

# **DETERMINACIÓN DE LOS FACTORES IMPLICADOS Y ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DE LA DEMANDA DE PRODUCTOS AGRÍCOLAS**

*DETERMINATION OF THE INVOLVED FACTORS AND AUTOMATIC LEARNING ALGORITHMS FOR THE PREDICTION OF THE DEMAND OF AGRICULTURAL PRODUCTS*

**Omar Guillermo García Salas**

Tecnológico Nacional de México / IT de Celaya, México  
*17031273@itcelaya.edu.mx*

**Norma Verónica Ramírez Pérez**

Tecnológico Nacional de México / IT de Celaya, México  
*norma.ramirez@itcelaya.edu.mx*

**Mariana Mosqueda León**

Tecnológico Nacional de México / IT de Celaya, México  
*18030921@celaya.tecnm.mx*

**José Guillermo Fierro Mendoza**

Tecnológico Nacional de México / IT de Celaya, México  
*guillermo.fierro@itcelaya.edu.mx*

**Julio Armando Asato España**

Tecnológico Nacional de México / IT de Celaya, México  
*julio.asato@itcelaya.edu.mx*

**Recepción:** 17/diciembre/2021

**Aceptación:** 22/febrero/2022

## **Resumen**

El presente artículo trata sobre la primera etapa de un proyecto para predecir la demanda/precio de productos agroalimentarios. Este acercamiento busca encontrar los principales factores que afectan la demanda de estos productos, además, la evaluación de diversos algoritmos de inteligencia artificial, que aplicando aprendizaje automático ayuden a obtener valores de predicción de una manera eficiente y con niveles de precisión adecuados. En esta ocasión se consideró el producto del brócoli fresco de exportación como sujeto de estudio. Mediante el cruce

de información de diversas fuentes, se llegó a la identificación de los factores que se considera, afectan el comportamiento de su precio en el mercado. Con base a los registros históricos obtenidos de fuentes oficiales, se elaboró un dataset con los factores que se identifican como relevantes en todo el proceso de la producción del brócoli y como primer acercamiento se implementó un modelo predictivo utilizando Python para predecir el comportamiento de la demanda con base en los registros históricos con un nivel de confianza aceptable. En parte introductoria se hace un análisis documental relacionado con la identificación y contextualización de la necesidad de la predicción de la demanda. En la sección relativa al método se describe la forma de organización de la investigación y conceptos relacionados. Se presentan los resultados donde se muestra el análisis de la información que condujo a la identificación de los factores y la implementación del modelo de predicción en esta etapa preliminar y posteriormente se presenta la discusión donde se establecen precisiones relacionadas con los resultados obtenidos y las tecnologías utilizadas y finalmente en conclusiones se presentan los logros de la primera etapa del proyecto de investigación.

**Palabras Clave:** Agroalimentarios, aprendizaje automático, demanda/precio, inteligencia artificial, predicción.

### **Abstract**

*This article deals with the first stage of a project to predict the demand / price of agri-food products. This approach seeks to find the main factors that affect the demand for these products, in addition to the evaluation of various artificial intelligence algorithms, which applying machine learning help to obtain prediction values efficiently and with adequate levels of precision. On this occasion, the fresh broccoli product for export was considered as a study subject. By crossing information from various sources, it was possible to identify the factors that are considered to affect the behavior of its price in the market. Based on the historical records obtained from official sources, a dataset was elaborated with the factors that are identified as relevant in the entire broccoli production process and as a first approach, a predictive model was implemented using Python to predict the behavior*

*of demand in based on historical records with an acceptable level of confidence. In an introductory part, a documentary analysis related to the identification and contextualization of the need for demand prediction is made. The method section describes how to organize the research and related concepts. The results are presented where the analysis of the information that led to the identification of the factors and the implementation of the prediction model in this preliminary stage is shown and later the discussion is presented where details related to the results obtained and the technologies used are presented, and finally in conclusions the achievements of the first stage of the research project are presented.*

**Keywords:** *Agri-food, artificial intelligence, demand/price, machine learning, prediction.*

## **1. Introducción**

Como expresa Boada [2017], existe una necesidad creciente de análisis de datos corporativos relacionada con la predicción del comportamiento de la demanda y contar con aplicaciones que optimicen su estimación en función de la dinámica del negocio. Esta necesidad se reiteró con base en información obtenida de empresarios dedicados a la exportación de productos agrícolas como brócoli y coliflor en el estado de Guanajuato, que manifiestan que la volatilidad de la demanda es uno de los principales problemas de los ejecutivos, lo que los motiva a intentar predecir la demanda considerando ciertos indicios, con la finalidad de encontrar los factores que inciden en la determinación de los precios de sus productos y analizar con certidumbre posible escenarios de venta y mejorar sus procesos logísticos, incluso prevenir situaciones de riesgo, apoyando la toma de decisiones. Los resultados de este artículo se derivan de un proyecto de investigación registrado ante el Programa para el Desarrollo Profesional Docente (PRODEP), presentado por el Cuerpo Académico de Tecnologías de la Información (ITCEL-CA-22), que se denomina “Sistema de predicción de demanda para la planificación de la producción con soporte en la toma de decisiones para la industria alimentaria basado en técnicas de Machine Learning”. En dicho proyecto se tienen consideradas varias etapas, una de las iniciales es precisamente la gestión de un banco de datos, motivo

