Article **Journal-Mathematical and Quantitative Methods**
December 2022, Vol.6 No.11 6-17

Implementation of models to predict avocado exports

Implementación de modelos para predecir exportaciones de aguacate

OCHOA-ORNELAS, Raquel†* & GUDIÑO-OCHOA, Alberto

Tecnológico Nacional de México/Instituto Tecnológico de Ciudad Guzmán, México.

ID 1st Author: *Raquel, Ochoa-Ornelas* / **ORC ID:** 0000-0003-1824-5789, **Researcher ID Thomson:** S-4687-2018, **arXiv Author ID:** Raquel Ochoa, **SNI-CONACYT ID:** 668976

ID 1st Co-author: *Alberto, Gudiño-Ochoa* / **ORC ID:** 0000-0002-2366-7452, **Researcher ID Thomson:** HDN-0235-2022, **arXiv Author ID:** AlbertoGO, **PubMed Autor ID:** albertogudo@orcid, **SNI-CONACYT ID:** 1135216

DOI: 10.35429/JMQM.2022.11.6.6.17 Received July 15, 2022; Accepted December 30, 2022

Abstract

Business intelligence (BI) integrates and analyzes the behavior of historical data streams, obtaining predictions. The project allows companies in the agricultural sector dedicated to the cultivation and export of avocados to make decisions based on artificial intelligence, promoting growth and competitiveness in the market. Models that apply data analytics are implemented through simple linear regression and recurrent neural networks (RNN). To carry out the project, data extraction, transformation, and loading (ETL) processes were used. The coding was developed in Python with the Django framework, using the sklearn, linear_model, LinearRegression, seaborn, stastsmodels and tensorflow libraries, among others. In the predictions for the years 2016 to 2021, greater precision was verified in the linear regression model. When making the export projection for the next six years, the results coincide with a minimal difference between the two models.

Export, Models, Neural, Regression

Resumen

La inteligencia de negocios (BI) integra y analiza el comportamiento de las secuencias de datos históricos obteniendo predicciones. El proyecto permite a las empresas del sector agrícola dedicadas al cultivo y exportación del aguacate tomar decisiones basadas en la inteligencia artificial impulsando el crecimiento y competitividad en el mercado. Se implementan modelos que aplican la analítica de datos a través de la regresión lineal simple y redes neuronales recurrentes (RNN). Para llevar a cabo el proyecto se emplearon procesos de extracción de datos, transformación y carga (ETL). La codificación se desarrolló en Python con el framework Django, empleando las librerías sklearn, linear_model, LinearRegression, seaborn, stastsmodels y tensorflow, entre otras. En las predicciones de los años 2016 al 2021, se comprobó mayor precisión en el modelo de regresión lineal. Al realizar la proyección de exportación de los próximos seis años, los resultados coinciden con una mínima diferencia entre los dos modelos.

Exportación, Modelos, Neuronal, Regresión

Citation: OCHOA-ORNELAS, Raquel & GUDIÑO-OCHOA, Alberto. Implementation of models to predict avocado exports. Journal-Mathematical and Quantitative Methods. 2022. 6-11:6-17.

* Author Correspondence (E-mail: raquel.oo@cdguzman.tecnm.mx)
† Researcher contributing as first author.

Introduction

Esguerra (2022) he comments in his research on the risks in the implementation of the FTA with the United States, which will culminate in the year 2030, where the commercialization of rice will be affected when the production of that country enters Colombia, without tariffs and competitive prices, since the productivity and agrarian policies favor the northern country more. Therefore, the global competition in this and other crops generates the essential need to implement predictive tools to reduce uncertainty through the Strategic Marketing Plan model, which structures the objectives, context, plan, strategies, programs, also including stages measurement, action and reaction and continuous improvement. This will allow reaching the sales targets in 2025, since the model is a useful tool to expand and diversify markets to other areas.

Martínez et al. (2022) they state that today the challenges are greater in organizations due to globalization, deregulation, changes in the environment, and technological acceleration, which are factors that require strategies to adapt to changes, creating innovative models to create advantages competitive.

Gomez (2021) they present a strategic business plan to promote commercial statistics linking avocado producers, and in this way, guarantee economic development, keeping an administrative scheme in force that allows overcoming any difficulty that prevents companies from achieving goals, developing tools that promote a level of competitiveness, increasing efficiency and control in their activities and achievement of objectives.

Amaya (2022) mentions that Colombia is one of the main Hass avocado producers, therefore, they consider that during avocado exports an effort must be made to position itself competitively and achieve high standards with strict control in the production and use of technologies. In relation to the contribution of water, it is necessary to implement sensors to monitor humidity, so through a study of dynamic evaluation of the soil, a simulation is implemented through the Hidrus-1D software.

