***Start-Up Agrícola***

1. ***Introducción***

Con la necesidad de satisfacer la demanda de productos agrícolas, la compañía dentro de sus prioridades de negocio busca distribuir de manera eficiente una variedad de productos a usuarios finales.

Por consiguiente, para minimizar perdidas de productos no comercializados y/o estropeados tenemos la necesidad de diseñar, construir e implementar un modelo predictivo de demanda de productos agrícolas nativos.

Así, gestionar de manera eficiente los niveles de inventario de cada producto y disminuir los niveles de almacenamiento reduciendo el desperdicio alimenticio e incrementando los niveles de frescura de dichos productos.

En adición, se considera evaluar dentro de la gestión de inventario hábitos de consumo para optimizar la operación y distribución de productos como estrategia de negocio.

1. ***Pregunta de negocio***

Por lo tanto, desde un enfoque analítico se desea una aproximación a nuestra investigación respondiendo ¿Existirá información histórica suficiente para identificar estacionariedad estadística, identificar correlación entre eventos, encontrar patrones de consumo, identificar estacionalidad, que permitan desarrollar un modelo para pronosticar los niveles de demanda por producto a nivel semanal?

1. ***Metodología propuesta***

En esta sección desglosaremos un poco el enfoque que se desarrolló en la construcción de esta solución. Inicialmente la metodología a seguir está estrechamente relacionada con los hallazgos encontrados en la parte de exploración.

Es necesario implementar actividades de identificación de valores faltantes para evaluar alguna metodología de imputación. Seguidamente se ejecutan gráficas y procesos de identificación de observaciones atípicas en la serie de datos a nivel diaria, semanal y mensual para evitar un sesgo mayor o inferior a algún nivel de predicción. Ejecutamos el método *IsolationForest* de *sklearn*, algoritmo de detección de valores atípicos basado en árboles. Este algoritmo es adecuado a el problema que estamos analizando ya que evalúa minoría de observaciones e identifica observaciones con atributos diferentes a instancias normales.

Posteriormente partimos de un análisis descriptivo para intentar conocer a nivel general o por cliente, patrones de consumo y priorizar los niveles de predicción a productos y/o clientes con mayor participación en la compañía.

Realizamos una generación de variables sintéticas para añadir un poco más de contexto, demarcando los fines de semanas, días de semanas, días del mes, fin de mes, y para un enfoque inicial, se realizó un scalling sobre las variables Pedidos y Precio, junto con un *labelling* sobre los productos todo esto para ser incluido como entrada a las Red LSTM.

Por consiguiente, evaluamos el desempeño de un modelo ARIMA y posteriormente un modelo utilizando el algoritmo Prophet.

Entrenamos los modelos con el 90% de las semanas y estimaremos el 10% aplicando un ciclo de Rolling forecast.

En el proceso de Rolling forecast se estima la primera semana y se incluye lo observado de dicha semana en nuestra base de entrenamiento para la estimación de la segunda semana, se itera de la misma manera hasta obtener la predicción total del 10% de las semanas a estimar.

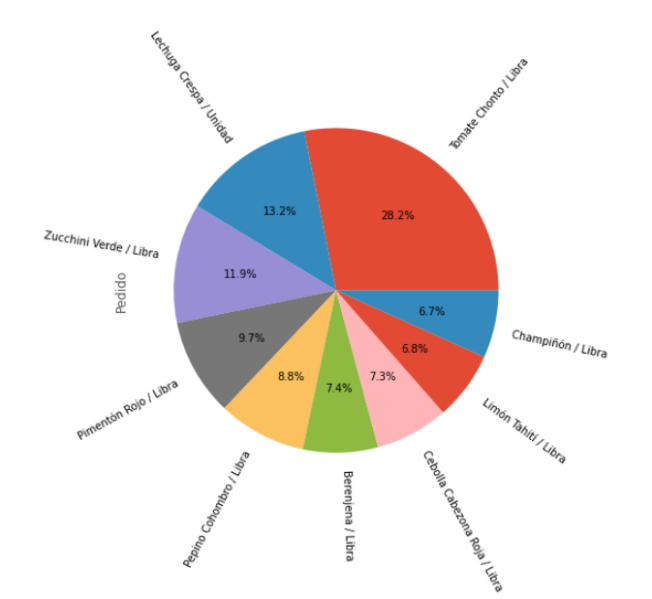
Debido al alto número de productos a predecir, inicialmente la predicción se realizará para la totalidad de productos sin identificación del tipo de producto ni el cliente, esto con el objetivo de tener lecturas totales y a partir de los resultados tomar decisiones que prioricen futuras predicciones por producto y permita identificar el modelo con mejor desempeño.