del presente artículo, y sobre el cual posteriormente se aplicaron variados procesos de clasificación para la predicción. Dentro de las actividades de esta etapa, se analizaron distintas propuestas sobre técnicas predicción aplicadas en diferentes áreas, de esta manera, poder contar con un amplio repertorio de conocimiento generado por otros proyectos para la generación de un modelo predictivo propio, enfocado en la demanda en organizaciones orientadas a la producción, distribución y venta de productos del ramo alimenticio. Dentro de la revisión de literatura se consideraron diferentes trabajos, el primero de ellos presentado por Bermúdez [2011], consiste en un proyecto que busca que los bancos de alimentos puedan contar con un modelo que les permita modificar su sistema de distribución actual, todo esto gracias al desarrollo de un sistema PLEM (Programación lineal entera-mixta) los cuales sirven para formular modelos que contribuyen eficientemente en la toma de decisiones, aplicables a todos los niveles organizacionales.

El siguiente trabajo analizado fue el de Valencia y Correa [2013], que proponen un diseño de un algoritmo que permita predecir la demanda de energía diaria en Colombia, el cual fue creado haciendo uso de un modelo dinámico bayesiano basado en el filtro del Kalman. Esto se hizo con la intención de prevenir fallas en sistemas que regularmente lo hacen o zonas en donde se inició el servicio eléctrico. Por último, en esta reseña, se analiza la tesis de maestría en estadística matemática de Quezada [2017], que trata de la construcción de un modelo, haciendo uso del método de k-vecinos para estimar los valores atípicos dentro de las 31 cortes superiores de Justicia, durante la tesis se muestra el análisis descriptivo de los datos el cual permite llegar a dicho objetivo.

Ahora bien, en la agricultura la inteligencia artificial permite escalar los modelos de producción sin comprometer la calidad de los procesos, lo que es esencial dada la creciente competitividad en el mercado. Entonces, se pretende que el modelo predictivo a formular permita anticiparse al valor en el mercado y tomar ventaja de este para maximizar ganancias, así como de reducir pérdidas tanto monetarias como de alimentos perecederos. En general los empresarios registran muchos datos, pero no tienen certeza de que sean los adecuados o como utilizarlos para que la predicción de la demanda sea la más aproximada a la real.

El objetivo de este avance se relaciona con dos líneas de trabajo relativamente paralelas, una es la determinación de los factores que determinan el conjunto de datos (data set) que deberá considerarse para el proceso predictivo, y la segunda es el desarrollo de un producto de software que permita la ejecución de los algoritmos de predicción, el cual sea susceptible de ser transferido a las empresas interesadas para la puesta en operación de la solución.

## **2. Métodos**

Los proyectos como al que se refiere el presente artículo requieren de diferentes recursos tecnológicos y metodológicos. Respecto a los últimos, se aplicó el método de investigación documental para la revisión de la literatura publicada sobre el tema, aunado a prácticas de investigación de campo, ya que hubo que realizar entrevistas y visitas a plantas productoras, para poder afinar e integrar adecuadamente los hallazgos encontrados en fuentes documentales. De acuerdo con García [2012], la investigación documental consta de las siguientes tres etapas:

- Elaboración de un esquema de trabajo.
- Búsqueda y selección de información.
- Integración de registros de información.

Las cuales fueron desarrolladas integrando un banco de fuentes de información de tipo cuantitativo y cualitativo, provenientes principalmente de México y Estados Unidos de América, que son las naciones de mayor interés para el caso de estudio. Sin embargo, para el discernimiento de tan abundante información, se complementó este proceso con trabajo de campo, dentro del cual se destaca el uso de la técnica de entrevista, para obtener información de tipo cualitativo que permita integrar, ponderar y depurar la información obtenida por medios documentales.

Respecto a recursos tecnológicos, algunas de las técnicas de predicción que serán mencionadas a lo largo de este documento serán las siguientes:

- Redes neuronales: son sistemas de aprendizaje que emulan el funcionamiento del cerebro humano, con lo que es posible establecer relaciones no lineales entre las variables de entrada y las de salida. Su principal ventaja que consiste

en procesar información en paralelo en tiempo real ha permitido su aplicación en la clasificación y reconocimiento de patrones en sistemas complejos [Martínez, 2014].