Piñeros (2022) proposes the implementation of technology through the development of an artificial neural network in Python that can recognize the level of maturity of the uncut Hass avocado, identifying and evaluating its characteristics through images of each of its development stages, achieving the implementation. with more than 90% accuracy. This project allows farmers to benefit and not generate fruit waste.

Today it is essential for companies to articulate tools, techniques and predictive models that provide reliable and timely information to be used in decision making (Valderrama, 2022). Predictive models provide estimates to users to be prepared and establish future strategies to increase the profitability of a business (Campos, 2018). Data warehouse (DW) systems involve a set of technologies to record transactions, preserve and collect data, constituting a robust source of information to be used later in the analysis of the company's performance. The DW allow assisting decision making using relational database managers among them (Lund et al., 2021). The Association of Avocado Producers and Exporters of Jalisco (APEAJAL) has a DW that allows the registration and monitoring of avocado export certificates through a relational database stored in a cloud server managed with MySQL. The platform also generates reports, as well as projections on a variety of statistics available at different user levels to support decision making, includes the option of sending stored historical data to a CSV file that through Machine Learning (ML) models and RNN generates predictions. The data was grouped in the destinations Canada, Europe, and Japan in general by year and by each packer, according to the requirements established by APEAJAL. In the development of the application, tools such as Anaconda, Jupyter, Python and Django were used, presenting comparative graphs of each model in previously defined destinations: Canada, Europe, Japan, among others.

ISSN-On line: 2531-2979

Theoretical framework

Data science

BI contains a set of methodologies, processes, and technologies to collect, process, analyze and transform information, which can be used for strategic decision making in an organization, to achieve better performance, producing new business opportunities while maintaining a competitive advantage (Benítez, 2019). Data science is a tool for the exploitation of data, one of its main objectives is to find models and behaviors from the information to achieve predictions. Data science encompasses different fields of research in which areas such as computing, statistics, mathematics, among others, are involved, developing new algorithms and computing techniques for the capture, storage, and processing of information (García et al, 2018).

ETL processes are widely used in business intelligence. In the extraction stage, relevant information is collected from the different areas of an organization, it is very important in this section to extract significant data. In the transformation stage, a series of methods are carried out for the purification of the data, they are cleaned and transformed, guaranteeing the quality of the information. Finally, the last stage is the upload, where the data is joined and analyzed for decision making (Aguilar, 2019).

Data mining is an analysis process that uses mathematical, statistical, and artificial intelligence techniques, which aims to discover useful or valuable information to generate or analyze patterns that help an organization make decisions (Joyanes, 2015). Currently, different experimental models are available, such as linear regression methods, decision tree, random forest, SVM method, xgboost, logistic regression, naive Bayes classification (Valderrama, 2022).

Used tools

Anaconda is free software that provides a set of tools designed for research and science. Its installation provides access to different environments allowing coding in Python languages, with thousands of packages and open-source libraries (García et al., 2018).

Jupyter Notebook is a web-based integrated development environment (IDE) that uses a default browser, integrates text and code output, and visualizations in an immediate, organized, and clear manner (Rolon et al., 2016).

Machine learning

Machine learning has been used for two decades in the analysis and modeling of data through the implementation of algorithms using different types of programming languages. The Python scikit-learn machine learning package, used in data science, includes implementations with methods of modeling procedures and unified data, applicable for educational and behavioral statistics (Hao & Ho, 2019).

Other Python libraries

The matplotlib library is essential to obtain information about the underlying structure of the data set through its visualization, since it allows the scientist to detect the learning algorithm that is appropriate for the data set under study. In addition, it generates ideas of the transformations to be applied to the data set in an appropriate way (Bisong, 2019a). The matplotlib library is considered for the creation and visualization of graphs based on data, which contains graphs of lines, bars, scatter boxes, among others. It has a great adaptation to styles and custom annotations allowing developers to create any type of visualization (Yim & Yu, 2018). The numpy library facilitates scientific computing using multidimensional arrays, offers an efficient and fast data structure since it uses cache memory (Harris et al., 2020). The pandas library allows the import of data from xlsx files to a DataFrame, storing the columns in a list (Llerena, 2020).

Linear regression

Linear regression analysis is a statistical technique used to study and model relationships between variables. Regression analysis is widely used in data mining and has become an indispensable tool in data science and analysis because it can be applied to a wide variety of problems (Montgomery et al., 2021).

Jas et al. (2020) comment that generalized linear models (GLM) are tools for regression and classification applied in science, economics, business, and finance. To avoid overfitting the data, it is essential to regularize the model by adding penalty terms. The tools available in the Python data science realm do not offer all these capabilities:

- statsmodels offers a wide range of distributions but no regularization.
- scikit-learn provides elastic network regularization but only limited noise distribution options.
- lightning provides regularization, but only for linear (Gaussian) and logistic (binomial) regression.