Evaluamos el desempeño de los modelos utilizando la métrica MSE.

Con el objetivo de acotar la cantidad de productos a predecir identificamos los clientes con mayor participación en el volumen de pedidos histórico, y así identificar los productos con mayor número de pedidos para los clientes identificados anteriormente con mayor participación.

Finalmente, ejecutamos el proceso de predicción de demanda con los modelos ARIMA y Prophet por producto.

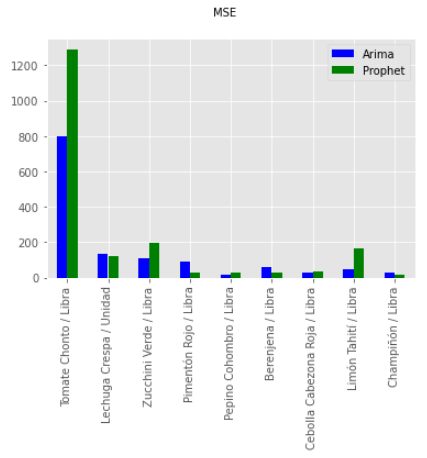
1. ***Resultados y análisis de los resultados***

Se seleccionaron 9 productos con mayor participación, representan el 53.4% del total de pedidos en 12 clientes con participación del 90.7%

*Figura 1. Productos que recogen el mayor impacto hacia la compañía.*

En el Anexo 3 se incluyen gráficas *Boxplot* por semana correspondiente a los 9 productos donde se analiza la distribución histórica de pedidos. Adicionalmente se incluyen los valores predichos con ARIMA y Prophet versus valores observados por producto, para el 10% de los periodos aplicando Rolling forecast.

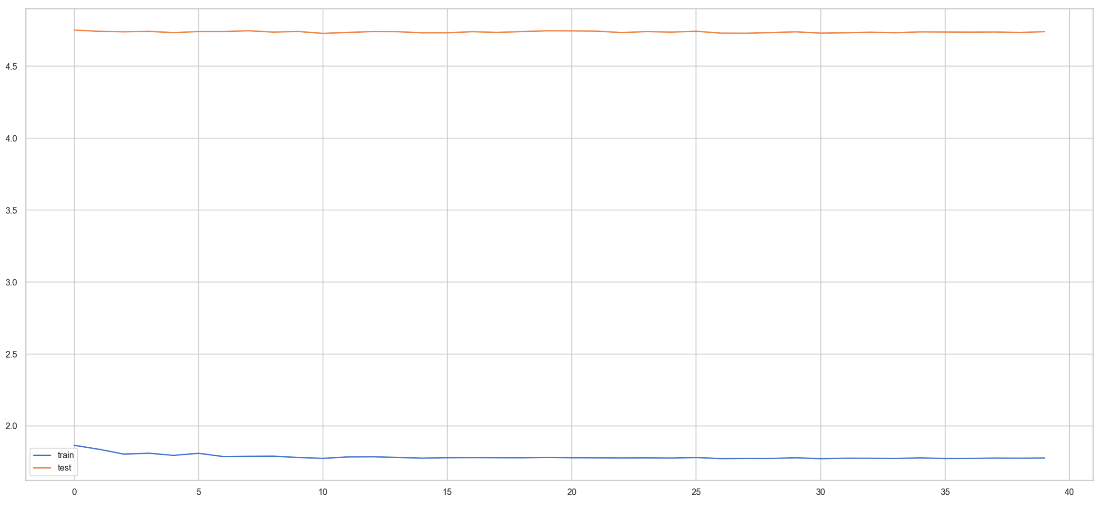
Se observa que para el producto “Tomate chonto” el modelo ARIMA tiene mejor desempeño, para los productos restantes las diferencias no son tan marcadas, con Prophet se logra menor MSE en 4 productos.



