

UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA
CARRERA DE INGENIERÍA INDUSTRIAL



PRONÓSTICO DE DEMANDA Y
ESTABLECIMIENTO DE INVENTARIOS DE
PEDIDOS POR CAMPAÑA DE JOYAS DE FANTASÍA
EN COMPAÑÍA DEDICADA A LA VENTA DIRECTA

TESIS

Para optar el título profesional de Ingeniero Industrial

AUTOR

Miguel Angel Guillermo Quispe (ORCID: 0009-0006-9474-5135)

ASESOR

Claudia Fausta Maura Antonini Bova (ORCID: 0000-0003-0008-5879)

José Antonio Larco Martinelli (ORCID: 0000-0002-0258-8368)

Lima – Perú

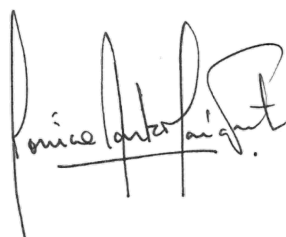
2023

DECLARACIÓN JURADA

Yo, Mónica Cecilia Santa María Fuster identificada con DNI No 18226712 en mi condición de autoridad responsable de validar la autenticidad de los trabajos de investigación y tesis de la UNIVERSIDAD DE INGENIERIA Y TECNOLOGIA, DECLARO BAJO JURAMENTO:

Que la tesis denominada “PRONÓSTICO DE DEMANDA Y ESTABLECIMIENTO DE INVENTARIOS DE PEDIDOS POR CAMPAÑA DE JOYAS DE FANTASÍA EN COMPAÑÍA DEDICADA A LA VENTA DIRECTA” ha sido elaborada por el señor Miguel Angel Guillermo Quispe, con la asesoría de Claudia Fausta Maura Antonini Bova identificada con CE 001549712 y de José Antonio Larco Martinelli identificado con DNI 40239630, y que se presenta para obtener el título profesional de Ingeniero industrial, ha sido sometida a los mecanismos de control y sanciones anti plagio previstos en la normativa interna de la universidad, encontrándose un porcentaje de similitud de 0%.

En fe de lo cual firmo la presente.



Dra. Mónica Santa María Fuster
Directora de Investigación

En Barranco, el 8 de febrero 2024

Dedicatoria:

La dedicatoria en esta oportunidad va dirigida a mis padres, familiares, amigos, compañeros y todos aquellos que compartieron momentos con mi persona. De todos ellos aprendí siempre lo mejor.

Agradecimientos:

Primero agradezco a Dios y mi patrón San Miguel De Arcángel. Después, agradecer a mis padres por todo lo que me han brindado en mis estudios como en la vida. Además, agradecer a mi asesor de tesis 1 y a mi asesora de 2 por todos los servicios brindados, al igual que a todos mis profesores. También agradecer a la señora Soila quien desde cuando yo era muy pequeño me brindaba sus consejos a mí y a mis padres.

Ella viene siendo un factor importante en lo personal, profesional y académico. Y estoy seguro que a pasar de su deceso, ella lo seguirá siendo porque sus consejos siempre los tengo presente.

TABLA DE CONTENIDO

	Pág.
RESUMEN	13
ABSTRACT	14
INTRODUCCIÓN.....	15
CAPÍTULO I.....	21
MARCO TEÓRICO	21
1.1. Venta directa.....	21
1.1.1. Definición de venta directa	21
1.1.2. Categorías de productos.....	21
1.1.3. Ventajas y desventajas de la venta directa.....	22
1.2. Variables de mercadotecnia en la venta directa.....	23
1.3. Pronóstico de demanda.....	25
1.3.1. Pronóstico de demanda desagregado (Bottom up)	27
1.3.2. Modelos de pronósticos desagregado en la venta directa	27
1.3.3. Modelo dinámico bayesiano	27
1.3.4. Arboles de Decisión	28
1.3.5. Random forest.....	30
1.3.6. Diferencia de resultados de pronósticos entre métodos tradicionales y ML.....	35
1.3.7. Pronóstico de demanda agregado (Top down).....	35
1.3.7.1. Modelos de pronósticos agregado en la venta directa	35
1.3.7.2. Regresión múltiple.....	35

1.3.7.3. Modelo Holt-Winter	37
1.3.8. Diferencia entre top down y bottom up	38
1.3.9. Pronósticos de demanda de productos en el sector de venta directa	39
1.3.10. Pronósticos de demanda y gestión de inventarios para productos de un solo periodo.....	41
1.3.11. Pronósticos de demanda de ventas de productos por catálogo mediante juicio de experto	46
1.7. Medidas de exactitud de pronósticos de demanda de productos.....	47
CAPÍTULO II.....	49
METODOLOGÍA.....	49
2.1. Justificación.....	49
2.2. Reproducibilidad	49
2.3. Validez.....	50
2.4. Definición del objeto de estudio.....	50
2.5. Variables independientes.....	50
2.6. Variable dependiente	51
2.7. Variables contextuales.....	51
2.8. Pasos de la metodología	51
CAPÍTULO III	59
RESULTADOS	59
CONCLUSIONES Y RECOMENDACIONES	75
REFERENCIAS BIBLIOGRÁFICAS	77
ANEXOS	81

ÍNDICE DE TABLAS

	Pág.
Tabla 1. Categoría de productos en la venta directa	22
Tabla 2. Ventajas y desventajas de la venta directa.....	22
Tabla 3. Variables de mercadotecnia en la venta directa.....	23
Tabla 4. Actividades relacionadas a la generación de pronósticos	25
Tabla 5. Métodos cuantitativos, cualitativos e impredecibles.	26
Tabla 6. Modelos de pronósticos de series de tiempo con sus componentes respectivos	27
Tabla 7. Ventajas y desventajas de CART.....	29
Tabla 8. Ventajas y desventajas de Random Forest.....	33
Tabla 9. Parámetros de RF.....	34
Tabla 10. Funcionalidades por tipo de pronóstico Top down y Bottom up.....	39
Tabla 11. Métodos de pronóstico para productos de lanzamiento por campañas.....	40
Tabla 12. Datos de pronósticos, demanda real, error y A/F ratios de productos	42
Tabla 13. Datos de A/F ratios clasificados por productos de una temporada anterior	43
Tabla 14. Tabla de función de distribución empírica para un producto de la temporada actual	44
Tabla 15. Campañas usadas como muestras	53
Tabla 16. Resultados para selección de pronóstico por campaña.....	59
Tabla 17. Valores de mtry óptimo para cada campaña.....	62

Tabla 18. Error tipo MAPE por campaña del 2019	64
Tabla 19. Resultados de pruebas de hipótesis por campañas pronosticas para validación de modelo.....	67
Tabla 20. Resultados de factores críticos de algunos productos de la campaña n°1	68
Tabla 21. Resultados de A/F ratios de la campaña n° 18 del 2018.....	69
Tabla 22. Resultados de costos de oportunidad del 2018 y 2019	73
Tabla 23. Mejoras financieras.....	73

INDICE DE FIGURAS

	Pág.
Figura 1. Número de cortes en una campaña.....	17
Figura 2. Envío de pedidos y pre pedidos por corte correspondiente a cada zona	17
Figura 3. Momentos de compra para línea de Joyería	19
Figura 4. Construcción de árbol de decisión.....	28
Figura 5. Conjunto de entrenamiento y de test para CART.....	30
Figura 6. Equilibrio entre error de sesgo y varianza.....	31
Figura 7. Construcción de RF	32
Figura 8. Conjuntos de entrenamiento y de test para Random Forest	34
Figura 9. Función de distribución empírica del nuevo producto de la temporada entrante	45
Figura 10. Pasos de la metodología	51
Figura 11. Sub-pasos adicionales de ejecución de newsvendor.....	55
Figura 12. Grafica Error vs Numero de árboles de la campaña 3.....	60
Figura 13. Grafica Error vs Numero de árboles de la campaña 5.....	61
Figura 14. Grafica Error vs Mtry de la campaña 4	63
Figura 15. Grafica de importancia de variables de la campaña 10	63
Figura 16. Grafica Demanda real vs Pronóstico de la campaña 17	65
Figura 17. Grafica Demanda real vs Pronóstico de la campaña 16	66

Figura 18. Distribución empírica acumulada del producto n° 1 de la campaña 1 del 2019	70
Figura 19. Relación A/F ratio versus probabilidad para la campaña 1 del 2019	71
Figura 20. Relación A/F ratio versus probabilidad para la campaña 2 del 2019	71
Figura 21. Comparación de cantidades de sobrantes y faltantes en el 2018 y 2019.....	72

ÍNDICE DE ANEXOS

	Pág.
Anexo 1: Número de árboles vs Error del modelo. Campaña 1	81
Anexo 2: Número de árboles vs Error del modelo. Campaña 2	81
Anexo 3: Número de árboles vs Error del modelo. Campaña 3	82
Anexo 4: Número de árboles vs Error del modelo. Campaña 4	82
Anexo 5: Número de árboles vs Error del modelo. Campaña 5	83
Anexo 6: Número de árboles vs Error del modelo. Campaña 6	83
Anexo 7: Número de árboles vs Error del modelo. Campaña 7	84
Anexo 8: Número de árboles vs Error del modelo. Campaña 8	84
Anexo 9: Número de árboles vs Error del modelo. Campaña 9	85
Anexo 10: Número de árboles vs Error del modelo. Campaña 10	85
Anexo 11: Número de árboles vs Error del modelo. Campaña 11	86
Anexo 12: Número de árboles vs Error del modelo. Campaña 12	86
Anexo 13: Número de árboles vs Error del modelo. Campaña 13	87
Anexo 14: Número de árboles vs Error del modelo. Campaña 14	87
Anexo 15: Número de árboles vs Error del modelo. Campaña 15	88
Anexo 16: Número de árboles vs Error del modelo. Campaña 16	89
Anexo 17: Número de árboles vs Error del modelo. Campaña 17	89
Anexo 18: Número de árboles vs Error del modelo. Campaña 18	90
Anexo 19: Demanda real vs pronostico. Campaña 1	90

Anexo 20: Demanda real vs pronostico. Campaña 2.....	91
Anexo 21: Demanda real vs pronostico. Campaña 3.....	91
Anexo 22: Demanda real vs pronostico. Campaña 4.....	92
Anexo 23: Demanda real vs pronostico. Campaña 5.....	92
Anexo 24: Demanda real vs pronostico. Campaña 6.....	93
Anexo 25: Demanda real vs pronostico. Campaña 7.....	93
Anexo 26: Demanda real vs pronostico. Campaña 8.....	94
Anexo 27: Demanda real vs pronostico. Campaña 9.....	95
Anexo 28: Demanda real vs pronostico. Campaña 10.....	95
Anexo 29: Demanda real vs pronostico. Campaña 11.....	95
Anexo 30: Demanda real vs pronostico. Campaña 12.....	96
Anexo 31: Demanda real vs pronostico. Campaña 13.....	96
Anexo 32: Demanda real vs pronostico. Campaña 14.....	97
Anexo 33: Demanda real vs pronostico. Campaña 15.....	97
Anexo 34: Demanda real vs pronostico. Campaña 16.....	98
Anexo 35: Demanda real vs pronostico. Campaña 17.....	98
Anexo 36: Demanda real vs pronostico. Campaña 18.....	99
Anexo 37: Relación A/F ratio versus probabilidad para la campaña 1 del 2019.....	99
Anexo 38: Relación A/F ratio versus probabilidad para la campaña 2 del 2019.....	100
Anexo 39: Relación A/F ratio versus probabilidad para la campaña 3 del 2019.....	100
Anexo 40: Relación A/F ratio versus probabilidad para la campaña 4 del 2019.....	101
Anexo 41: Relación A/F ratio versus probabilidad para la campaña 5 del 2019.....	101
Anexo 42: Relación A/F ratio versus probabilidad para la campaña 6 del 2019.....	102

Anexo 43: Relación A/F ratio versus probabilidad para la campaña 7 del 2019.....	102
Anexo 44: Relación A/F ratio versus probabilidad para la campaña 8 del 2019.....	103
Anexo 45: Relación A/F ratio versus probabilidad para la campaña 9 del 2019.....	103
Anexo 46: Relación A/F ratio versus probabilidad para la campaña 10 del 2019.....	104
Anexo 47: Relación A/F ratio versus probabilidad para la campaña 11 del 2019.....	104
Anexo 48: Relación A/F ratio versus probabilidad para la campaña 12 del 2019.....	105
Anexo 49: Relación A/F ratio versus probabilidad para la campaña 13 del 2019.....	105
Anexo 50: Relación A/F ratio versus probabilidad para la campaña 14 del 2019.....	106
Anexo 51: Relación A/F ratio versus probabilidad para la campaña 15 del 2019.....	106
Anexo 52: Relación A/F ratio versus probabilidad para la campaña 16 del 2019.....	107
Anexo 53: Relación A/F ratio versus probabilidad para la campaña 17 del 2019.....	107
Anexo 54: Relación A/F ratio versus probabilidad para la campaña 18 del 2019.....	108

RESUMEN

DML es una compañía dedicada a la venta directa. Esta se encuentra en Perú, Colombia, Ecuador, Bolivia y Guatemala. En el 2018 en Perú se obtuvo el 45% de ventas de todos los países donde se encuentran, quedando así en el primer lugar del nivel de ventas. En Perú DML cuenta con 7 líneas de productos. Para el 2018 las 3 líneas que más ventas tuvieron fueron: Ropa exterior, hogar y joyería respectivamente con 40%, 32% y 10%. Las dos primeras líneas cuentan con modelos de pronósticos por campaña, sin embargo, para la línea de joyería no se tiene ninguna. Por ello también para esta categoría, en el mismo año se presentaron costos por pedir de más de 7 millones de soles y costos por pedir de menos de 4.4 millones de soles. DML vende los productos por campañas. Se tienen 18 campañas por año. La mayoría de los productos ofrecidos por campaña son nuevos y solo el dos por ciento se vuelve a vender. En este contexto usar modelos de pronósticos de series de tiempo no es correcto ya que no se cuenta con suficiente información histórica de la demanda de los productos ofrecidos por campaña. Por otro lado, sí se podría usar algoritmos de machine learning como random forest. Para validar los resultados del modelo de pronóstico se hace uso de MAPE y se valida también mediante pruebas de hipótesis si la media o mediana de la diferencia en los valores pronosticados y reales son iguales o diferentes de cero. Mediante la implementación del nuevo modelo de pronóstico y la ejecución de newsvendor se demostró mejoras en los costos por pedir de más y de menos del año 2018 al 2019. El costo de oportunidad total se redujo en 7.9 millones de soles. La utilidad bruta para el 2019 mejoró en 665,105 soles con un capital inversión en los productos de 11 millones de soles menos y teniendo para el 2019 casi la misma de cantidad de SKU's ofrecidos al mercado.

PALABRAS CLAVES:

Venta directa; Joyas; Pronósticos; Machine learning; Random Forest; Productos de lanzamiento; Newsvendor

ABSTRACT

FORECAST OF DEMAND FOR PRE-ORDERS AND ORDERS OF CAMPAIGN FOR COSTUME JEWELRY IN DML

DML is a company dedicated to direct sales. They are in Peru, Colombia, Ecuador, Bolivia and Guatemala. In 2018, Peru obtained 45% of sales of all the countries where they are located, thus placing it in first place in terms of sales. In Peru DML has 7 product lines. For 2018, the 3 lines that had the most sales were: Outerwear, home and jewelry respectively with 40%, 32% and 10%. The first two lines have forecast models per campaign, however, for the jewelry line there is none. Therefore, also for this category, in the same year there were ordering costs of more than 7 million soles and ordering costs of less than 4.4 million soles. DML sells products by campaigns. There are 18 campaigns per year. Most of the products offered per campaign are new and only two percent are resold. In this context, using time series forecasting models is not correct since there is not enough historical information on the demand for the products offered per campaign. On the other hand, machine learning algorithms such as Random Forest could be used. To validate the results of the forecast model, MAPE (mean absolute percentage error) is used and it is also validated through hypothesis tests if the mean or median of the difference in the predicted and actual values are equal to or different from zero. Through the implementation of the new forecasting model and the execution of newsvendor, improvements were demonstrated in the costs of over and under-ordering from 2018 to 2019. The total opportunity cost was reduced by 7.9 million soles. The gross profit for 2019 improved by 665,105 soles with a capital investment in the products of 11 million soles less and having for 2019 almost the same number of SKUs offered to the market.

KEYWORDS:

Direct sale; Jewelry; Forecast; Machine learning; Random Forest; Launch products; Newsvendor

INTRODUCCIÓN

En Perú, a setiembre del 2019, se tienen 19 empresas que emplean el sistema de venta directa para comercializar sus productos [1]. Este rubro de la industria se define por la Cámara Peruana de Venta Directa (CAPEVEDI) como un modelo de negocio que ofrece oportunidades a personas independientes para comercializar y realizar la venta de productos y servicios, por lo general fuera de un establecimiento comercial fijo [1]. Entonces, los que tienen el contacto final con los clientes son estas personas independientes, que a su vez son conocidos como asesores de ventas. Ellos acuerdan y realizan la venta de los productos que la empresa produce o compra a terceros.

Según una publicación del diario Gestion.pe en el 2019 [2] durante los tres primeros meses de ese año el rubro prosperó por encima de la economía nacional, adquiriendo un crecimiento de 4.9% en ventas en relación a semejante periodo del 2018, con lo cual este negocio facturó 1,071 millones de soles en los tres primeros meses del 2019 y además se reportó una ampliación de 2.5% en la cantidad de emprendedores independientes en este rubro, teniendo al 2019 a 613,936 emprendedores.

Una de las empresas pertenecientes a este rubro es DML, como se le llamará para la presente investigación a esta empresa dedicada a la venta directa. La empresa divide la gama de sus productos en 7 líneas: Ropa exterior, Ropa interior, Fragancias, Johnson & Johnson, Joyería, L'Oréal y Hogar. La empresa está presente en cinco países Perú, Colombia, Bolivia, Ecuador, y Guatemala. Las ventas registradas de DML en los países antes mencionados durante el 2018 con respecto al total fueron de 45 %, 32 %, 8 %, 10 %, y 5 % respectivamente. La presente investigación se desarrollará en DML Perú.

Los asesores de DML generan ventas mediante catálogos físicos por períodos denominados campañas, las cuales tienen un tiempo de duración de 18 días. El tiempo de pasar de una campaña a otra es de 2 a 3 días. Los productos que se venden en cada campaña son diferentes y en su mayoría no se repiten.

DML tiene cobertura a nivel nacional. La compañía referencia a sus asesores de ventas agrupándolos por zonas. En Perú se tienen 113 zonas. Una asesora solo pertenece a una zona.

En una campaña, los asesores de venta de DML envían sus órdenes de pedidos de la presente campaña e intenciones de compra (pre-pedidos) de la siguiente campaña en una fecha específica. El primero hace referencia a las órdenes de ventas que los asesores generan de aquellos productos que son de la presente campaña y tienen pactado con los clientes su venta. El segundo hace referencia a las intenciones de ventas de los productos de la siguiente campaña.

Para recibir las órdenes de pedidos e intenciones de compra de todas estas zonas, ya mencionadas, de forma ordenada durante una campaña la empresa ha establecido días de corte. En la empresa se definen 11 cortes por campaña, como se muestra en la figura 1. Cada corte de campaña está presente en promedio después de un día.

En cada uno de estos cortes se tiene asignado a cierto número de zonas. Los asesores de ventas que pertenezcan a una zona, y a su vez esta zona esta asignada a un corte, tienen que enviar sus pedidos e intenciones de compra en esa fecha de corte como se muestra en la figura 1.

Las asesoras de venta están agrupadas por zonas a nivel nacional. Estas zonas en la figura 1 están representadas por los puntos rojos. Las zonas que están dentro del círculo 1 deben enviar su solicitud de pedidos e intenciones de compra en la fecha correspondiente al corte 1 y las otras zonas numeradas por los círculos en el mapa deben realizarlo en los cortes correspondientes al número que pertenecen.

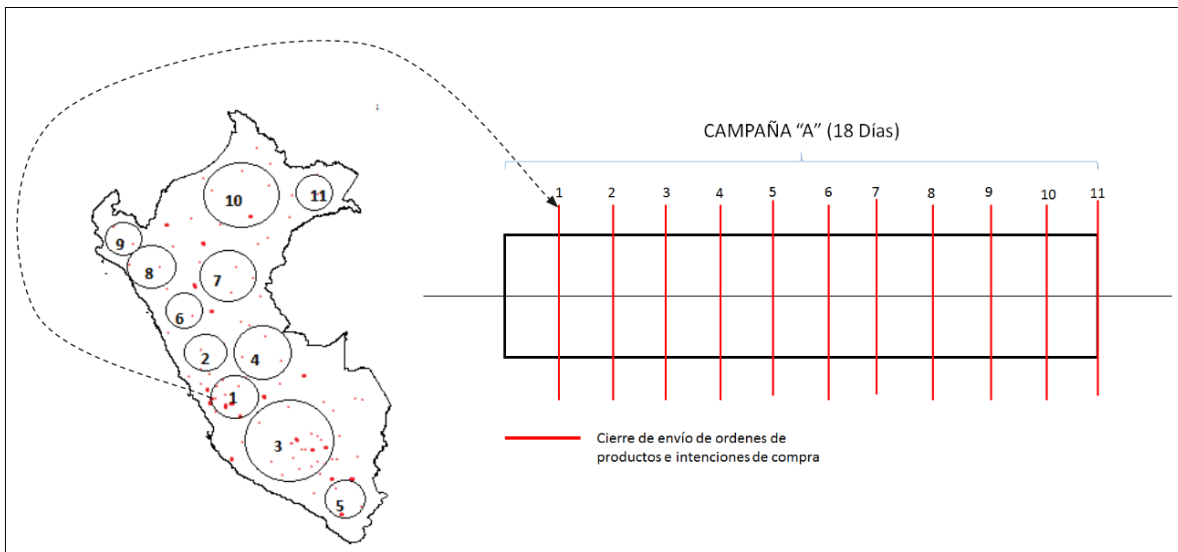


Figura 1. Número de cortes en una campaña

Entonces, como se explicó líneas anteriores un asesor hasta antes de la fecha de corte que le corresponde tiene para realizar ventas de la campaña en la cual se encuentra y recabar información de las intenciones de compra de los clientes para la siguiente campaña. Ya en la fecha de corte que le corresponde envía sus pedidos de la presente campaña e intenciones de compra de la siguiente a la empresa. Pasado la fecha de corte el asesor trabajará los pedidos de la siguiente campaña y las intenciones de compra de la tras próxima así como se muestra en la figura 2.

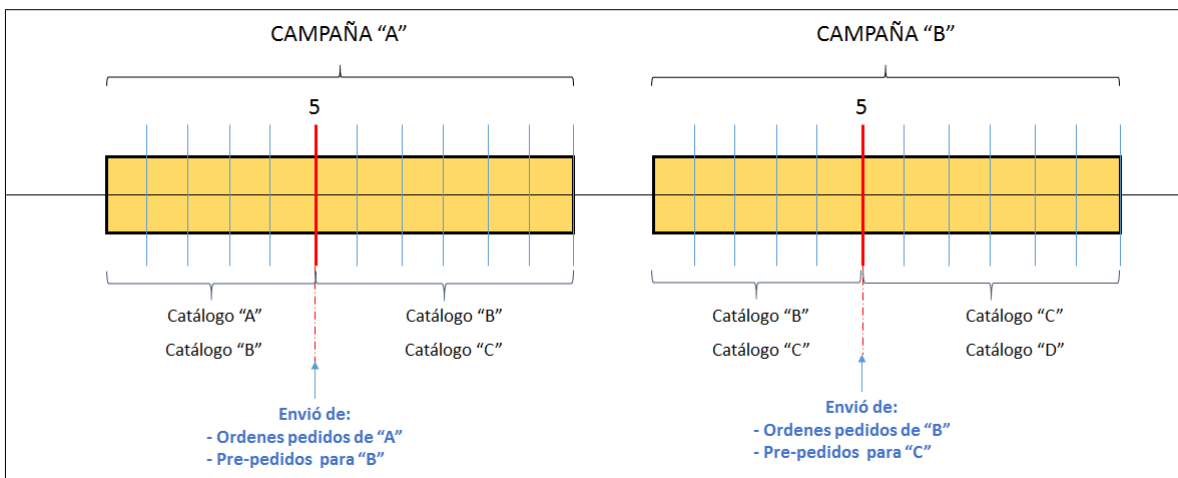


Figura 2. Envío de pedidos y pre pedidos por corte correspondiente a cada zona.

En el 2018 las tres líneas de productos que más ventas tuvieron fueron Ropa Exterior, Hogar y Joyería las cuales representaron el 40%, 32% y 10% de la venta total, lo que viene a ser 139,107,455; 111,578,380 y 34,279,455 soles respectivamente. Estos números ya consideran el desembolso de dinero de las comisiones a las asesoras. La presente investigación se desarrollará en base a la línea de Joyería ya que de las tres es la única que no presenta un modelo de pronósticos.

Las áreas que manejan las operaciones de DML Perú son Ventas, Administración y finanzas, Marketing, Operaciones, Logística y Recursos humanos. Dentro del área de Operaciones se encuentra el grupo de Modelos Estadísticos quienes se encargan de predecir las ventas, en cantidades, de todas las categorías de productos para cada campaña. Las predicciones calculadas son derivadas al área de Compras quienes finalmente las convierten en órdenes de pedidos para los proveedores. El área de Modelos Estadísticos evalúa su rendimiento predictivo por temporada mediante el siguiente indicador: ventas reales en unidades entre ventas pronosticadas en unidades.

En el 2018 para línea de Joyería la cantidad de productos vendidos fue de 1,228,329 y los artículos que sobraron en el almacén y faltaron abastecer a los asesores de venta después de una campaña acumularon un total de 514,000 y 290,000 respectivamente. El precio de venta promedio de los productos a las asesoras es de 29 soles y el costo de compra promedio es de 13.63 soles. Por lo tanto, el costo de pedir de más por campaña acumuló un total de 7,005,820 soles y el costo pedir de menos en cada campaña acumuló un total de 4,441,930. La suma de ambos montos representó el 63% del margen de contribución total de la línea de Joyería en el 2018.

El área de compras define tres momentos para colocar las órdenes de requerimientos de productos de la línea de Joyería a sus proveedores y así abastecer al almacén de la empresa de productos para una campaña. A continuación, se explicará los tres momentos de colocación de órdenes para la línea de Joyería con la ayuda de la figura número 3.

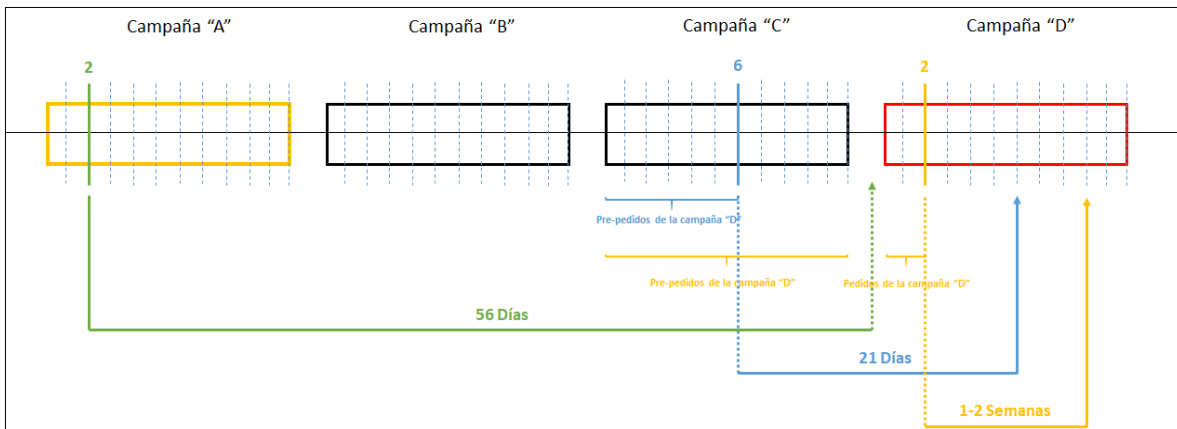


Figura 3. Momentos de compra para línea de Joyería

El primer momento, anticipos, el área de Compras envía la orden de requerimiento de productos con 56 días de anticipación al inicio de la campaña que se desea abastecer. Para este momento no se tiene información de los pre-pedidos de la campaña que se desea abastecer. La intención de este momento es abastecer al almacén a lo mucho con el 65% del total de productos que se venderá durante esa campaña.

