

Micro-proyecto I - Start Up Agrícola

Daniela Carolina Sanabria Guerrero

Juan Camilo Marín Cala

Jorge Enrique González García

Introducción

La compañía del Start-up agrícola busca implementar un sistema de recomendación sobre el manejo del inventario de sus productos, con el fin de mejorar la operatividad del negocio y volverse mucho más eficiente. Actualmente la compañía distribuye 121 productos agrícolas a distintos clientes, y cada uno de ellos tiene un nivel de demanda diferente, lo cual hace que las necesidades de inventario de cada uno de estos productos varíen. Por la corta vida que tienen los productos de este negocio, es de vital importancia tener claramente definido un sistema que le facilite a la gerencia gestionar y administrar adecuadamente sus productos minimizando el costo de inventario.

Definición del problema

Actualmente la empresa cuenta con un problema operativo con respecto al manejo de sus productos, lo cual puede desembocar en menores retornos del negocio. En primer lugar, al tener un número elevado de SKU en su portafolio, le es difícil identificar cuáles de ellos son los más demandados y que están generando un verdadero impacto en sus ventas.

En segundo lugar, cada uno de estos productos tiene una demanda con periodicidad diferente, en consecuencia, la política de pedidos e inventarios debe manejarse de manera diferencial por cada uno de ellos si se desea hacer de una manera efectiva. Al no contar con un pronóstico certero por cada producto, las ineficiencias en el inventario están a la orden del día, incrementando el riesgo que la compañía sufra fuertes pérdidas económicas por desperdicios en productos que no se pidieron en la cantidad adecuada.

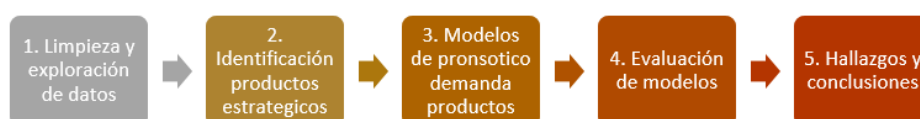
Pregunta de investigación

A raíz de las necesidades de la compañía y los problemas que se identificaron en una fase diagnóstico, se plantean las siguientes preguntas de investigación:

¿Cuáles son los productos más representativos en las ventas totales de la compañía? ¿Cuál es el mejor pronóstico para cada uno de los productos clave de forma tal que se pueda tener un manejo mucho más certero de los pedidos de productos?

Metodología propuesta

Para lograr generar un procedimiento que genere valor para la compañía se propone la siguiente metodología, compuesta de 5 diferentes etapas.



1. Limpieza y exploración de datos:

En primer lugar, se calcula la demanda de cada uno de los productos, definida como el número de unidades por el precio, y así mismo se realiza la agrupación de ventas por producto en todas las semanas.

Nombre_producto		Tomate Chonto / Libra	Champiñón / Libra	Pimentón Rojo / Libra	Lechuga Crespa / Unidad	Zucchini Verde / Libra	Cebolla Cabezona Roja / Libra	Pepino Cohombro / Libra	Cebolla Cabezona Blanca / Libra	Zanahoria / Libra	Limon Tahiti / Libra	Berenjena / Libra
Año	Semana											
2016	44	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	46	11000.0	11200.0	8000.0	0.0	3000.0	8000.0	12600.0	0.0	1200.0	0.0	4480.0
	47	97900.0	39200.0	73600.0	2400.0	25800.0	29000.0	35910.0	9400.0	9000.0	36800.0	50400.0
	48	100100.0	29400.0	79800.0	4800.0	27200.0	21000.0	40320.0	16920.0	4500.0	32000.0	41920.0
	49	94600.0	14000.0	34500.0	4800.0	40000.0	22000.0	29400.0	3000.0	6000.0	48100.0	34800.0
	50	111100.0	14000.0	61500.0	32400.0	30400.0	36000.0	50400.0	1000.0	0.0	22100.0	38400.0
	51	72700.0	0.0	24100.0	0.0	38200.0	18000.0	38200.0	0.0	0.0	27300.0	24900.0
	52	75400.0	0.0	29900.0	0.0	24700.0	23000.0	41000.0	0.0	0.0	18200.0	24000.0
2017	1	71500.0	0.0	40300.0	0.0	26000.0	12000.0	23000.0	0.0	0.0	24700.0	33000.0
	2	120900.0	15000.0	75400.0	0.0	57200.0	28000.0	38000.0	1000.0	600.0	22100.0	40500.0

Como se observa el resultado final de la primera etapa explotaría, los productos han quedado agrupados en una serie de tiempo semanal que muestra el total de demanda que ha sido reportada para cada una de las referencias que maneja la compañía.

2. Identificación de los productos estratégicos:

Producto de la primera etapa exploratoria de limpieza, se evidencia claramente como hay referencias que tienen una periodicidad muy baja y así mismo que hay productos que no son representativos dentro de las ventas de la compañía. Con el fin de hacer un proceso de selección adecuado se busca calcular el total de demanda de la compañía en todo el periodo analizado y determinar cuáles productos son aquellos que representan el mayor porcentaje de ventas.

Un factor adicional que se considera a la hora de seleccionar los productos clave de la compañía es la cantidad de datos que se tiene disponible. Este motivo es crucial para la siguiente etapa debido a que habrá productos con muy baja periodicidad que no tendrán los suficientes registros para hacer uso de metodologías de pronóstico basadas en series de tiempo o en redes neuronales.

