

Cindy Zulima Alzate Román	Código 201920019
Jahir Stevens Rodríguez Riveros	Código 201819361
Sergio Alberto Mora Pardo	Código 201920547

1. Introduccion con la definición del problema y la pregunta de investigación (20%)

La start-up agrícola debe gestionar los inventarios de manera inteligente, disminuyendo la ventana de tiempo entre cosecha y consumo. Los problemas de inventarios en producción usualmente se resuelven con el modelo de Cantidad Economica de Pedido (EOQ). El cual funciona bastante bien con productos no perecederos porque no hay un impacto en el valor del producto entre su producción y consumo, caso contrario el de los productos agrícolas. Donde el producto cuenta con una vida limita al tal punto que su valor es igual a cero o negativo por lo que gran parte del éxito depende de la capacidad que tenga la start-up para predecir la demanda y de esa forma pueda desarrollar acciones que le ayuden a mejorar la gestión de la cadena productiva.

El principal reto en el ciclo de vida de la start-up agrícola se encuentra en desarrollar un modelo predictivo de demanda para los productos más relevantes con los que trabaja, obteniendo una predicción a nivel semanal (una semana adelante).

La pregunta de investigación es ¿Es posible desarrollar un modelo predictivo de la para los productos más relevantes con los que trabaja la start-up agrícola a partir de los datos disponibles con los que cuenta?

2. Metodología propuesta (10%)

Análisis Exploratorio de Datos

- ¿Cuántos productos tenemos?
- ¿Cuántos clientes tenemos?
- ¿Cuáles son los principales clientes?
- ¿Qué es lo que más compran los principales clientes?
- ¿Cuántos días tenemos?
- ¿Que día de la semana más venden?
- ¿Que más piden más productos?
- ¿Qué clientes son los que compran los productos más vendidos? ¿Hay relación con los clientes que más compran?
- ¿Cuál es la distribución de precios que compran los clientes? ¿Entre más caro más se compra o viceversa?
- ¿Los precios fluctúan por producto o siempre manejan los mismos? ¿sí fluctúan, porqué lo hacen: depende del cliente o de la fecha?

Ingeniería de características

- ¿Valdrá la pena sacar el valor total de la compra? (precio x cantidad)
- ¿Valdrá la pena sacar el día de la semana de las fechas?
- ¿Valdrá la pena sacar el mes de la fecha?
- ¿Valdrá la pena sacar el año de la fecha?
- ¿Valdrá la pena sacar la semana del año de la fecha?
- ¿Valdrá la pena dividir el nombre del producto entre '/'? -> se podría sacar unidad por producto (gramos, libras,...) -> si tienen número estas unidades ¿es posible sacar el número?
- ¿De la división anterior, es posible sacar el nombre del producto?

Los ejercicios se desarrollaron para los 6 productos que mayor demanda presentaron en la base de datos recopilada. Como un ejercicio inicial se utilizó un ajuste de pronóstico basado en Redes Neuronales Recurrentes testeado sobre un *rolling forecast* con un ajuste en sus hiperparámetros automatizado con Optuna.

La métrica de evaluación utilizada fue el RMSE ponderado por la demanda de cada producto, Si Rolling Forecast es un pronóstico que se realiza de forma iterativa semana tras semana, es decir que pronostica una semana y incluye ese dato para utilizarlo como un dato de entrada para predecir la siguiente semana el proceso lo repite hasta completar todo el periodo. No Rolling Forecast pronostica todo el periodo de tiempo y no uno a uno como lo hace el otro modelo.

3. Resultados (20%) y análisis de los resultados (25%)

Primero se ejecutó el código de redes neuronales, obteniendo los siguientes resultados

Producto	SI Rolling Forecast	NO Rolling Forecast
FRU0046-Tomate chonto	3966.543	15733469.800
VER0035-Pepino cohombro	2524.238	6371780.000
VER0038-Pimentón rojo	2765.918	7650306.000
VER0010-Zucchini verde	1475.261	2176395.200
VER0028-Lechuga crespá	1431.103	2048057.800
VER0052-Plátano maduro ¹	0	0

Posteriormente, se buscó una mejora con la utilización de Optuna, que realizó una correcta optimización de los hiperparámetros para las redes, sin embargo, no se tuvo ningún progreso en los resultados de los RMSE calculados.