- Series de tiempo: estas tienen como propósito estudiar el comportamiento de una o varias variables, cuyos datos son registrados de manera cronológica en el tiempo y con ello es posible predecir sus valores futuros, lo que apoya a la toma de decisiones en una empresa u organización.
- Máquina de Soporte Vectorial (SVM): son algoritmos que realizan su proceso de análisis sobre dos clases distintas de los puntos de entrada. La descripción dada por los datos de los vectores de soporte obtenidos es capaz de formar una frontera de decisión dentro del dominio de los datos de aprendizaje de una manera muy efectiva. Esta función de frontera, cuando es traída de regreso al espacio de entrada, puede separar los datos en todas las clases distintas, cada una formando un agrupamiento [Betancourt, 2005].
- K-Vecinos: es un algoritmo basado en instancia de tipo supervisado de Machine Learning. Puede usarse para clasificar nuevas muestras (valores discretos) o para predecir (regresión, valores continuos). Al ser un método sencillo, es ideal para introducirse en el mundo del Aprendizaje Automático. Sirve esencialmente para clasificar valores buscando los puntos de datos “más similares” (por cercanía) aprendidos en la etapa de entrenamiento y haciendo conjeturas de nuevos puntos basado en esa clasificación [Aprendemachinelearning,2018].
- Modelo Bayesiano: esta teoría aporta una herramienta global que, mediante una visión de conjunto, es capaz de abordar diferentes tipos de situaciones de elección bajo condiciones de incertidumbre. Los métodos estadísticos bayesianos proporcionan la forma más general y potente de inferencia estadística; permite utilizar la información inicial proporcionada por expertos, si se dispone, y aplica la teoría de la información para determinar distribuciones no informativas que producen resultados objetivos, es decir, resultados que sólo dependen del modelo probabilístico asumido y de los datos experimentales obtenidos [Bernardo, 2004].

Adicionalmente a estos recursos de tipo algorítmico, se utilizaron diferentes herramientas tecnológicas o software, algunos de ellos para efecto de desarrollo de aplicaciones especializadas susceptibles de ser aplicadas en procesos de transferencia tecnológica. Dentro de esta se destaca el uso de la tecnología de desarrollo de software Python. Esta herramienta cuenta con librerías como Keras y Tensorflow, que permiten la creación y entrenamiento de redes neuronales. Ahora bien, para verificar la validez de las predicciones obtenidas es necesario contar con un valor de contraste, para lo cual podemos aplicar diferentes alternativas, un primer mecanismo, es el cual será obtenido mediante el software libre Weka, que es considerado como una compilación de diferentes algoritmos de aprendizaje automático para trabajos relacionados a IA, deeplearning, clasificación, entre otros. Una de las ventajas de esta herramienta es que es de uso fácil, pero cuenta con bastante profundidad en sus análisis. Otra alternativa es dividir el conjunto de datos históricos en dos diferentes marcos, datos de entrenamiento y datos de validación del modelo, utilizando un esquema de 80-20 o 70-30 que nos permitirá identificar la precisión del modelo. Adicionalmente se usará el software propietario Matlab® de la empresa MathWorks. Este es un entorno de desarrollo que ofrece un gran número de toolboxes que pueden estar especializadas en cumplir tareas relacionadas al deeplearning, redes neuronales e incluso series de tiempo. Otra ventaja del producto es el soporte y apoyo metodológico que ofrece la organización desarrolladora, por el hecho de contar con licencia autorizada.

El proceso en la creación de un sistema predictivo es complejo, en especial cuando no hay un banco de datos estructurado, por lo que se comenzó por seguir el modelo planteado por MathWorks. De esta manera se identifican los pasos a seguir presentados en la figura 1.

Los pasos indicados en el método para la creación de un sistema predictivo, son:

- Determinar los recursos para la obtención de datos: este paso consiste en determinar si el origen de la información será de empresas, registros almacenados, o ambos, como fue en este caso.
- Acceso a información: es importante delimitar las fuentes de datos significativos, base de datos y archivos.

- Explorar la información: identificar los picos de datos o valores atípicos para comenzar a tener una mejor visualización de los datos.
- Procesamiento de la información: aquí es donde se aplica la técnica o las técnicas que sirven para poder manejar la información, clasificarla y analizarla.
- Desarrollo de los modelos predictivos: obtener los parámetros de optimización, modelos de creación como Machine Learning al igual que los modelos de validación.
- Integración a un sistema: hacer la integración a sistemas que permitan visualizar la aplicación de modelo obtenido y probar su funcionalidad.



Fuente: Adaptada de [MathWorks, 2021]

Figura 1 Proceso para la creación de un sistema predictivo.

Es importante destacar que estas actividades fueron aplicadas de acuerdo a las necesidades propias que el proceso investigativo requería.