En el segundo momento, pre-pedidos, se genera la orden de requerimiento de productos en el corte 6 de una campaña antes a la que se quiere predecir, como se ve en la figura 3. Para este segundo momento se tiene información de las intenciones de compra de la campaña que se quiere pronosticar. La intención para este segundo momento es abastecer 25% adicional al primer momento.

En el tercer momento, nombrado pedidos, se envía la última orden de requerimientos de productos a los proveedores durante los primeros días de la campaña, en el corte número 2. Para este momento se tiene información de todas las intenciones de compra para la campaña que se quiere abastecer y también de los pedidos de esa campaña hasta el corte 2. La finalidad de este momento es abastecer lo faltante para cumplir con la demanda de la campaña que en promedio es el 10% de toda la campaña. A la fecha, junio del 2023, este tercer momento no se usa en todas las campañas del año según lo indica el área de Supply Chain de la compañía ya que con los dos primeros momentos se logra regularmente abastecer.

La presente investigación se desarrollará en base al momento de pre pedidos y se determinará los pronósticos de demanda por campaña para la línea de Joyería ya que en el momento de pre pedidos se tiene información de las intenciones de compra para la campaña a la cual se desea abastecer. Con la información de intenciones de compra y demás variables de mercadotecnia se puede pronosticar la demanda real por campaña de los productos, lo cual es uno de los objetivos principales de la presente investigación que se explican en el siguiente párrafo.

La presente investigación tiene como objetivo principal disminuir la cantidad faltantes y sobrantes anuales. Asimismo, la presente investigación para cumplir con tal objetivo tiene como propósito seleccionar variables de catálogo para usar en el modelo de pronóstico, definir un modelo de pronóstico en base a las características del producto, implementar un modelo de pronósticos por campaña para la línea de Joyería, validar el modelo de pronóstico para la línea de Joyería y determinar las cantidades a solicitar a los proveedores mediante newsvendor.

CAPÍTULO I

MARCO TEÓRICO

1.1. Venta directa

1.1.1. Definición de la venta directa

La venta directa es definida como la venta de productos y prestaciones de servicios fuera de locales comerciales físicos, generalmente en hogares, reuniones o lugares de trabajo donde las transacciones son iniciadas y concluidas por un vendedor [3], [4]. Estos productos son ofrecidos mediante catálogos físicos [1]. Asimismo, estos asesores de ventas obtienen una comisión por las ventas que realizan [5].

1.1.2. Categorías de productos

En la siguiente tabla se muestra las categorías de productos que son ofrecidos mediante la venta directa.

Categoría	Productos
Personal	Cosméticos, fragancias, cuidado de la piel, joyas, ropa para damas y niños y otros artículos personales
Alimentos	Alimentos congelados, suplementos nutricionales y planes de dieta
Bienestar	Suplementos nutricionales, planes de dieta, aromaterapia y otros productos.
Hogar	Decoración del hogar, muebles, utensilios de cocina, limpieza del hogar y del automóvil.
Familia	Libros, juguetes, juegos, audio, video, productos financieros, otros productos para el hogar y ayudas comerciales.
Servicios	Telecomunicaciones y utilidades.

Tabla 1. Categoría de productos en la venta directa [3]

1.1.3. Ventajas y desventajas de la venta directa

En la siguiente tabla se muestran las ventajas y desventajas por parte de las compañías que han adoptado este tipo de canal de distribución a través de dos fuentes [4], [6].

Ventajas	Desventajas
<ul style="list-style-type: none"> - Mayor cobertura de mercado. - Genera alto valor a sus clientes. - Reducción de gastos en publicidad. - Precios de productos más bajos. En comparación si el producto se vendería en un centro comercial. - Eliminación de empresas intermediarias para la distribución y venta de los productos. 	<ul style="list-style-type: none"> - Alta variabilidad de la forma de venta por cada asesora. - Gestión de la demanda difícil de manejar. - Gasto en capacitaciones a las asesoras de ventas.

Tabla 2. Ventajas y desventajas de la venta directa [4], [6].

En la presente tabla se destaca la eliminación de los intermediarios para llegar al usuario final en la cadena de suministros de la compañía. Al tener este beneficio la empresa puede generar mayores márgenes de ganancia por producto ya que sus costos de venta se verán reducidos [6].

1.2. Variables de mercadotecnia en la venta directa

Como se explicó líneas anteriores los productos en la venta directa son vendidos a través de catálogos y por campañas. Por cada campaña se ofrecen nuevos productos que no tienen historial de ventas entonces para realizar los pronósticos de demanda por campaña las compañías toman en consideración variables de mercadotecnia [5]. A continuación, se muestran la mayoría de las variables de mercadotecnia que se usan en la industria para realizar pronóstico por producto.

Variable	Descripción
Precio	Indica el precio de venta del producto en la campaña.
Descuento	Indica si el producto tiene o no descuento.
Número de asesoras	Indica el número de asesoras inscritas para la presente campaña.
Exposición	Indica si el producto es mostrado en página entera o parcialmente a ella.
Familia de producto	Grupo de productos con características comunes.
Tonalidades	Representa la tonalidad del producto en el catálogo.
Características	Según [5] son atributos de los productos, por lo general en diseño y su uso supedita del tipo de empresa en el rubro.
Número de catálogo	Es el número que tiene asignado en el año el catálogo.
Semana	Corresponde a la semana del año en el que se encuentra el producto en venta.
Mes	Hace referencia al mes en el cual se encuentra el producto en venta.
Oferta	Hace referencia a si el producto tiene una oferta o no. En ocasiones también puede ser expresada en porcentajes.
Variable promoción canibalizadora	Explica si la demanda de dos o más productos en un mismo catálogo son afectados por tener relación en sus promociones.

Tabla 3. Variables de mercadotecnia en la venta directa [5], [7]

En la tabla anterior se mostró la mayoría de las variables de mercadotecnia usadas en la industria. La selección de uso de ellas dependerá de cada compañía y sus estrategias correspondientes. Para el caso del presente proyecto, se elegirán algunas de ellas y otras que

no se encuentran en la tabla anterior, debido al contexto del problema que se desea resolver. Líneas adelante se explica el uso y significado de estas variables elegidas.

Con respecto a los modelos frecuentes usados para la predicción en la industria de venta directa que hacen uso de variables de mercadotecnia, dos de las más usadas son: Regresión Lineal Múltiple y Modelo Lineal Dinámico Bayesiano. Además, el modelo de regresión se considera como la columna vertebral para los demás modelos posteriores. [7]

1.3. Pronósticos de demanda

Los pronósticos en las industrias son utilizados para predicciones climáticas, generación de presupuesto, pronósticos financieros, predicciones de precios de activos, etc. [8]. En la presente investigación se dará enfoque a los pronósticos de demanda de productos.

Los pronósticos de demanda de productos juegan un papel importante dentro de una compañía dado que un pronóstico bueno o malo puede afectar a toda la organización [9]–[11].

A continuación, se muestra una tabla con algunas áreas en las cuales el pronóstico cumple un papel importante.

Áreas	Actividades
Programación	Programación de producción, transporte, personal, presupuestos, etc.
Adquisición de recursos	Compras de materias primas e insumos, contratación de personal, compra y alquiler de maquinaria, etc.
Determinación de requisitos de recursos	Determinación de recursos a corto, mediano y largo plazo. Ejemplos: recursos financieros, humanos, de productos y tecnológicos.

Tabla 4. Actividades relacionadas a la generación de pronósticos [9]

Ahora bien, para determinar los pronósticos de demanda existen métodos de pronósticos que están divididos en cuantitativos y cualitativos [9]. En la tabla 5, se resume este esquema de categorización.

<p>Cuantitativo: Se tiene suficiente información cuantitativa</p> <ul style="list-style-type: none"> - Series temporales: Predicen la continuación de patrones históricos como el crecimiento de las ventas - Explicativo: comprender cómo las variables explicativas como los precios y la publicidad afectan las ventas.
<p>Cualitativo: Hay poca o ninguna información cuantitativa disponible, pero existe suficiente conocimiento cualitativo.</p>
<p>Impredecible: poca o ninguna información está disponible.</p>

Tabla 5. Métodos cuantitativos, cualitativos e impredecibles [9]

La literatura indica que los modelos de series temporales están compuestos por cinco componentes: Nivel, tendencia, variaciones estacionales, movimientos cíclicos y fluctuaciones aleatorias [8]–[11].

A continuación, se resume en una tabla algunos modelos de series temporales con los componentes que cada uno de ellos toman en consideración.

Modelos	Nivel	Tendencia	Estacionalidad	Fluctuaciones aleatorias	Movimientos cíclicos
Promedio móvil simple	x			x	
Suavización exponencial simple	x			x	
Suavización exponencial con tendencia	x	x		x	
Holt-Winters	x	x	x	x	

Tabla 6. Modelos de pronósticos de series de tiempo con sus componentes respectivos [8], [9]

Tanto [8], [9] coinciden que el modelo Holt-Winters puede ser usado para pronósticos agregados de productos, mientras que los otros mostrados en la tabla están orientados a pronósticos individuales de productos. En la sección 1.3.7.3. se explicará con más detalle el modelo Holt-Winters.

1.3.1. Pronósticos de demanda desagregado (bottom-up)

Un método simple de pronóstico ascendente que implica primero generar pronósticos para cada producto (SKU) y luego sumarlos para producir pronósticos de una familia o grupo de productos. Una de las ventajas de este enfoque es que se pronostica en el nivel inferior de una familia de productos y, por lo tanto, no se pierde información debido a la agregación [12].

1.3.2. Modelos de pronósticos desagregado en la venta directa

En esta sección se mostrarán tres modelos de pronóstico desagregado que tienen funcionalidad cuando se tiene poca información de la demanda histórica de un producto [13].

1.3.3. Modelo dinámico bayesiano

En resumen, el modelo de pronóstico dinámico bayesiano inicia de la determinación de una distribución de probabilidad a priori para el o los parámetros y de otra distribución

para los datos, adquiriendo con la multiplicación de ambos la distribución de probabilidad a posteriori, con ello se discierne la función predictiva [13].

1.3.4. Árboles de decisión.

El enfoque de árboles de decisión (CART) fue definido en 1984 por Breiman [14]. Este es un algoritmo de aprendizaje supervisado, por lo general de uso en propósitos de clasificación donde las variables independientes y dependiente pueden ser continuas o categóricas [14]. Este algoritmo fracciona la muestra o población en conjuntos homogéneos en función a la variable independiente más significativa [14]. Asimismo, según Orellana [14] “la construcción del árbol sigue un enfoque de división binaria recursiva (top-down greedy approach)”.

En la siguiente imagen se muestra la construcción de un árbol de decisión.

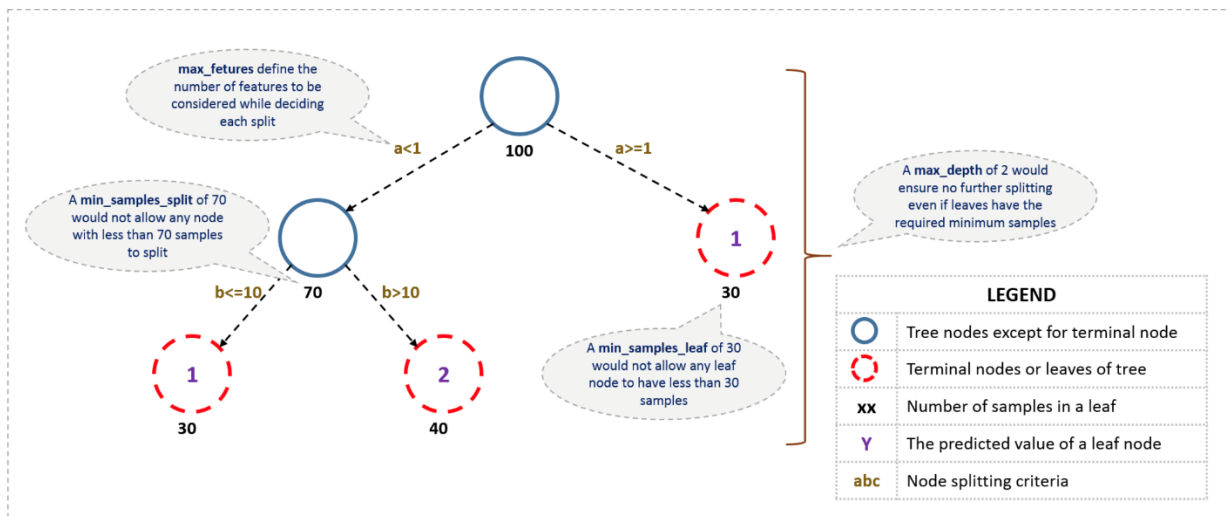


Figura 4. Construcción de árbol de decisión [14]

En la presente imagen se puede observar que el árbol fue construido con un set de datos de 100 muestras donde la variable predictiva (de entrada) elegida para el primer nodo fue “a” y para el segundo, “b”, asimismo, de color amarillo se muestra los criterios de ramificación para cada uno de ellos y los círculos de color azul son los nodos elegidos. Las variables predichas (Ys) se muestran en los nodos finales con sus valores correspondientes,

los cuales están representados de color morado dentro de los círculos rojos con líneas discontinuas. [14]

A continuación, también, se muestra una tabla con las ventajas y desventajas de este algoritmo.

Ventajas	Desventajas
<ul style="list-style-type: none"> - Sencillo de interpretar [14]. - Conveniente en determinar importancia de variables y exploración de datos viendo cientos de variables [14]. - Menor purificación de datos: datos faltantes y outliers no influyen en el modelo [14]. - Se puede usar variables categóricas y continuas. - Es un método no paramétrico. 	<ul style="list-style-type: none"> - Fallas en la predicción por sobreajuste. - Extravío de información al categorizar variables de tipo continuo (tiempo, precio, peso, etc.) [14]. - Volubilidad: un cambio minúsculo en los datos puede cambiar considerablemente la configuración del árbol, por ello, la explicación no es tan clara como parece. - No muy bueno para predecir. Modelo como SVM y clasificadores tipo ensamblador por lo general poseen tasas de exactitud más altas [14].

Tabla 7. Ventajas y desventajas de CART [14]

En el anterior recuadro se pudo observar las ventajas y desventajas de CART. Entre la que más se destacan son: es fácil de entender y es un algoritmo no paramétrico. Tomando en consideración este último, no es necesario realizar transformaciones a la variable dependiente para una posible validación del método.

También, en la anterior tabla se hizo mención del término ensamblador. Este término significa grupo. Los métodos o modelos de este tipo están formados por un grupo de modelos predictivos los cuales brindan un mejor performance en estabilidad y precisión del método [14]. Por ejemplo, ellos ofrecen una mejora determinante al modelo de CART [14]. Más adelante también se menciona este término ya se explicará un modelo que pertenece a este tipo.

Con respecto a los métodos por los cuales el árbol decide donde ramificarse, existen distintos tipos de ellos. Algunos son: Índice Gini, Chi Cuadrado, Ganancia de la información y Reducción en la varianza [14]. En el caso de uso del paquete “rpart” en RStudio, este usa el método de Gini por defecto.

El modelo de árbol de decisión divide el set de datos en una parte de entrenamiento y otra de testeo. Este set de datos por lo general contiene a las variables independientes o causales y la variable dependiente u objetivo. En la siguiente imagen se tiene una representación gráfica de esto.

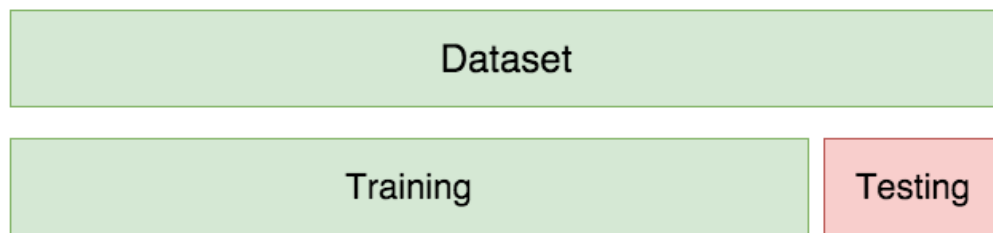


Figura 5. Conjunto de entrenamiento y de test para CART [14]

Por lo general según las literaturas el conjunto de entrenamiento corresponde al 80% de toda la información y el 20% se usa en la fase de testeo. La parte de entrenamiento tiene la intención de encontrar patrones (efectos de las variables independientes a la variable dependiente) que generalicen el modelo y con el conjunto de testeo se prueba el modelo ajustado para determinar el performance de este. [14]

Finalmente, es preciso agregar, también, que CART es una fuente base para el algoritmo de machine learning, Random Forest [15].

1.3.5. Random forest.

Random forest (RF) es un algoritmo de machine learning que se ha convertido en una herramienta valiosa para la predicción en distintos campos de la industria [15]. Así como se explicó en la sección anterior, RF es la consecuencia de generar varios árboles de decisión.

Esta herramienta tiene la capacidad de aprendizaje automático a través de poca información previa [16].

Adicionalmente, RF es un método de tipo ensamblador los cuales están compuestos de un conjunto de modelos predictivos que facilitan lograr una mejor estabilidad y precisión del modelo, lo cual provee un progreso significativo a los modelos de CART [14].

Los modelos de árboles de decisión, así como todos, tienen de problemas de sesgo y varianza. Por ejemplo, al elaborar un árbol corto se conseguirá un modelo con alto sesgo y baja varianza [14]. Tradicionalmente, al aumentar la dificultad del método, se notará una disminución en el error de predicción gracias a un sesgo menor en el modelo [14]. Pero, en cierto punto el modelo se tornará muy complejo y se generará un sobreajuste del modelo el cual comenzará a experimentar de varianza alta [14]. Es por ello por lo que con el uso de métodos de tipo ensamblador se construyen modelos óptimos que buscan mantener un equilibrio entre estos dos tipos de inconvenientes [14].

Mediante la siguiente figura se explica el equilibrio que se debería mantener para evitar estos dos tipos de errores (sesgo y varianza).

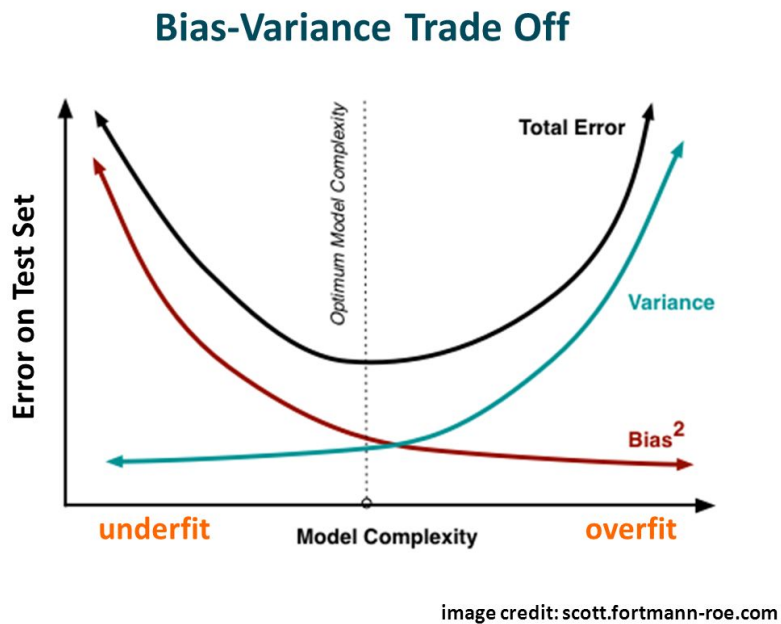


Figura 6. Equilibrio entre error de sesgo y varianza [14]

En la imagen anterior se muestra el “trade off” entre los errores de sesgo y varianza. El uso de Random Forest al ser de tipo ensamblador ofrece este “trade off”. [14]

RF es útil para regresión y clasificación. Este método construye varios árboles de decisión. Lo cual lo diferencia de CART por que genera múltiples árboles. Para el caso de clasificación se toma la moda de la salida de los árboles y para regresión se realiza un promedio de ellas. [14], [15]

En la siguiente imagen se explica la construcción de RF a partir de varios árboles de decisión.

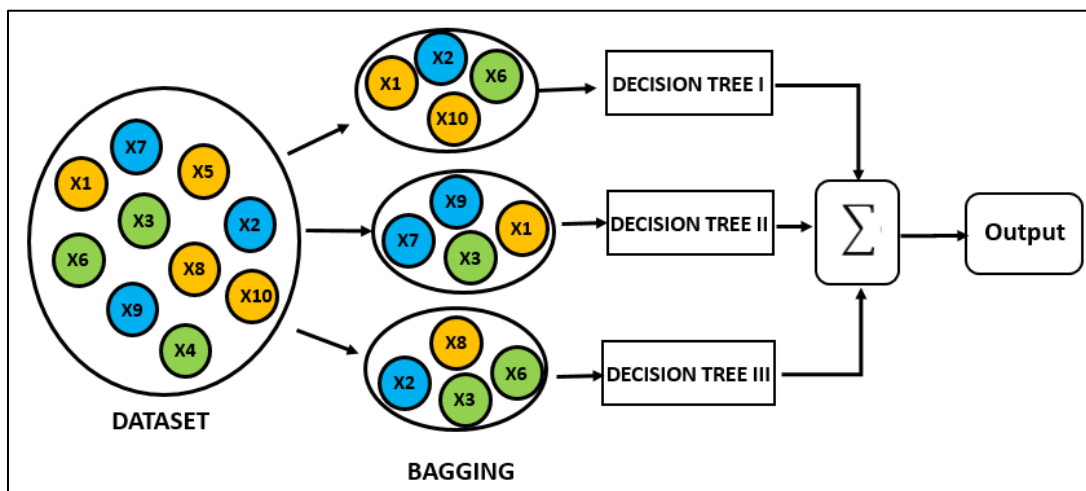


Figura 7. Construcción de RF [14]

Se muestra un conjunto de datos y a partir de ellos se crean subconjuntos aleatorios para generar varios árboles de decisión y múltiples modelos para después combinar los resultados de estos clasificadores con el objetivo de reducir la varianza de las predicciones y generar un mejor modelo. Para el caso de RF para regresión, donde la variable “target” es continua, se toma el promedio de las predicciones de todos los árboles y en el caso de clasificación se toma en cuenta la moda. [14], [15], [17]

A continuación, también, se muestra una tabla con algunas ventajas y desventajas de RF.

Ventajas	Desventajas
<ul style="list-style-type: none"> - Tienen la posibilidad de gestionar hasta miles de variables de entrada y denotar las más significativas [14]. Método de reducción de dimensionalidad [14]. - Es posible usarlo como método no supervisado (clustering) y detección de outliers. - También al igual que CART es un algoritmo no paramétrico. - Puede trabajar con registros faltantes. 	<ul style="list-style-type: none"> - Difícil de interpretar. - Según Orellana [14] “en regresión no puede predecir más allá del rango de valores del conjunto de entrenamiento”. - Bajo control en lo que realiza (modelo caja negra para modeladores estadísticos) [14].

Tabla 8. Ventajas y desventajas de Random Forest [14]

En la tabla anterior se mostraron las ventajas y desventajas de RF. Entre las que más se destacan están: primero que se puede trabajar con base de datos que contengan celdas vacías y segundo que es un método no paramétrico, no requiere validación sobre la normalidad de los datos.

Para la construcción de los árboles de RF, primero debe dividirse aleatoriamente la base de datos en dos partes. Una de las mismas se usará como datos de entrenamiento para el aprendizaje y la otra como validación para aprobar el nivel de aprendizaje. Todo ello se realiza con el propósito de conseguir los parámetros que mayor eficiencia de pronóstico pueda entregar el modelo. Ahora bien, en la literatura sugieren separar la parte de entrenamiento y validación en 2/3 y 1/3 o 80% y 20% respectivamente. [14], [18]

Como se explica en líneas anteriores, al igual que en Árboles de Decisión, para RF también el conjunto de datos se divide en dos partes (entrenamiento y testeo) como se muestra en la siguiente imagen.



Figura 8. Conjuntos de entrenamiento y de test para Random Forest [14]

Esta división en el conjunto de datos ayudara a que en la parte de entrenamientos se generen varios árboles de decisión y así el modelo pueda aprender a partir de ellos para luego validar el aprendizaje con el conjunto de testeo. A partir de esta validación se podrán definir mejores parámetros de aprendizaje para el modelo de RF. [14], [15], [17]

En la siguiente tabla se describen algunos de los parámetros que se usan en RF.

Parámetro	Descripción
Ntree	Número de árboles en el bosque.
Mtry	Número de variables aleatorias como candidatas en cada ramificación.
Sampsize	El número de muestras sobre las cuales entrenar. Valores más pequeños causarían la introducción de sesgo y reducción en el tiempo de cómputo y valores más grandes incrementarían el performance del modelo, pero con riesgo de originar sobreajuste [14]. Por lo general se encuentra en el rango de 60-80% [14].
Nodesize	Mínimo número de muestras dentro de los nodos terminales.
Maxnodes	Máximo número de nodos terminales.

Tabla 9. Parámetros de Random Forest [14], [17]

En la tabla anterior se observaron algunos de los parámetros de RF. En la presente investigación se considerará al menos dos de ellos.

1.3.6. Diferencia de resultados de pronósticos entre métodos tradicionales y machine learning.

Los siguientes autores [19]–[22] explican que no siempre los resultados de los métodos de Machine Learning han demostrado ser superiores a los métodos tradicionales como regresión lineal y series de tiempo.

Además, las comparaciones entre los modelos de machine learning y los métodos tradicionales no siempre son posibles en algunos casos: por ejemplo, uso de conjuntos de datos con registros faltantes, que pueden ser utilizados por algunos métodos de Machine learning, pero no por los métodos tradicionales de regresión [23].

1.3.7. Pronóstico de demanda agregado (top-down)

El enfoque de pronóstico de demanda de arriba hacia abajo indica primero generar pronósticos de toda la familia de productos y luego desglosarlo por SKU. Una de las principales desventajas de este tipo de pronóstico es que no se puede capturar y aprovechar las características de series individuales como la dinámica del tiempo, eventos especiales, etc. [12].

1.3.7.1. Modelos de pronósticos agregado en la venta directa

En esta sección se mostrarán dos métodos de pronóstico agregado en el sector de Venta directa. Estos métodos tienen funcionalidad a partir de contar con suficiente información histórica de demandas agregadas [8], [9].

1.3.7.2. Regresión múltiple

La principal ventaja de este modelo de pronóstico para el sector de venta de productos por catálogo es que hace uso de las variables de mercadotecnia. Por tal motivo este modelo es considerado como la columna vertebral de los pronósticos en el sector [7]. Adicionalmente, en otros campos como medio ambiental (“Green Mining”) y predicción de

concentrado de concreto este modelo es considerado como básico e inicial para los demás modelos complejos [24], [25].

El modelo de regresión múltiple está compuesto por una variable dependiente o también llamada variable “target” u objetivo, la cual se identifica como Y, y demás variables independientes identificadas como X_S . El modelo trata de determinar la relación entre la variable objetivo y las demás variables independientes. Se tienen varios objetivos de estudio para estas relaciones y algunas de ellas son: Analizar los efectos de los cambios de las X_S en Y, predecir el valor de Y dado un conjunto de X_S y examinar si las X_S tienen efecto significativo en la Y. Para el presente estudio se centrará el uso de regresión múltiple en el segundo caso. [26]

Un modelo de regresión múltiple puede ser representado por la siguiente ecuación:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots \dots \dots + \beta_k x_k + \varepsilon$$

Donde la variable dependiente es relacionada con k variables independientes a través de los parámetros $\beta_1, \beta_2, \beta_3 \dots \beta_k$. Los parámetros $\beta_1, \beta_2, \beta_3 \dots \beta_k$ son los coeficientes de la regresión asociados respectivamente a cada variable independiente y ε es el componente de error aleatorio que indica la diferencia entre la relación observada y ajustada. El error aleatorio debe tener una distribución normal con media 0 y varianza (σ). [18]

En el libro [26] también se explica que, en el caso habitual, cada coeficiente de regresión mide el cambio en la variable dependiente por incremento unitario en esa variable independiente, manteniendo constantes las otras variables X_S .

También, en ocasiones se requiere de realizar transformaciones a la ecuación del modelo de regresión. Esto se observa cuando se desea realizar inferencias sobre el modelo. A continuación, se muestra un modelo básico con una transformación en la variable dependiente.

$$T(x_t) = \ln x_t$$

$T(x_t)$, a diferencia de x_t , sería modelado y es la variable dependiente. Entonces, para cualquier valor dado de $T(x_t)$, el x_t correspondiente se obtendría invirtiendo la transformación, es decir, resolviendo la ecuación [8].