*Figura 2. Gráfica comparativa de rendimiento para los dos modelos con mejores resultados*

Es importante resaltar que ambos algoritmos tienen predicciones similares al analizar los resultados por producto.

En el anexo 4 se incluyen graficas de resultados de predicción para un nuevo periodo. Las cuales corresponden a valores predichos utilizando igualmente el modelo Arima y el algoritmo Prophet, las predicciones corresponden al mismo periodo.

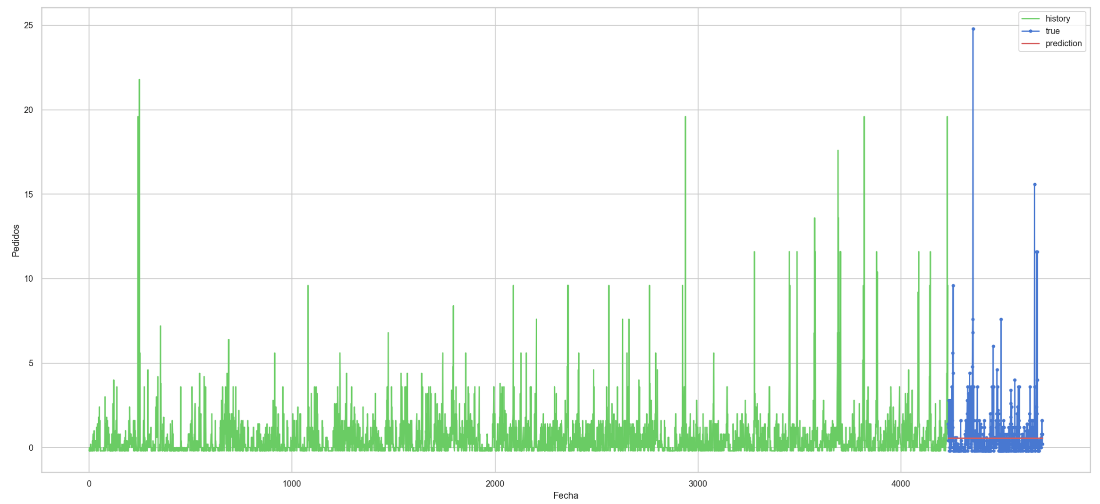


*Figura 3. Resultado de entrenamiento de una LSTM Bidirectional*

Intentando determinar que, si una red neuronal LSTM tenía un comportamiento aceptable sobre los datos de pruebas, se implementó una red con la siguiente configuración.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| Capa Bidireccional LSTM | Units = 10 | Input Shape= (4232,130) |  |
| Capa DropOut |  |  | Rate=0.2 |
| Capa Dense | Units = 1 |  |  |
| Loss = mean\_squared\_error |  | Optimizer = adam |  |

Los resultados obtenidos no fueron concluyentes, la red no logró incorporar valor explicativo sobre los datos expuestos, el resultado se refleja claramente sobre la predicción.



*Figura 4. Predicción de la LSTM Bidireccional*

1. ***Conclusiones***

Los dos algoritmos (ARIMA & Prophet) tienen predicciones similares, para la mayoría de los casos la predicción de Prophet se encuentra por debajo que la predicción usando Arima. Con lo que se interpretaría una posición estratégica conservadora que soporta la toma de decisiones para la compañía agrícola.

En todo caso no se da preferencia a ningún algoritmo en especial, se recomienda a la compañía agrícola el uso de los dos para estimar 12 periodos siguientes, que equivalen a 3 meses aproximadamente y nuevamente evaluar su desempeño.

Posteriormente a una posible o no preferencia entre los algoritmos, como recomendación y posibles extensiones, se sugiere la inclusión de nuevos productos de acuerdo con priorización a productos con mayor participación.

