



**TRABAJO DE GRADO**  
**Opción Seminario-Diplomado.**

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES  
EN COMPRA Y PRODUCCION DE HILOS EN FABRICATO, UTILIZANDO  
ESTRATEGIAS DE MACHINE LEARNING**

Corporación Universitaria Remington.  
Facultad de Ingeniería  
Ingeniería de Sistemas

Estudiantes:  
**RAFAEL ANDRES ROJAS OROZCO**  
Tutor: Juan Carlos Briñez de León  
Opción de Trabajo de grado Seminario-Diplomado.  
2024.

### **Dedicatoria**

A mi esposa, mi madre, mi hijo y mi hermano, quienes me han acompañado en este recorrido académico, su apoyo incondicional ha sido mi mayor fuerza y motivación. Sin ustedes, nada de esto tendría el mismo significado ni propósito. Cada paso dado, cada obstáculo superado, ha estado marcado por su amor y aliento, iluminando mi camino en los momentos más difíciles. Gracias por ser mi pilar y fuente de inspiración. Juntos hemos transformado esta lucha en una bella historia de crecimiento, sacrificio y victoria. ¡Por ustedes, por nuestro amor, por un futuro lleno de logros compartidos!

### **Agradecimientos**

Quiero expresar mi más sincero agradecimiento a cada uno de ustedes que han sido pilares fundamentales en mi trayecto académico. Agradezco a Dios por ser mi guía y fuerza en cada paso de este recorrido. A la institución educativa Uniremington por proporcionarme los conocimientos y recursos necesarios para alcanzar este logro. A mis queridos compañeros de estudio, por su colaboración, respaldo y amistad que han hecho este camino más llevadero. Al profesor Juan Carlos Briñez de León por su invaluable apoyo y estímulo en la culminación de este proyecto. Su contribución ha sido inestimable en mi proceso de aprendizaje y crecimiento académico.

A todos ustedes, les doy las gracias por ser parte integral de este capítulo tan relevante en mi vida académica. Su presencia y respaldo han posibilitado este logro.

## Tabla de Contenidos

### Contenido

1. Resumen.....	5
2. Palabras clave.....	6
3. Marco conceptual y contextual .....	6
3.1 Acceso a Tecnologías Avanzadas .....	8
3.2 Factores del mercado y sociales.....	8
3.3 Tendencias .....	9
3.4 ¿Qué es Machine Learning?.....	9
3.5 ¿Cómo funciona el Machine Learning? .....	10
4. Pregunta problema .....	11
5. Acercamiento a los datos .....	11
6. Descripción de variables .....	11
7. Posibles aplicaciones. ....	13
8. Aproximaciones con gráficos. ....	14
9. Objetivos.....	15
9.1. Objetivo general.....	15
9.2 Objetivos específicos. ....	16
10. Desarrollo e implementación del aprendizaje.....	16
11. Procesamiento de los datos .....	24
12. Modelo de toma de decisiones .....	26
12.1 Análisis de datos .....	26
12.2 Implementación en contextos reales .....	26
13. Resultados adicionales .....	27
14. Conclusiones .....	28
15. Referencias.....	30

## 1. Resumen

Este proyecto busca mejorar la gestión de la demanda de hilos en la empresa textil Fabricato mediante el uso de métodos de machine learning. El objetivo principal es optimizar la cadena de suministro, reducir costos y mejorar la planificación de compras y producción a través de una predicción precisa de las cantidades necesarias.

Este proyecto implica recopilar y depurar una base de datos que consta aproximadamente de 6,335 registros internos de Fabricato, los cuales incluyen información sobre explosión de hilos, adquisiciones de materias primas y pedidos de clientes. La calidad y consistencia estos datos son esenciales para un análisis posterior. Una vez recopilados y procesados, se emplearán algoritmos de aprendizaje automático, como regresión para analizar los datos y lograr extraer información relevante sobre tendencias, estas técnicas abarcan normalización de los datos, eliminación de valores atípicos e imputación valores de faltantes. Asimismo, se realizarán pruebas de validación y ajustes para garantizar la robustez del modelo desarrollado.

La aplicación de modelos predictivos permitirá a Fabricato prever la demanda de hilos en los próximos meses, teniendo en cuenta las fluctuaciones del mercado y los plazos de entrega de la materia prima. Esta anticipación es vital para mantener un flujo óptimo en el proceso productivo, evitar excesos de inventario y disminuir los costos asociados a la gestión de materiales innecesarios.