1.3.7.3. Método estacional de Holt-Winters

Holt y Winter amplificaron el método Holt para captar también estacionalidad, lo cual se ve en productos que tienen demandas estacionales como los helados o productos navideños. Este método o modelo Holt-Winters tiene función de pronóstico y 3 funciones de suavización [12], [27]. Una se tiene para el nivel (o constante) ℓ_t , una que describe la tendencia b_t y otra que explica el elemento o factor estacional S_t , con sus correspondientes parámetros α , β y γ [12], [27]. También, se hace uso de m para indicar la periodicidad de la estacionalidad, lo cual quiere indicar la cantidad de estaciones en un año [12], [27]. Por ejemplo, para datos trimestrales se usa m igual a 4 y para los mensuales m igual a 12.

Existen dos modificaciones para este modelo que cambian debido a la naturaleza del elemento estacional [12], [27]. Se hace uso del modelo aditivo si las modificaciones estacionales son cercanamente constantes a lo largo de la serie, por otro lado, se usa el modelo multiplicativo cuando las modificaciones estacionales cambian conforme al nivel de la serie [12], [27]. En el modelo aditivo, el termino estacional se expone medidas absolutas en la escala de la serie observada, y en la función de nivel, la serie se desestacionaliza disminuyendo el elemento estacional [12], [27]. Por cada año, el elemento estacional sumará aproximadamente cero [12], [27]. Para el modelo multiplicativo, el elemento estacional se muestra en términos de porcentajes, y la serie se desestacionaliza dividiéndola por el componente estacional y a diferencia del otro método, dentro de cada año, el componente estacional sumará aproximadamente m [12], [27].

- Método aditivo:

La ecuación para este método es:

$$\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$$

Donde:

$$\begin{aligned}\ell_t &= \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m},\end{aligned}$$

K es la parte entera de $(h-1)/m$, lo que indica que las estimaciones de los índices estacionales utilizados para el pronóstico provienen del último año de la muestra. La ecuación de nivel (ℓ_t) muestra un promedio ponderado entre la observación ajustada estacionalmente ($y_t - s_{t-m}$) y el pronóstico no estacional ($\ell_{t-1} + b_{t-1}$) para el tiempo t. Cabe señalar que la ecuación de tendencia es idéntica al método lineal de Holt. También, la formula estacional muestra un promedio ponderado entre el índice estacional actual ($y_t - \ell_{t-1} - b_{t-1}$) y el índice estacional de la misma temporada del año pasado (hace m periodos de tiempo). [12]

A menudo se expresa la ecuación para el componente estacional como:

$$s_t = \gamma^*(y_t - \ell_t) + (1 - \gamma^*)s_{t-m}$$

Si se sustituye ℓ_t de la ecuación de suavización por el nivel del componente anterior, se tiene:

$$s_t = \gamma^*(1 - \alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1 - \alpha)]s_{t-m}$$

Como se puede observar que esta última ecuación es idéntica a la ecuación de suavización para el componente estacional que se especifica como $\gamma = \gamma^*(1 - \alpha)$. El parámetro γ se encuentra entre 0 y 1 pudiendo tomar esos dos valores, lo cual se traslada a $0 \leq \gamma \leq 1 - \alpha$.

1.3.8. Diferencias entre top down y bottom up

El desarrollo de pronósticos de demanda tipo de arriba hacia abajo como de abajo hacia arriba dependen básicamente del objetivo que tenga la organización al obtener el pronóstico [28], [29]. En la siguiente tabla se muestran algunas de las principales funcionalidades para cada uno de estos tipos de pronósticos.

Actividades	Top down	Bottom up
Cronogramas de producción y distribución por SKU		x
Capacidad de flota de camiones para envíos a nivel nacional	x	
Contratación de mano de obra anual	x	
Estrategias de marketing por producto		x
Planes de venta por producto		x

Tabla 10. Funcionalidades por tipo de pronóstico Top down y Bottom up [28]

Es preciso agregar que según [29] los pronósticos de demanda de abajo hacia arriba tienden a ser más precisos.

1.3.9. Pronósticos de demanda de productos en el sector de venta directa

Los productos que se vende en el sector de venta directa se caracterizan por: ser productos de lanzamiento (nuevos productos), tener tiempos de reabastecimientos largos (de 30 a 60 días) y tener un ciclo de vida corto (de 15 a 30 días). [30], [31]

En la siguiente tabla se expone 3 documentos que sugieren sus propios métodos de pronósticos. También, se tiene dos métodos de pronósticos para productos de lanzamiento. Con base en las características del producto a pronosticar se toma en consideración una de ellas.

Referencia	Paper / Método	Características del producto				
		Productos nuevos	Tiempo de reabastecimiento largos	Ciclo de vida corto	Uso de variables de catalogo	Pronóstico desagregado
[32]	Designing a decision-support system for new product sales forecasting	✓				✓
[33]	Product Growth Models for Medium-Term	✓	✓			✓
[34]	Top-flop	✓	✓	✓		
[35]	Accurate Retail Testing of Fashion Merchandise - Methodology and Application	✓	✓	✓		
[16], [36]	Random Forest	✓	✓	✓	✓	✓

Tabla 11. Métodos de pronóstico para productos de lanzamiento por campañas [16], [32]–[36].

De los métodos y documentos expuestos en la tabla, para la presente investigación se considerará el método de Random Forest. Como se puede notar este método es el que más se ajusta al problema y a las características de los productos.

Adicionalmente, como se explicó en secciones anteriores Regresión Lineal Múltiple es considerada como el modelo inicial para pronósticos en esta industria [7] y Árboles de Decisión es el modelo previo y menos sofisticado que Random Forest [14], [15]. Por ello, antes de realizar los pronósticos por campaña se evaluarán entre estos tres métodos cuál de

ellos tiene el menor error en la parte de entrenamiento y se verificará y constatará el uso de Random Forest para los pronósticos por campaña.

1.3.10. Pronósticos de demanda y gestión de inventarios para productos de un solo periodo.

Dentro de los métodos de pronósticos de demanda y gestión de inventario de productos con un solo periodo de vida se tiene a “News Vendor”, el cual tiene como principal objetivo maximizar el beneficio esperado cuando se tiene una demanda aleatoria. Este método determina las cantidades a vender en una campaña en función a la distribución de la demanda del producto, el precio y el costo de esta. [37]–[39]

Una vez obtenido el pronóstico de demanda de un producto, el siguiente paso es obtener más información acerca del pronóstico. Se debe obtener una imagen completa del pronóstico y de la incertidumbre de este. Ello se obtiene con una función de distribución del pronóstico. [39]

Esta función de distribución empírica se construirá de acuerdo con el error de pronóstico histórico y relación histórica de demanda real sobre pronóstico de productos de campañas anteriores. Entonces en una tabla se tendrá datos de los productos de una campaña anterior con sus pronósticos, demanda real, error y A/F ratios (demanda real entre pronostico) como se puede ver en la siguiente tabla mostrada en el libro de Cachon [39].

Product Description	Forecast	Actual Demand	Error*	A/F Ratio**
JR ZEN FL 3/2	90	140	-50	1.56
EPIC 5/3 W/HD	120	83	37	0.69
JR ZEN 3/2	140	143	-3	1.02
WMS ZEN-ZIP 4/3	170	163	7	0.96
HEATWAVE 3/2	170	212	-42	1.25
JR EPIC 3/2	180	175	5	0.97
WMS ZEN 3/2	180	195	-15	1.08
ZEN-ZIP 5/4/3 W/HOOD	270	317	-47	1.17
WMS EPIC 5/3 W/HD	320	369	-49	1.15
EVO 3/2	380	587	-207	1.54
JR EPIC 4/3	380	571	-191	1.50
WMS EPIC 2MM FULL	390	311	79	0.80
HEATWAVE 4/3	430	274	156	0.64
ZEN 4/3	430	239	191	0.56
EVO 4/3	440	623	-183	1.42
ZEN FL 3/2	450	365	85	0.81
HEAT 4/3	460	450	10	0.98
ZEN-ZIP 2MM FULL	470	116	354	0.25
HEAT 3/2	500	635	-135	1.27
WMS EPIC 3/2	610	830	-220	1.36
WMS ELITE 3/2	650	364	286	0.56
ZEN-ZIP 3/2	660	788	-128	1.19
ZEN 2MM S/S FULL	680	453	227	0.67
EPIC 2MM S/S FULL	740	607	133	0.82
EPIC 4/3	1,020	732	288	0.72
WMS EPIC 4/3	1,060	1,552	-492	1.46
JR HAMMER 3/2	1,220	721	499	0.59
HAMMER 3/2	1,300	1,696	-396	1.30
HAMMER S/S FULL	1,490	1,832	-342	1.23
EPIC 3/2	2,190	3,504	-1,314	1.60
ZEN 3/2	3,190	1,195	1,995	0.37
ZEN-ZIP 4/3	3,810	3,289	521	0.86
WMS HAMMER 3/2 FULL	6,490	3,673	2,817	0.57

Tabla 12. Datos de pronósticos, demanda real, error y A/F ratios de productos [39]

Como se puede apreciar en la tabla anterior la relación A/F también expresa la exactitud de pronóstico de los productos de una campaña anterior, donde un valor de 1 expresa un pronóstico exacto. Dado ello la precisión de pronóstico de una campaña actual puede ser comparable a una anterior para aquellos productos los cuales la tarea de pronósticos sea comparable al producto de interés [39]. Luego, se ordena los datos en orden ascendente de la relación A/F y se clasifica del 1 hasta el último elemento del conjunto de datos. Después,

en una columna adicional se coloca el percentil correspondiente para cada producto como se puede observar en la siguiente tabla.

Product Description	Forecast	Actual Demand	A/F Ratio*	Rank	Percentile**
ZEN-ZIP 2MM FULL	470	116	0.25	1	3.0
ZEN 3/2	3,190	1,195	0.37	2	6.1
ZEN 4/3	430	239	0.56	3	9.1
WMS ELITE 3/2	650	364	0.56	4	12.1
WMS HAMMER 3/2 FULL	6,490	3,673	0.57	5	15.2
JR HAMMER 3/2	1,220	721	0.59	6	18.2
HEATWAVE 4/3	430	274	0.64	7	21.2
ZEN 2MM S/S FULL	680	453	0.67	8	24.2
EPIC 5/3 W/HD	120	83	0.69	9	27.3
EPIC 4/3	1,020	732	0.72	10	30.3
WMS EPIC 2MM FULL	390	311	0.80	11	33.3
ZEN FL 3/2	450	365	0.81	12	36.4
EPIC 2MM S/S FULL	740	607	0.82	13	39.4
ZEN-ZIP 4/3	3,810	3,289	0.86	14	42.4
WMS ZEN-ZIP 4/3	170	163	0.96	15	45.5
JR EPIC 3/2	180	175	0.97	16	48.5
HEAT 4/3	460	450	0.98	17	51.5
JR ZEN 3/2	140	143	1.02	18	54.5
WMS ZEN 3/2	180	195	1.08	19	57.6
WMS EPIC 5/3 W/HD	320	369	1.15	20	60.6
ZEN-ZIP 5/4/3 W/HOOD	270	317	1.17	21	63.6
ZEN-ZIP 3/2	660	788	1.19	22	66.7
HAMMER S/S FULL	1,490	1,832	1.23	23	69.7
HEATWAVE 3/2	170	212	1.25	24	72.7
HEAT 3/2	500	635	1.27	25	75.8
HAMMER 3/2	1,300	1,696	1.30	26	78.8
WMS EPIC 3/2	610	830	1.36	27	81.8
EVO 4/3	440	623	1.42	28	84.8
WMS EPIC 4/3	1,060	1,552	1.46	29	87.9
JR EPIC 4/3	380	571	1.50	30	90.9
EVO 3/2	380	587	1.54	31	93.9
JR ZEN FL 3/2	90	140	1.56	32	97.0
EPIC 3/2	2,190	3,504	1.60	33	100.0

Tabla 13. Datos de A/F ratios clasificados por productos de una temporada anterior [39]

Después, de acuerdo con el ejemplo en el libro de Cachon [39] la distribución de demanda empírica para un producto de la campaña actual se construye con la multiplicación del pronóstico de este por cada una de las relaciones A/F de la tabla anterior. Donde la probabilidad de cada cantidad (A/F por pronóstico) es igual a la posición de clasificación

entre el total de elementos de la tabla. En la siguiente tabla se puede observar el resultado de este paso.

A/F Ratio	Q	F(Q)	A/F Ratio	Q	F(Q)	A/F Ratio	Q	F(Q)
0.25	800	0.0303	0.81	2,592	0.3636	1.23	3,936	0.6970
0.37	1,184	0.0606	0.82	2,624	0.3939	1.25	4,000	0.7273
0.56	1,792	0.0909	0.86	2,752	0.4242	1.27	4,064	0.7576
0.56	1,792	0.1212	0.96	3,072	0.4545	1.30	4,160	0.7879
0.57	1,824	0.1515	0.97	3,104	0.4848	1.36	4,352	0.8182
0.59	1,888	0.1818	0.98	3,136	0.5152	1.42	4,544	0.8485
0.64	2,048	0.2121	1.02	3,264	0.5455	1.46	4,672	0.8788
0.67	2,144	0.2424	1.08	3,456	0.5758	1.50	4,800	0.9091
0.69	2,208	0.2727	1.15	3,680	0.6061	1.54	4,928	0.9394
0.72	2,304	0.3030	1.17	3,744	0.6364	1.56	4,992	0.9697
0.80	2,560	0.3333	1.19	3,808	0.6667	1.60	5,120	1.0000

Tabla 14. Tabla de función de distribución empírica para un producto de la temporada actual [39]

En esta tabla se observa las cantidades correspondientes, Q , del producto de la campaña entrante con su respectiva probabilidad acumulada, $F(Q)$. $F(Q)$ es la probabilidad de que la demanda sea igual o menor a Q . Entonces, la anterior tabla expresa el pronóstico de demanda del producto de una campaña entrante en función de distribución empírica. Empírica porque esta función de distribución fue construida con los datos observados empíricamente. [39]

Adicionalmente, en la siguiente imagen se muestra gráficamente la distribución empírica de la tabla número 13 (probabilidad vs cantidad).

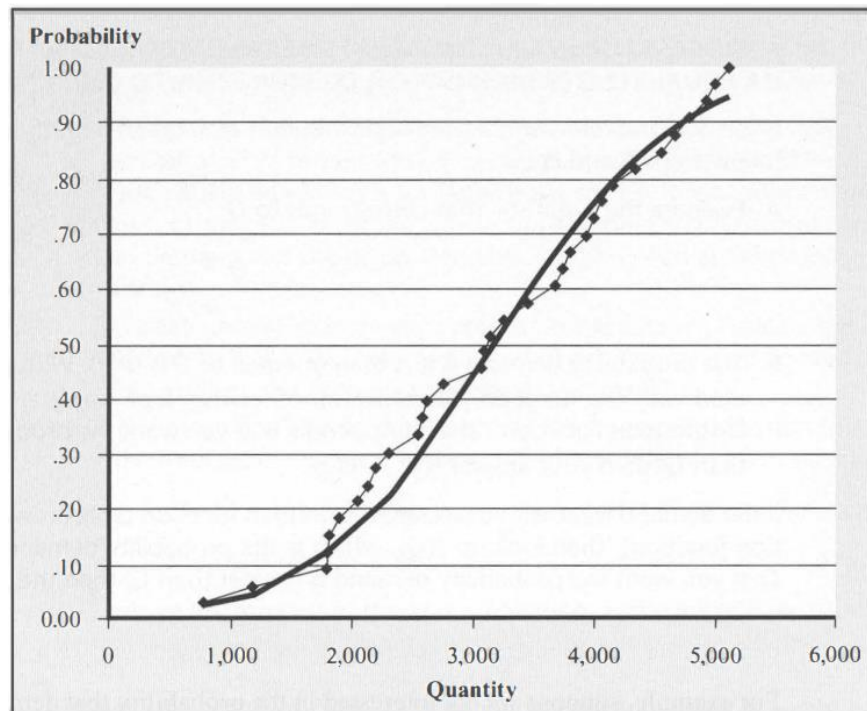


Figura 9. Función de distribución empírica del nuevo producto de la temporada entrante [39]

La anterior grafica muestra también como la distribución empírica se ajusta a una distribución normal por lo que esta distribución se podría considerar como una buena representación de la demanda del producto. [39]

Entonces, ahora se establecerá como se determinará, mediante la implementación de Newsvendor, la cantidad a solicitar dado las nuevas informaciones que se tiene: Precio de producto, costo de producto y distribución de probabilidad de demanda del producto. Como se indicó líneas anteriores Newsvendor busca maximizar el beneficio esperado cuanto tienen problemas de establecer inventarios de productos nuevos al mercado [39]. En ese sentido cuando se ofrecen productos por encima de la cantidad de demanda real se tiene costos por pedir de más o también llamado en inglés “overage cost” (C_o) y cuando se ofrecen cantidades por debajo de la demanda real se incurre en costos por de menos, “underage cost” (C_u). C_o y C_u indican los costos incurridos por unidad. Ante esta problemática, newsvendor trata de establecer una cantidad de inventario que equilibre estos dos tipos de costos para así maximizar el beneficio esperado del producto. Mediante la siguiente ecuación se busca lograr

este objetivo donde la cantidad Q determinada igualará la pérdida esperada en la unidad Q (parte izquierda de la ecuación) con la ganancia esperada en la unidad Q (parte derecha de la ecuación). [39]

$$C_o \times F(Q) = C_u \times (1 - F(Q))$$

$F(Q)$ como se indicó antes es la probabilidad de que la demanda sea menor o igual a Q . Resolviendo la ecuación anterior se tiene la siguiente relación.

$$F(Q) = \frac{C_u}{C_o + C_u}$$

La parte derecha de la ecuación se le denomina factor crítico. Entonces la cantidad que maximiza el beneficio esperado es la cantidad del pedido Q tal que la demanda es menor o igual a Q con probabilidad $C_u/(C_o + C_u)$. Dado ello, ahora se tiene una explicación más de por qué el pronóstico tiene que ser una función de distribución. [39]

Continuando con el ejemplo del mismo producto se establecerá la cantidad a solicitar que maximiza el beneficio esperado. El precio del producto es de 180 dólares, el costo es de 110 dólares y precio de salvamento de 90 dólares. Con estos datos se tiene: $C_u = 70$ y $C_o = 20$, entonces el factor crítico es de 0.7778. Buscando este valor de factor crítico en la tabla de la distribución empírica se tiene un Q de 4064 unidades. [39]

Mediante el mismo procedimiento mostrado en esta sección se buscará las cantidades para cada producto de todas las campañas del 2019 que maximicen el beneficio (utilidad bruta) esperado.

1.3.11. Pronósticos de demanda de ventas de productos por catálogo mediante juicio de experto.

En la literatura de pronóstico de demanda de productos nuevos mediante juicio de experto Michael [40], en su investigación, simula las decisiones de un experto mediante programas heurísticos y pone a prueba el éxito de este ante el modelo AHAF el cual es un modelo de pronóstico para productos por catálogo. El modelo AHAF también hace uso de

los programas heurísticos para optimizar las tomas de decisiones del experto. Finalmente, el autor compara los resultados y observa que, para su investigación, el AHAF tiene mejores resultados que el comprador experto.

1.7. Medidas de exactitud de pronósticos de demanda de productos.

En esta sección se mostrarán tres tipos de cálculos de exactitud de pronósticos de productos sugeridos por [8]. Los autores muestran MSE, “Mean Square Error”; MAD, “Mean Absolute Error” y MAPE, “Mean Absolute Percent Error”. Para el primero, a continuación, se presenta la formula correspondiente de cálculo.

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_{t-1,t})^2 = \sum_{t=1}^n (e_{t-1,t})^2$$

n representa la cantidad de productos. Entonces la formula anterior calcula la sumatoria de las diferencias al cuadrado del valor real menos lo pronosticado de los productos y todo ello lo divide entre la cantidad total de productos.

La fórmula de MAD se muestra a continuación.

$$\text{MAD} = \frac{1}{n} \sum_{t=1}^n |x_t - \hat{x}_{t-1,t}| = \frac{1}{n} \sum_{t=1}^n |e_{t-1,t}|$$

Al igual que en la anterior formula también se considera a n como la cantidad de productos. El cálculo de MAD es igual a la sumatoria de las diferencias absolutas entre el valor real de la demanda menos el valor pronosticado de los productos y todo ello lo divide entre n.

Ahora se muestra la fórmula de MAPE.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_{t-1,t}}{x_t} \right| = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{x_t} \right|$$

Aquí, también, se considera a n como la cantidad total de productos. El cálculo de MAPE es igual a la sumatoria de la división del error absoluto entre valor real de la demanda de todos los productos y todo ello se divide entre n .

CAPÍTULO II

METODOLOGÍA

2.1 Justificación

En relación con la tabla 11 se pudo argumentar que para la presente investigación se hará uso del método de clasificación computacional Random Forest (RF). Este método genera arboles de decisiones con base en las variables que son seleccionadas por el usuario. Al no tener información de la demanda de los productos de lanzamientos, debido a que son nuevos, este método realiza pronósticos con base en variables independientes seleccionadas. Es decir, realiza pronósticos para los productos por campaña tomando en cuenta a las características similares de productos anteriores. Estas características están definidas en las variables independientes X_s .

Como se indicó, además en el marco teórico antes de realizar los pronósticos finales por cada campaña se verificará entre los modelos de random forest, regresión lineal y arboles de decisión cuál de ellos tiene el mejor performance para ser usado en la campaña a predecir. En los resultados se mostrará una tabla con la exactitud de pronóstico de estos tres modelos y se escogerá el mejor para predecir en la campaña.

También, en los resultados se mostrará la exactitud de los pronósticos a través del uso de MAPE (“mean absolute percent error”). Este indicador de exactitud, que se definió en la última sección del marco teórico, el cual viene siendo usado para validación de exactitud de pronósticos generados por algoritmos de machine learning [18].

2.2 Reproducibilidad

Los pronósticos realizados en RF ya vienen siendo usados en áreas como la industria alimenticia [28]. A través de los pasos de la metodología de la presente investigación, este

método puede ser reproducible en otros campos y a su vez en compañías con contextos similares al que se está presentado.

2.3 Validez

En la presente investigación se determinará que variables son más significativas para la demanda cuando se realicen los modelos de RF para cada campaña del 2019.

También, se digitalará la variable Modelo, la cual indica si la joya mostrada en el catálogo esta con una modelo o no. La definición de esta variable se expuso en el marco teórico y al igual que las otras variables, que ya se encuentran digitadas en la base de datos de la empresa. Las definiciones de las variables se pueden apreciar en el artículo [4].

Asimismo, para determinar el “performance” de los modelos de RF por campaña se hará uso de MAPE. Adicionalmente, a ello se validará la exactitud de los modelos por campaña mediante pruebas de hipótesis sobre los errores. La implementación y justificación de estas validaciones se explican líneas adelante.

2.4 Definición del objeto de estudio

En la presente investigación se tiene como objeto de estudio a la base de datos de los productos vendidos en campañas anteriores, las cuales serán usadas para realizar pronósticos de productos de lanzamiento en la línea de joyas de la empresa DML. A continuación, se muestran las variables que serán utilizadas en el presente trabajo.

2.5 Variables independientes

Para el presente proyecto se toma en consideración las siguientes variables independientes de mercadotecnia (X_s). Estas buscan explicar la variable dependiente (Y), los cuales son las demandas por campaña de los productos de la línea de joyería.

- Subgrupo o familia de joyas: Grupo o familia a la cual pertenece el producto.
- Pre-pedidos del producto i : Cantidad en unidades de intenciones de venta del producto para una campaña.
- Número de campaña: Número de campaña a la cual pertenece el producto en el año.

- Precio del producto: Valor de precio unitario del producto en el catálogo.
- Número de página: Número de página donde se encuentra el producto en el catálogo.
- Modelo: Si el producto esta exhibido con una/un modelo o no.

2.6 Variable dependiente

A partir de las variables independientes explicadas en la anterior sección se busca la relación y efecto en la presente variable dependiente (Y).

- Demanda de producto: Demanda de un producto en unidades, la cual pertenece a una campaña.

2.7 Variables contextuales

Se presenta algunas variables contextuales que también podrían influir en el variable dependiente.

- Número de competidores: Hace referencia a la cantidad de competidores que se tiene en la industria local.
- Experiencia de la asesora: Se refiere a los años y experticia que tiene cada asesora en la industria.

2.8 Pasos de la metodología

A continuación, se explicarán los pasos de la metodología de la presente investigación.



Figura 10. Pasos de la metodología

- Selección de variables independientes:

A inicio de la metodología se deben seleccionar que variables serán usadas para la presente investigación. Las variables independientes son: Familia de producto, prepedidos, número de campaña, precio del producto, número de página y modelos, mientras que la variable dependiente es la demanda del producto por campaña. Asimismo, se va a digitar la variable “modelo” para los años 2018 y 2019. La cual indica si la joya esta presentada por una modelo o no en el catálogo.

- Entrenamiento de modelo de pronóstico:

Para el presente paso se toma en consideración como data de entrenamiento el 80% del tamaño de muestra y el otro restante será de validación. Todo ello con el objetivo de determinar los parámetros óptimos para el modelo de RF y validar el método/modelo a usar para predecir las demandas de la campaña en acción.

A continuación, se muestra un cuadro donde explica que muestras se van a usar para realizar los pronósticos de cada una de las campañas del 2019.

Campañas usadas como muestras	Campaña a pronosticar
Del 2018-1 al 2018-18	2019-1
Del 2018-1 al 2019-1	2019-2
Del 2018-1 al 2019-2	2019-3
Del 2018-1 al 2019-3	2019-4
Del 2018-1 al 2019-4	2019-5
Del 2018-1 al 2019-5	2019-6
Del 2018-1 al 2019-6	2019-7
Del 2018-1 al 2019-7	2019-8
Del 2018-1 al 2019-8	2019-9
Del 2018-1 al 2019-9	2019-10
Del 2018-1 al 2019-10	2019-11
Del 2018-1 al 2019-11	2019-12
Del 2018-1 al 2019-12	2019-13
Del 2018-1 al 2019-13	2019-14
Del 2018-1 al 2019-14	2019-15
Del 2018-1 al 2019-15	2019-16
Del 2018-1 al 2019-16	2019-17
Del 2018-1 al 2019-17	2019-18

Tabla 15. Campañas usadas como muestras

- Selección de modelo de pronóstico a usar:

Una vez definido el conjunto de entrenamiento y testeo por campaña, se procederá a validar el modelo a usar para pronosticar por campaña. En este caso, se definirá el modelo a usar entre: Random Forest, Regresión lineal Multiple y arboles de decisión de regresión. Si bien es cierto, en el marco teórico se explica que random forest debido a sus propiedades computacionales a priori es mejor para pronósticos, en esta sección también se validará mediante el indicador de exactitud MAPE que modelo es mejor para realizar los pronósticos

por campaña. En la sección de resultados se mostrará una tabla por campaña y modelo los valores de exactitud de pronóstico mediante MAPE.

- Pronóstico de demanda:

Una vez seleccionado el modelo a usar y obtenido los parámetros óptimos para RF, a través de iteraciones en RStudio, se procederá a realizar las predicciones. Cabe resaltar, que para el número de árboles que se va a tomar en consideración es 70. A partir de 70 árboles el error del modelo se mantiene constante. Esto último se observará cuando se realicen los modelos de RF gracias al comando “VarImPlot” en RStudio. Además, como se explica en [41] para los datasets que se trabajaron ahí, en ocasiones, una mayor cantidad de árboles en un bosque solo aumenta su costo computacional y no tiene una ganancia de rendimiento significativo. También, en esa investigación sugieren usar entre 64 y 128 árboles, para lo cual con esas cantidades indican que es posible obtener un buen balance entre el área promedio ponderada bajo la curva ROC (AUC – siglas en ingles), tiempo de procesamiento y uso de memoria.

- Validación de la exactitud de pronóstico:

Para validar y hallar los errores de los pronósticos desarrollados por cada campaña se hará uso de MAPE (“mean absolute percent error”). Silver [8] indica que MAPE es usado debido a que no se ve afectado por la magnitud de los valores de la demanda, ya que se expresa como un porcentaje. Sin embargo, el mismo autor añade que no es apropiado si los valores de demanda son muy bajos. Por ejemplo, una previsión de 1 unidad de demanda combinada con un valor real de 2 unidades muestra un error del 100%. El gerente debe decidir si es más deseable un error informado del 100% o de 1 unidad.