Se muestra a continuación el resumen del producto donde nos indica la importancia del producto medida como el porcentaje del total de demanda, la demanda y el número de faltantes en la serie.

Nombre_producto	NaNs	Demanda	Importancia
Tomate Chonto / Libra	1	5981700.0	0.155343
Champiñón / Libra	4	4029800.0	0.104653
Pimentón Rojo / Libra	5	2203150.0	0.057215
Lechuga Crespa / Unidad	6	1718400.0	0.044626
Zucchini Verde / Libra	8	1628100.0	0.042281
...
Pera Nacional / Libra	44	3000.0	0.000078
Toronjil	45	2500.0	0.000065
Manzana Nacional / Libra	45	2000.0	0.000052
Zucchini Amarillo / Libra	45	1500.0	0.000039
Cubio / Libra	45	800.0	0.000021

Finalmente se propone seleccionar los productos que acumulen el 80% del total de la demanda como productos representativos para la compañía, posterior a esta selección se hace una segunda revisión para estos productos tengan por lo menos 20 datos, lo cual lo cumple y de esta manera se pueda hacer uso de modelos.

Posterior a la selección la base de 121 productos queda reducida a un total de 38 referencias, que generan el 80% del total de la demanda de la compañía, se espera que haciendo un pronóstico certero de la dinámica de este grupo de productos se pueda mejorar el manejo operativo de los pedidos por parte del start-up.

3. Modelos de pronóstico de demanda

Una vez se tiene la demanda por semana, se procede a realizar el pronóstico de la demanda por cada una de las 38 referencias seleccionadas de las cuáles sabemos que tienen información por lo menos en 20 de las 46 semanas estudiadas. Para realizar este pronóstico se propone realizar 4 metamodelos basados en redes neuronales, donde cada uno calcula los 38 MSE por referencia, basados en dos arquitecturas:

	RMSE General
1-LSTM	20510.537645
2-Redes Recurrentes Simples	20623.025064
3-Convulacionales LSTM	28693.415804
4-Redes Recurrentes Simples 2	27468.216909

- **LSTM:** Este modelo usa una arquitectura bidireccional LSTM (Long Short Term Memory) que conecta dos capas ocultas de direcciones opuestas a la misma salida, en este caso el MSE promedio alcanzado fue de 20510 en el primer modelo y 28693 en el tercer modelo.
- **Redes Recurrentes Simples:** Se estimaron dos modelos con arquitectura de redes recurrentes simples, con un lookback de 3 y 4, que indica el número de pasos de tiempo previos que se deben utilizar como variables de entrada para predecir el siguiente periodo de tiempo (t , $t-1$, $t-2$ y $t-3$), en este caso el MSE promedio alcanzado fue de 20623 en el segundo modelo y 27468 en el cuarto modelo.

4. Evaluación de los modelos

Se evalúa el MSE de cada modelo para cada referencia y la comparación entre modelos está dada por el MSE promedio. Esta métrica es una de las más comunes para la evaluación de la regresión y como su nombre lo indica, mide el error cuadrático medio de las predicciones en test, cuanto menor sea su valor, mejor es el modelo teniendo en cuenta que nunca va a ser negativo, ya que se evalúa el cuadrado de los errores. En cuanto a los modelos evaluados, el que mejor MSE promedio presentó fue de 21759,69, es decir, el LSTM.

5. Hallazgos y conclusiones

Una primera conclusión que viene desde el preprocesamiento de los datos es que el 31% de las referencias (38 productos) explican el 80% de la demanda que va a tener y que las últimas 46 semanas ha vendido más de 10 veces, por lo que debería concentrarse más en estos productos a la hora de realizar su inventario para que siempre estén disponibles y no se desperdicie.

Igualmente se puede notar que las mejores temporadas son el tercer trimestre del año y el primero (en orden de importancia), sin embargo, se necesita información de otros años para comprobar esta estacionalidad; siendo el mejor mes agosto, donde hay una mayor demanda a nivel general, en cambio, en meses como junio se presenta una disminución de la demanda por lo que no sería necesario solicitar tanto en los pedidos. Adicionalmente, es importante registrar mejor los datos ya que existen muchos NaNs que entorpecen la labor de los modelos a la hora de hacer predicciones.

En cuanto a la predicción, en términos generales se recomienda utilizar un modelo LSTM que fue el que mejor se ajustó a los datos, lo ideal sería correr el modelo con la información actualizada semana tras semana verificando la calidad de los datos. Sin embargo, se evidencia que para cada referencia hay un modelo con el que se logra el mínimo MSE, por lo que se recomienda correr todos los modelos y ver las predicciones de los mejores por producto.

Nombre_producto	
Tomate Chonto / Libra	1-LSTM
Champiñón / Libra	2-Redes Recurrentes Simples
Pimentón Rojo / Libra	1-LSTM
Lechuga Crespa / Unidad	1-LSTM
Zucchini Verde / Libra	1-LSTM
Cebolla Cabezona Roja / Libra	2-Redes Recurrentes Simples
Pepino Cohombro / Libra	2-Redes Recurrentes Simples
Cebolla Cabezona Blanca / Libra	3-Convulacionales LSTM
Zanahoria / Libra	2-Redes Recurrentes Simples
Limón Tahití / Libra	4-Redes Recurrentes Simples 2