### 3. Resultados

El producto alimenticio en el que se centrará la atención en este proyecto es el brócoli fresco, esto debido a que regionalmente es un producto de suma importancia por su impacto en la actividad económica de la agroindustria del estado de

Guanajuato, dado que en varias plantas la producción es completamente para exportación.

Como un primer acercamiento a datos relevantes y cumpliendo el primer paso del modelo, se identificó la base de datos del Servicio de Información Agroalimentaria y Pesquera (SIAP). Esta organización gubernamental recolecta información sobre estadística básica del sector agrícola, aquí encontramos entre los atributos más importantes; la temporada, hectáreas cosechadas, sembradas, precio, y el valor de producción.

En el caso particular para predicción de la demanda se modeló el presente a partir de los datos históricos para extrapolar el futuro. Es importante acotar que a medida que se va retroalimentando con los datos reales el modelo va mejorando, es decir si se alimenta el modelo con los nuevos datos, se ajusta el modelo para generar cada vez un modelo más preciso. Utilizamos un modelo de Deep Learning, aplicando el modelado para una sola variable al principio, el histórico de la demanda en base de la temporada, dejando para la segunda fase añadir variables adicionales. Es de destacar que, de acuerdo a las necesidades detectadas, se requerirá que los datos sean registrados diariamente, lo cual es factible de acuerdo a las entrevistas realizadas con empresarios.

En el proceso de producción del brócoli se da en tres grandes etapas, la generación de la semilla, la generación de la plántula, trasplantado en los campos agrícolas, cosecha y preparación y empaque para distribución. Entre cada etapa se identifican una serie de rendimientos, dignos de analizarse en particular para cada una de las etapas, ya que son afectaciones del producto en cada etapa. En esta etapa preliminar solo nos enfocaremos a determinar el comportamiento de la demanda y establecer un modelo de predicción basado en datos históricos. Sin embargo, como parte del análisis de datos consideramos importante presentar los patrones de comportamiento analizados entre los diversos factores, tabla 1.

Cabe destacar la correlación presente entre lo sembrado y su valor de producción, que se representa gráficamente en la figura 2, donde se observa el mayor rendimiento sobre el valor, que se alcanza no superando las cien hectareas de

siembra, aunque bajo un razonamiento abductivo también podría ser un indicador sobre la capacidad de producción de muchos agricultores.

Tabla 1 Determinación de factores que inciden en la determinación de la demanda.

Factor	Descripción
Volumen de Producción	La cantidad de producto generada en un cierto espacio de tiempo. Ejemplo miles de unidades por mes, o por año.
Temporalidad	Una serie temporal es una sucesión de observaciones de una variable tomada en varios instantes de tiempo. En el caso de productos analizados, son las temporadas de enero-julio y agosto-diciembre. También se conocen como temporadas primavera-verano, otoño-invierno. Los datos son registrados y analizados por año y por mes.
Precio	Precio de la unidad de venta del producto, por ejemplo, cabezas de brócoli, normalmente vendidas en cajas de varios componentes.
Rendimiento	La medida de la afectación de la capacidad de producción entre las etapas.
Producto siniestrado	Producto no logrado por diversas situaciones.



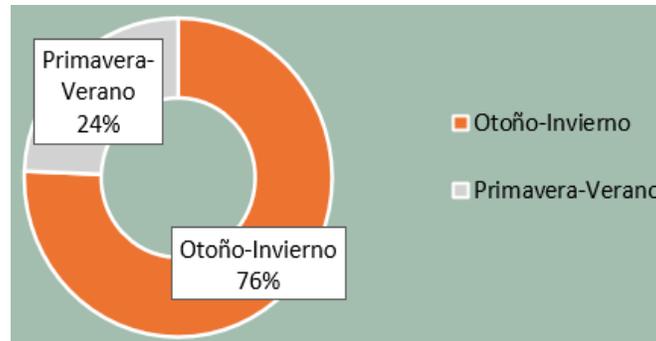
Fuente: [Elaboración propia].

Figura 2 Correlación entre la producción sembrada y el valor de la producción.

Los casos de producción siniestrada se presentan de manera más recurrente en la temporada de otoño-invierno, como se aprecia en la figura 3, lo cual podría ser un factor a tener en cuenta, ya que es evidente que este hecho afecta el precio.

La gráfica referente al cambio de precio de la figura 4, indica que el brócoli seguirá con mucha demanda en un futuro cercano, así que el precio seguirá a la alza, además, podemos notar como las temporadas no llegan afectar el precio del mismo. Estos datos sólo toman en cuenta los precios y consumo en México.

También se presenta la gráfica referente a la cantidad de hectareas sembradas de brócoli de los últimos años figura 5, aquí es posible seguir confirmando que la demanda del brócoli sigue en aumento.