También, para determinar la exactitud de los pronósticos por campaña se realizará pruebas de hipótesis donde se validarán los errores (positivos y negativos). Para las 18 campañas se usarán pruebas “one sample Z” o pruebas de “Wilcoxon” para una muestra. Se escogerá una de esas dos pruebas dependiendo si los errores de esa campaña se distribuyen normalmente o no. Si los errores de una campaña se distribuyen normalmente se usará “one

sample Z”, de otra forma se usará prueba de “Wilcoxon” de una muestra. Es preciso indicar también, que en este caso se utilizará la prueba “one sample Z” y no “one sample t” debido a la cantidad de muestras por campaña, las cuales son mayores a 30. [42], [43]

Además, a ello se le suma determinar la cantidad de faltantes y sobrantes por cada campaña del 2019 y con ello también se hallará el valor monetario. Recordemos que el objetivo del presente trabajo es disminuir la cantidad de sobrantes y faltante para el 2019.

- Ejecución de newsvendor:

En este paso final, una vez obtenido el pronóstico de demanda por producto se determinará la cantidad a solicitar la cual maximicé el beneficio (utilidad bruta) esperado. Como se explica líneas anteriores el pronóstico es solo el “input” para la ejecución de newsvendor. Por ello dentro de este paso para la ejecución de newsvendor se tienen sub-pasos adicionales: Construcción de distribución empírica por producto, Calculo de factor crítico por producto y cantidad a solicitar que maximicé el beneficio esperado. [39]

En la siguiente imagen se muestra de manera gráfica estos sub-pasos adicionales dentro de la ejecución de newsvendor.

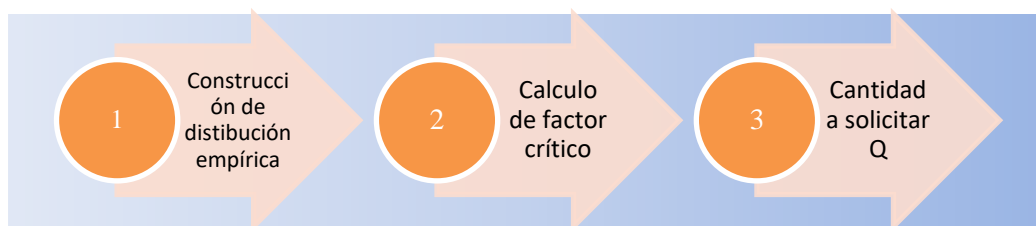


Figura 11. Sub-pasos adicionales de ejecución de newsvendor

1. Construcción de distribución empírica:

Para la ejecución de newsvendor no solo basta con el pronóstico determinado con Random Forest, sino se debe tener una imagen completa del pronóstico y de la

incertidumbre de este. Ello se logra obtener con una distribución de probabilidad del pronóstico. Como se explica en la parte de introducción de newsvendor se construirá una distribución empírica por producto para la ejecución.

La construcción de esta distribución por producto se genera a través de la relación A/F ratios (demanda real sobre pronósticos) de los productos de una campaña anterior a la actual. Por ejemplo, si se está trabajando en la campaña número 2, se hallará las relaciones A/F ratios de la campaña 1 para construcción de las distribuciones empíricas de los productos de la campaña 2. Después de ello, se ordenará de forma ascendente a los A/F ratios hallados de los productos para tener una tabla similar a la siguiente donde indica también la probabilidad acumulada por cada valor de A/F ratios y una columna adicional con el cálculo de A/F ratios por el pronóstico del producto en cuestión.

A/F Ratio	Q	F(Q)	A/F Ratio	Q	F(Q)	A/F Ratio	Q	F(Q)
0.25	800	0.0303	0.81	2,592	0.3636	1.23	3,936	0.6970
0.37	1,184	0.0606	0.82	2,624	0.3939	1.25	4,000	0.7273
0.56	1,792	0.0909	0.86	2,752	0.4242	1.27	4,064	0.7576
0.56	1,792	0.1212	0.96	3,072	0.4545	1.30	4,160	0.7879
0.57	1,824	0.1515	0.97	3,104	0.4848	1.36	4,352	0.8182
0.59	1,888	0.1818	0.98	3,136	0.5152	1.42	4,544	0.8485
0.64	2,048	0.2121	1.02	3,264	0.5455	1.46	4,672	0.8788
0.67	2,144	0.2424	1.08	3,456	0.5758	1.50	4,800	0.9091
0.69	2,208	0.2727	1.15	3,680	0.6061	1.54	4,928	0.9394
0.72	2,304	0.3030	1.17	3,744	0.6364	1.56	4,992	0.9697
0.80	2,560	0.3333	1.19	3,808	0.6667	1.60	5,120	1.0000

Tabla 14. Tabla de función de distribución empírica para un producto de la temporada actual [39]

Esta tabla también fue mostrada en el marco teórico de para explicar el procedimiento de construcción. Adicional a ello, se puede mostrar con la siguiente imagen como se ve gráficamente la distribución empírica acumulada.

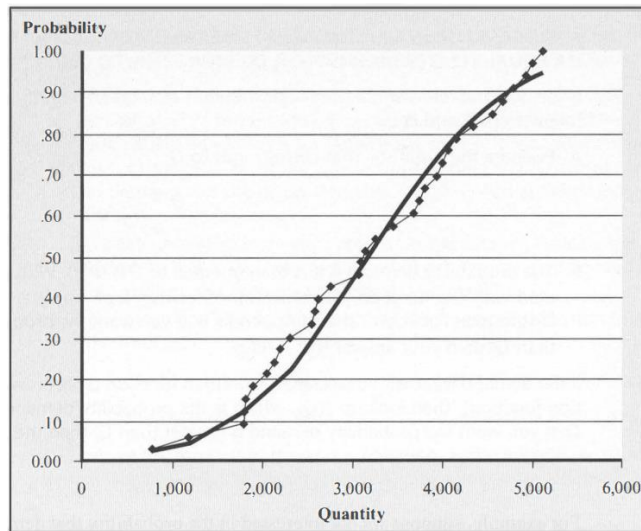


Figura 9. Función de distribución empírica del nuevo producto de la temporada entrante [39]

Ahora se tiene la distribución del pronóstico y el siguiente paso es definir que probabilidad acumulada de esta distribución se toma en consideración para determinar la cantidad a solicitar.

2. Cálculo de factor crítico por producto:

Como también se explicó en el marco teórico la cantidad a solicitar, Q , debe ser igual tal que la demanda sea menor o igual a Q con probabilidad $C_u / (C_o + C_u)$. A esta última relación se le denomina factor crítico. Entonces para determinar el factor crítico por producto se halla el C_u , costo unitario por pedir de menos, y el C_o , costo unitario por pedir de más. El C_u para este proyecto es igual al precio del producto menos el costo de este, ya que no se tiene algún costo de salvamento para los productos. El C_o es igual al costo del producto. La siguiente expresión muestra de manera más precisa el cálculo del factor crítico.

$$F(Q) = \frac{C_u}{C_o + C_u}$$

En la sección de resultados del presente trabajo se mostrarán los valores de factores críticos para algunos de los productos.

3. Cantidad a solicitar que maximiza el beneficio esperado:

Entonces una vez determinada la distribución de probabilidad del pronóstico del producto de la campaña actual y su respectivo factor crítico se procede a hallar la cantidad que maximiza el beneficio esperado.

En la tabla de distribución de probabilidad del producto se buscará la cantidad Q correspondiente, la cual tenga como probabilidad acumulada al factor crítico previamente hallado.

Estos pasos expuestos en la presente sección se siguen para todos los productos de todas las campañas del 2019. En la sección de resultados se muestran con tablas y gráficos calculados para productos del 2019.

CAPÍTULO III

RESULTADOS

Como se indicó en la sección de metodología, en esta parte se mostrará también los resultados de MAPE de la comparación de los modelos de Random Forest, Árboles de Decisión y Regresión Lineal Múltiple.

MAPE			
Campaña	Random Forest	Regresión Lineal	Árbol de desición
1	26.78%	47.42%	35.03%
2	27.41%	47.90%	33.92%
3	29.04%	48.19%	42.82%
4	33.85%	43.14%	39.76%
5	28.42%	54.59%	36.83%
6	27.15%	39.77%	34.41%
7	34.53%	48.95%	36.38%
8	27.75%	44.27%	36.28%
9	30.94%	47.91%	36.45%
10	27.96%	39.90%	35.42%
11	32.42%	45.77%	36.08%
12	27.97%	38.40%	35.04%
13	28.57%	40.21%	35.51%
14	30.40%	37.84%	36.78%
15	32.46%	44.31%	39.72%
16	29.11%	40.13%	36.53%
17	30.86%	39.33%	38.63%
18	35.16%	59.56%	37.53%

Tabla 16. Resultados para selección de pronostico por campaña

En la tabla se observa que los valores más bajos de MAPE por campaña los tiene Random Forest. También, debido a juicio de experto en la compañía se elige RF como modelo a usar en las campañas.

Asimismo, como se explicó anteriormente para los modelos de RF se hizo uso de 70 árboles. En los anexos se mostrarán gráficos Error vs número de árboles para cada campaña pronosticada. En la cual se observará lo que previamente se afirmó, que a partir de 70 árboles el error del modelo se mantiene constante.

A continuación, se muestran dos figuras (Error vs número de árboles) donde se puede notar que a partir de 70 árboles los errores de esos modelos se mantienen constantes. Estas figuras corresponden a las campañas 3 y 5 respectivamente.

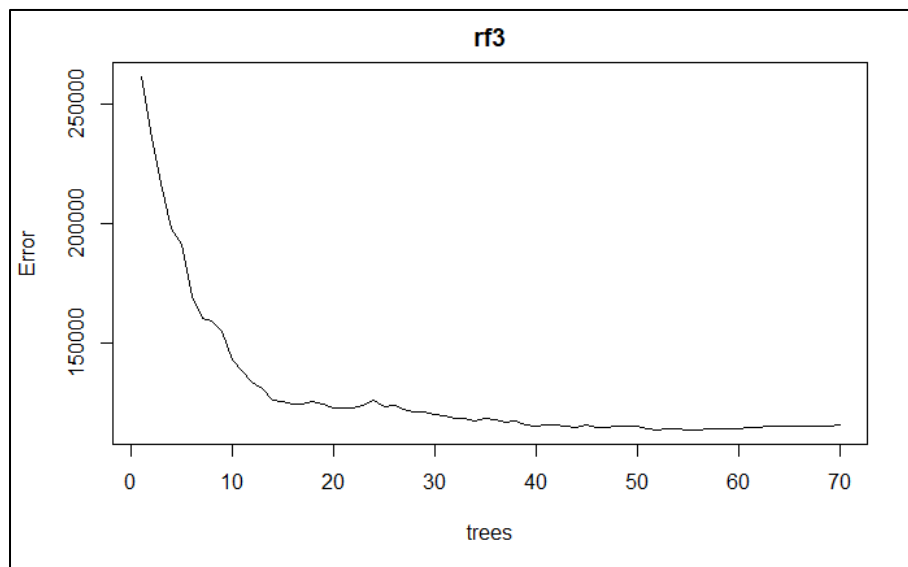


Figura 12. Grafica Error vs Número de árboles de la campaña 3.

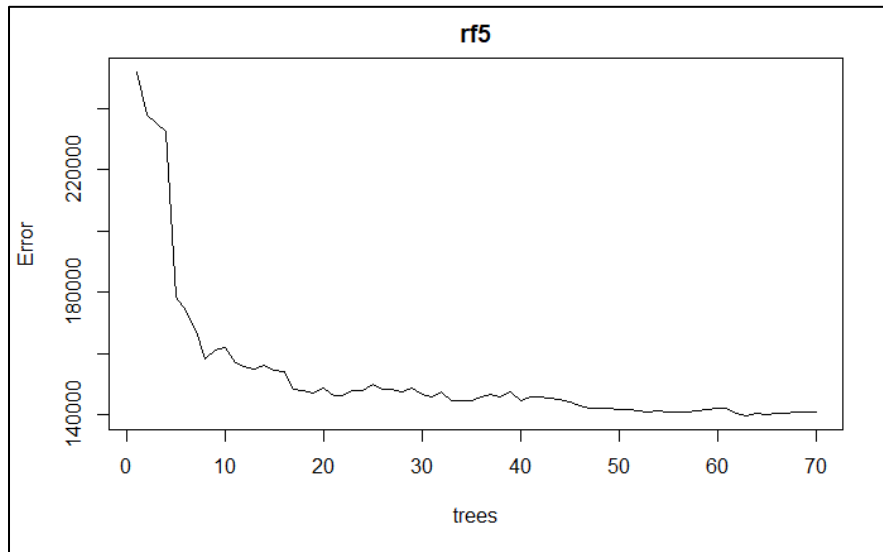


Figura 13. Grafica Error vs Número de árboles de la campaña 5.

Como se puede notar en las gráficas anteriores, cantidades de árboles poco antes a 70 comienzan a tener los errores mínimos y luego de ellos los errores se mantienen constantes. Entonces con ello se prueba que la cantidad de 70 árboles evita los sobreajustes de los modelos.

Asimismo, con ayuda de iteraciones en RStudio se optimizo el parámetro “mtry” para cada uno de los modelos. Ese valor hallado finalmente se usó para establecer las predicciones de cada campaña. En la siguiente tabla se muestra los valores de Mtry hallados para cada campaña. También, después de esta tabla se expone una gráfica Campaña versus Mtry de la campaña 4, donde se puede observar el valor óptimo de Mtry para dicha campaña.

Campaña	Mtry Optimo
2019-1	6
2019-2	6
2019-3	6
2019-4	3
2019-5	6
2019-6	6
2019-7	4
2019-8	6
2019-9	6
2019-10	6
2019-11	3
2019-12	6
2019-13	6
2019-14	6
2019-15	6
2019-16	6
2019-17	6
2019-18	4

Tabla 17. Valores de mtry óptimo para cada campaña

En la tabla anterior se observa que la cantidad de 6 para los valores de Mtry es la cantidad que se repite más en las campañas.

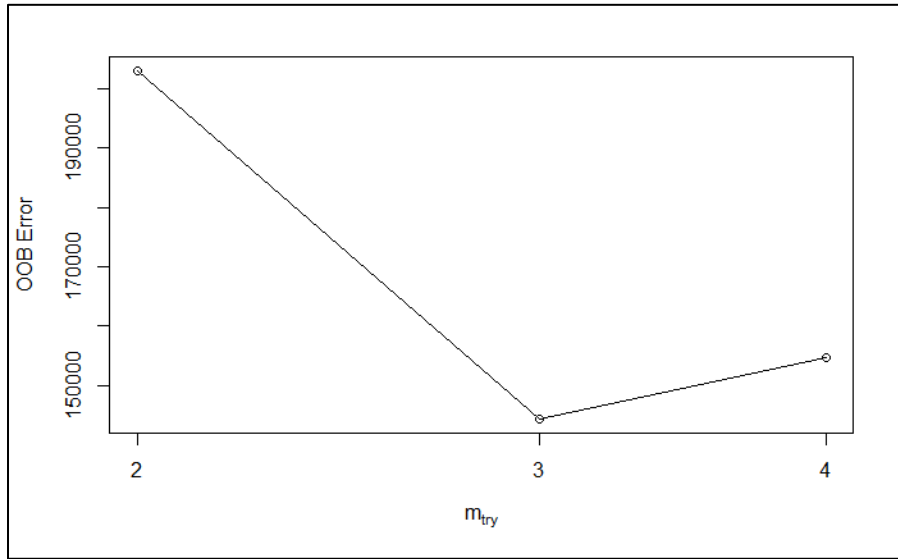


Figura 14. Grafica Error vs Mtry de la campaña 4

A manera de ejemplo en la figura anterior se observa que el valor óptimo de mtry para ese modelo de RF es de 3.

Con respecto a la importancia de variables de cada campaña, la variable que tiene mayor relevancia en todas es “Pre-pedido”. A continuación, se muestra una gráfica de la importancia de variables de la campaña 10 a manera de ejemplo. En la parte inferior de la gráfica se observa el error del modelo si esa variable es retirada [14].

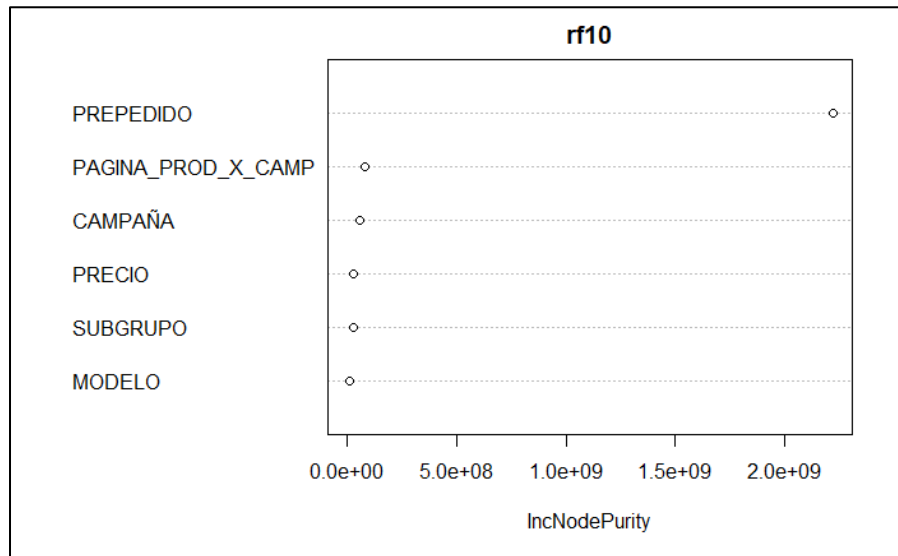


Figura 15. Grafica de importancia de variables de la campaña 10

En la figura 15 se nota que la variable pre-pedido tiene mayor importancia que las demás para el modelo de la campaña 10, luego está la variable número de página, después la variable campaña, después el precio, seguido de subgrupo y finalmente se tiene la variable modelo.

Seguido de ello se muestra una tabla con los errores de cada modelo por campaña que se obtuvo después de realizar RF. Los errores, como se explicó líneas antes son hallados mediante MAPE.

Campanña	MAPE
1	32.79%
2	31.47%
3	31.97%
4	51.55%
5	37.03%
6	35.49%
7	32.64%
8	30.12%
9	43.47%
10	42.74%
11	38.76%
12	52.29%
13	54.62%
14	33.49%
15	58.40%
16	67.28%
17	28.68%
18	43.96%

Tabla 18. Erro tipo MAPE por campaña del 2019

En la anterior tabla se puede notar los errores de cada pronostico. El menor error se tiene en la campaña 17 con un MAPE de 28.68%. Por otro lado, el mayor error se tiene en la campaña 16 con un MAPE de 67.28%. A manera de ejemplo, se muestran las gráficas de Demanda real vs pronósticos para estas dos campañas.

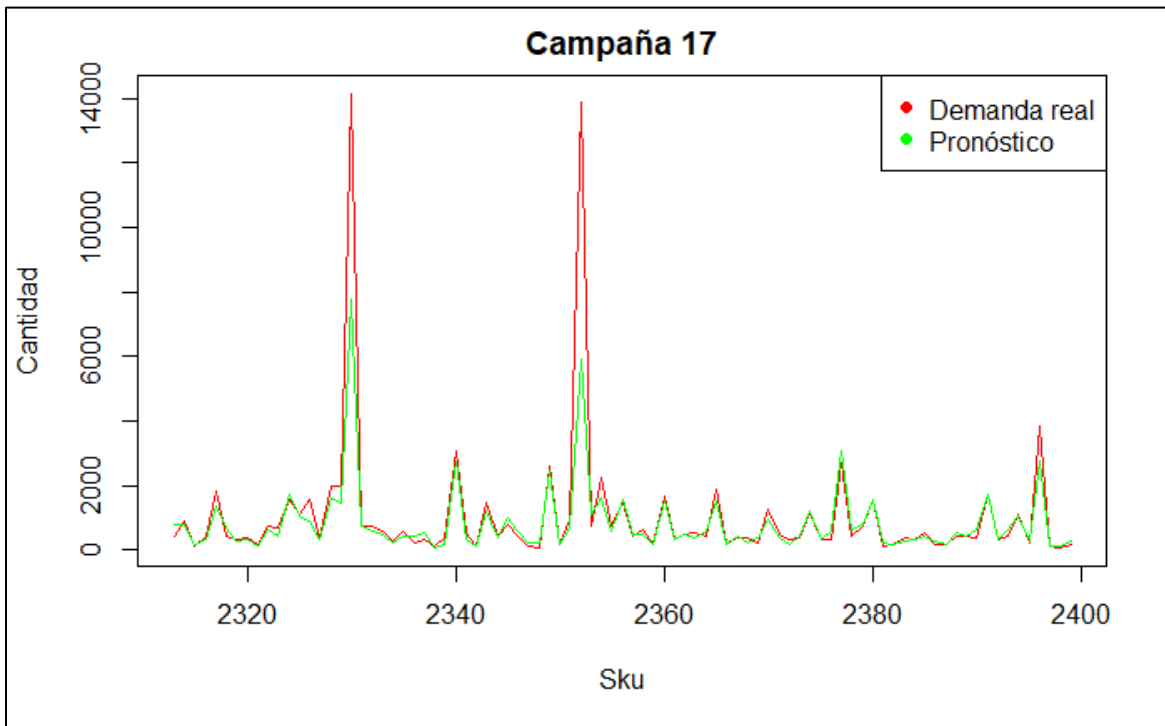


Figura 16. Grafica Demanda real vs Pronóstico de la campaña 17

La grafica 16 muestra la demanda real versus el pronóstico desarrollado por RF para la campaña 17. Se puede notar que ambas líneas están casi juntas debido al bajo error que presenta el pronóstico. Ello también se debe a que los parámetros número de árboles y mtry previamente fueron calculados. Lo cual ayuda a que el modelo no tenga problemas de sesgo y varianza.

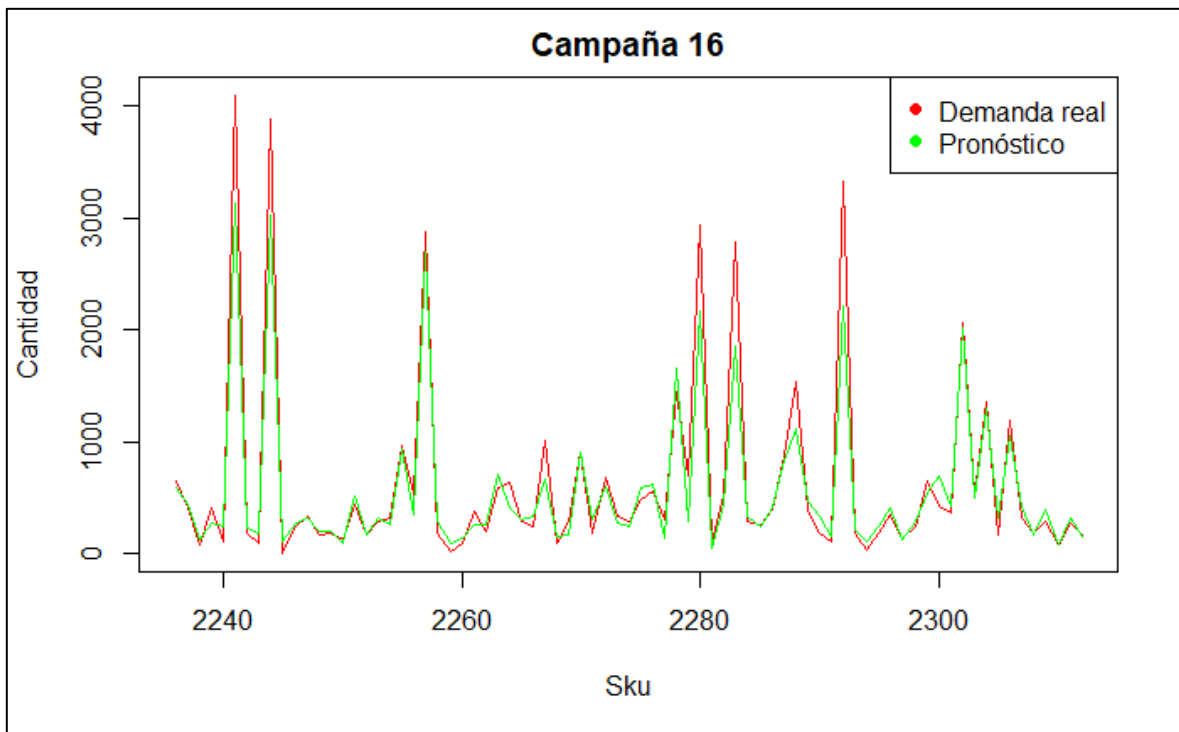


Figura 17. Gráfica Demanda real vs Pronóstico de la campaña 16

En esta gráfica se puede observar que hay picos de valores de demanda real que finalmente no son alcanzados por los pronósticos. Esto se debe principalmente a una de las desventajas de RF antes mencionada. Esta desventaja indica que, en regresión, RF no puede predecir más allá del rango de valores del conjunto de entrenamiento [14].

También, se mostrarán los resultados de las pruebas de hipótesis antes explicadas en el documento de las 18 campañas pronosticadas donde se evalúa si la media o mediana de los errores de los pronósticos son iguales o diferentes a cero.

Campañas	Valores de p-value por test		
	Anderson-Darling normality test	One Sample Z-test	Wilcoxon test
1	2.084e-05 (< 0.05)		< 2.2e-16 (< 0.05)
2	3.973e-08 (< 0.05)		7.057e-07 (< 0.05)
3	2.087e-15 (< 0.05)		1.284e-05 (< 0.05)
4	< 2.2e-16 (< 0.05)		0.4709
5	0.972	0.000	
6	4.561e-11 (< 0.05)		0.002187
7	2.218e-12 (< 0.05)		0.0009572
8	9.84e-05 (< 0.05)		0.4234
9	4e-10 (< 0.05)		0.8865
10	0.4686	0.058	
11	3.33e-14 (< 0.05)		0.7757
12	3.608e-09 (< 0.05)		1.499e-08 (< 0.05)
13	5.874e-11 (< 0.05)		0.0002087
14	6.799e-14 (< 0.05)		0.005346
15	< 2.2e-16 (< 0.05)		0.08319
16	< 2.2e-16 (< 0.05)		0.8281
17	< 2.2e-16 (< 0.05)		0.1813
18	< 2.2e-16 (< 0.05)		0.61

Tabla 19. Resultados de pruebas de hipótesis por campañas pronosticas para validación de modelo

Como se puede observar en la anterior tabla para 9 de las 18 campañas pronosticadas el modelo de Random Forest genera pronósticos con errores iguales a 0.

Finalmente, con la obtención de los costos por producto de todos los sku's del 2019 (brindados por la compañía), se procedió a realizar la ejecución de newsvendor por producto y campaña, ya que los pronosticos son un paso previo a la implementación de newsvendor y se consideran también como “inputs” para esta implementación como se explica en la metodología del presente trabajo [8], [39].

Gracias a las demandas pronosticadas por RF de los productos del 2019, datos de los precios y costos de estos se definió la cantidad a solicitar por producto a los proveedores a través de la implementación de newsvendor.

Para el presente proyecto se tiene el precio y costo de los productos, y en esta ocasión no se tiene algún costo de salvamento por ellos. Por ello el costo unitario de pedir de menos

(underage cost, cu) es igual al precio menos el costo del producto (margen contribución) y el costo de pedir de más (overage cost, co) es igual al costo de producto. Con estos últimos datos se determina el factor crítico por producto que es igual a la división entre el costo de pedir de menos y la suma del costo de pedir de menos más el costo de pedir de más [8].

Por ejemplo, el factor crítico del primer producto de la campaña número 1 del 2019 será el siguiente. El precio es de 32.43 soles y el costo es de 11.35 soles. Entonces el factor crítico de este producto es la división entre 21.08 y 43.78, cual resulta 0.65. De esta forma se determinó los factores críticos para todos los productos de todas las campañas del 2019. En el siguiente cuadro se muestra algunos productos de la campaña número 1 con sus factores críticos hallados.