***Referencias***

* [Isolation Forest](https://ieeexplore.ieee.org/abstract/document/4781136), 2008.

… our proposed method takes advantage of two anomalies’ quantitative properties: i) they are the minority consisting of fewer instances and ii) they have attribute-values that are very different from those of normal instances.

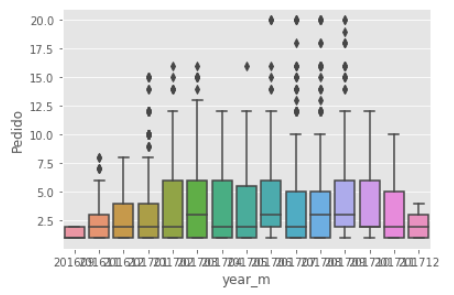
* 4 Automatic Outlier Detection Algorithms in Python by Jason Brownlee on July 8, 2020 in Data Preparation

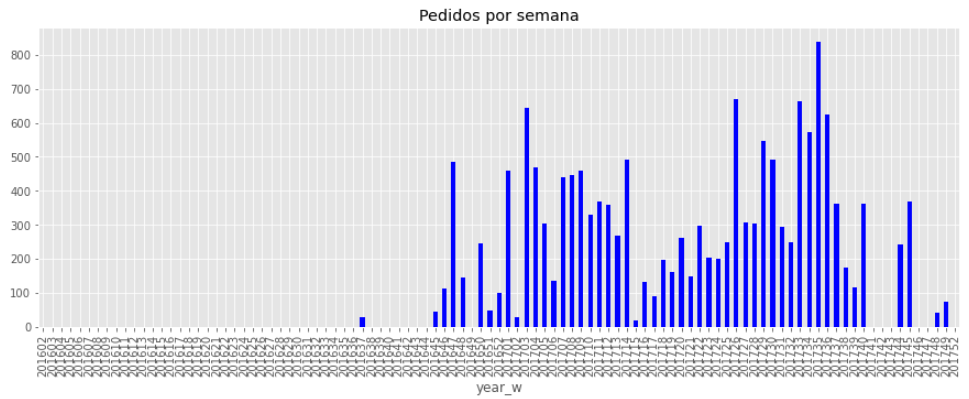
https://machinelearningmastery.com/model-based-outlier-detection-and-removal-in-python/

***Anexo 1.***

Descriptivos – consolidado total de pedidos para el total de productos .

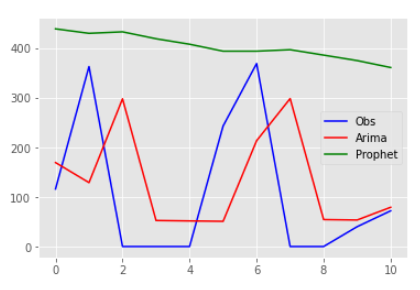






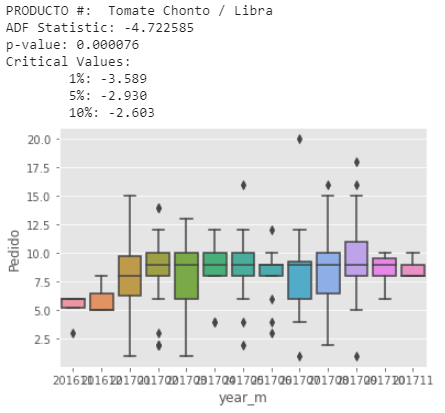
***Anexo 2.***

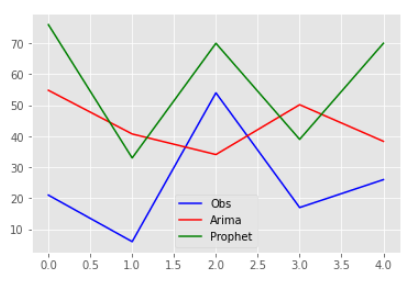
Predicción de demanda para la totalidad de productos, aplicando un modelo Arima y el algoritmo Prophet .