Montgomery et al. (2021) describes the linear regression model with the following equation:

$$y = \beta 0 + \beta 1 x \quad (1)$$

Where:

$\beta 0$ is the intercept,

$\beta 1$ is the slope,

y is the dependent variable and

x is the independent variable.

Molina (2021) explains that the minimum method consists of minimizing the residuals between the real values and those estimated by the linear regression line. The method allows us to calculate the line that is closest on average to all the points. The following equation considers a deterministic component and a stochastic component where the error estimate is included:

$$y_i = \beta_0 + \beta_1 X_i + e_i \quad (2)$$

Madroñero et al. (2021) explain that to validate the parameters with the method of least squares, a statistical analysis is carried out, it is important to consider the coefficient of determination, standard deviation, and the t-student test, among others.

In addition, they determine that to obtain the values of $\beta 0$ and $\beta 1$ it is necessary to use the calculation of the method of least squares to generate values that minimize the expression through the following formulas:

$$\beta_1 = \frac{\sum(x_i - \bar{x})(yi - \bar{y})}{\sum(xi - \bar{x})^2} \quad (3)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (4)$$

Where:

x_i = Value of the independent variable.

$\bar{x}$ = Mean of the independent variable.

$\bar{y}$ = Mean of the dependent variable.

n = Total number of observations.

Neural networks

Neural networks mimic the way neurons work in humans. In the 1940s, Warren McCulloch and Walter Pitts proposed electrical and mathematical models of neural networks. Donald Hebb, at the end of the decade, explained that it is precisely in the synapse or connection between neurons where learning takes place, since the information is represented in the brain in a set of active or inactive neurons. The previous rules are applied in the artificial models. In 1956, the first conference on artificial intelligence was presented in the city of Dartmouth where it was stated that, in the human brain, each of the neurons generates an electrical impulse by connecting with other neurons through their dendrites. Neurotransmitters regulate the impulse between the axonal branching and the dendrite, having a negative or positive effect on the neuron that receives several signals at the same time by combining them, generating a level of stimulation or activation, emitting electrical signals through impulses with intensity and frequency, called firing rate (Antona, 2017). Artificial neural networks are a class of machine learning models inspired by studies of the human central nervous system, which are made up of several interconnected neurons organized in layers. The neurons of one layer send information to the neurons of the next layer and in this way, the network calculates the results (Gulli et al., 2019).

Artificial neural networks are currently implemented in different sectors, since they are trained through machine learning algorithms to solve data prediction problems (Antona, 2017).

A neural network is a simplified model of how information processing works in the human brain. It considers simultaneously interconnected processing units that are organized in layers, similar to abstract versions of neurons. Layers are organized into an input layer, one or more hidden layers, and an output layer. The units are connected where the input data is in the first layer and the values are propagated to each neuron in the next layer, sending the result to the output layer. As individual records are examined, the network learns by generating a prediction for each record, adjusting. The process is repeated countless times, thus improving the predictions, achieving stopping criteria through training (IBM, 2021). Figure 1 shows the structure of a neural network:

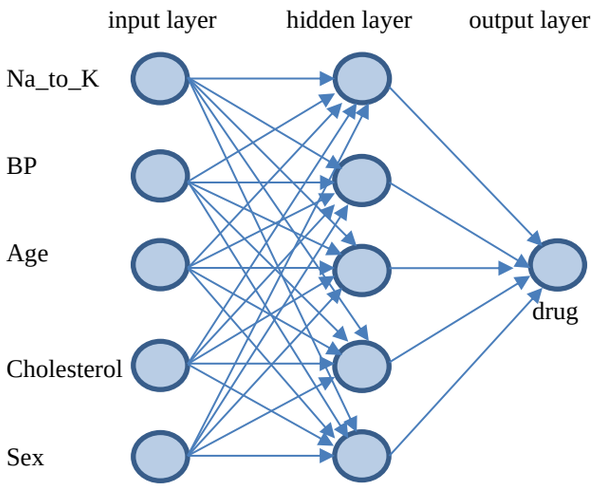

Figure 1 Structure of a neural network
Source: IBM 2021

Recurrent Neural Networks (RNNs) solve learning problems where historical events are related to make future predictions, such as language modeling, stock market prediction, stock prices, and future movements. RNNs fit a time series or task sequence, that is, feedback from an input sequence, allowing the network to independently incorporate this sequence knowledge when making the prediction. This type of recurrent neural network is completely different from other artificial neural network architectures, since it maintains a memory or state of previous calculations (Bisong, 2019b).

Caja (2020) comments that the most used libraries for machine learning are: tensorflow and keras. The tensorflow library is an open source deep learning library created by Google. Unlike other libraries dedicated to learning, it is designed to be used for research and development (Brownlee, 2016). Each data element is known as a tensor, which is a representation of vectors and matrices in higher dimensions. The ranks of each tensor indicate its dimensions (Moocarme, 2020).