INDICE	NOMBRE_SID	CAMP	PRECIO ING A DUPREE	Costo de producto	Factor crítico
1212	ESTUCHE TREBOL PEARL -P-	201901	32.435	11.35225	0.65
1213	TOBILLERA LOVE GOLD -D-	201901	16.185	5.66475	0.65
1214	ANILLO SOFIA T7 -D-	201901	16.185	5.66475	0.65
1215	TRIO DE ANILLOS EMILI -D-	201901	23.985	8.39475	0.65
1216	ESTUCHE MARISA GOLD -P-	201901	45.435	15.90225	0.65
1217	PULSERA ELIZABETH -P-	201901	30.485	10.66975	0.65

Tabla 20. Resultados de factores críticos de algunos productos de la campaña n° 1

El siguiente paso fue determinar la distribución de probabilidad de la demanda por producto. Para la distribución de probabilidad de la demanda se realizó el procedimiento mencionado en el libro de Cachon [39]. Un proceso que usa la relación histórica entre las demandas reales y pronósticos de los productos de campañas anteriores para construir una distribución emperica del producto de la campaña actual. A esta relación se le denomina A/F ratios. Este procedimiento también se explica en la parte de metodología del presente trabajo.

Entonces para determinar las distribuciones de los productos por ejemplo de la campaña número 1 se necesitó de las relaciones A/F de los productos de una campaña anterior. Luego de ello, se ordenó de forma ascendente los productos de acuerdo con sus relaciones A/F y se clasifíco por percentiles a cada uno. En la siguiente imagen se observará,

a forma de ejemplo, parte de la tabla de la campaña anterior a la número 1 con sus respectivas relaciones y percentiles.

INDICE	NOMBRE_SID	Demanda real	CAMP	Forecast	ERROR	A/F RATIO	Posición	F(Q)
1182	ARETES ZAIRA -D-	195	201818	300.8092857	105.809286	0.6483	1	0.0145
1175	MEDALLON LILIANA -D-	421	201818	534.5290476	113.529048	0.7876	2	0.0290
1197	ARETES LUCIA -P	476	201818	531.2814286	55.2814286	0.8959	3	0.0435
1207	ESTUCHE BELL CRYSTAL -P	132	201818	133.7716667	1.77166667	0.9868	4	0.0580
1143	ARETES NAILA -P	1015	201818	986.4835714	-28.5164286	1.0289	5	0.0725
1167	ESTUCHE KIRA GOLD -P-	253	201818	241.0307143	-11.9692857	1.0497	6	0.0870
1172	ARETES DAFNE -P-	344	201818	324.7816667	-19.2183333	1.0592	7	0.1014
1180	ARETES KIARA -P-	444	201818	416.2880952	-27.7119048	1.0666	8	0.1159
1173	ARETES DANIELA -P	1071	201818	980.2790476	-90.7209524	1.0925	9	0.1304
1181	PULSERA DULCE FLORCITA -P -	319	201818	287.4861905	-31.5138095	1.1096	10	0.1449
1165	ANILLO LAILA T7 -D-	47	201818	39.30071429	-7.69928571	1.1959	11	0.1594
1171	BRAZALETE ADARA GOLD-P-	175	201818	145.8847619	-29.1152381	1.1996	12	0.1739
1198	ANILLO UNICORNIO -P-	952	201818	785.922619	-166.077381	1.2113	13	0.1884
1202	ARETES SAMANTA FLOR -P	526	201818	429.9459524	-96.0540476	1.2234	14	0.2029
1187	PULSERA GRISEL PEARL -P-	274	201818	223.325	-50.675	1.2269	15	0.2174
1169	ARETES LILIANA -D-	971	201818	785.3345238	-185.665476	1.2364	16	0.2319
1195	PULSERA LAILA -D-	573	201818	452.0169048	-120.983095	1.2677	17	0.2464
1201	ARETES VALERIA -P	1252	201818	986.3380952	-265.661905	1.2693	18	0.2609
1193	ESTUCHE DULCE FLORCITA -P -	686	201818	539.5302381	-146.469762	1.2715	19	0.2754
1170	SET X 7 PARES DE ARETES BOHEM -P-	1427	201818	1108.359048	-318.640952	1.2875	20	0.2899
1164	ESTUCHE UNICORNIO -P-	1752	201818	1334.38381	-417.61619	1.3130	21	0.3043
1194	ESTUCHE ELENA DORE -P	201	201818	147.8440476	-53.1559524	1.3595	22	0.3188
1179	SET X 5 PARES DE ARETES ALESSIA -P-	2003	201818	1422.368333	-580.631667	1.4082	23	0.3333
1162	ARETES SAMARA -D-	211	201818	148.3930952	-62.6069048	1.4219	24	0.3478

Tabla 21. Resultados de A/F ratios de la campaña n° 18 del 2018.

Tomando en consideración el primer producto de la campaña número 1 del 2019, el cual tiene un pronóstico de 300 unidades se puede construir su distribución empírica a través de la multiplicación de cada una de las relaciones A/F mostradas y su pronóstico respectivo. Quedando así la siguiente grafica de distribución acumulada para este producto.

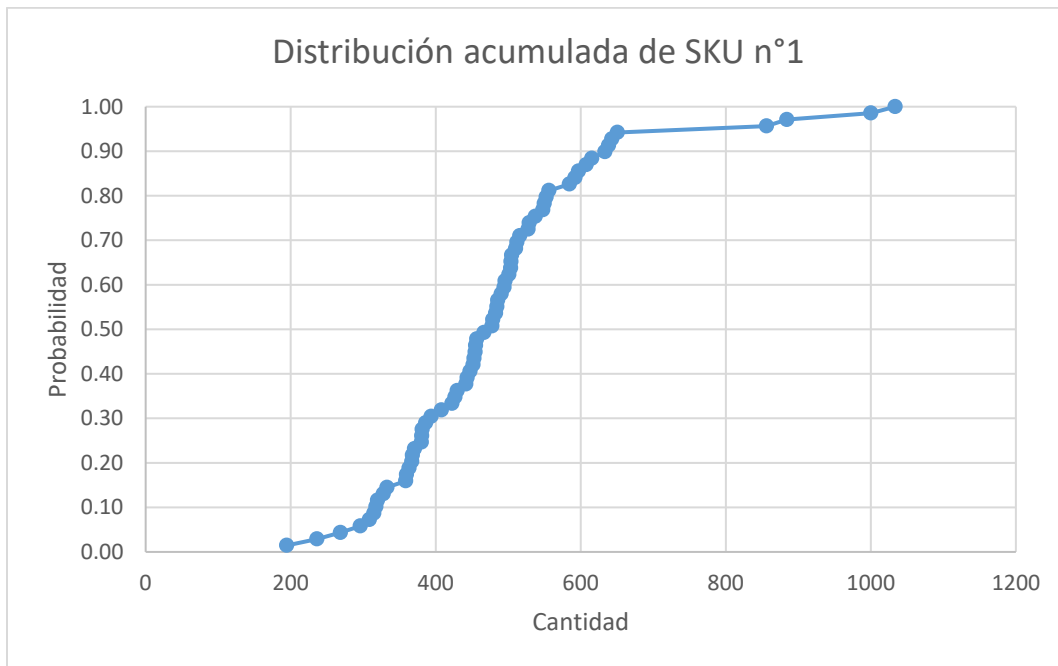


Figura 18. Distribución empírica acumulada del producto n° 1 de la campaña 1 del 2019.

La grafica anterior de doble entrada muestra la relación de Cantidad a solicitar versus la probabilidad. Continuando con el ejemplo del mismo producto, se tiene un factor crítico de 0.65 que se mostró previamente. Dado este factor crítico según la distribución mostrada correspondería solicitar al proveedor 504 unidades de este producto. De esta misma forma y procedimiento se determinó las cantidades a solicitar al proveedor para todos los productos de todas las campañas del 2019.

En las siguientes dos graficas se mostrarán las relaciones de A/F ratios versus la probabilidad para las campañas 1 y 2 del 2019. Las gráficas de las campañas restantes para el 2019 se expondrán en los anexos del documento.

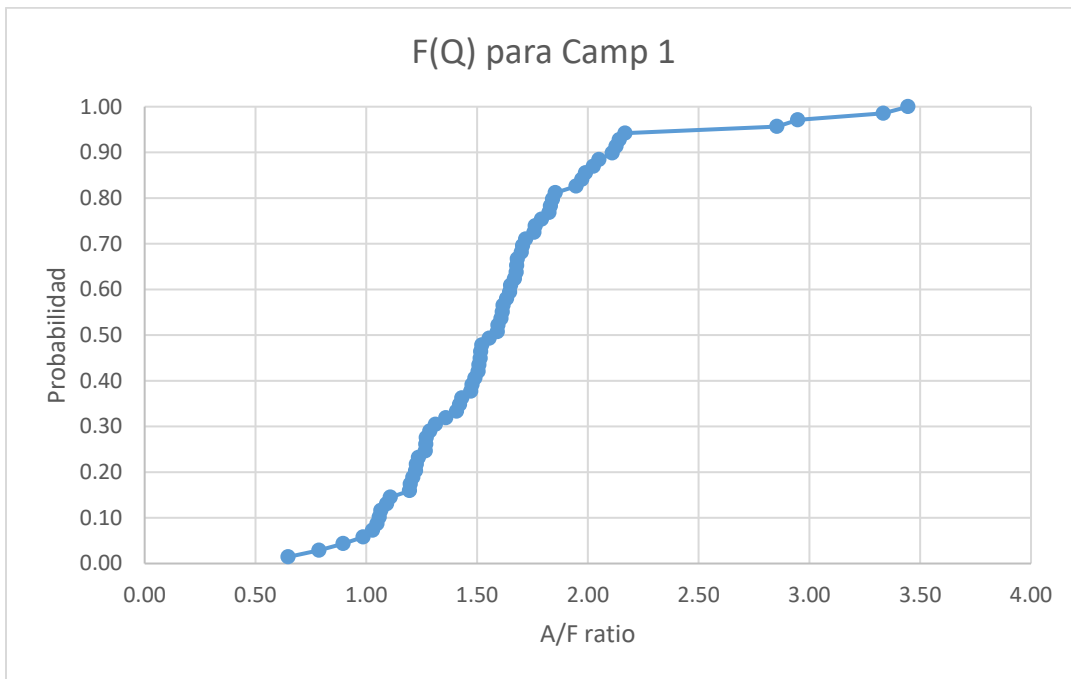


Figura 19. Relación A/F ratio versus probabilidad para la campaña 1 del 2019.

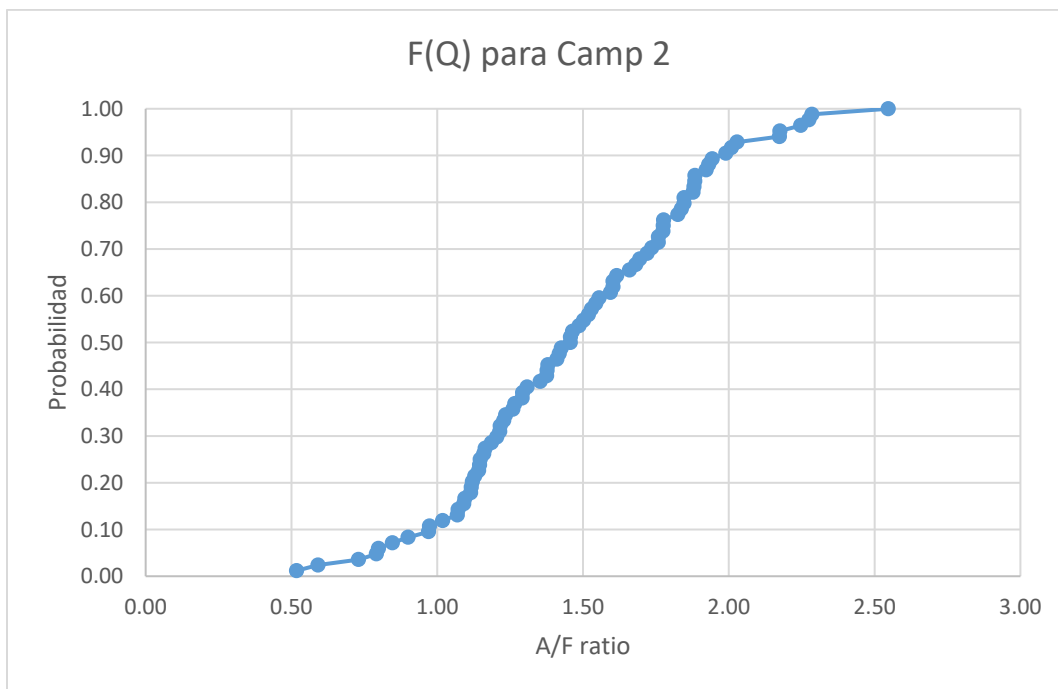


Figura 20. Relación A/F ratio versus probabilidad para la campaña 2 del 2019.

Las figuras anteriores muestran que ellas se acercan aproximadamente a una distribución normal acumulada, lo cual también indicaría una buena representación para las demandas reales de esas campañas.

Ahora una vez definido las cantidades a solicitar mediante newsvendor, en la siguiente figura se muestra la comparación de resultados del 2018 y la nueva implementación para el 2019.

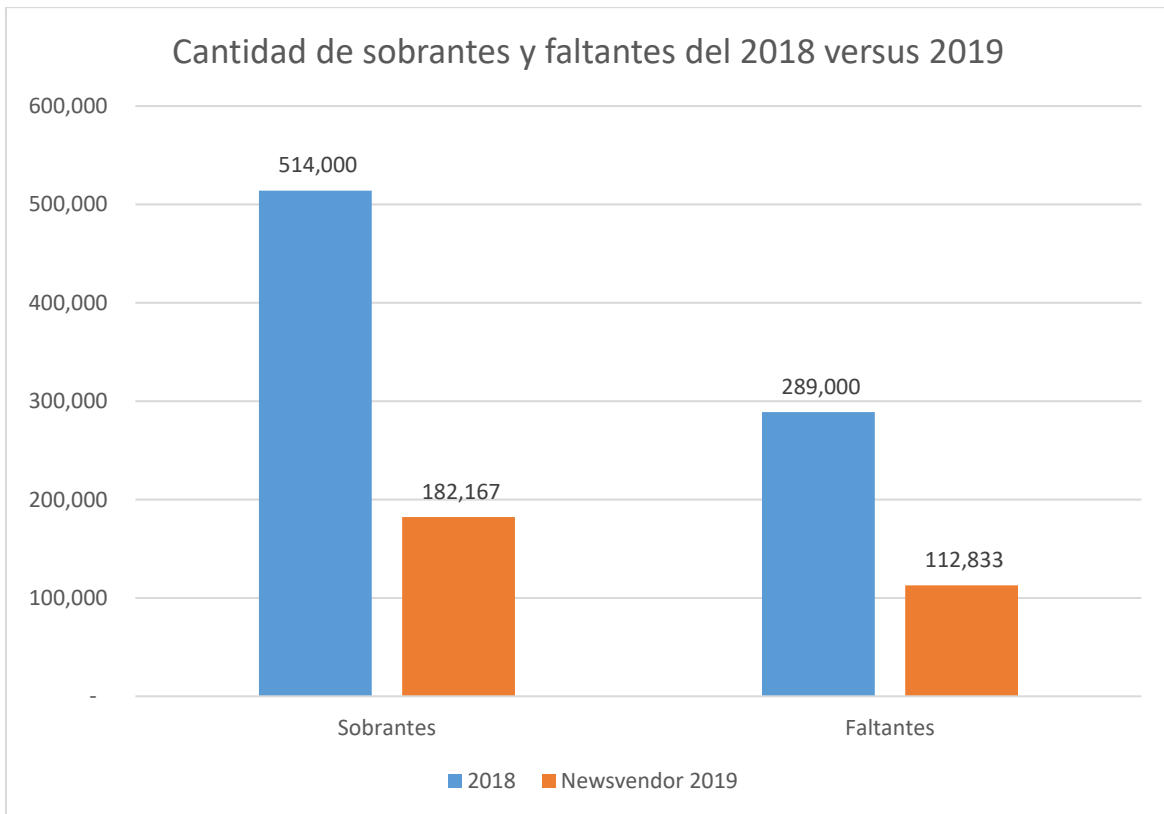


Figura 21. Comparación de cantidades de sobrantes y faltantes en el 2018 y 2019.

La implantación de ejecutar news vendor, como se puede apreciar en la imagen anterior trae mejoras en los resultados de cantidades de sobrantes y faltantes. Con esta nueva implementación la cantidad de sobrantes disminuye en 331,833 unidades y la cantidad de faltantes en 176,167 unidades. Teniendo así mayor cantidad de demanda satisfecha. También en la siguiente tabla, se muestra los valores en soles de los costos por pedir de más, costos

por pedir de menos y costo total de oportunidad entre el 2018 y la implementación para el 2019.

	Costo de pedir de más	Costo de pedir de menos	Costo de oportunidad
2018	S/ 7,005,829.00	S/ 4,441,930.00	S/ 11,447,759.00
Newsvendor 2019	S/ 1,678,322.07	S/ 1,830,306.21	S/ 3,508,628.28

Tabla 22. Resultados de costos de oportunidad del 2018 y 2019.

La implementación de newsvendor genera una mejora de 7,939,131 soles en los costos de oportunidad. Y en los costos de pedir de más y solicitar de menos se consiguen mejoras de 5,327,507 soles y 2,611,624 soles respectivamente.

Con respecto, a la mejora en los estados de resultados que genera la implementación a comparación de los resultados obtenidos en el 2018, se muestra la siguiente tabla de doble entrada.

	2018	Implementación 2019
N° de SKU's	1211	1265
Capital de inversión	S/ 20,681,194.00	S/ 9,333,642.00
Magen total - Costo por pedir de más	S/ 11,873,596.73	S/ 12,538,702.00

Tabla 23. Mejoras financieras

El capital de inversión para el 2018 y 2019 se determinó a través de las unidades adquiridas de DML por campaña y multiplicadas por el costo de cada producto. Se observa una mejora en el capital de inversión de 11,347,552 soles ya que se requirieron menos productos para satisfacer la demanda del consumidor. Además, la utilidad bruta para el 2018 fue de

11,873,596 soles y para el 2019 es de 12,538,702 soles teniendo así una mejora de 665,106 soles. Esto se dio ofreciendo casi la misma cantidad de SKU's para ambos años.

Con respecto a la cantidad de asesoras que desertaría trabajar con DML, esto disminuirá debido a los pronostico y cantidades nuevas solicitadas a los proveedores. Ya que como se puede apreciar en la figura 21 la cantidad de sobrantes disminuye del 2018 al 2019 en 331,833 unidades y esto incentivará a las asesoras a vender sus productos campaña a campaña con más certeza y menos temor a quedarse con los productos.

CONCLUSIONES Y RECOMENDACIONES

1. Se disminuyó la cantidad de faltantes y sobrantes del 2018 al 2019 a través la implementación de un modelo de machine learning, Random Forest, y ejecución de newsvendor. Con lo cual se disminuyó el costo de oportunidad de 11.4 millones a 3.5 millones de soles.
2. Con la nueva implementación se mejoró la utilidad bruta para el 2019 en 665,106 soles con un menor capital de inversión en productos el cual se redujo en 11.3 millones de soles, ofreciendo al mercado casi la misma cantidad de SKU's que en el 2018.
3. Se definió un modelo de pronóstico para la línea de joyería en base a las características logísticas de los productos.
4. Se implementó y validó el modelo de pronóstico Random Forests y se comparó con otros modelos de pronósticos supervisados.
5. Se determinó la variable más significativa para el modelo implementado con respecto a las que se trabajó en la presente investigación, la cual es Pre pedidos para todas las campañas del 2019.
6. Se determinó los parámetros óptimos para los modelos de RF para cada campaña del 2019.
7. 9 de las 18 campañas pronosticas tienen errores iguales a 0.
8. Para implementaciones posteriores de este modelo de pronósticos se recomienda adicionar más variables para observar cómo se desenvolvería las nuevas predicciones.

REFERENCIAS BIBLIOGRÁFICAS

- [1] C. P. de V. D. | CAPEVEDI, “Cámara Peruana de Venta Directa | CAPEVEDI.” Accessed: Dec. 03, 2020. [Online]. Available: <https://capevedi.com/>
- [2] Gestión.pe, “Venta directa creció por encima de la economía en 4.9% en el primer trimestre.” Accessed: Dec. 02, 2023. [Online]. Available: <https://gestion.pe/economia/venta-directa-crecio-economia-peruana-4-9-primer-trimestre-ano-272445-noticia/?ref=gesr>
- [3] S. Brodie *et al.*, “Public perceptions of direct selling: an international perspective,” *Westminster Business*, vol. 0, no. 0, pp. 3–27, 2004.
- [4] L. Kinyua Kathuni and N. Galo Mugenda, “Direct Sales Strategy Applied by Commercial Banks in Kenya,” *international Journal of Business, Humanities and Technology*, vol. 2, no. 7, pp. 132–138, 2012.
- [5] A. J. Boada and R. Mayorca, “Planificación de demanda, en empresas con estilo de venta por catálogo,” *Rev Lasallista Investig*, vol. 8, no. 2, pp. 124–135, 2011.
- [6] T. Hendershott and J. Zhang, “A model of direct and intermediated sales,” *J Econ Manag Strategy*, vol. 15, no. 2, pp. 279–316, 2006, doi: 10.1111/j.1530-9134.2006.00101.x.
- [7] A. J. Boada and D. De Vasconcelos, “Modelo estadístico de regresión múltiple, columna vertebral para predecir en empresas multinacionales con estilo de venta por catálogo,” *Rev Lasallista Investig*, vol. 10, no. 1, pp. 1794–4449, 2013, Accessed: Jun. 02, 2023. [Online]. Available: https://www.researchgate.net/publication/350787697_Modelo_estadistico_de_regresion_multiple_columna_vertebral_para_predecir_en_empresas_multinacionales_con_estilo_de_venta_por_catologo
- [8] E. A. Silver, D. F. Pyke, and D. J. Thomas, *Inventory and Production Management in Supply Chains*. 2016.
- [9] S. G. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting methods and applications*. 1997.
- [10] H. Ando, “Demand forecasting method, demand forecasting system, and recording medium,” 2000
- [11] T. Aykin, “System and method for integrated supply chain and contact center management,” 2012

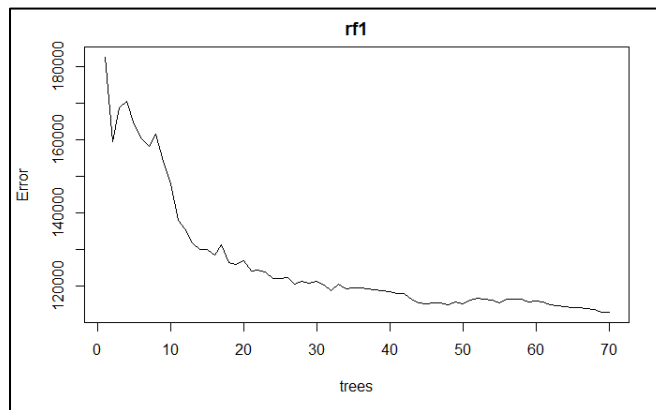
- [12] R. J. Hyndman and G. Athanasopoulos, *Chapter 10 Forecasting hierarchical or grouped time series | Forecasting: Principles and Practice*. Accessed: Dec. 06, 2020. [Online]. Available: <https://otexts.com/fpp2/hierarchical.html>
- [13] M. Valencia Cárdenas, J. C. Correa Morales, F. J. Díaz Serna, and S. Ramírez Agudelo, "Aplicación de modelación bayesiana y optimización para pronósticos de demanda," *Ingeniería y Desarrollo*, vol. 32, no. 2, pp. 179–199, 2014, doi: <http://dx.doi.org/10.14482/inde.32.2.5403>.
- [14] J. Orellana Alvear, "Arboles de decision y Random Forest." Accessed: Dec. 11, 2020. [Online]. Available: <https://bookdown.org/content/2031/referencias.html>
- [15] C. M. YEŞİLKANAT, "Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm," *Chaos Solitons Fractals*, vol. 140, 2020, doi: [10.1016/j.chaos.2020.110210](https://doi.org/10.1016/j.chaos.2020.110210).
- [16] E. M. M. van der Heide, R. F. Veerkamp, M. L. van Pelt, C. Kamphuis, I. Athanasiadis, and B. J. Ducro, "Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle," *J Dairy Sci*, vol. 102, no. 10, pp. 9409–9421, 2019, doi: [10.3168/jds.2019-16295](https://doi.org/10.3168/jds.2019-16295).
- [17] Leo. Breiman, "Random forests," *Random Forests*, vol. 45, pp. 1–122, 2001, doi: [10.1201/9780429469275-8](https://doi.org/10.1201/9780429469275-8).
- [18] S. Sharma, R. Gupta, R. Bhatia, A. P. Toor, and H. Setia, "Predicting microbial response to anthropogenic environmental disturbances using artificial neural network and multiple linear regression," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 65–70, Jun. 2021, doi: [10.1016/j.ijcce.2021.03.001](https://doi.org/10.1016/j.ijcce.2021.03.001).
- [19] P. Cortez, M. Portelinha, S. Rodrigues, V. Cadavez, and A. Teixeira, "Lamb meat quality assessment by Support Vector Machines," *Neural Process Lett*, vol. 24, no. 1, pp. 41–51, 2006, doi: [10.1007/s11063-006-9009-6](https://doi.org/10.1007/s11063-006-9009-6).
- [20] T. Van Hertem *et al.*, "Automatic lameness detection based on consecutive 3D-video recordings," *Biosyst Eng*, vol. 119, pp. 108–116, 2014, doi: [10.1016/j.biosystemseng.2014.01.009](https://doi.org/10.1016/j.biosystemseng.2014.01.009).
- [21] K. Hempstalk, S. McParland, and D. P. Berry, "Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows," *J Dairy Sci*, vol. 98, no. 8, pp. 5262–5273, 2015, doi: [10.3168/jds.2014-8984](https://doi.org/10.3168/jds.2014-8984).
- [22] F. Ghafouri-Kesbi, G. Rahimi-Mianji, M. Honarvar, and A. Nejati-Javaremi, "Predictive ability of Random Forests, Boosting, Support Vector Machines and Genomic Best Linear Unbiased

- Prediction in different scenarios of genomic evaluation,” *Anim Prod Sci*, vol. 57, no. 2, pp. 229–236, 2017, doi: <https://doi.org/10.1071/AN15538>.
- [23] D. A. Bennett, “How can I deal with missing data in my study?,” *Aust N Z J Public Health*, vol. 25, no. 5, pp. 464–469, 2001, doi: [10.1111/j.1467-842X.2001.tb00294.x](https://doi.org/10.1111/j.1467-842X.2001.tb00294.x).
- [24] Z. Wang *et al.*, “Selection of optimal models for predicting growth stress in *Artemisia desertorum* by comparison of linear regression and multiple neural networks: Take the construction of a green mine in the Bayan Obo mine as an example,” *Ecotoxicol Environ Saf*, vol. 235, Apr. 2022, doi: [10.1016/j.ecoenv.2022.113400](https://doi.org/10.1016/j.ecoenv.2022.113400).
- [25] A. M. al-Swaidani, W. T. Khwies, M. al-Baly, and T. Lala, “Development of multiple linear regression, artificial neural networks and fuzzy logic models to predict the efficiency factor and durability indicator of nano natural pozzolana as cement additive,” *Journal of Building Engineering*, vol. 52, p. 104475, 2022, doi: <https://doi.org/10.1016/j.jobbe.2022.104475>.
- [26] G. S. Maddala, *Introduction to Econometrics*, vol. 2. New York: Macmillan, 1992.
- [27] M. Morocho and J. Catalina, “Desarrollo de un modelo de flujos de datos que permita identificar el comportamiento de la morbimortalidad del adulto mayor ecuatoriano”.
- [28] B. D. Williams and M. A. Waller, “Top-Down Versus Bottom-Up Demand Forecasts: The Value of Shared Point-of-Sale Data in the Retail Supply Chain,” *Journal of Business Logistics*, vol. 32, no. 1, pp. 17–26, 2011, doi: <https://doi.org/10.1111/j.2158-1592.2011.01002.x>.
- [29] B. J. Dangerfield and J. S. Morris, “Top-down or bottom-up: Aggregate versus disaggregate extrapolations,” *Int J Forecast*, vol. 8, no. 2, pp. 233–241, 1992, doi: [https://doi.org/10.1016/0169-2070\(92\)90121-O](https://doi.org/10.1016/0169-2070(92)90121-O).
- [30] K. B. Kahn, “Solving the problems of new product forecasting,” *Bus Horiz*, vol. 57, no. 5, pp. 607–615, 2014, doi: <https://doi.org/10.1016/j.bushor.2014.05.003>.
- [31] M. E. Nenni, L. Giustiniano, and L. Pirolo, “Demand Forecasting in the Fashion Industry: A Review,” *Special Collection: Innovations in Fashion Industry (2013)*, vol. 5, 2013, doi: <https://doi.org/10.5772/56840>.
- [32] C. Ching-Chin, A. I. Ka Ieng, K. Ieng, W. Ling-Ling, and K. Ling-Chieh, “Designing a decision-support system for new product sales forecasting,” *Expert Syst Appl*, vol. 37, no. 2, pp. 1654–1665, 2010, doi: <https://doi.org/10.1016/j.eswa.2009.06.087>.
- [33] A. A. Kurawarwala and H. Matsuo, “Product Growth Models for Medium-Term Forecasting of Short Life Cycle Products,” *Technol Forecast Soc Change*, vol. 57, no. 3, pp. 169–196, 1998, doi: [https://doi.org/10.1016/S0040-1625\(97\)00102-9](https://doi.org/10.1016/S0040-1625(97)00102-9).

