- Almache Almache Mayra Alexandra, & Villacres Escudero Andrea Estefania. (2021, August). *EVALUACIÓN DE LOS PARÁMETROS DE CALIDAD EN DOS VARIEDADES DE NARANJA, VALENCIA (Citrus x sinensis) Y NARANJA AGRIA (Citrus x aurantium) PARA LA ELABORACIÓN DE JUGO*.
<http://repositorio.utc.edu.ec/bitstream/27000/8160/1/PC-002077.pdf>
- Ashok, V., & Vinod, D. S. (2014). Automatic quality evaluation of fruits using Probabilistic Neural Network approach. *Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014*, 308–311.
<https://doi.org/10.1109/IC3I.2014.7019807>
- ASTUDILLO, S. Y. (2018). *Importancia de la producción de naranja en Caluma y el impacto que tiene en los festivales del cantón* [UNIVERSIDAD SAN FRANCISCO DE QUITO].
<https://repositorio.usfq.edu.ec/bitstream/23000/7935/1/141127.pdf>
- Atecnea. (2021, April 15). (31) *Machine Learning con MATLAB (Tutorial BASICO)* - YouTube.
<https://www.youtube.com/watch?v=torhYoh10uU&list=LL&index=3>
- Bazalar Pereda, M. S., Tejerina, M. A., Paganini, J. H., & Gonzalez, S. C. (n.d.). *Selección de una Arquitectura de Red Neuronal Artificial eficiente para predecir el coeficiente de difusividad másica del aguaymanto (Physalis peruviana L.) deshidratado osmoconvectivamente*.
- Betancourt, D. F. (2016, March 7). *Medición del error en pronósticos de demanda*.
<https://www.ingenioempresa.com/medicion-error-pronostico/>
- Cantón Rioverde* - Wikipedia, la enciclopedia libre. (n.d.). Retrieved April 26, 2022, from
https://es.wikipedia.org/wiki/Cant%C3%B3n_Rioverde

- CODEX STAN 245-2004. (n.d.). *NORMA PARA LA NARANJA*. Retrieved April 26, 2022, from https://www.fao.org/fao-who-codexalimentarius/sh-proxy/ar/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252Fstandards%252FCXS%2B245-2004%252FCXS_245s.pdf
- Comisión Europea. (2011, June 11). *Reglamento de Ejecución (UE) N° 543/2011*. <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32011R0543&from=ES>
- Durán Barón, R., & Villa, A. L. (n.d.). *EVOLUCIÓN DE LOS PARÁMETROS DE CALIDAD DE NARANJA VALENCIA PRODUCIDA EN EL MUNICIPIO DE CHIMICHAGUA, CESAR-COLOMBIA EVOLUTION OF THE PARAMETERS OF QUALITY ON VALENCIA ORANGE GROWTH ON THE MUNICIPALITY OF CHIMICHAGUA, CESAR-COLOMBIA*.
- EFSA. (2013). *Frutas 270*. https://www.mapa.gob.es/es/ministerio/servicios/informacion/naranja_tcm30-102369.pdf
- El Telégrafo. (2017). *La sobreproducción de naranja provoca que el precio de la fruta en finca se desplome*. <https://www.eltelegrafo.com.ec/noticias/88/8/la-sobreproduccion-de-naranja-provoca-que-el-precio-de-la-fruta-en-finca-se-desplome>
- Entrenar modelos de regresión para predecir datos mediante aprendizaje automático supervisado - MATLAB - MathWorks América Latina*. (n.d.). Retrieved April 15, 2022, from https://la.mathworks.com/help/stats/regressionlearner-app.html?searchHighlight=Regression%20Learner&s_tid=srchtitle_Regression%20Learner_1
- Estefanía Romero Rodríguez, Á. H. G. R. C. V. J. A. A. (2020). *Determinación de la Calidad de los Frutos Cítricos en Laboratorio*.
- Figueredo-Ávila, G. A., & Ballesteros-Ricaurte, J. A. (2016). Fruit ripeness identification with artificial neural networks - A review. *Ciencia y Agricultura*, 13(1), 117–132. <https://doi.org/10.19053/01228420.4811>