In the process, tensorflow uses keras as an API, commonly known as tf.keras. In the tf.keras library, the layers are connected to each other and, as a result, it provides a clean and easy to understand model. The keras library is open source, developed in Python in 2015 by Google Francois Chollet. It has a simple application programming interface (API) that allows modeling of neural systems. Google's tensorflow library integrates its keras core as a frontend API for defining and modeling neural networks. In tensorflow version 2.0, keras becomes a high-level API. One of the major drawbacks is that the keras library contains a complex level of abstraction that makes it difficult to learn. The advantage lies mainly in the fact that the development is light in lines of code and very intuitive (Caja, 2020).

The keras library makes developing deep learning models as fast and easy as possible (Brownlee, 2016). It also helps to create a neural network, implementing models that contain input data to obtain output data (Torres, 2020).

Using keras, productivity is increased by saving time in code implementation, focusing its efforts on deep learning algorithms (Atienza, 2020).

Epoch

It is the total number of times that the forward propagation and backpropagation algorithm is executed. The training data goes through each cycle to the learning neural network (Ibañez, 2020).

Methodology

The project was carried out in APEAJAL, the data source was obtained from a MySQL database stored in the cloud and generated through a monitoring platform for export certificates from the state of Jalisco. Through ETL processes, the input values were obtained, classifying the data for each model in the groups of agreed destinations. In this way, once the models were implemented, the predictions for the years 2016 to 2021 were obtained, which were compared with real data, as well as the prediction for the years 2022 to 2027.

In the first instance, the algorithms of each model were executed in Anaconda and Jupyter using the Python programming language. Once the results were verified, a Python application was developed in the Django framework to display the predictions in an organized way in a user interface that must be authenticated through a login.

For the linear regression model, the prediction from 2016 to 2021 was obtained using the libraries in Figure 2:

```
from sklearn.linear_model import LinearRegression
import numpy as np
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
```

Figure 2 Regression model Libraries

Figure 3 shows the Python implementation for the linear regression model:

```
objetivo="VENTAS_REALES"
independientes=datos_numericos.drop(columns=objetivo).columns
# Crear el modelo y ajustarlo
modelo= LinearRegression()
modelo.fit(X=datos_numericos[independientes], y=datos_numericos[objetivo])
# Agregamos las predicciones como una nueva columna del dataset original
datos["REGRESION_LINEAL"]=modelo.predict(datos_numericos[independientes])
```

Figure 3 Linear regression 2016 to 2021

To obtain the prediction from 2022 to 2027 in linear regression, the libraries of Figure 4 were used.

```
from sklearn import linear_model
from sklearn.metrics import
mean_squared_error, r2_score
```

Figure 4 Linear regression prediction

Figure 5 shows the implementation of the libraries for the 2022 to 2027 prediction with the linear regression model:

```
regresion_lineal = linear_model.LinearRegression()
regresion_lineal.fit(x.reshape(-1,1), y)
print('\nParámetros del modelo de regresión')
print('b (Pendiente) = ' + str(regresion_lineal.coef_) + ',
a (Punto de corte) = '
                    + str(regresion_lineal.intercept_))
pronostico = regresion_lineal.predict(z.reshape(-1,1))
```

Figure 5 Linear regression 2022 to 2027

For the neural network model, the library of Figure 6 was used:

```
import tensorflow as tf
```

Figure 6 Neural network

Figure 7 shows the training of the neural network model assigning 70,000 epochs:

```
print ("Seleccionando las columnas")
X_train = ventas_df['PERIODO']
y_train = ventas_df['VENTAS']#Creando el Modelo
print ("Creando el modelo")
model = tf.keras.Sequential()
model.add(tf.keras.layers.Dense(units=1,
input_shape=[1]))
#Compilado
print ("Compilando el modelo")
model.compile(optimizer=tf.keras.optimizers.Adam(1),
loss='mean_squared_error')
#Entrenando el modelo
print ("Entrenando el modelo")
epochs_hist = model.fit(X_train, y_train, epochs =70000);
#Evaluando modelo
print ("Evaluando el modelo entrenado")
print ("Keys:")
print(epochs_hist.history.keys())
```

Figure 7 Training the neural network model

During the implementation, several trainings were carried out with different values in epochs, using a fixed set of values, to measure the precision, learning curve and training time, and thus be able to determine the optimal variation for the model. By increasing the number of units in the layers of the neural network, better results were obtained, however, the training time was noticeably affected. The learning curve shows the level of learning success over time. Figure 8 shows the training graph generated with 70,000 epochs:

Figure 8 Training curve 70,000 epochs

Figure 9 presents a partial code to obtain the 2016 to 2021 forecast through the neural network model:

```
for contador in range (6):
        #Realizar una predicción utilizando el modelo entrenado
        Venta_C = contador+1
        Venta_F = model.predict([Venta_C])
          #red neuronal
```