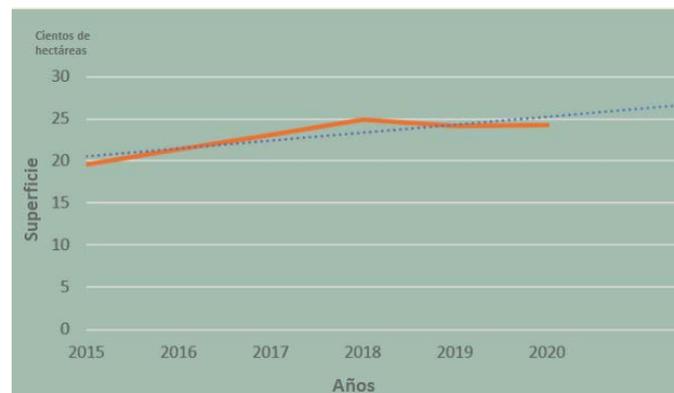
Fuente: [Elaboración propia].

Figura 3 Casos de producción siniestrada por temporada de siembra.



Fuente: [Elaboración propia].

Figura 4 Precio promedio por temporada.

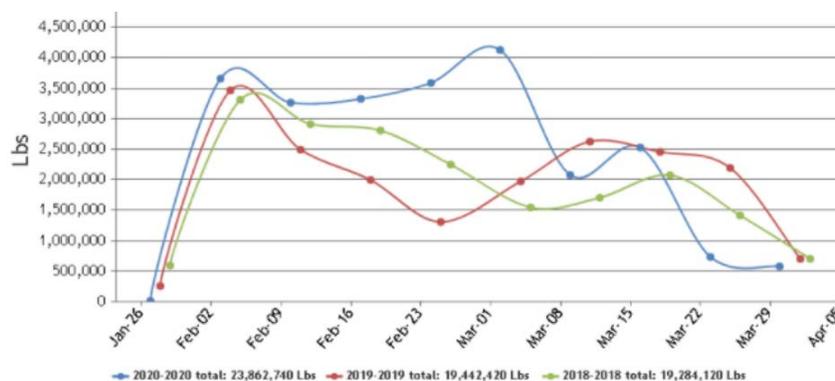


Fuente: [Elaboración propia].

Figura 5 Histórico de superficie sembrada.

Con las gráficas presentadas se pone en evidencia la importancia de variables como el volumen de producción, la cosecha siniestrada, sembrado-cosechado, precios, etc. Se decidió buscar en otras fuentes sobre otras variables que podrían afectar la

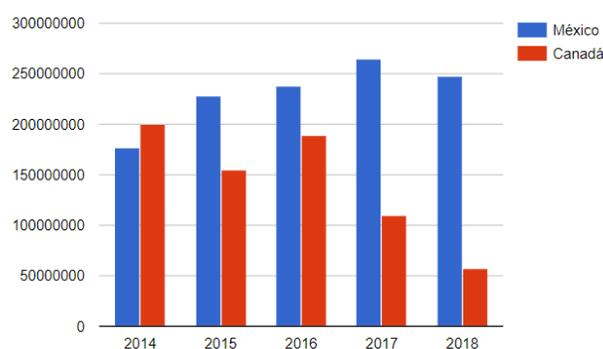
demanda-precio del brócoli para tomarlas en cuenta en nuestro sistema. Se encontraron datos sobre el efecto de la pandemia en la demanda del brócoli, que se observan en la figura 6.



Fuente: [Produce Blue Book, 2020].

Figura 6 Efecto de la demanda del brócoli en pandemia.

Esto será un factor importante a tener en cuenta debido a que ha afectado la demanda a futuro del producto. Otra característica importante a tener en cuenta ha sido sobre las exportaciones de brócoli en México, la figura 7 muestra la exportación entre México y EU, donde es evidente que México ha sido el principal exportador de brócoli en Estados Unidos, su exportación se mantiene constante y en crecimiento, a comparación con Canada, año con año ha disminuido su exportación a EU.

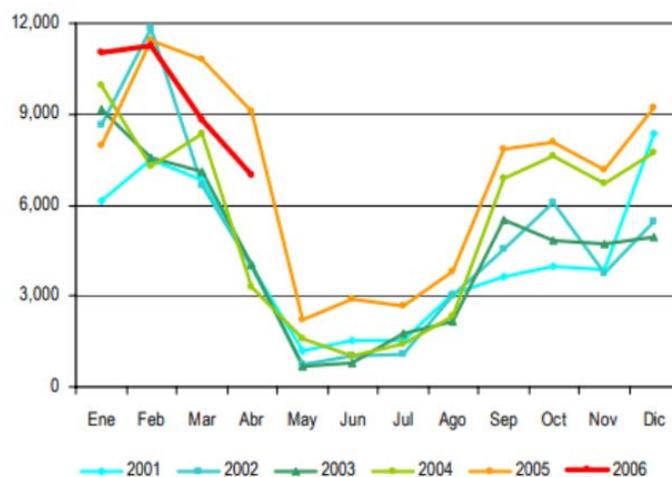


Fuente: [Negocios y Mercados, 2020].

