



# **Entrega microproyecto 1**

## **Modelos Avanzados de Análisis de Datos 2**

**Agosto 19 de 2020**

**Camilo Andres Suarez Trillos  
200321493**

**Carlos Francisco Silva Ortiz  
201920463**

**Juan Camilo Florez Caro  
201620135**



## 1. Introducción con la definición del problema y la pregunta de investigación (20%):

Una start-up agrícola quiere construir una aplicación online que conecte el campo con la ciudad de manera directa y sin intermediarios. De esta manera, los agricultores ponen a la disposición de la empresa sus distintos productos, y la start-up se encarga de atender las necesidades de demanda distribuyendo los productos hasta los clientes finales en la ciudad.

Para optimizar la gestión logística y desarrollar esta labor de manera eficiente, los dueños de la start-up han visto la necesidad de diseñar e implementar un sistema de gestión y predicción que les permita planear con la debida anticipación la demanda esperada. Así podrán coordinar las actividades de producción, junto con los agricultores, y atender la demanda mientras se minimizan las pérdidas por unidades de producto no vendido o estropeado.

Este es un reto altamente relevante para la sostenibilidad de la industria alimentaria, donde gran cantidad de las pérdidas se refieren a grupos alimenticios donde la frescura del producto es un importante criterio para el consumo. El desperdicio de los alimentos se relaciona con el sobre-pedido o el sobre-almacenamiento, lo cual puede a su vez ser consecuencia de dificultades en el pronóstico de la demanda. Este problema es aún más marcado a medida que uno sube en la cadena de producción y se aleja de los patrones de comportamiento del consumidor final.

Por lo tanto, la start-up agrícola debe gestionar los inventarios de manera inteligente, disminuyendo la ventana de tiempo entre cosecha y consumo. Para ello, necesita desarrollar modelos de predicción de demanda que apoyen su operación, logrando la disponibilidad de los productos para atender a los clientes del día siguiente, minimizando los desperdicios de alimentos e identificando hábitos de consumo que le permita optimizar sus operaciones logísticas y de negocio.

Por ello se pide desarrollar un modelo predictivo de demanda para los productos más relevantes con los que trabaja la start-up agrícola, obteniendo una predicción a nivel semanal (una semana adelante).

## 2. Metodología Propuesta:

- Realizar la identificación de los productos más relevantes (aquellos que tienen una demanda de al menos el 0.5% del total de los pedidos de la base)
- Realizar 4 modelos de predicción con Facebook Prophet:
  - Datos agrupados de forma semanal y estimar 1 observación.
  - Datos originales agrupados por días y estimar 7 días.
  - Datos posteriores a '2016-11-13' de forma semanal y estimar 1 observación.
  - Datos posteriores a '2016-11-13' agrupados por días y estimar 7 días.

- 
- Calcular el RMSE de los dos modelos con rolling forecast de las 4 últimas semanas (o 28 días) y seleccionar el modelo con mejor desempeño

### 3. Resultados:

Al realizar los cuatro modelos a través de FB Prophet y sacar el RMSE de cada modelo encontramos que el modelo de mejor performance es el modelo semanal sin datos filtrados al tener un menor RMSE vs el modelo diario, (14,77 pedidos, 38,78 pedidos, 25,76 pedidos, 66,01 pedidos)

A continuación, se adjunta la tabla de pedidos para la próxima semana de los principales productos de la start up:

Nombre	Pedidos
Tomate Chonto / Libra	115
Plátano Maduro / Libra	50
Pepino Cohombro / Libra	44
Cebolla Cabezona Blanca / Libra	39
Pimentón Rojo / Libra	38
Cebolla Cabezona Roja / Libra	32
Zucchini Verde / Libra	32
Zanahoria / Libra	25
Lechuga Crespa / Unidad	23
Limón Tahití / Libra	18
Berenjena / Libra	16
Champiñón / Libra	10
Tomate Larga Vida / Libra	10
Papa Criolla / Libra	6
Mazorca Desgranada / Libra	5
Cebolla Larga / Libra	5
Rúgula / Libra	5
Manzana Roja / Unidad	5
Papa Pastusa / Libra	5
Cilantro / 100 gramos	5
Brócoli / Libra	4
Habichuela / Libra	4
Banano Criollo / Libra	4
Mango Tommy Atkins / Libra	4
Ahuyama / Libra	4
Mora Castilla / Libra	3
Papa Sabanera / Libra	3
Yerbabuena / 100 gramos	3

<b>Perejil Crespo / 100 gramos</b>	<b>3</b>
<b>Aguacate Papelillo / Libra</b>	<b>3</b>
<b>Apio / Libra</b>	<b>2</b>
<b>Perejil Liso / 100 gramos</b>	<b>2</b>
<b>Manzana Verde / Unidad</b>	<b>1</b>
<b>Espinaca / Libra</b>	<b>1</b>
<b>Fresa Pareja / Libra</b>	<b>1</b>

#### 4. Conclusiones:

- De los cuatro modelos realizados, el de mejor performance fue el semanal contemplando todo el dataset, arrojando un RMSE de 14,77 pedidos.
- Para la realización de un forecast de pedidos por productos, encontramos que la serie de datos antes de '2016-11-13' no tiene la regularidad necesaria para que el modelo pueda interpretar las estacionalidades correctamente y realizar un forecast con un error bajo.
- Se recomienda a la empresa emergente realizar una recolección rigurosa de pedidos diaria, de esta manera se podrá a futuro mejorar el performance de las predicciones y disminuir el desperdicio alimentario.
- Para poder gestionar adecuadamente la demanda es necesario contar con más información de días previos, debido a que en el análisis exploratorio realizado inicialmente se evidencia que la regularidad de recolección de la información no es estable.
- En la medida que se tenga más información constante, el algoritmo podrá fácilmente establecer la demanda diaria con menor error que la semanal, permitiendo tener al startup una data más granular, que le permitirá tener aún menos pérdidas de alimentos cada semana.