Figure 9 Neural network prediction 2016 to 2021

Subsequently, the code was implemented to obtain the prediction for 2022 to 2027, see Figure 10:

```
for contador in range (7,13):
        #Realizar una predicción utilizando el modelo entrenado
        Venta_C = contador
        Venta_F = model.predict([Venta_C])
          #red neuronal
```

Figure 10 Neural network prediction 2022 to 2017

Results

To implement and test the prediction models, data on exports from the destination Japan were obtained during the years 2016 to 2021. Comparing the results of the predictions of the linear regression model and the neural network model, it was verified that in the first and fourth year the predictions are close to the real data of exports, having a variation of approximately two tons of difference in the last two years. See Figure 11:

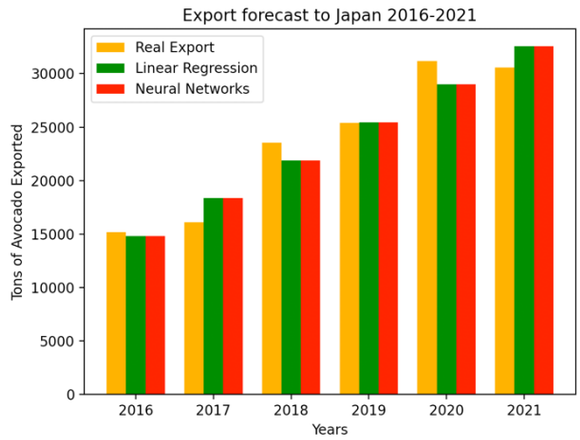

Figure 11 Predictions 2016 to 2021 Japan

In relation to the results obtained from the predictions of the two models, results with minimal amounts of difference are perceived. Table 1 shows the calculated predictions.

Years	Real export tons	Linear regression	Neural network
2016	15182	14807.333333	14807.748047
2017	16115	18355.133333	18355.619141
2018	23556	21902.933333	21903.488281
2019	25417	25450.73333	25451.359375
2020	31194	28998.533333	28999.230469
2021	30597	32546.333333	32547.101562

Table 1 Comparative Real Data – Prediction 2016 to 2021

The results of the 2022 to 2027 prediction obtained from the linear regression models and neural networks are presented in Figure 12:

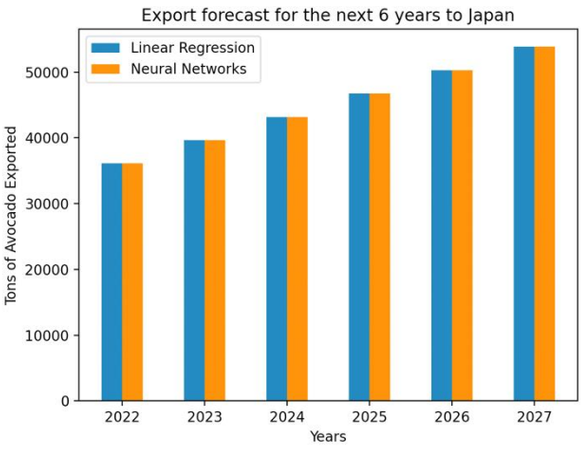

Figure 12 Prediction 2022 to 2027 Japan

The amounts generated by each of the models in the period 2022 to 2027 are presented in Table 2:

Years	Linear regression	Neural network
2022	36094.133333	36094.968750
2023	39641.933333	39642.839844
2024	43189.733333	43190.710938
2025	46737.533333	46738.582031
2026	50285.333333	50286.453125
2027	53833.133333	53834.324219

Table 2 Forecast results 2022 to 2027 linear regression - neural networks

Acknowledgements

Thanks to APEAJAL for the support in the development of the project, providing the necessary information for the implementation of the models.

Financing

This work has been financed by TecNM [15269.22-P].

Conclusions

Training time is an important factor when implementing an artificial neural network, since the amounts of units in the layers, the number of layers, and the value of the epochs play an important role. Regarding the precision of the model, a comparison is made between the linear regression models and neural networks, where a minimum difference can be observed in the results obtained, and an approximation in the comparison with the real data.

The project meets the requirements proposed by the association, proposing the results of two prediction models that were compared with real data and manual calculations, in the case of the linear regression model. The project offers APEAJAL the automatic generation of future predictions, obtaining the input data automatically through ETL processes from a reliable and constantly fed database during the registration and monitoring of export certificates. The application offers the generation of predictions for the next six years in a general way or by packer in the different destinations proposed by the association: Canada, Europe, Japan, and others.