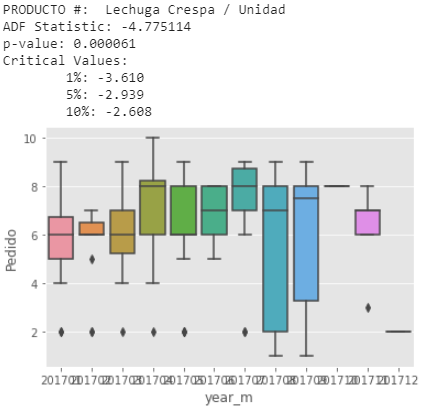


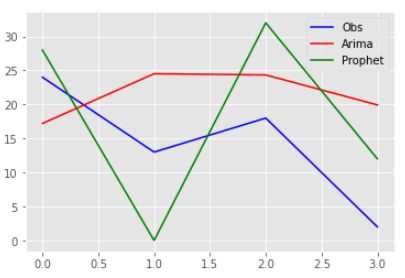
***Anexo 3.***

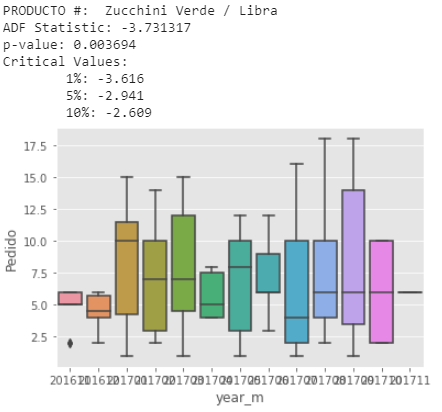
Boxplot y valores predichos versus observados por producto (9 productos más relevantes), para el 10% de los periodos aplicando Rolling forecast.