Figura 7 Exportación de México y Canada a EU.

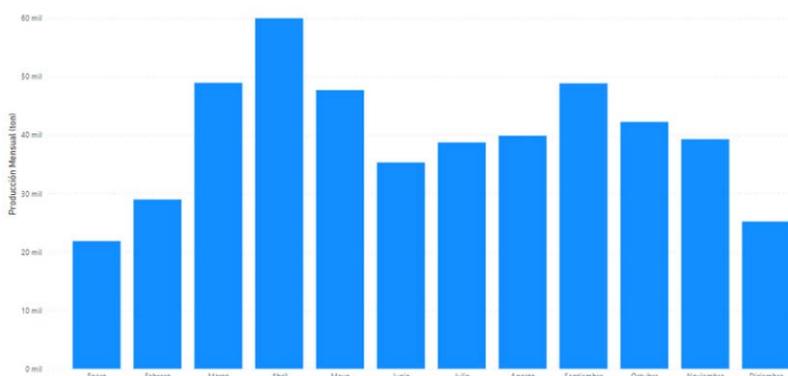
En la figura 8 se muestra la temporalidad del brócoli, ahí se puede apreciar como su demanda cambia en los meses, teniendo dos temporadas de alta demanda

coincidiendo con los datos previamente analizados, aunque los datos no sean recientes, comparandola con la producción del año 2018 de la figura 9, lo cual coincide con los meses de mayor demanda.



Fuente: [Ministerio de Agricultura y Desarrollo Rural, 2006].

Figura 8 Registro histórico de la temporalidad del brócoli fresco.



Fuente: [Negocios y Mercados, 2020].

Figura 9 Producción de Brócoli en 2018.

En resumen, de acuerdo a la evidencia recabada, los factores que se incorporaron en el data set para el análisis de las variables como posibles factores para la predicción de la demanda son: la temporada, el volumen de producción histórica, el precio de venta, la superficie sembrada y cosechada, la superficie siniestrada, el estado y municipio de localización, la capacidad de producción, la calidad del producto y el volumen del mercado de exportación, como se muestra en figura 10.

Anio	Idestado	Idddr	Idcader	Idmunicipio	Idciclo	Idmodalidad	Idunidadmedida	Idcultivo	Sown	Harvested	Damaged	Volumenproduccion	Yield	Price
2008-03-01	22	100	6	63	4	2	400402	11340000	70.0	70.0	0	1007.20	28.70	6350.00
2008-09-01	22	100	6	63	2	2	400402	11340000	273.0	273.0	0	4095.00	30.00	6300.00
2009-03-01	33	150	9	78	6	3	600603	17010000	122.5	122.5	0	1775.20	43.20	9450.00
2009-09-01	33	150	9	78	3	3	600603	17010000	410.0	410.0	0	6352.50	46.50	10250.00
2010-03-01	33	150	9	78	6	3	600603	17010000	46.0	46.0	0	787.00	48.50	10900.00
2010-09-01	33	150	9	78	3	3	600603	17010000	315.0	315.0	0	4862.50	46.00	12500.00
2011-03-01	33	150	9	78	6	3	600603	17010000	55.0	55.0	0	931.20	46.80	11353.85
2011-09-01	33	150	9	78	3	3	600603	17010000	315.0	315.0	0	5112.00	48.30	11700.00
2012-03-01	22	100	6	63	4	2	400402	11340000	119.0	119.0	0	1784.30	29.98	8728.57
2012-09-01	22	100	6	63	2	2	400402	11340000	250.0	250.0	0	3570.50	28.90	8170.61

Fuente: [Elaboración propia].

Figura 10 Fragmento del dataset utilizado.

Para efectos de la determinación del modelo inicial sólo es de interés para esta primera etapa el volumen de la predicción y la temporabilidad se incluye en una variable de marca de tiempo. También se identificó el ciclo de vida del producto en cuestión, así como las tendencias o comportamientos del costo en el mercado, al igual que elementos que determinan el costo de producción, los cuales eventualmente podrán ser considerados como variables complementarias del modelo. Respecto a la determinación o predicción de la demanda mediante un modelo en Python, se llegó a la conclusión que su comportamiento está determinado por la temporalidad de la producción y es totalmente factible establecer un modelo Deep Learning, ya que responde a patrón de un conjunto de muestras tomadas a intervalos de tiempo regulares, dependiente del tiempo. Se han realizado algunos modelos y se está analizando su comportamiento al mediano y largo plazo, intentando detectar patrones y poder hacer pronósticos de cómo será su comportamiento futuro.