References

Antona Cortés, C. (2017). *Herramientas modernas en redes neuronales: la librería Keras* (Bachelor's thesis). URL: http://hdl.handle.net/10486/677854

Aguilar, L. J. (2019). Inteligencia de negocios y analítica de datos: una visión global de business intelligence & analytics. Alpha Editorial. URL: https://books.google.es/books?hl=es&lr=&id=ifR5EAAAQBAJ&oi=fnd&pg=PR7&dq=Inteligencia+de+negocios+y+anali%CC%81tica+de+datos:+una+visio%CC%81n+global+de+business+intelligence+%26+analytics&ots=bBjkKLrG-b&sig=7L5GuDH2yqWhzNYQ2uyC3IRLl-w#v=onepage&q=Inteligencia%20de%20negocios%20y%20anali%CC%81tica%20de%20datos%3A%20una%20visio%CC%81n%20global%20de%20business%20intelligence%20%26%20analytics&f=false

Amaya, K. L. L. (2022). Presentan evaluación de la dinámica de agua en Hydrus-1D para tres suelos de ladera cultivados con aguacate Hass (persea americana) en el Valle del Cauca, Colombia. Investigación e Innovación en Ingenierías, 10(1), 75-98. URL: https://revistas.unisimon.edu.co/index.php/innovacioning/article/view/4811 DOI: 10.17081/invinno

Atienza, R. (2020). Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, deep RL, unsupervised learning, object detection and segmentation, and more. Packt Publishing Ltd. URL: https://books.google.es/books?hl=es&lr=&id=68rTDwAAQBAJ&oi=fnd&pg=PP1&dq=Advanced+Deep+Learning+with+TensorFlow+2+and+Keras:+Apply+DL,+GANs,+VAEs,+deep+RL,+unsupervised+learning,+object+detection+and+segmentation,+and+more&ots=ttaE9FhhGI&sig=XXTcWhBNT55aEqElkh4EwbATxN4#v=onepage&q=Advanced%20Deep%20Learning%20with%20TensorFlow%202%20and%20Keras%3A%20Apply%20DL%2C%20GANs%2C%20VAEs%2C%20deep%20RL%2C%20unsupervised%20learning%2C%20object%20detection%20and%20segmentation%2C%20and%20more&f=false

Benítez, Y. L. (2019). Business intelligence. ADGG102PO. IC Editorial. URL: https://books.google.es/books?hl=es&lr=&id=oEwpEAAAQBAJ&oi=fnd&pg=PT5&dq=Ben%C3%ADtez,+Y.+L.+(2019).+Business+intelligence.+ADGG102PO.+IC+Editorial.&ots=ga_1Z0DxLC&sig=Dt0kanGjRGoR0LfsIlNnUZk6dII#v=onepage&q&f=false

Bisong, E. (2019a). Matplotlib and seaborn. In *Building machine learning and deep learning models on google cloud platform* (pp. 151-165). Apress, Berkeley, CA. URL: https://link.springer.com/chapter/10.1007/978-1-4842-4470-8_12. DOI: 10.1007/978-1-4842-4470-8_12

Bisong, E. (2019b). *Building machine learning and deep learning models on Google cloud platform* (pp. 59-64). Berkeley, CA: Apress. URL: https://www.oreilly.com/library/view/building-machine-learning/9781484244708/

Brownlee, J. (2016). Deep learning with Python: develop deep learning models on Theano and TensorFlow using Keras. Machine Learning Mastery. URL: https://books.google.es/books?hl=es&lr=&id=K-ipDwAAQBAJ&oi=fnd&pg=PP1&dq=Brownlee,+J.+(2016).+Deep+learning+with+Python:+develop+deep+learning+models+on+Theano+and+TensorFlow+using+Keras.+Machine+Learning+Mastery.&ots=oqVwUQYrxu&sig=sIut9naUAkswILH0N4b0t5tLg8Y#v=onepage&q=Brownlee%2C%20J.%20(2016).%20Deep%20learning%20with%20Python%3A%20develop%20deep%20learning%20models%20on%20Theano%20and%20TensorFlow%20using%20Keras.%20Machine%20Learning%20Mastery.&f=false

Caja García, O. (2020). *Librería Python para el aprendizaje y la implementación de redes neuronales* (Doctoral dissertation, Universitat Politècnica de València). URL: https://riunet.upv.es/handle/10251/152226

Campos Guillen, A. L. (2018). Modelamiento matemático predictivo del comportamiento de la exportación de aguaymanto (physalis peruviana L.) fresco del Perú a un país destino frente a factores externos e internos. URL: https://hdl.handle.net/20.500.12759/4374

Esguerra Álvarez, J. C. (2022). Plan de marketing para la distribución de insumos agrícolas de Sucampo en el departamento del Cesar (Doctoral dissertation, Universidad del Rosario). URL: https://repository.urosario.edu.co/handle/10336/34445 DOI: https://doi.org/10.48713/10336_34445