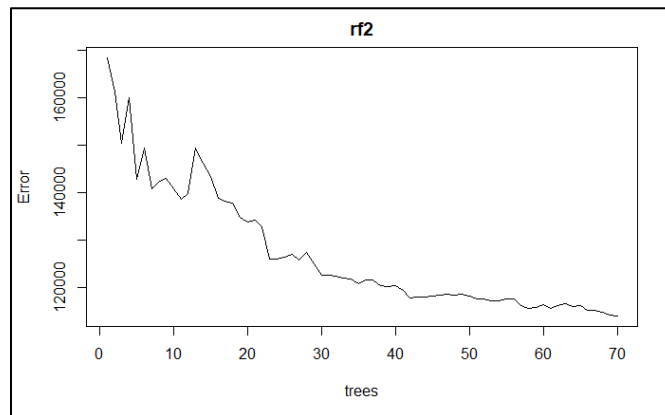
- [34] J. Mostard, R. Teunter, and R. De Koster, "Forecasting demand for single-period products: A case study in the apparel industry," *Eur J Oper Res*, vol. 211, no. 1, pp. 139–147, 2011, doi: <https://doi.org/10.1016/j.ejor.2010.11.001>.
- [35] M. Fisher and K. Rajaram, "Accurate Retail Testing of Fashion Merchandise: Methodology and Application," *Marketing Science*, vol. 19, no. 3, pp. 203–296, 2000, doi: <https://doi.org/10.1287/mksc.19.3.266.11800>.
- [36] S. A. Gyamerah, P. Ngare, and D. Ikpe, "Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov Kernel function," *Agric For Meteorol*, vol. 280, p. 107808, Jan. 2020, doi: 10.1016/j.agrformet.2019.107808.
- [37] M. Khouja, "The single-period (news-vendor) problem: literature review and suggestions for future research," *Omega (Westport)*, vol. 27, no. 5, pp. 537–553, 1999, doi: [https://doi.org/10.1016/S0305-0483\(99\)00017-1](https://doi.org/10.1016/S0305-0483(99)00017-1).
- [38] P. Dutta, D. Chakraborty, and A. R. Roy, "An inventory model for single-period products with reordering opportunities under fuzzy demand," *Computers & Mathematics with Applications*, vol. 53, no. 10, pp. 1502–1517, 2007, doi: <https://doi.org/10.1016/j.camwa.2006.04.029>.
- [39] G. Cachon and C. Terwiesch, *Matching Supply with Demand: An Introduction to Operations Management*, 2nd ed. 2008.
- [40] G. C. Michael, "A Computer Simulation Model for Forecasting Catalog Sales," *Journal of Marketing Research*, vol. 8, no. 2, pp. 224–229, 1971, doi: <https://doi.org/10.1177/002224377100800211>.
- [41] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How many trees in a random forest?," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 154–168, 2012, doi: 10.1007/978-3-642-31537-4_13.
- [42] A. E. Cisneros cisneros and Tecnológico de Monterrey, *Módulo III : Analizar*.
- [43] O. Ramírez Alán, "Aprendiendo Bioestadística con R." Accessed: Sep. 02, 2022. [Online]. Available: https://osoramirez.github.io/R_Para_Biologos/test-de-hipotesis-y-pruebas-no-parametricas.html

ANEXOS

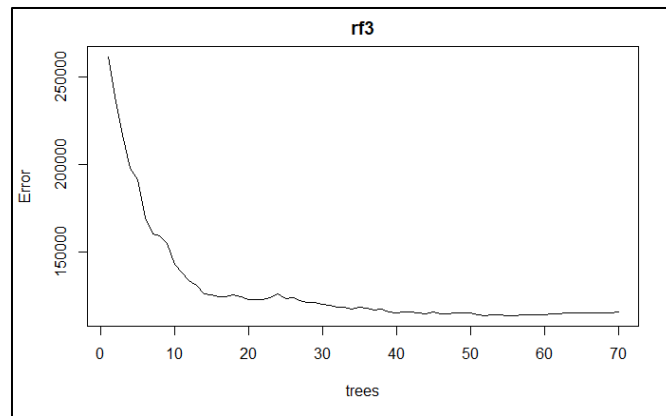
ANEXO 1: Número de árboles vs Error del modelo. Campaña 1



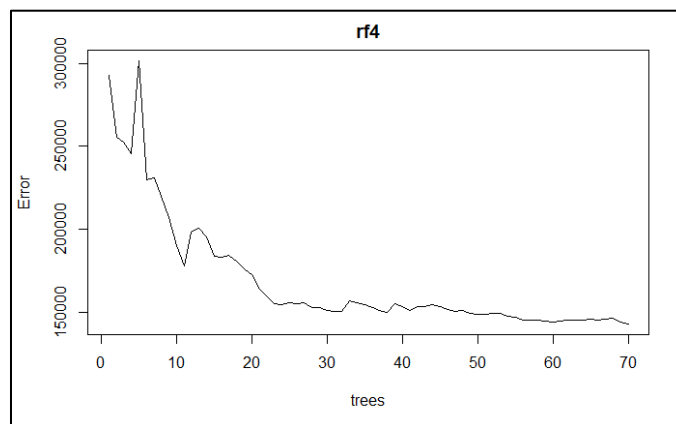
ANEXO 2: Número de árboles vs Error del modelo. Campaña 2



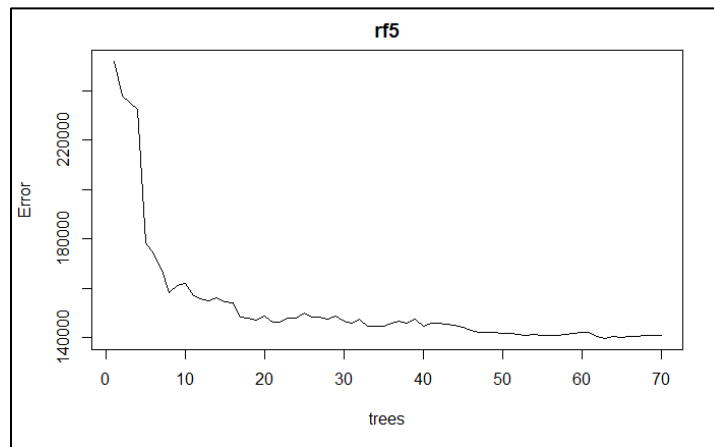
ANEXO 3: Número de árboles vs Error del modelo. Campaña 3



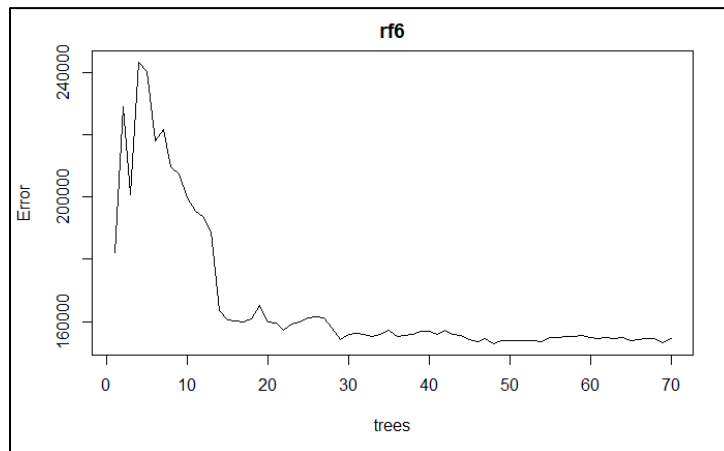
ANEXO 4: Número de árboles vs Error del modelo. Campaña 4



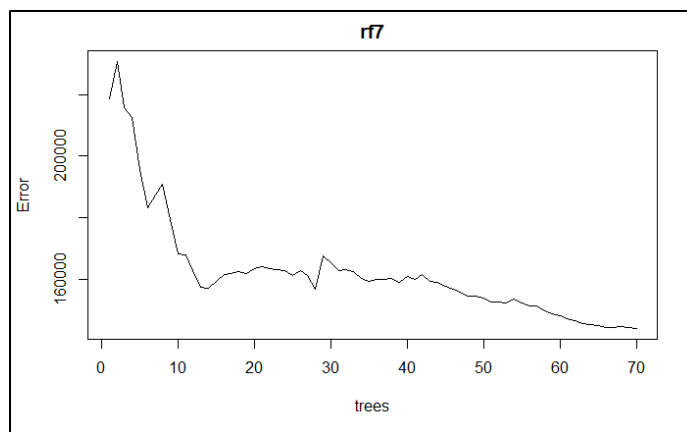
ANEXO 5: Número de árboles vs Error del modelo. Campaña 5



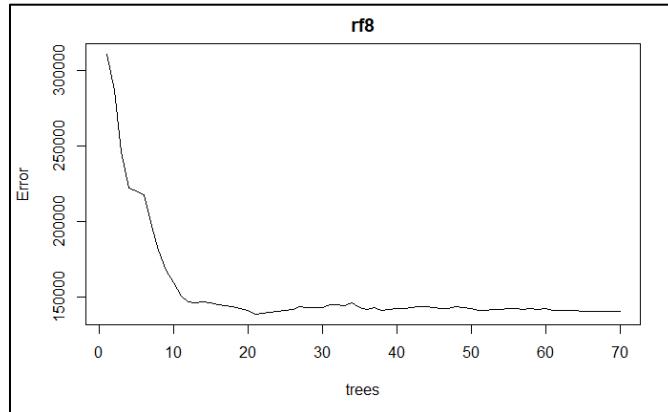
ANEXO 6: Número de árboles vs Error del modelo. Campaña 6



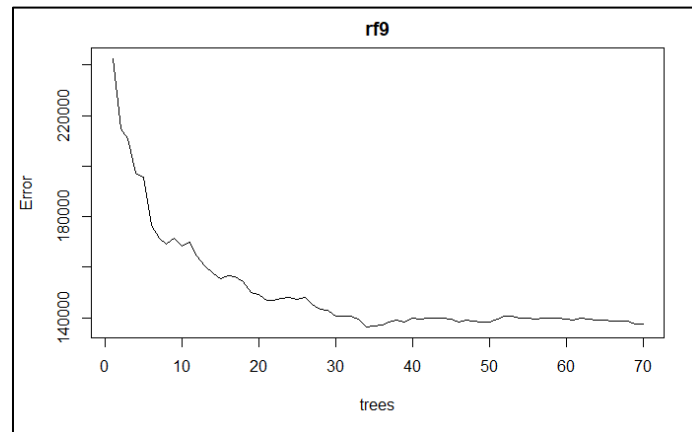
ANEXO 7: Número de árboles vs Error del modelo. Campaña 7



ANEXO 8: Número de árboles vs Error del modelo. Campaña 8

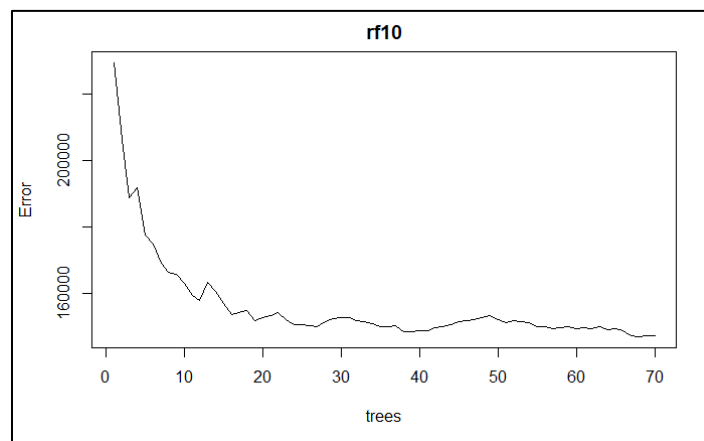


ANEXO 9: Número de árboles vs Error del modelo. Campaña 9



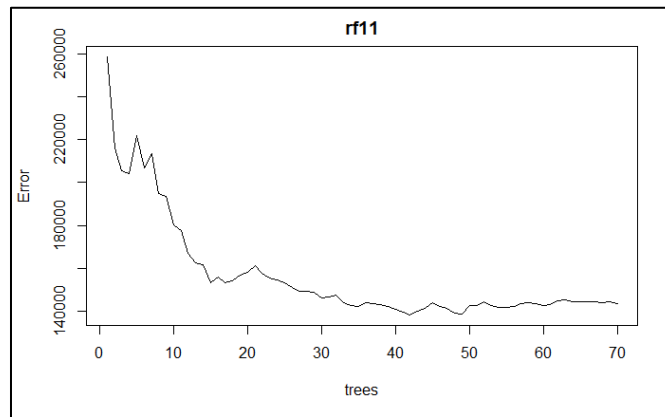
ANEXO 10: Número de árboles vs Error del modelo. Campaña

10



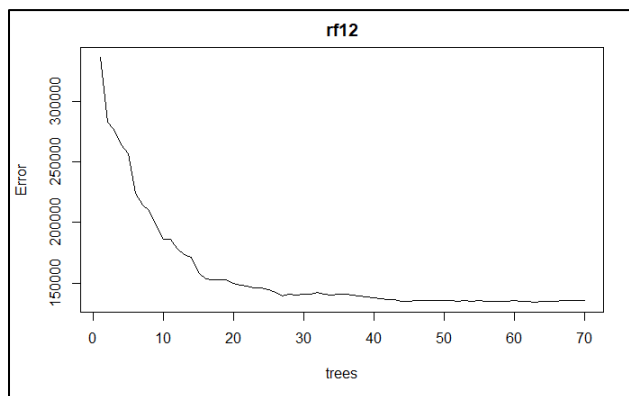
ANEXO 11: Número de árboles vs Error del modelo. Campaña

11



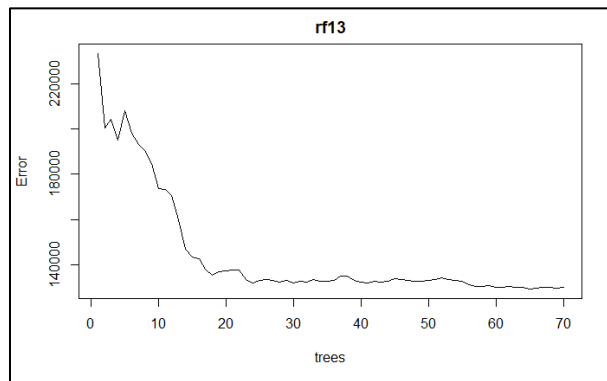
ANEXO 12: Número de árboles vs Error del modelo. Campaña

12



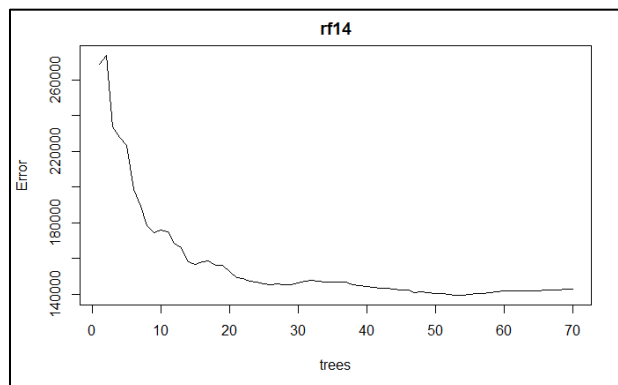
ANEXO 13: Número de árboles vs Error del modelo. Campaña

13



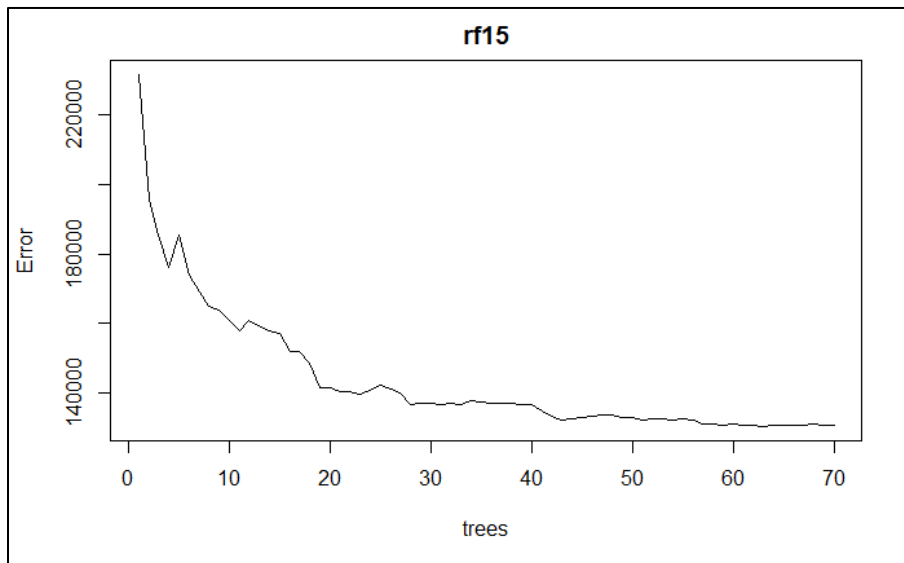
ANEXO 14: Número de árboles vs Error del modelo. Campaña

14



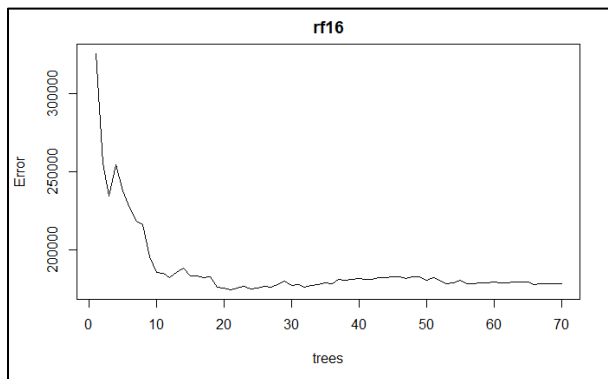
ANEXO 15: Número de árboles vs Error del modelo. Campaña

15



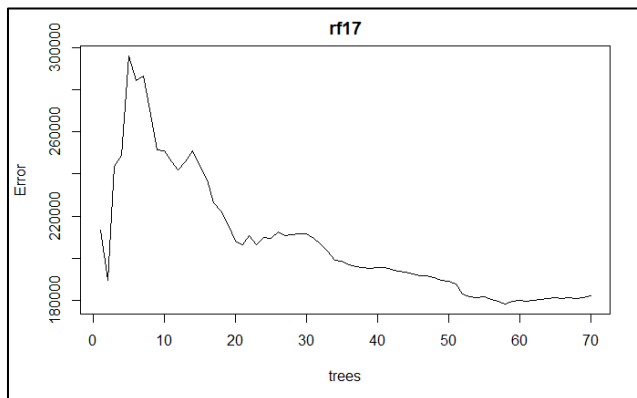
ANEXO 16: Número de árboles vs Error del modelo. Campaña

16



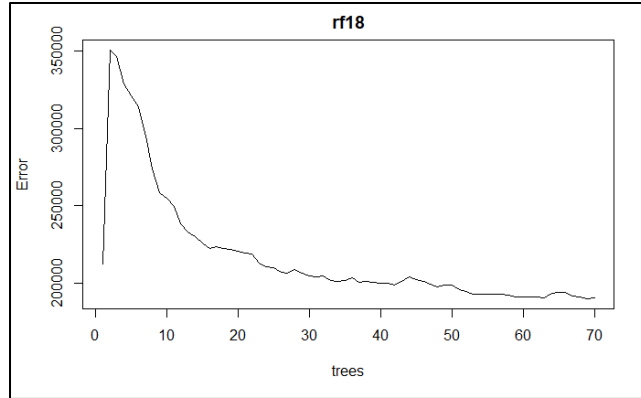
ANEXO 17: Número de árboles vs Error del modelo. Campaña

17

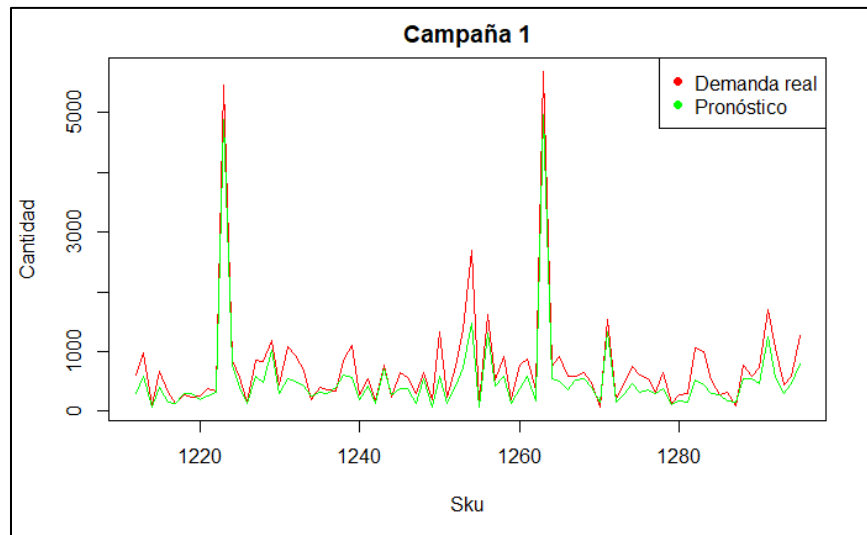


ANEXO 18: Número de árboles vs Error del modelo. Campaña

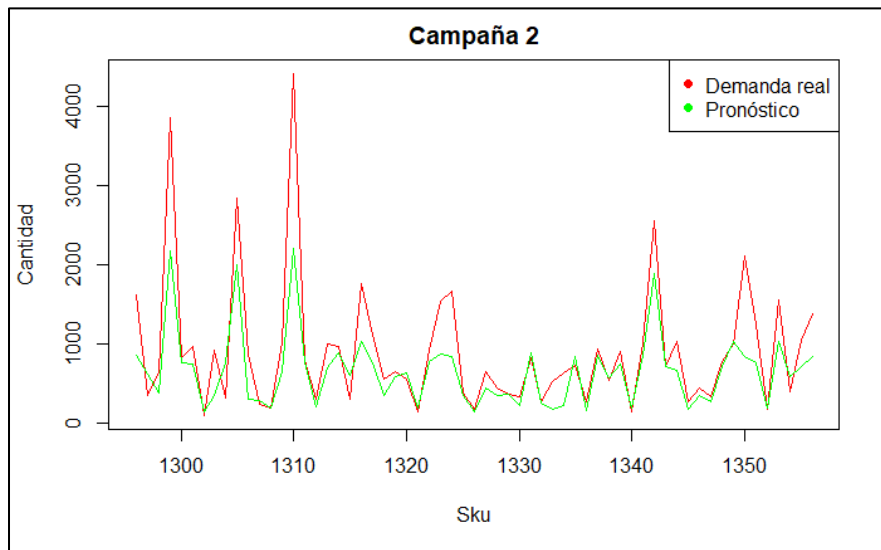
18



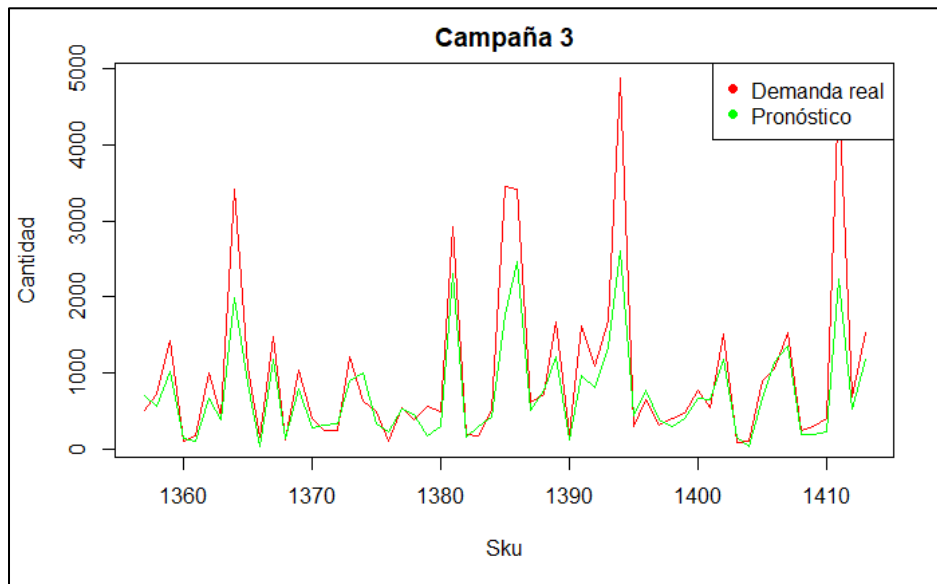
ANEXO 19: Demanda real vs pronostico. Campaña 1



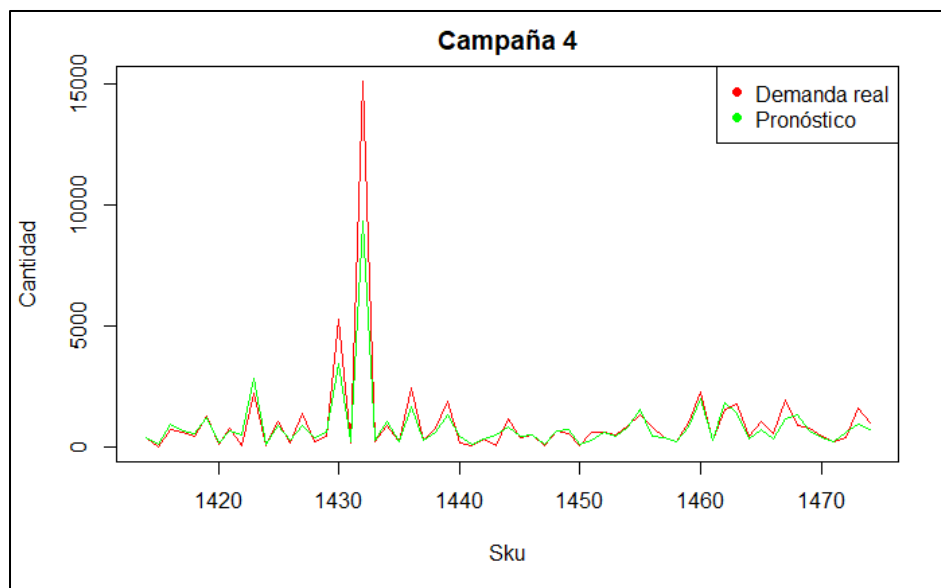
ANEXO 20: Demanda real vs pronostico. Campaña 2



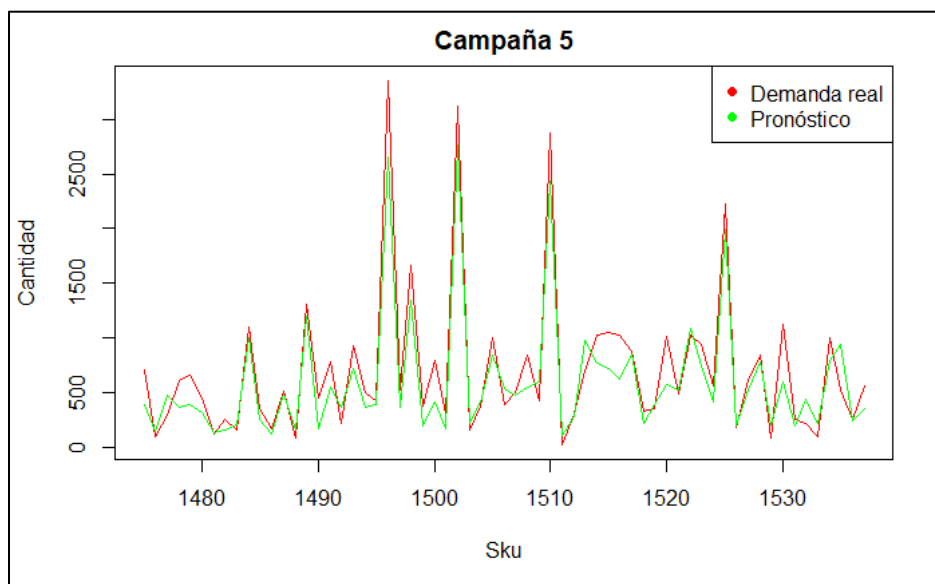
ANEXO 21: Demanda real vs pronostico. Campaña 3