- Flores, F. , C. S. D. , E. U. R. , D. S. J. , & V. M. S. (2021). *Vista de Diseño e implementación de una red neuronal artificial para predecir el rendimiento académico en estudiantes de Ingeniería Civil de la UNIFSLB*. <https://doi.org/10.47796/ves.v10i1.464>
- Function fitting neural network - MATLAB fitnet - MathWorks América Latina*. (n.d.). Retrieved April 15, 2022, from https://la.mathworks.com/help/deeplearning/ref/fitnet.html?s_tid=doc_ta
- Gan, X., Chen, L., Yang, D., & Liu, G. (2011). The research of rainfall prediction models based on Matlab neural network. *CCIS2011 - Proceedings: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, 45–48. <https://doi.org/10.1109/CCIS.2011.6045029>
- Georgi Alexander Campelo Valle. (2020a). *Situación actual de los productores de naranja (Citrus sinensis) en el Ecuador*”.
- Georgi Alexander Campelo Valle. (2020b). *UNIVERSIDAD TÉCNICA DE BABAHOYO FACULTAD DE CIENCIAS AGROPECUARIAS CARRERA DE INGENIERIA AGROPECUARIA*.
- Gustavo Andrés Figueredo-Ávila, J. A. B.-R. (2015, October 12). *Fruit ripeness identification with artificial neural networks*. <https://www.redalyc.org/journal/5600/560062814010/html/>
- Herrera, J. C. (2015, February 26). *Vista de Diseño de un sistema automático de selección de frutos de café mediante técnicas de visión artificial | Revista UIS Ingenierías*. <https://revistas.uis.edu.co/index.php/revistauisingenierias/article/view/7-14/5758>
- INEC. (2020). *Estadísticas Agropecuarias* /. <https://www.ecuadorencifras.gob.ec/estadisticas-agropecuarias-2/>
- INEN 380. (n.d.). *Determinacion de Solidos Solubles* . Retrieved April 25, 2022, from <https://www.normalizacion.gob.ec/buzon/normas/380.pdf>
- INEN 381. (n.d.). *Norma Técnica Ecuatoriana CONSERVAS VEGETALES DETERMINACIÓN DE ACIDEZ TITULABLE METODO POTENCIOMETRICO DE REFERENCIA*. Retrieved April 25, 2022, from <https://www.normalizacion.gob.ec/buzon/normas/381.pdf>

- Ing. Vicente Gutiérrez Rico. (2017, August). *Cosecha, postcosecha y comercialización de la de la naranja*.
https://formaciontecnicabolivia.org/sites/default/files/publicaciones/guia_cosecha_postcosecha_y_transformacion_de_la_naranja.pdf
- ISOTools. (2015). *Ventajas de la estandarización de procesos*.
<https://www.isotools.org/2018/08/23/que-aportan-los-procesos/>
- Kevin Javier, & Vergara Valenzuela. (2021). *DISEÑO E IMPLEMENTACIÓN DE MODELOS PREDICTIVOS DE CLASIFICACIÓN Y REGRESIÓN DE PROTEÍNAS BASADOS EN LA DIGITALIZACIÓN DE PROPIEDADES FÍSICOQUÍMICAS* [UNIVERSIDAD DE CHILE FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS DEPARTAMENTO DE INGENIERÍA QUÍMICA, BIOTECNOLOGÍA Y MATERIALES].
<https://repositorio.uchile.cl/bitstream/handle/2250/180185/Dise%c3%bl0-e-implementation-de-modelos-predictivos-de-clasificacion-y-regresion-de-proteinas-basadas-en-la-digitalizacion-de-propiedades-fisicoquimicas.pdf?sequence=1&isAllowed=y>
- La Madrid Távara, B., Eduardo ASESOR, L., Oblitas Vera, I., & Leonardo, C. (2019). *IMPLEMENTACIÓN DE UN ALGORITMO DE CONTROL DE CALIDAD PARA LA SELECCIÓN DE PRODUCTOS AGRICOLAS UTILIZANDO VISIÓN ARTIFICIAL*.
- Lalinde, H., Diego, J., Castro, E., Rangel, C., Gerardo, J., Sierra, T., Andrés, C., Torrado, A., Karina, M., Sierra, C., Milena, S., Pirela, B., & José, V. (2018). *Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones*.
<https://www.redalyc.org/articulo.oa?>
- Lucía, O., Vargas, T., Andrés, Á., & Pérez, N. (2018). *Implementación de un Sistema de Visión Artificial para la clasificación de naranja producida en el departamento del Quindío* (Vol. 26, Issue 45).
- Madeleine Hanco Condori. (2017). *ELABORACIÓN DE UN RECUBRIMIENTO COMESTIBLE Y SU COMPARACIÓN CON RECUBRIMIENTO COMERCIAL APLICADO EN NARANJA (Citrus sinensis) VALENCIA* [UNIVERSIDAD NACIONAL DEL ALTIPLANO].

http://repositorio.unap.edu.pe/bitstream/handle/UNAP/4215/Hancco_Condori_Madeleine.pdf?sequence=1&isAllowed=y

Martínez-González, M. E., Balois-Morales, R., Alia-Tejacal, I., Cortes-Cruz, M. A., Palomino-Hermosillo, Y. A., López-Gúzman, G. G., Martínez-González, M. E., Balois-Morales, R., Alia-Tejacal, I., Cortes-Cruz, M. A., Palomino-Hermosillo, Y. A., & López-Gúzman, G. G. (2017). Poscosecha de frutos: maduración y cambios bioquímicos. *Revista Mexicana de Ciencias Agrícolas*, 8(SPE19), 4075–4087. <https://doi.org/10.29312/REMEXCA.V0I19.674>