García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, A., & Padilla, W. (2018). Ciencia de datos. Técnicas Analíticas y Aprendizaje Estadístico. Bogotá, Colombia. Publicaciones Altaria, SL. URL: https://d1wqtxts1xzle7.cloudfront.net/64031156/Ciencia_de_datos_2018-with-cover-page-v2.pdf?Expires=1669010617&Signature=GdmnvC~BiRt~jybQkBEgwalJ69Br6JrmrjJJ1TdB1FHqoJ5~hg5oRMJhdnGp3YvJUb0xnu2VZr-tE80fxpzPx~wnUdFx6A0VuOH0FkhR7d51IoCbjcd8JLxIkPjbohzS50BD~ftZ53oqQW6thN5HSQBT50Fi3MxXrqDNIw7-5qosFXFaQWWJCTrvLVw2XHebOaOoFktCwwXLMp9mA7GNijQ2lO2t1y-ccXIvqSfmNLFdcssGGUgyyafT~04VdMs~2uyo4ir-M~6fySV8hkPNfKGyl7G6T0t4hy7zkZtfsmgx3zmRHG84ZDmNzQ1HJx~add8k0PvtnGorsxIx3NdaPg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Gomez Bolivar, S. P. (2021) Plan Estratégico para el Fortalecimiento de la Estructura Organizacional de la Asociación ASPAVIC y así Mejorar la Comercialización del Aguacate en el Municipio de la Victoria Caldas en el Año 2021. URL: https://repository.unad.edu.co/handle/10596/48127

Gulli, A., Kapoor, A., & Pal, S. (2019). Deep learning with TensorFlow 2 and Keras: regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API. Packt Publishing Ltd. URL: https://books.google.es/books?hl=es&lr=&id=BVnHDwAAQBAJ&oi=fnd&pg=PP1&dq=Gulli,+A.,+Kapoor,+A.,+%26+Pal,+S.+(2019).+Deep+learning+with+TensorFlow+2+and+Keras:+regression,+ConvNets,+GANs,+RNNs,+NLP,+and+more+with+TensorFlow+2+and+the+Keras+API.+Packt+Publishing+Ltd.&ots=K_xc5sTxV_&sig=VMM9qdiCKJwgNMsYuBqlCVCT_TA#v=onepage&q=Gulli%2C%20A.%2C%20Kapoor%2C%20A.%2C%20%26%20Pal%2C%20S.%20(2019).%20Deep%20learning%20with%20TensorFlow%202%20and%20Keras%3A%20regression%2C%20ConvNets%2C%20GANs%2C%20RNNs%2C%20NLP%2C%20and%20more%20with%20TensorFlow%202%20and%20the%20Keras%20API.%20Packt%20Publishing%20Ltd.&f=false

Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics*, *44*(3), 348-361. URL: https://journals.sagepub.com/doi/abs/10.3102/1076998619832248. DOI: https://doi.org/10.3102/1076998619832248

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357-362. URL: https://www.nature.com/articles/s41586-020-2649-2. DOI: https://doi.org/10.1038/s41586-020-2649-2

Ibáñez, L. J. (2020). Arquitectura de Red Neuronal para el Desarrollo de Agentes Conversacionales destinados a la Atención al Cliente en las Redes Sociales. *Ciencia y tecnología*, (20), 37-53. URL: https://dialnet.unirioja.es/servlet/articulo?codigo=7763843

IBM (2021). El modelo de redes neuronales. Tomado de URL: https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model el 17 de Noviembre de 2022

Jas, M., Achakulvisut, T., Idrizović, A., Acuna, D., Antalek, M., Marques, V., ... & Ramkumar, P. (2020). Pyglmnet: Python implementation of elastic-net regularized generalized linear models. *Journal of Open Source Software*, *5*(47). URL: http://hdl.handle.net/1854/LU-8650816. DOI: 10.21105/joss.01959

Joyanes, L. (2015). Sistemas de información en la empresa. Alpha Editorial. URL: https://books.google.es/books?hl=es&lr=&id=oHNxEAAAQBAJ&oi=fnd&pg=PR5&dq=Joyanes,+L.+(2015).+Sistemas+de+informaci%C3%B3n+en+la+empresa.+Alpha+Editorial&ots=dcH4u8n4QL&sig=W4AMQQox6mo3VuTcYP7rp0RE8O4#v=onepage&q&f=false

Llerena Izquierdo, J. (2020). Codifica en Python. URL: https://dspace.ups.edu.ec/handle/123456789/19346.