¹ El RMSE es igual a 0 porque Optuna no logro converger con los datos de entrenamiento.

Finalmente, se usó un enfoque econométrico para poder abordar la pregunta de investigación usando los modelos ARIMA, SARIMA, Prophet los cuales se compararon con un LSTM.

Resultados y análisis de los resultados

Las series para los 6 productos con mayor demanda no cumplían la condición de estacionariedad por lo que se determinó aplicar diferencias a las series originales y de esa forma se logró cumplir esta condición. Con las series diferenciadas se pudieron correr los siguientes modelos:

Cálculo RMSE Log						
Modelo	Tomate	Pepino	Pimentón	Zucchini	Lechuga	Plátano
ARIMA	2.448	1.298	2.270	1.797	1.630	1.668
SARIMA	2.762	1.317	2.167	1.660	1.788	1.973
Prophet	2.968	2.029	1.991	2.249	2.200	1.756
LSTM	2.530	1.097	2.295	1.569	3.217	2.094

Donde se pudo identificar que cada uno de los productos debería modelarse de forma independiente porque no se encuentra un mismo modelo con el mejor ajuste respecto al RMSE Log para todos los productos. De acuerdo con lo anterior se tomó la decisión de implementar el modelo con mejor ajuste para cada uno de los productos.

Con la implementación de los modelos econométricos se reduce significativamente el error del pronóstico frente al error obtenido con la red neuronal con el ajuste de los hiperparámetros automatizados con el framework de Optuna. A continuación, el RMSE:

	Tomate	Pepino	Pimentón	Zucchini	Lechuga	Plátano
ARIMA	11.565193	3.6619654	9.6794008	6.0315257	5.1038747	5.3015540
SARIMA	15.831474	3.7322079	8.7320485	5.2593108	5.9774855	7.1922208
Prophet	19.452974	7.6064760	7.3228529	9.4782528	9.0250135	5.7892340
LSTM	12.553506	2.9951670	9.9244360	4.8018439	24.953148	8.1173195

En la siguiente tabla se puede ver el comparativo entre el desempeño obtenido entre las redes neuronales:

Producto	SI Rolling Forecast	LSTM
FRU0046-Tomate chonto	3966.543	12.5535061
VER0035-Pepino cohombro	2524.238	2.99516703
VER0038-Pimentón rojo	2765.918	9.92443601
VER0010-Zucchini verde	1475.261	4.80184395
VER0028-Lechuga crespa	1431.103	24.9531483
VER0052-Plátano maduro	0	8.11731959

4. Conclusiones (25%)

- Se concluyo que los modelos econometricos generan un mejor ajuste al pronostico demanda que la red neuronal obtenida con el ajuste de los hiperparámetros automatizados con el framework de Optuna.
- La red neuronal ajustada con TimeseriesGenerator (LSTM) obtuvo un mejor desempeño que la red autoajustada con Optuna, para el LTSM fue igual a 12.5535061 y con Optuna fue igual a 3966.543 para el Tomate y ese comportamiento se repitió con todos los productos.
- Para mejorar el desempeño de las redes neuronales se pueden utilizar realizar pruebas con un numero mayor de epocas y ver si con esto logran tener un mejor ajuste.
- La red con Optuna puede mejorar su desempeño al probarla con más trials.

5. Puede incluir referencias y anexos además de las 4 paginas.

El notebook se encuentra en la siguiente ubicación: https://nbviewer.jupyter.org/github/sergiomora03/deep-learning-intermediate/blob/master/Micro-Proyecto1_Start-Up%20Agricola/notebooks/MicroProyecto1_StartUp.ipynb