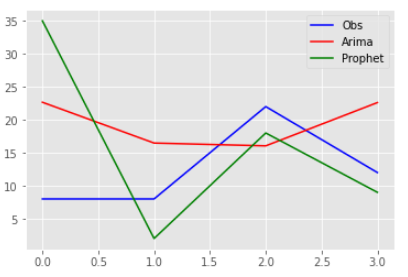


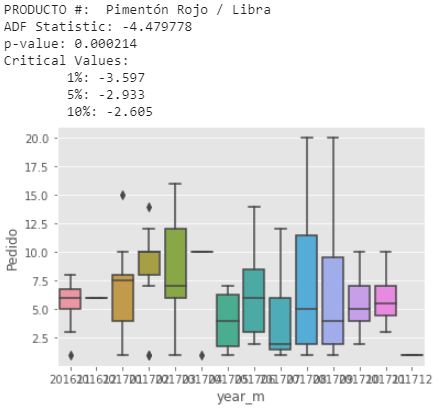


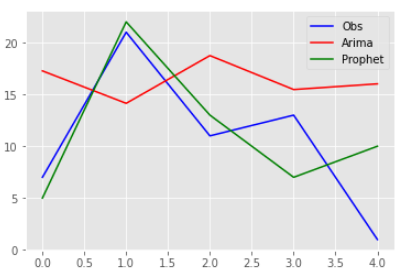


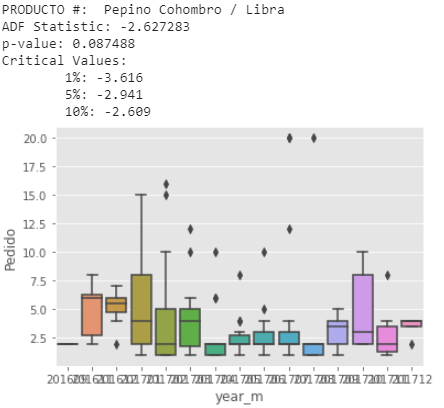


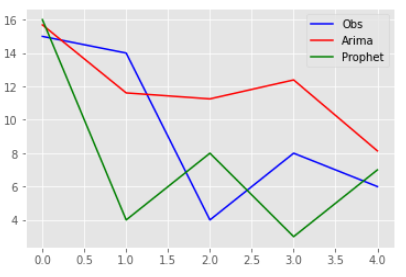


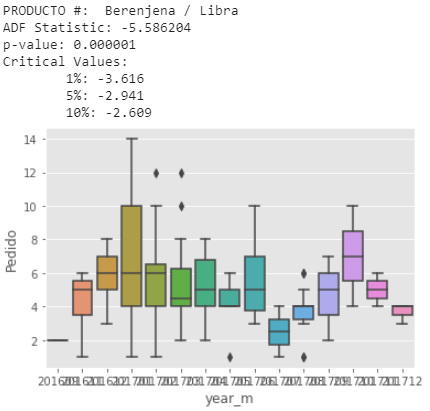


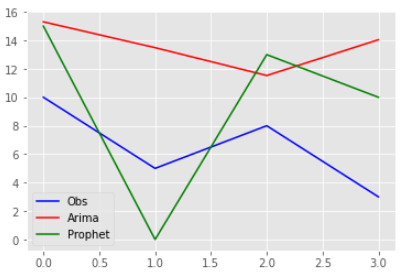


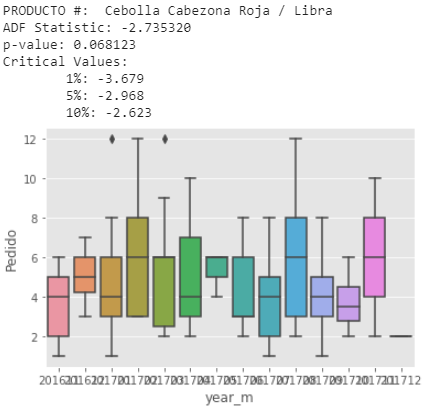


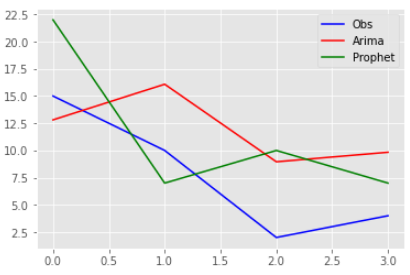


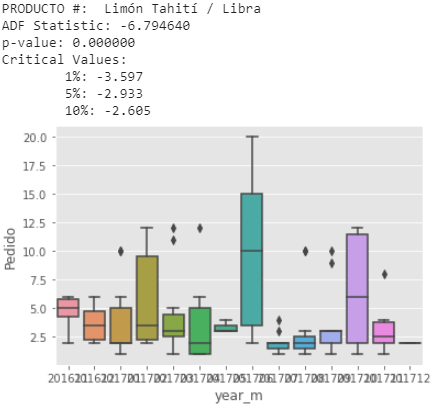


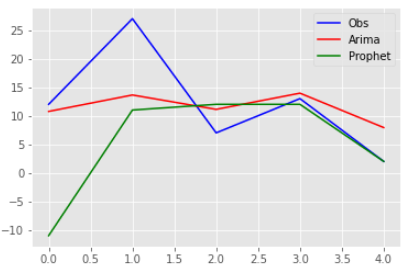


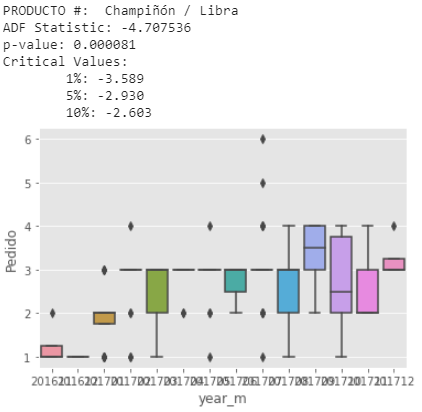


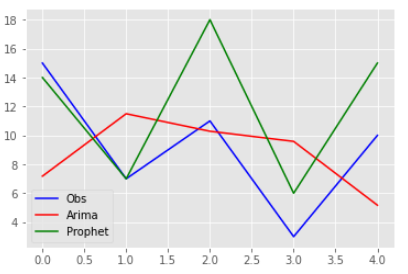












***Anexo 4.***

Predicción para un nuevo periodo. Valores predichos utilizando un modelo Arima y el algoritmo Prophet, las predicciones corresponden al mismo periodo.