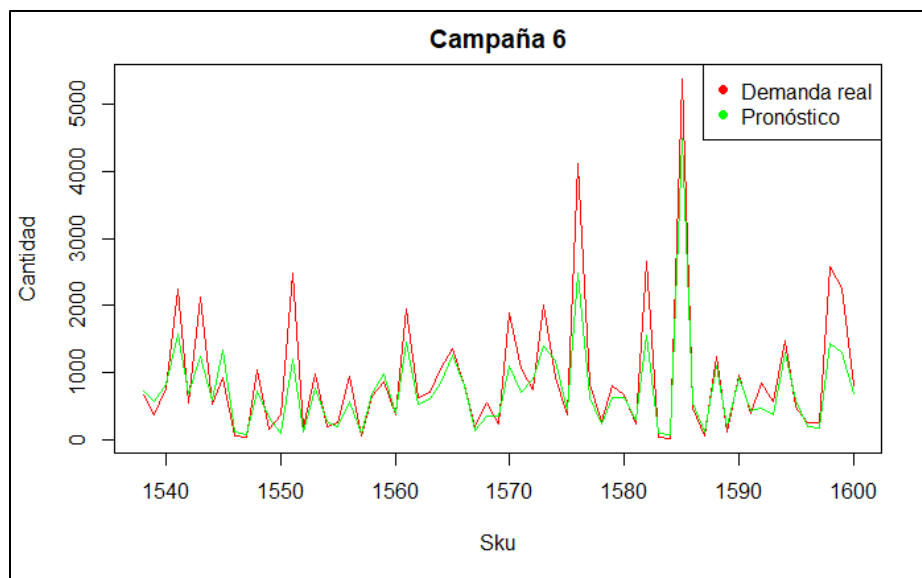
ANEXO 22: Demanda real vs Pronostico. Campaña 4



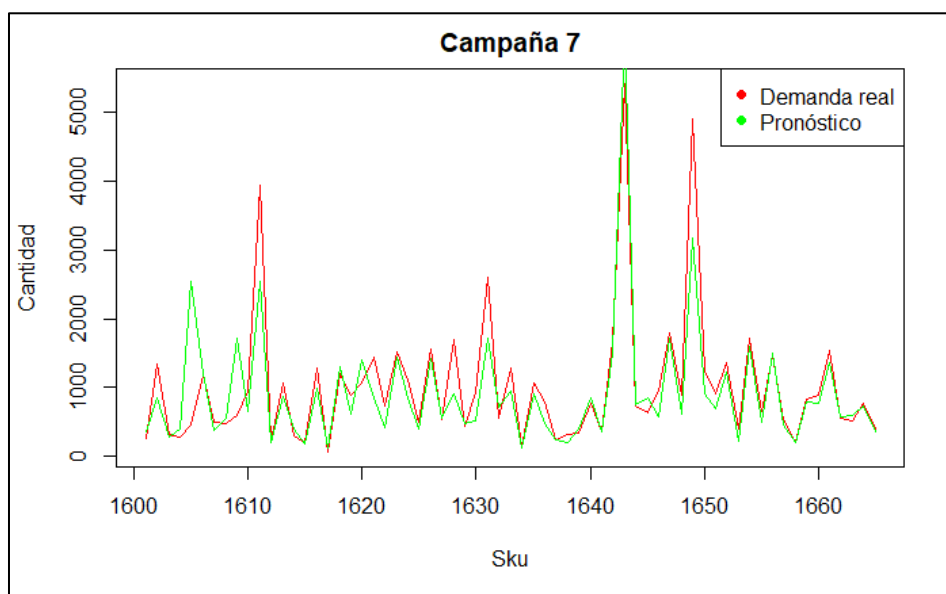
ANEXO 23: Demanda real vs pronostico. Campaña 5



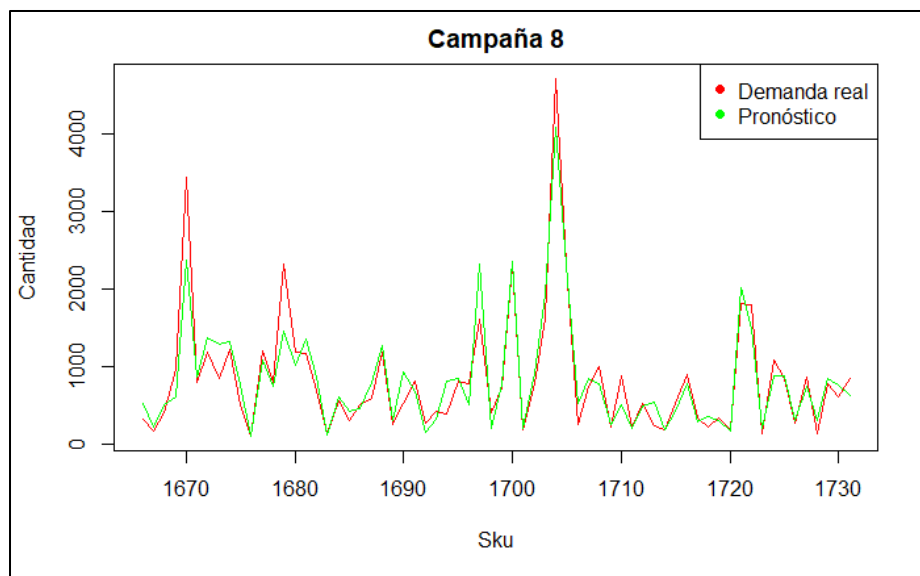
ANEXO 24: Demanda real vs pronóstico. Campaña 6



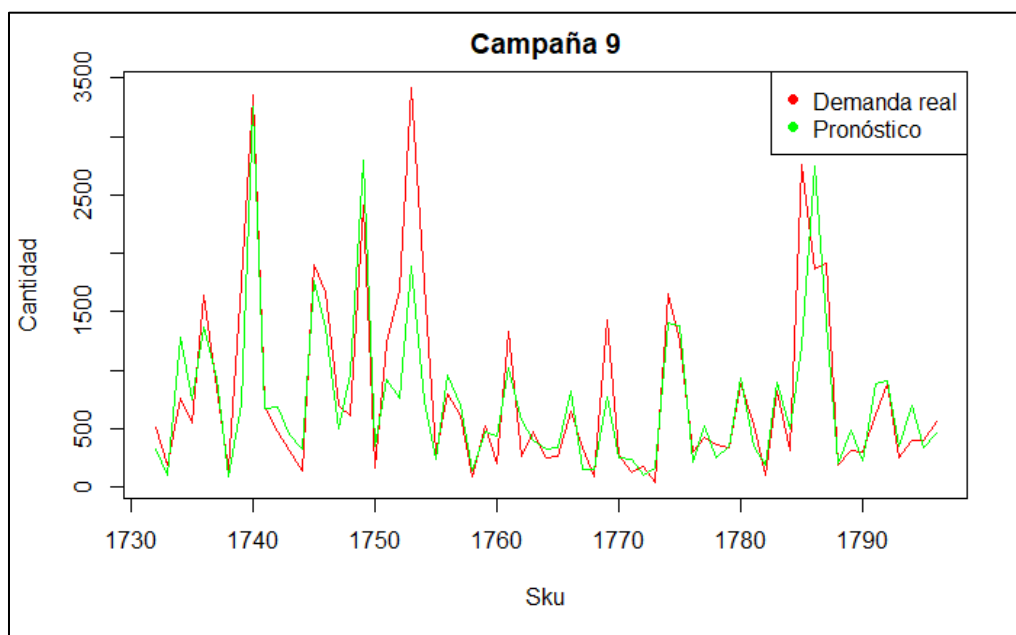
ANEXO 25: Demanda real vs pronóstico. Campaña 7



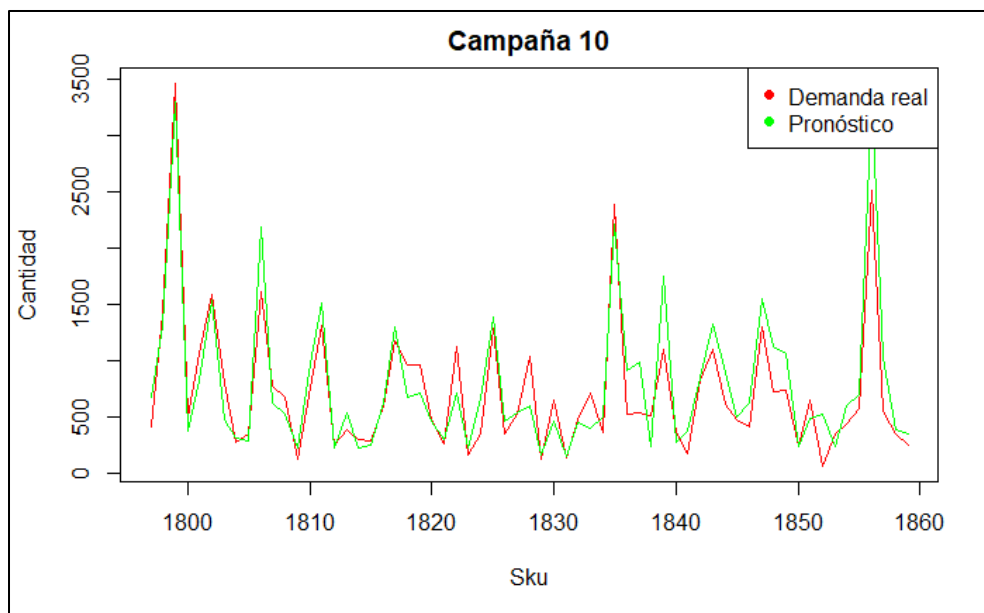
ANEXO 26: Demanda real vs pronostico. Campaña 8



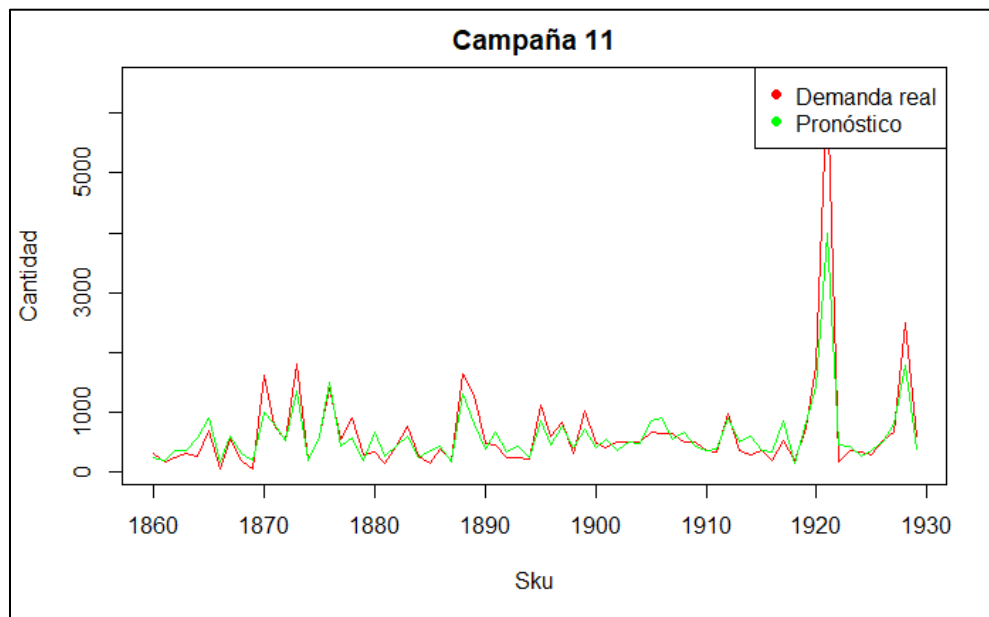
ANEXO 27: Demanda real vs pronostico. Campaña 9



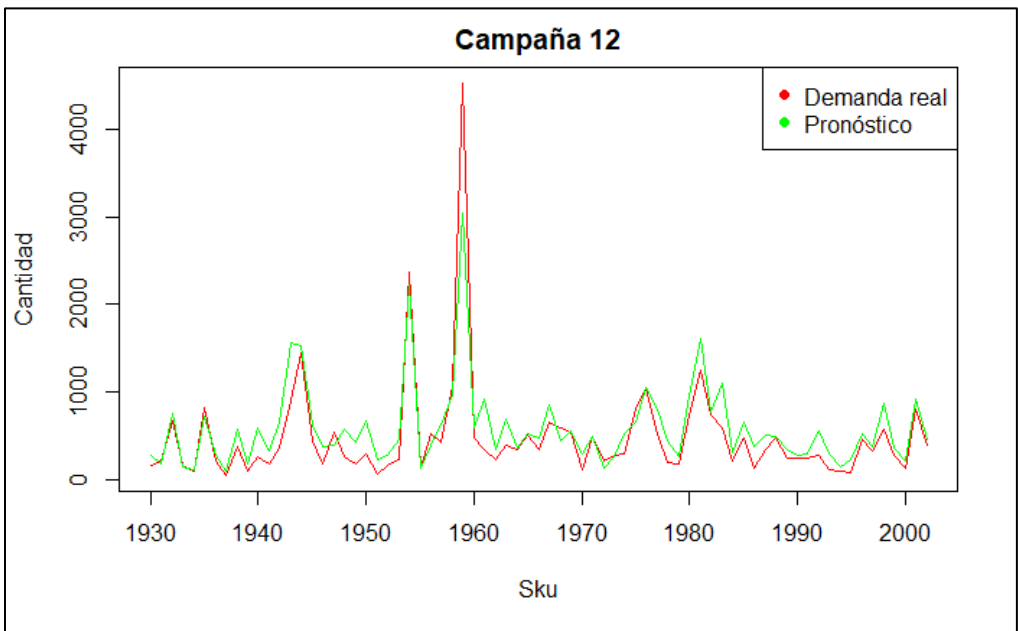
ANEXO 28: Demanda real vs pronóstico. Campaña 10.



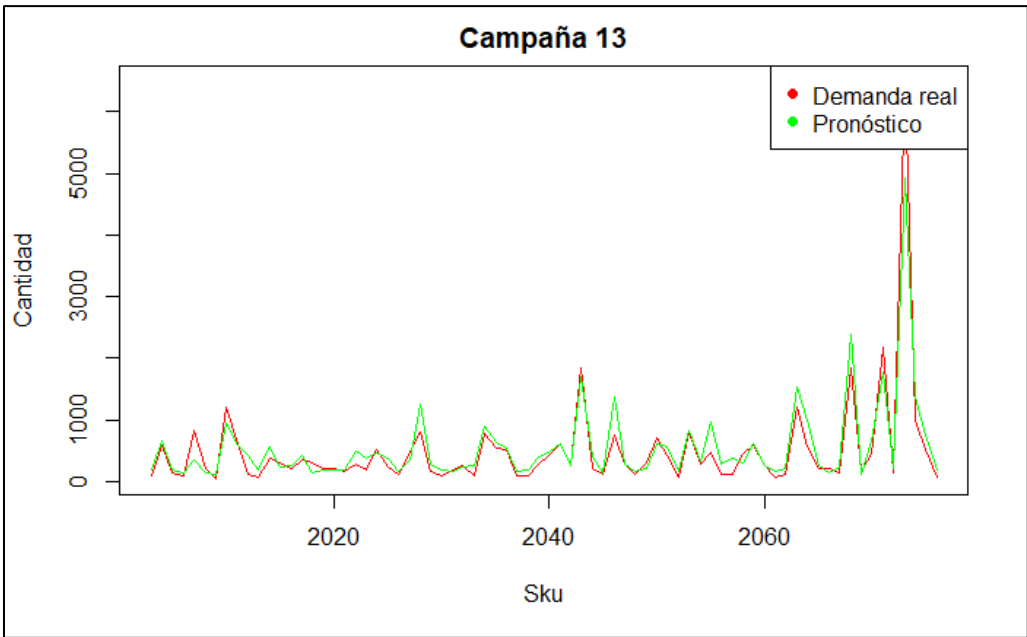
ANEXO 29: Demanda real vs pronóstico. Campaña 11



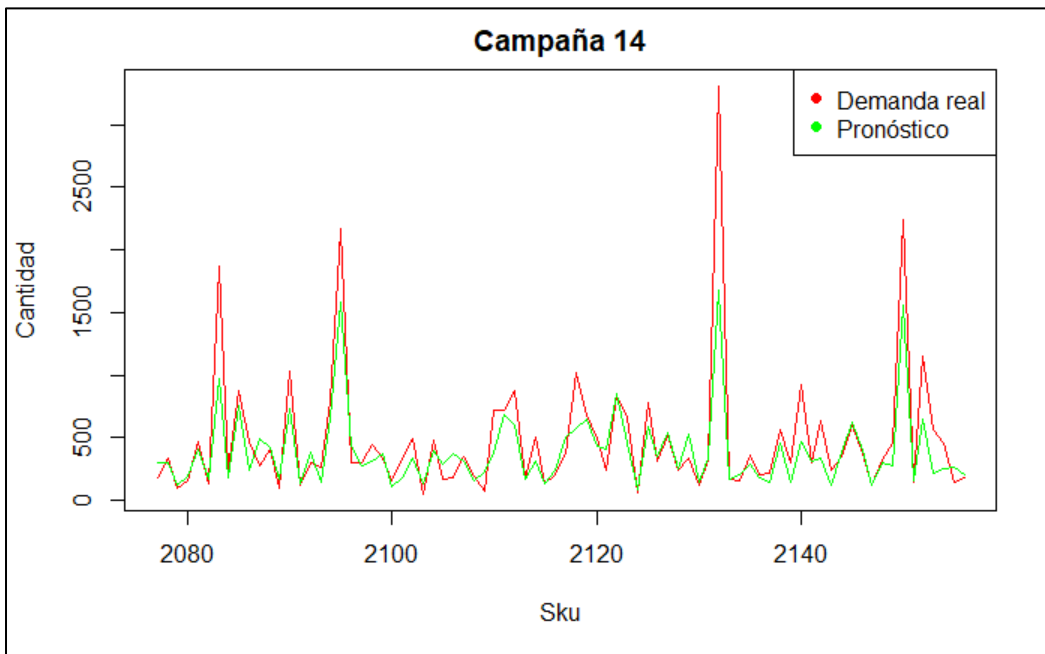
ANEXO 30: Demanda real vs pronostico. Campaña 12



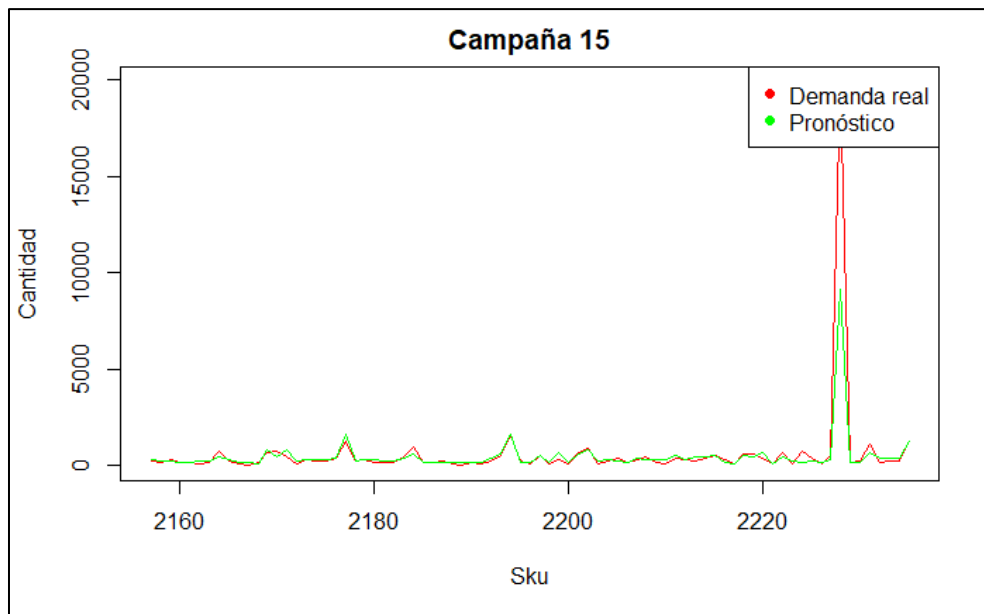
ANEXO 31: Demanda real vs pronostico. Campaña 13.



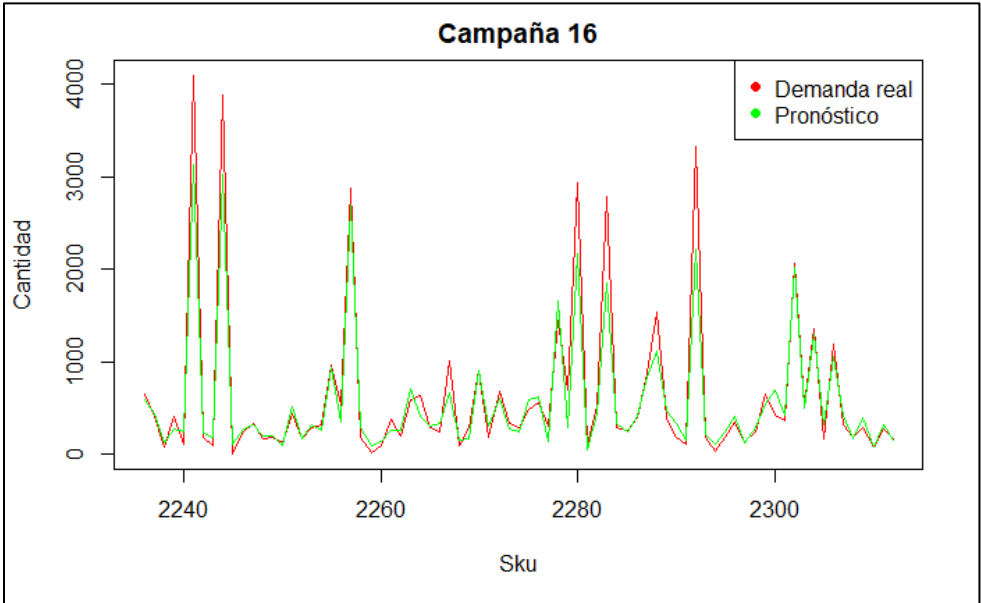
ANEXO 32: Demanda real vs pronostico. Campaña 14.



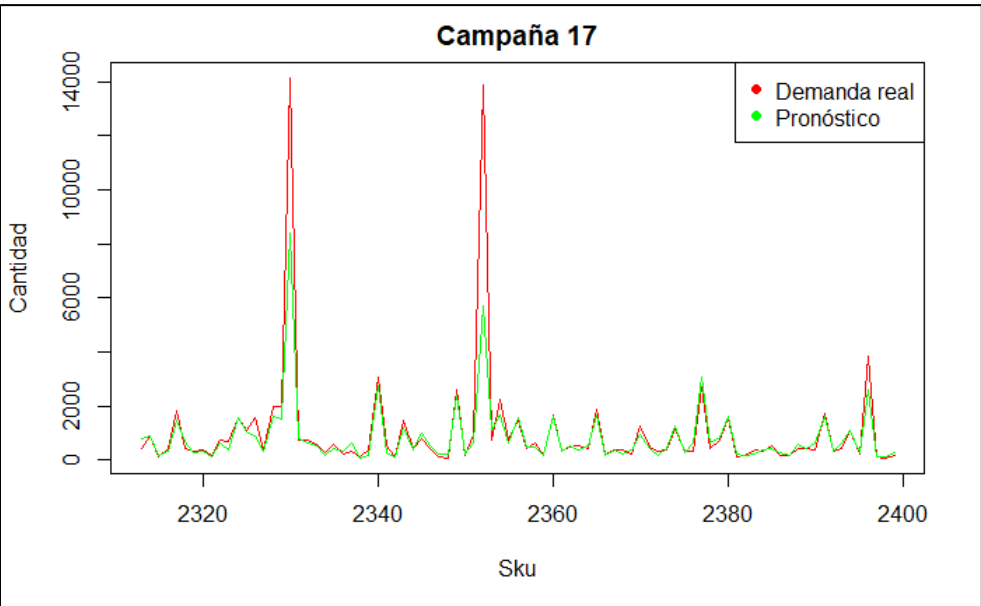
ANEXO 33: Demanda real vs pronostico. Campaña 15



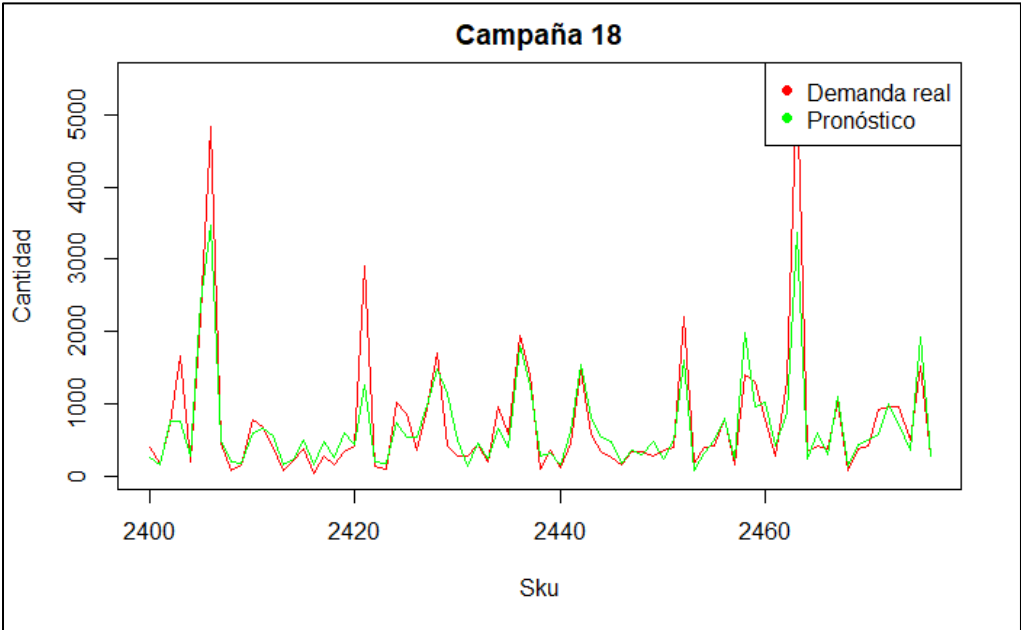
ANEXO 34: Demanda real vs pronostico. Campaña 16



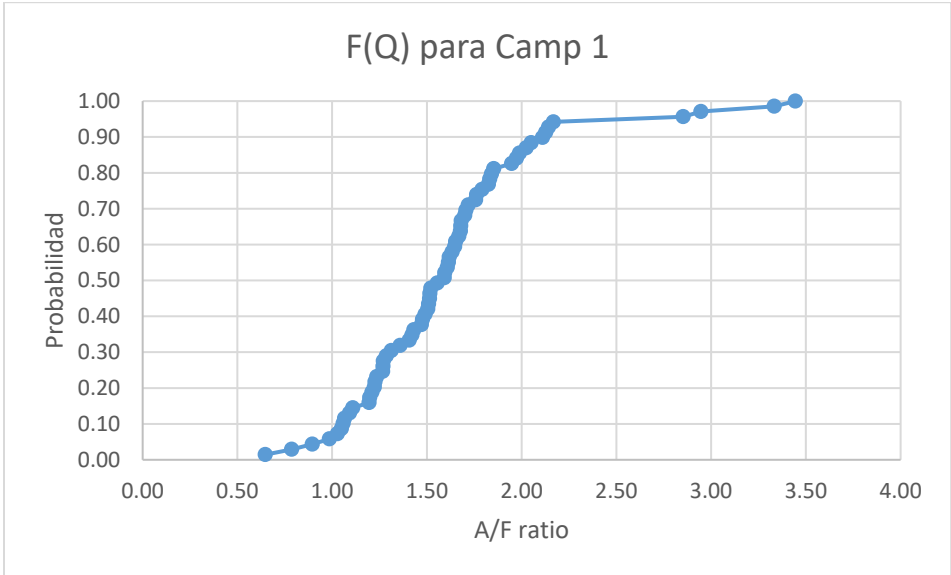
ANEXO 35: Demanda real vs pronostico. Campaña 17