Para modelar el comportamiento se toman los datos mensuales que son los datos registrados hasta el momento y se dividen en marcos previos de entrenamiento y validación. De esta manera, el modelo resultante y validado puede aplicarse al siguiente periodo. Se considera entrenar la red con diferentes periodos con datos de campo proporcionados por una empresa de la región de Guanajuato, dedicada al procesamiento y exportación del producto. Como paso siguiente se separaron los

datos en datos de entrenamiento y datos de prueba. Dado que el dataset cuenta con 28 datos, de los cuales 24 son de entrenamiento y 4 como parte del conjunto de datos de prueba y validación. Se crea y entrena un modelotipo ForecasterAutoreg a partir de un regresor RandomForestRegressor y una ventana temporal de seis lags, significando con ello que estamos utilizando los seis registros temporales anteriores. Por el momento para la fase instrumental del proyecto se está utilizando una arquitectura sencilla de red neuronal FeedForward, llamada también MLP por sus siglas Multi-Layered Perceptron, con pocas neuronas y como método de activación tangente hiperbólica se entregarán valores normalizados o transformados entre -1 y 1. Se crea y entrena un modelotipo ForecasterAutoreg a partir de un regresor RandomForestRegressor y una ventana temporal de seis lags, significando con ello que estamos utilizando los seis registros temporales anteriores. Una vez entrenado el modelo, se predicen los datos de test (cuatro meses que corresponden a dos temporadas a futuro).

Finalmente se cuantifica el error que resulta del modelo en sus predicciones. En este caso, se emplea como métrica el mean squared error (mse). Sin embargo, para identificar la mejor combinación de lags e hiperparámetros, la librería Skforecast dispone de la función `grid_search_forecaster` con la que comparar los resultados obtenidos con cada configuración del modelo.

Por otra parte, se llegó a la conclusión de la necesidad de gestionar los datos con periodicidad más corta, días sería lo más adecuado, para entrenar el modelo y tener una salida, intentando predecir más de un “día futuro” y de esta manera poder establecer un comportamiento de la demanda y anticipar la planeación al procesamiento de producción de acuerdo con el marco de tiempo de elaboración del producto y apoyarle a la empresa a adecuar su planeación.

#### **4. Discusión**

En la actualidad, la inteligencia artificial ha demostrado ser una herramienta útil en el uso del análisis de datos, a través del aprendizaje automático empleando algoritmos de clasificación y predicción, como: RNA, SVM, NAIVEBAYES, KNN y J48, todos ellos han logrado grandes resultados para resolver problemas de una

manera eficiente, pero al revisar la literatura las redes neuronales resultaron ser más mencionadas en diferentes aplicaciones con las características de los procesos predictivos que se están desarrollando. Por lo que, para implementar este modelo las RNA resultan ser más adecuadas, debido a su gran potencial de detectar no linealidades en series de tiempo y con esto poder hacer predicciones certeras. Los datos, al ser ordenados de manera cronológica, se adaptan a las características de una serie de tiempo, por esta razón, el algoritmo SVM no es viable, debido que no generan un buen desempeño en los resultados; esto último se dedujo de una investigación, acerca de la implementación de este algoritmo en series de tiempo, aunado a la experiencia obtenida dentro del grupo de trabajo participante en la presente investigación.

## **5. Conclusiones**

Como conclusión de este primer estudio para la creación del modelo se dio a la tarea de leer la literatura de los diferentes algoritmos para la creación del modelo aquí propuesto, y estimar su eficacia y eficiencia para los datasets de productos alimenticios. Después de hacer una revisión exhaustiva del comportamiento de los datos se concluye que la predicción de la demanda debe ser modelada con algoritmos correspondientes a series temporales, en particular las redes neuronales multicapa. Lo anterior permitió evaluar dichos algoritmos, en cuanto a su entrenamiento y validación, lo que en resultados preliminares arrojó mejores predicciones con datasets de prueba, formulados a partir de datos del estado de Guanajuato obtenidos de páginas gubernamentales. Es importante mencionar que esta primera parte es la base para aplicar al caso de estudio que corresponde a una empresa exportadora en el estado Guanajuato. De manera complementaria se definirán los mecanismos para la sistematización para la adquisición de datos en las instancias involucradas directamente en el proceso, como son productores, distribuidores, competencia y mercado. Por último, la generalización del proceso de determinación de la demanda que se ajuste a los datasets particulares de cada organización mediante un software desarrollado en Python, sería el siguiente paso por realizar y comprobar su nivel de precisión respecto a implementaciones

similares usando los productos de software especializados como Matlab® y Weka. Además, se tendrán que seleccionar las herramientas a usar para el desarrollo, lo cual podría estar sujeto a ajustes de acuerdo con los resultados preliminares que se obtengan y con el cotejo de los resultados del mercado, sin embargo, hay que tomar en cuenta que estos entornos de desarrollo permiten la integración en diferentes áreas o tecnologías de una manera flexible para generar modelos dinámicos y autoajustables, que representen la esencia del aprendizaje automático.