Meteored. (n.d.). *Tiempo en Ibarra. Clima*. Retrieved April 25, 2022, from https://www.meteored.com.ec/tiempo-en_Ibarra-America+Sur-Ecuador-Imbabura--1-20221.html

Niño, H., Aular, J., Cásares, M., & Torres, Y. G. (2020). Bioagro. *Bioagro*, 32(2), 131–138. <https://revistas.uclave.org/index.php/bioagro/article/view/2697>

Normas APA. (2016, November 20). *Fórmula para calcular la muestra de una población*. <https://normasapa.net/formula-muestra-poblacion/>

NTE INEN 2844 NORMA PARA LA NARANJA (CODEX STAN 245-2004, MOD). (2004).

¿Qué es una red neuronal? - MATLAB & Simulink. (n.d.). Retrieved April 16, 2022, from <https://la.mathworks.com/discovery/neural-network.html>

Red de trenes con funciones numéricas - MATLAB & Simulink - MathWorks América Latina. (n.d.). Retrieved April 15, 2022, from <https://la.mathworks.com/help/deeplearning/ug/train-network-on-data-set-of-numeric-features.html>

Retropropagación Levenberg-Marquardt - MATLAB trainlm - MathWorks América Latina. (n.d.). Retrieved April 15, 2022, from https://la.mathworks.com/help/deeplearning/ref/trainlm.html?s_tid=doc_ta#bvmliaj6

Robalino, A. (2010, February 5). *Naranja. Estudio agroindustrial en el Ecuador by MIPRO Sistema de Inteligencia de Mercados - Issuu*. <https://issuu.com/mipro/docs/naranja>

- Rojas-Lara, P. C., Pérez-Grajales, ; M, Colinas-León, ; M T B, Sahagún-Castellanos, ; J, & Avitia-García, ; E. (2008). MODELOS MATEMÁTICOS PARA ESTIMAR EL CRECIMIENTO DEL FRUTO DE CHILE MANZANO (*Capsicum pubescens* R y P). In *Revista Chapingo Serie Horticultura* (Vol. 14, Issue 3).
- Solano Robinzon Hugo Roberto. (2018). *CADENA COMERCIAL DE NARANJA (Citrus sinensis L.)*.
<http://repositorio.ug.edu.ec/bitstream/redug/36382/1/Solano%20Robinzon%20Hugo%20Roberto.pdf>
- Spanish Fruits & Delicacies. (2019, December 31). *TODO SOBRE LA NARANJA*.
<https://www.spanishfruitsanddelicacies.com/blogs/news/la-naranja-fuente-de-salud>
- Tomar, M. S., & Pradhan, R. C. (2022). Prediction of mass-based process designing parameters of amla fruit using different modeling techniques. *Journal of Food Process Engineering*, e14039. <https://doi.org/10.1111/JFPE.14039>
- Torres, R., Montes, E. J., Pérez, O. A., & Andrade, R. D. (2013a). Relación del Color y del Estado de Madurez con las Propiedades Fisicoquímicas de Frutas Tropicales. *Información Tecnológica*, 24(3), 51–56. <https://doi.org/10.4067/S0718-07642013000300007>
- Torres, R., Montes, E. J., Pérez, O. A., & Andrade, R. D. (2013b). Relación del Color y del Estado de Madurez con las Propiedades Fisicoquímicas de Frutas Tropicales. *Información Tecnológica*, 24(3), 51–56. <https://doi.org/10.4067/S0718-07642013000300007>
- VALENTÍN TASSILE. (2020). *Application of Non-Linear Mixed Models in the Growth of Pear Fruits*.
- Victor Aguilar-Alvarado, J., & Alfredo Campoverde-Molina, M. (2019). *Classification of fruits based on convolutional neural networks Classificação de frutos com base em redes neurais convolucionais Ciências de la ingeniería Artículo de investigación*. 5(01), 3–22. <https://doi.org/10.23857/pc.v5i01.1210>