Lund, M. I., Migani, S. I., Vera, C., Orellana Vasallo, A., Gómez, A. M., Pinto, S. E., ... & Molinari, M. L. (2021). Inteligencia y analítica de negocios para la toma de decisiones en diferentes contextos. In *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja)*. URL: http://sedici.unlp.edu.ar/handle/10915/120107

Madroñero, D. M., Mondragón, E. I., & Vergel-Ortega, M. (2021). Análisis estadístico para validar parámetros de modelos matemáticos por medio método de mínimos cuadrados. *Revista Boletín Redipe*, *10*(5), 343-359. URL: https://revista.redipe.org/index.php/1/article/view/1309/1221. DOI: DOI**:** https://doi.org/10.36260/rbr.v10i5.1309

Martínez-Arroyo, J. A., Valenzo-Jiménez, M. A., & Kido-Cruz, A. (2022). Gestión Ambiental, Innovación y Ventaja Competitiva en Empresas Exportadoras de Aguacate. Investigación administrativa, 51(129). URL: https://www.ipn.mx/assets/files/investigacion-administrativa/docs/revistas/129/art1.pdf DOI: https://doi.org/10.35426/iav51n129.01

Molina, M. (2021). La distancia más corta. El método de los mínimos cuadrados. *Revista electrónica AnestesiaR*, *13*(1), 3.

Montgomery D. C., Peck E. A., & Vining G. G. (2021). Introduction to linear regression analysis. John Wiley & Sons. URL: https://books.google.es/books?hl=es&lr=&id=tCIgEAAAQBAJ&oi=fnd&pg=PR13&dq=Montgomery+D.+C.,+Peck+E.+A.,+%26+Vining+G.+G.+(2021).+Introduction+to+linear+regression+analysis.+John+Wiley+%26+Sons.&ots=lfxgSpgYNq&sig=fSUJ2oiWVIBUBMS2adq7o2-XDZQ#v=onepage&q=Montgomery%20D.%20C.%2C%20Peck%20E.%20A.%2C%20%26%20Vining%20G.%20G.%20(2021).%20Introduction%20to%20linear%20regression%20analysis.%20John%20Wiley%20%26%20Sons.&f=false

Moocarme, M., Abdolahnejad, M., & Bhagwat, R. (2020). The Deep Learning with Keras Workshop: Learn how to define and train neural network models with just a few lines of code. Packt Publishing Ltd. URL: https://books.google.es/books?hl=es&lr=&id=XxL0DwAAQBAJ&oi=fnd&pg=PP1&dq=Moocarme,+M.,+Abdolahnejad,+M.,+%26+Bhagwat,+R.+(2020).+The+Deep+Learning+with+Keras+Workshop:+Learn+how+to+define+and+train+neural+network+models+with+just+a+few+lines+of+code.+Packt+Publishing+Ltd.&ots=tTJVBp-qqS&sig=Lmt6O45XKAXTR4Cvj1gBf1jiA2U#v=onepage&q&f=false

Piñeros Guzmán, J. (2022). Clasificación de la madurez del aguacate Hass para su cosecha mediante una red ANN. Universidad de los Andes. URL: http://hdl.handle.net/1992/58801

Rolon-Mérette, D., Ross, M., Rolon-Mérette, T., & Church, K. (2016). Introduction to Anaconda and Python: Installation and setup. Python for research in psychology, 16(5), S5 S11. URL: https://www.tqmp.org/SpecialIssues/vol16-5/S003/

Torres, J. (2020). Python deep learning: introducción práctica con Keras y TensorFlow 2. Marcombo. URL: https://books.google.es/books?hl=es&lr=&id=aExOEAAAQBAJ&oi=fnd&pg=PT14&dq=Torres,+J.+(2020).+Python+deep+learning:+introducci%C3%B3n+pr%C3%A1ctica+con+Keras+y+TensorFlow+2.+Marcombo.&ots=5HM1Q7WAsz&sig=AOi8_mO0dH83sdvm3L1-WvAZAj4#v=onepage&q=Torres%2C%20J.%20(2020).%20Python%20deep%20learning%3A%20introducci%C3%B3n%20pr%C3%A1ctica%20con%20Keras%20y%20TensorFlow%202.%20Marcombo.&f=false

Yim, A., Chung, C., & Yu, A. (2018). Matplotlib for Python Developers: Effective techniques for data visualization with Python. Packt Publishing Ltd. URL: https://books.google.es/books?hl=es&lr=&id=G99YDwAAQBAJ&oi=fnd&pg=PP1&dq=Yim,+A.,+Chung,+C.,+%26+Yu,+A.+(2018).+Matplotlib+for+Python+Developers:+Effective+techniques+for+data+visualization+with+Python.+Packt+Publishing+Ltd.&ots=tyb8pv7M2g&sig=Xfp_R_MvAvQ05xXploZhVc5ls64#v=onepage&q=Yim%2C%20A.%2C%20Chung%2C%20C.%2C%20%26%20Yu%2C%20A.%20(2018).%20Matplotlib%20for%20Python%20Developers%3A%20Effective%20techniques%20for%20data%20visualization%20with%20Python.%20Packt%20Publishing%20Ltd.&f=false

Valderrama Mundaca, O. E. (2022). Solución de inteligencia de negocios basada en técnicas de minería de datos, para apoyar la toma de decisiones, en la Gerencia Regional de Agricultura-Lambayeque. URL: http://hdl.handle.net/20.500.12423/5047.