## **7. Bibliografía y Referencias**

- [1] Aprendemachinelearning (2018, julio 10). Algoritmo k-Nearest Neighbor. <https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>.
- [2] Bagnato, J. I. (2018, 10 de julio). Clasificar con K-Nearest-Neighbor ejemplo en Python: <https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>
- [3] Bermúdez, Y. (2011). Aplicaciones de programación lineal, entera y mixta. Ingeniería Industrial. Actualidad y Nuevas Tendencias. II (7), 85-104: <https://www.redalyc.org/articulo.oa?id=215024822007>.
- [4] Bernardo, J. M. (2004). Aplicaciones en el Control Ambiental de Actividades Portuarias. Informe Metodológico, Universidad de Valencia, España. <https://www.puertos.es/es-es/medioambiente/Documents/pro14.pdf>.
- [5] Betancourt, G. (2005). Las máquinas de soporte vectorial (SVMs). Scientia Et Technica. [https://www.researchgate.net/publication/49588125\\_LAS\\_MQUINAS\\_DE\\_SOPORTE\\_VECTORIAL\\_SVMs](https://www.researchgate.net/publication/49588125_LAS_MQUINAS_DE_SOPORTE_VECTORIAL_SVMs).
- [6] Cuevas, J. y Gómez, A. (2013). Un modelo de asignación- empaque de despensas personalizadas para bancos de alimentos: un sistema sujeto a condiciones nutricionales y logísticas. *Dyna*, 88(3), 560–573.
- [7] Boada, A., (2017). Sistema de proyección de la demanda. Caso práctico de predicción automatizada en empresas de venta por catálogo. *Revista Perspectiva Empresarial*: [https://www.researchgate.net/profile/Antonio-Boada/publication/282293885\\_Uso\\_optimo\\_del\\_big\\_data\\_empresarial\\_a\\_fi](https://www.researchgate.net/profile/Antonio-Boada/publication/282293885_Uso_optimo_del_big_data_empresarial_a_fi)

- n\_de\_potenciar\_la\_prediccion\_de\_demanda\_de\_productos\_en\_el\_tiempo/links/560ae8b308ae576ce640f262/Uso-optimo-del-big-data-empresarial-a-fin-de-potenciar-la-prediccion-de-demanda-de-productos-en-el-tiempo.pdf.
- [8] García, L. (2012). Técnicas de investigación de campo y documental: basado en competencias. Grupo Editorial Éxodo.
- [9] MathWorks (s/f). Análisis predictivo, tres cosas que es necesario saber Mathworks.com. [Sitio Web]. <https://la.mathworks.com/discovery/predictive-analytics.html>
- [10] Melo, E. (2016). Modelo de predicción mensual de mortalidad general intrahospitalaria en el hospital regional Manuel Núñez Butrón-Puno. [Tesis de licenciatura]. Universidad Nacional del Altiplano, Puno, Perú. [http://repositorio.unap.edu.pe/bitstream/handle/UNAP/2092/Melo\\_Mayta\\_Ed\\_w%20in.pdf?sequence=1&isAllowed](http://repositorio.unap.edu.pe/bitstream/handle/UNAP/2092/Melo_Mayta_Ed_w%20in.pdf?sequence=1&isAllowed)
- [11] Ministerio de Agricultura y Desarrollo Rural (2006). Información de monitoreo internacional: [http://bibliotecadigital.agronet.gov.co/bitstream/1134\\_8/5270/2/2006101110206\\_Br%C3%B3coli%20Fresco%20-%20Junio.pdf](http://bibliotecadigital.agronet.gov.co/bitstream/1134_8/5270/2/2006101110206_Br%C3%B3coli%20Fresco%20-%20Junio.pdf).
- [12] Negocios y Mercados. (2020, marzo 31). Brócoli: <https://amhpac.org/negociosymercados/socios/v2/broccoli/>.
- [13] Produce Blue Book (2020, Abril 6). Broccoli, cauliflower see demand fall during pandemic. <https://www.producebluebook.com/2020/04/06/broccoli-cauliflower-see-demand-fall-during-pandemic/>.
- [14] Quezada, L. (2017). K-vecino más próximos en una aplicación de clasificación y predicción en el Poder Judicial del Perú. [Tesis de Maestría]. Universidad Nacional Mayor de San Marcos, Lima, Perú.
- [15] Reséndiz, J. A. (2006). Las máquinas de vectores de soporte para identificación en línea. [Tesis de maestría]. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México. <https://www.ctrl.cinvestav.mx/~yuw/pdf/MaTesJAR.pdf>
- [16] Valencia, M., y Correa J. C. (2013). Un Modelo Dinámico Bayesiano Para Pronóstico de Energía Diaria. *Revista Ingeniería Industrial*. Año. 12, N°2. <http://revistas.ubiobio.cl/index.php/RI/article/view/9/3253>.