Anexo

```
function [trainedModel, validationRMSE] = trainRegressionModel(thesis(trainingData))

% [trainedModel, validationRMSE] = trainRegressionModel(trainingData)

% Returns a trained regression model and its RMSE. This code recreates the
% model trained in Regression Learner app. Use the generated code to
% automate training the same model with new data, or to learn how to
% programmatically train models.

%
% Variables have been created in the base workspace.
% Structure 'trainedModelregre' exported from Regression Learner.
% To make predictions on a new table, T:
% yfit = trainedModelregre.predictFcn(T)

% For more information, see How to predict using an exported model.

% Input:
%   trainingData: A table containing the same predictor and response
%   columns as those imported into the app.

%
% Output:
%   trainedModel: A struct containing the trained regression model. The
%   struct contains various fields with information about the trained
%   model.
%
%   trainedModel.predictFcn: A function to make predictions on new data.
%
```

```
% validationRMSE: A double containing the RMSE. In the app, the
% History list displays the RMSE for each model.
%
% Use the code to train the model with new data. To retrain your model,
% call the function from the command line with your original data or new
% data as the input argument trainingData.
%
% For example, to retrain a regression model trained with the original data
% set T, enter:
% [trainedModel, validationRMSE] = trainRegressionModel(T)
%
% To make predictions with the returned 'trainedModel' on new data T2, use
% yfit = trainedModel.predictFcn(T2)
%
% T2 must be a table containing at least the same predictor columns as used
% during training. For details, enter:
% trainedModel.HowToPredict
%
% Auto-generated by MATLAB on 19-Oct-2022 20:21:45
%
% Extract predictors and response
% This code processes the data into the right shape for training the
% model.
```

```
inputTable = trainingData;

predictorNames = {'Variedad', 'peso', 'diametro', 'altura'};

predictors = inputTable(:, predictorNames);

response = inputTable.IM;

isCategoricalPredictor = [false, false, false, false];

% Train a regression model

% This code specifies all the model options and trains the model.

concatenatedPredictorsAndResponse = predictors;

concatenatedPredictorsAndResponse.IM = response;

linearModel = fitlm(...

    concatenatedPredictorsAndResponse, ...

    'linear', ...

    'RobustOpts', 'off');

% Create the result struct with predict function

predictorExtractionFcn = @(t) t(:, predictorNames);

linearModelPredictFcn = @(x) predict(linearModel, x);

trainedModel.predictFcn = @(x) linearModelPredictFcn(predictorExtractionFcn(x));

% Add additional fields to the result struct

trainedModel.RequiredVariables = {'Variedad', 'altura', 'diametro', 'peso'};

trainedModel.LinearModel = linearModel;
```

```
trainedModel.About = 'This struct is a trained model exported from Regression Learner
R2020a.';
```

```
trainedModel.HowToPredict = sprintf("To make predictions on a new table, T, use: \n yfit
= c.predictFcn(T) \nreplacing "c" with the name of the variable that is this struct, e.g.
"trainedModel". \n \nThe table, T, must contain the variables returned by: \n
c.RequiredVariables \nVariable formats (e.g. matrix/vector, datatype) must match the original
training data. \nAdditional variables are ignored. \n \nFor more information, see <a
href="matlab:helpview(fullfile(docroot, "stats", "stats.map"),
"appregression_exportmodeltoworkspace")">How to predict using an exported model</a>.'.);
```

```
% Extract predictors and response
```

```
% This code processes the data into the right shape for training the
```

```
% model.
```

```
inputTable = trainingData;
```

```
predictorNames = {'Variedad', 'peso', 'diametro', 'altura'};
```

```
predictors = inputTable(:, predictorNames);
```

```
response = inputTable.IM;
```

```
isCategoricalPredictor = [false, false, false, false];
```

```
% Perform cross-validation
```

```
KFolds = 5;
```

```
cvp = cvpartition(size(response, 1), 'KFold', KFolds);
```

```
% Initialize the predictions to the proper sizes
```

```
validationPredictions = response;
```

```
for fold = 1:KFolds
```

```
    trainingPredictors = predictors(cvp.training(fold), :);
```

```
trainingResponse = response(cvp.training(fold), :);  
foldIsCategoricalPredictor = isCategoricalPredictor;  
  
% Train a regression model  
% This code specifies all the model options and trains the model.  
concatenatedPredictorsAndResponse = trainingPredictors;  
concatenatedPredictorsAndResponse.IM = trainingResponse;  
linearModel = fitlm(...  
    concatenatedPredictorsAndResponse, ...  
    'linear', ...  
    'RobustOpts', 'off');  
  
% Create the result struct with predict function  
linearModelPredictFcn = @(x) predict(linearModel, x);  
validationPredictFcn = @(x) linearModelPredictFcn(x);  
  
% Add additional fields to the result struct  
  
% Compute validation predictions  
validationPredictors = predictors(cvp.test(fold), :);  
foldPredictions = validationPredictFcn(validationPredictors);  
  
% Store predictions in the original order  
validationPredictions(cvp.test(fold), :) = foldPredictions;
```

```
end
```

```
% Compute validation RMSE
```

```
isNotMissing = ~isnan(validationPredictions) & ~isnan(response);
```

```
validationRMSE = sqrt(nansum(( validationPredictions - response ).^2) /  
numel(response(isNotMissing) ));
```