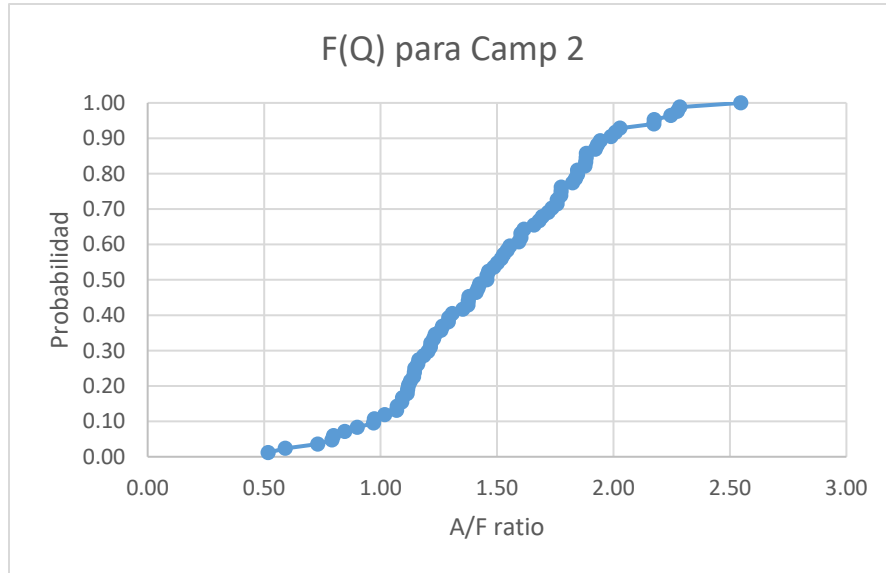
ANEXO 36: Demanda real vs pronostico. Campaña 18



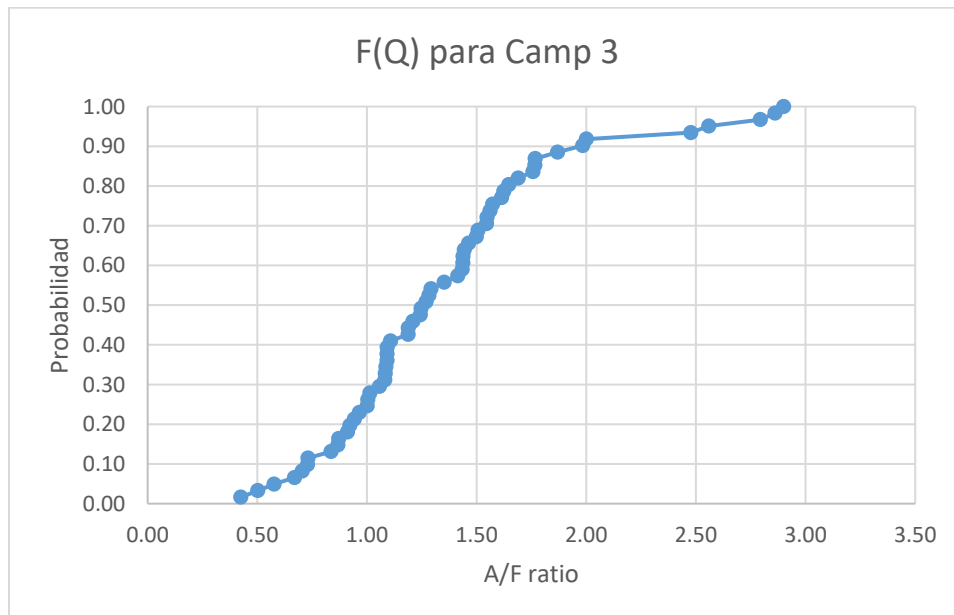
ANEXO 37: Relación A/F ratio versus probabilidad para la campaña 1 del 2019



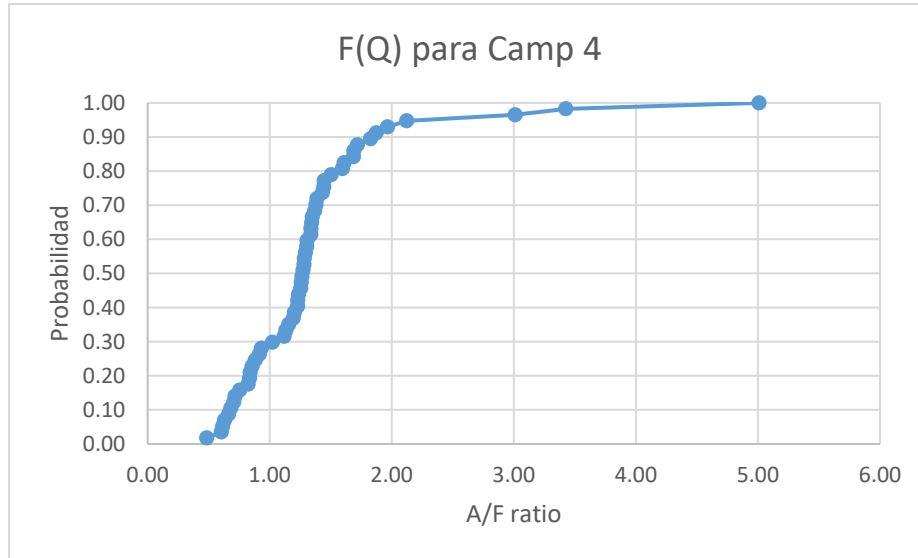
ANEXO 38: Relación A/F ratio versus probabilidad para la campaña 2 del 2019



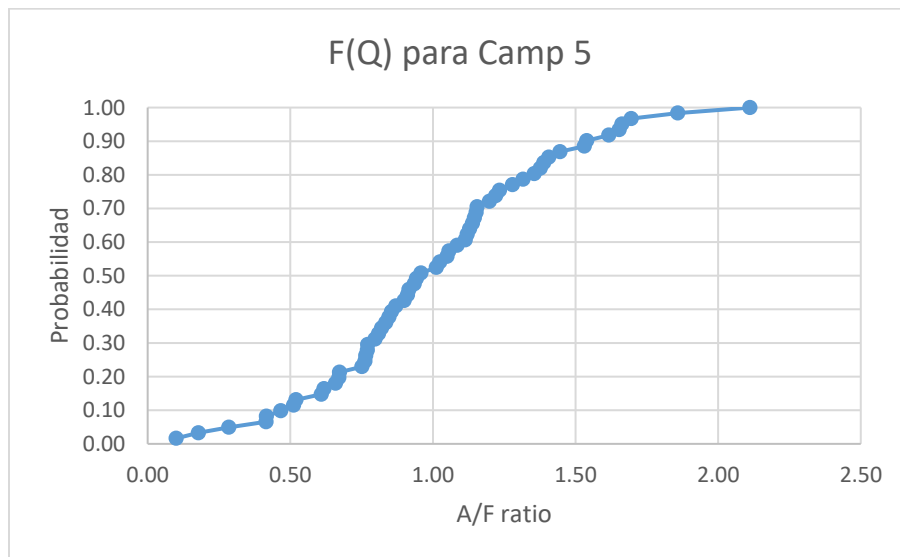
ANEXO 39: Relación A/F ratio versus probabilidad para la campaña 3 del 2019



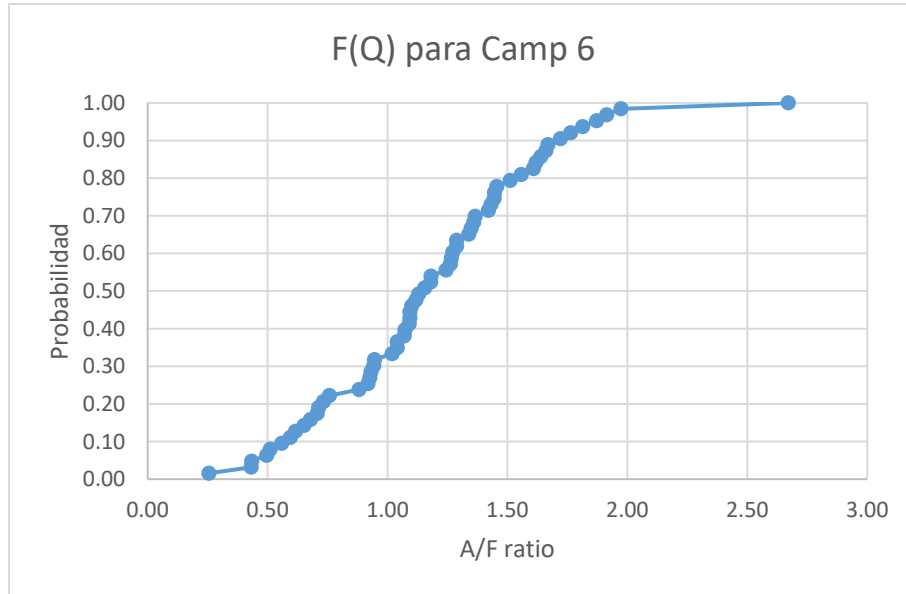
ANEXO 40: Relación A/F ratio versus probabilidad para la campaña 4 del 2019



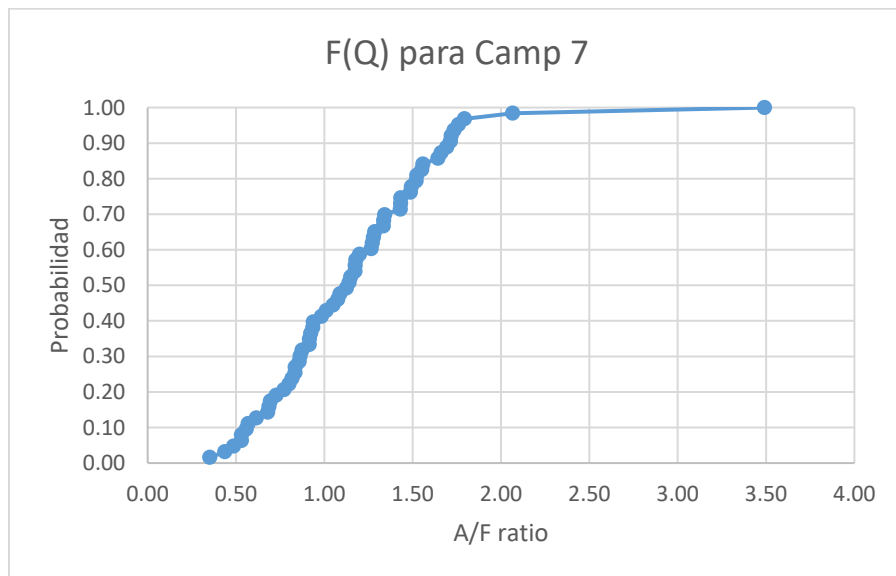
ANEXO 41: Relación A/F ratio versus probabilidad para la campaña 5 del 2019



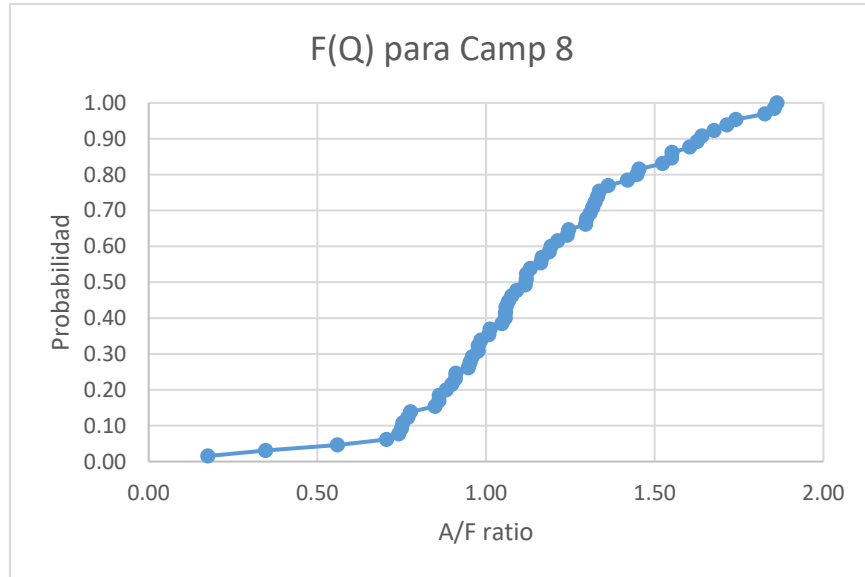
ANEXO 42: Relación A/F ratio versus probabilidad para la campaña 6 del 2019



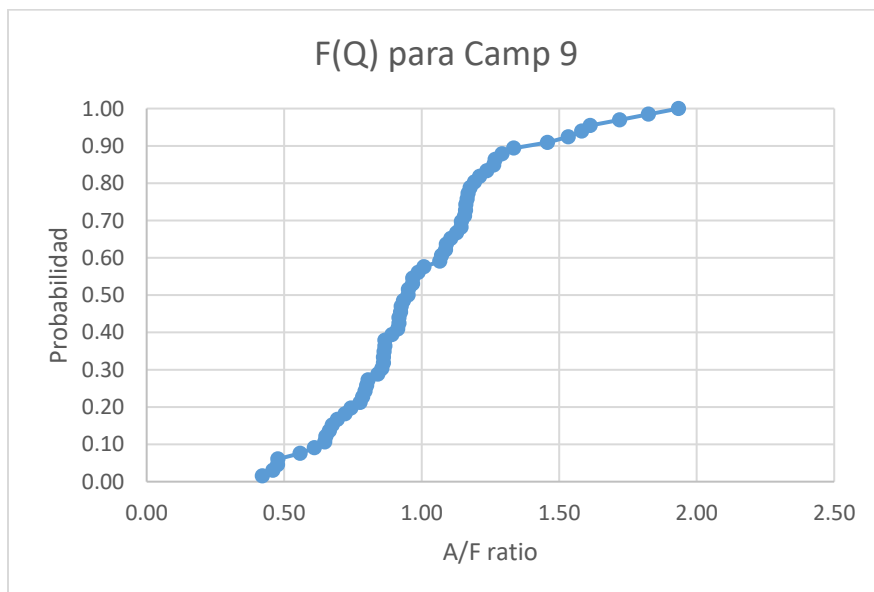
ANEXO 43: Relación A/F ratio versus probabilidad para la campaña 7 del 2019



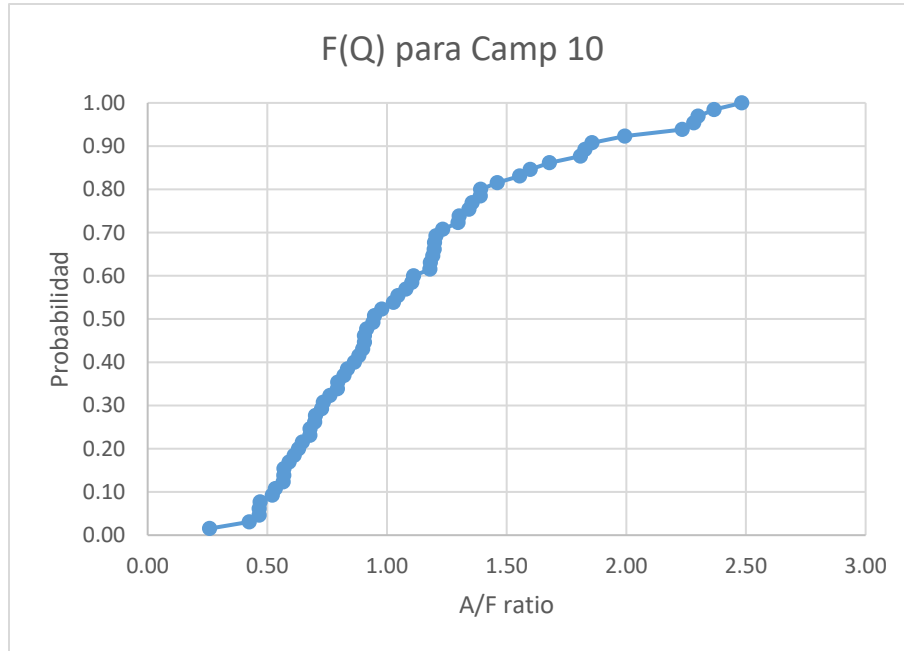
ANEXO 44: Relación A/F ratio versus probabilidad para la campaña 8 del 2019



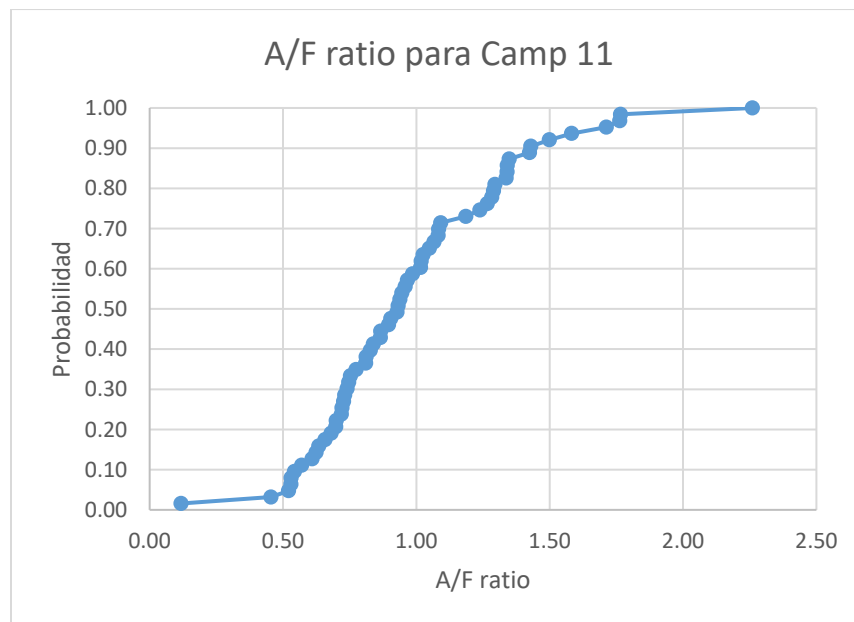
ANEXO 45: Relación A/F ratio versus probabilidad para la campaña 9 del 2019



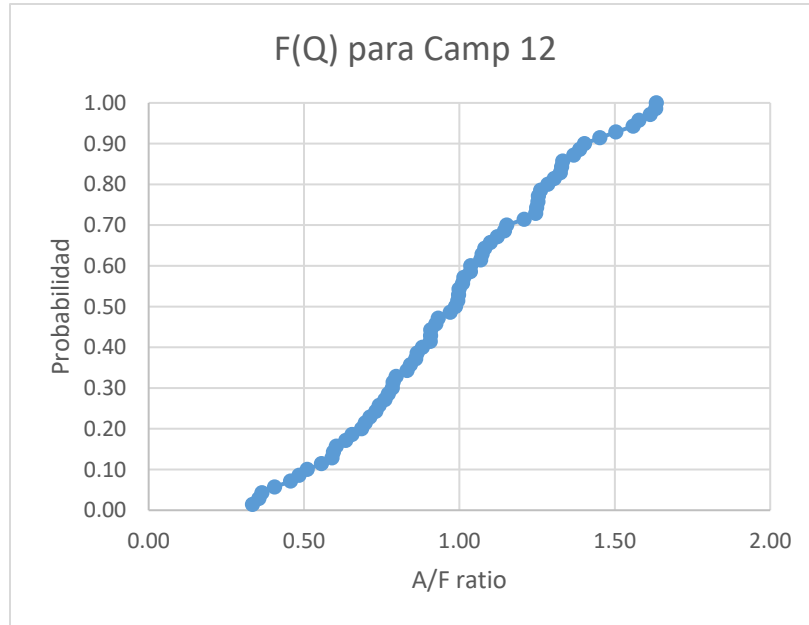
ANEXO 46: Relación A/F ratio versus probabilidad para la campaña 10 del 2019



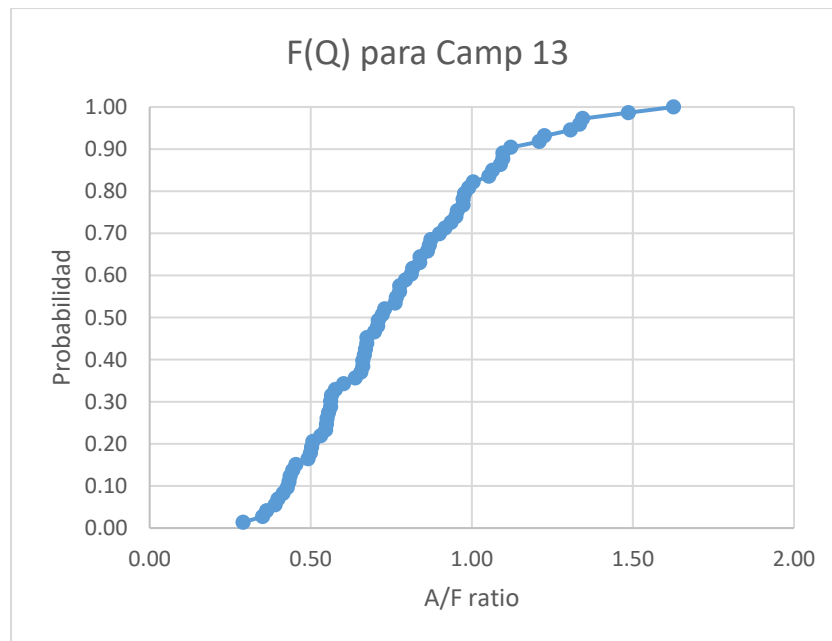
ANEXO 47: Relación A/F ratio versus probabilidad para la campaña 11 del 2019



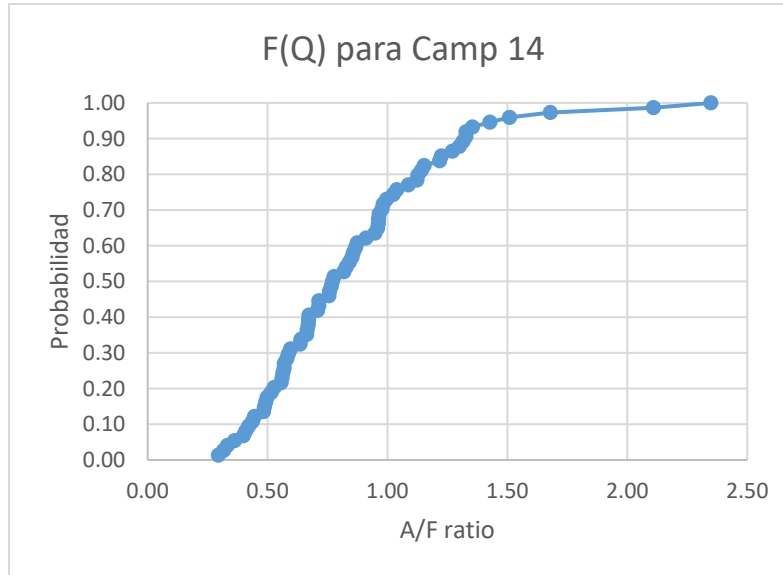
ANEXO 48: Relación A/F ratio versus probabilidad para la campaña 12 del 2019



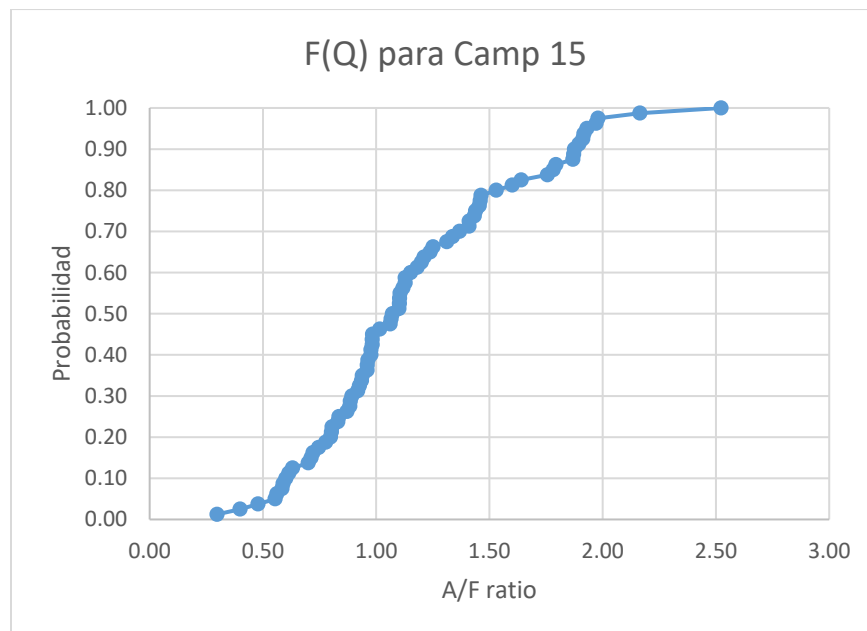
ANEXO 49: Relación A/F ratio versus probabilidad para la campaña 13 del 2019



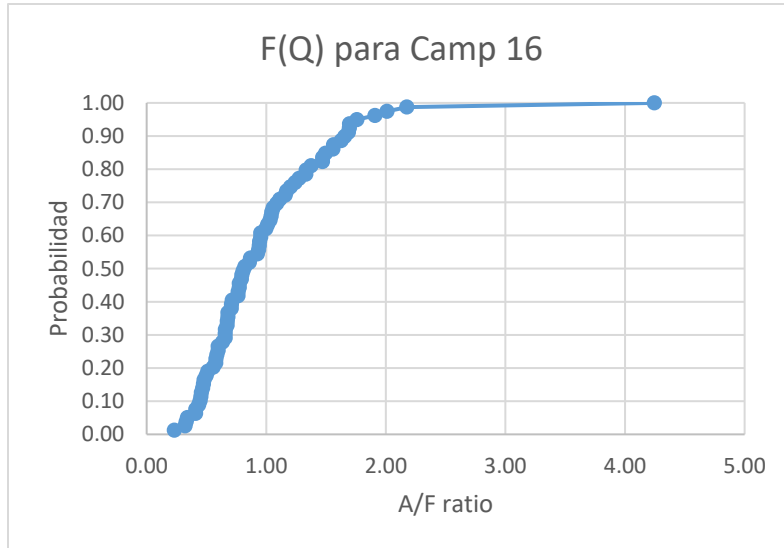
ANEXO 50: Relación A/F ratio versus probabilidad para la campaña 14 del 2019



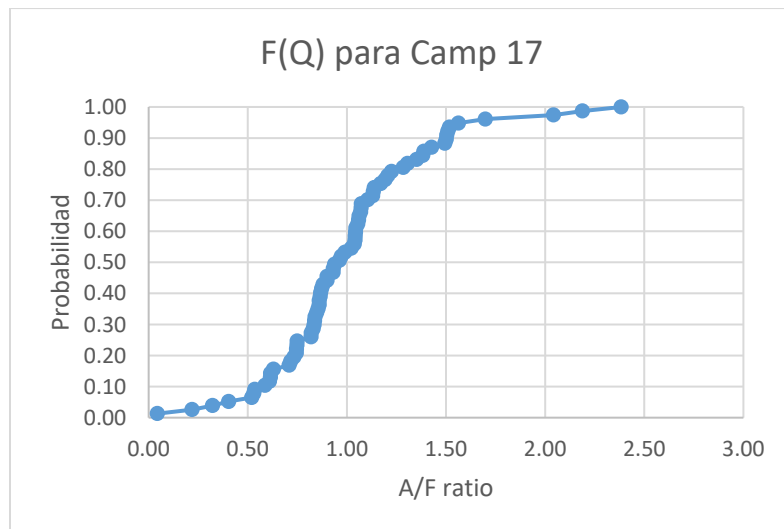
ANEXO 51: Relación A/F ratio versus probabilidad para la campaña 15 del 2019



ANEXO 52: Relación A/F ratio versus probabilidad para la campaña 16 del 2019



ANEXO 53: Relación A/F ratio versus probabilidad para la campaña 17 del 2019



ANEXO 54: Relación A/F ratio versus probabilidad para la campaña 18 del 2019