Además de pronosticar la demanda, el proyecto también se enfocará en evaluar la efectividad y precisión del software desarrollado mediante indicadores clave de rendimiento. Estas métricas abarcarán la reducción de gastos por excedentes en

inventario, el perfeccionamiento en la planificación de compras y la optimización de la cadena de suministro de hilos.

La digitalización y aplicación de algoritmos basados en aprendizaje automático han sido cruciales para optimizar los procesos y facilitar la toma de decisiones dentro del ámbito del suministro textil. La recopilación detallada y análisis exhaustivo de los datos, junto con estrategias avanzadas de regresión y clasificación, han contribuido significativamente a mejorar el desempeño global de Fabricato.

## **2. Palabras clave**

Gestión de la demanda de hilos, Dataset de Fabricato, Análisis de datos, Machine learning, Clasificación, Regresión lineal, Cadena de suministro, Optimización de inventarios, Planificación de compras, Industria textil.

## **3. Marco conceptual y contextual**

En la actualidad, el análisis de datos relacionados con la gestión de la demanda de hilos ha adquirido una importancia significativa debido a la necesidad de optimizar los procesos industriales y desarrollar estrategias efectivas de gestión de inventarios. La combinación de algoritmos informáticos y técnicas de aprendizaje automático brinda nuevas oportunidades para analizar y tomar decisiones fundamentadas en datos en este sector crítico de la cadena de suministro.

Explorar los datos vinculados a la demanda de hilos puede ser sumamente beneficioso para comprender más a fondo este fenómeno complejo y diseñar estrategias gerenciales más eficientes. El aprendizaje automático ofrece un conjunto diverso de herramientas y técnicas que posibilitan el análisis exhaustivo de extensas bases de datos sobre la demanda del hilo, detectar pautas en el consumo y construir modelos predictivos. (Hastie, 2009).

La gestión del abastecimiento textil representa un desafío importante en el ámbito industrial. En Fabricato, esta labor ha ganado un papel crucial debido a su impacto en el funcionamiento eficiente y los beneficios económicos corporativos. De acuerdo con sucesos recientes, una administración inadecuada del abastecimiento puede resultar en excedentes o escaseces en inventario, ambos con consecuencias negativas en términos económicos y en la capacidad empresarial para satisfacer las exigencias del mercado.

En Fabricato, se han llevado a cabo algunos estudios sobre la gestión de la demanda, pero aún se necesita más investigación para comprender mejor este fenómeno. El Aprendizaje Automático podría ser una herramienta valiosa para analizar los datos existentes sobre la demanda de hilos e identificar patrones específicos para la operación de Fabricato.

De acuerdo con datos internos de Fabricato, la demanda de hilos ha experimentado variaciones significativas en los últimos años, influenciada por factores como las tendencias del mercado, las temporadas de producción y las características específicas de los pedidos de los clientes. La capacidad de prever estas variaciones con precisión puede tener un impacto importante en la planificación y eficiencia de la cadena de suministro empresarial.

La estructura socioeconómica del mercado textil abarca temas como el nivel de demanda de productos, el ciclo de vida textil y las condiciones del mercado. Elementos como cambios en precios, competencia y disponibilidad de materias primas pueden contribuir al desafío que implica gestionar eficazmente la demanda por hilos.

El uso de algoritmos de ensamble, como los Random Forests, mejora significativamente la precisión de los modelos (Breiman, 2001).

### **3.1 Acceso a Tecnologías Avanzadas**

La importancia y la habilidad de aplicar tecnologías de vanguardia en Fabricato son esenciales para mejorar la gestión de la demanda. Es necesario explorar el uso de herramientas de análisis de datos, plataformas de Aprendizaje Automático y sistemas integrados de información para respaldar la toma de decisiones basada en datos.

### **3.2 Factores del mercado y sociales**

Los factores del mercado y sociales, como las preferencias de compra, la variabilidad estacional de los productos y las necesidades de los consumidores, también juegan un papel en la demanda de hilos. Es esencial considerar la interacción entre estos elementos y su impacto en la planificación de la producción y el control de inventarios.

### **3.3 Tendencias**

Detectar las tendencias de demanda de hilos, como los cambios estacionales, los momentos de mayor demanda y las variaciones en las preferencias de los consumidores, proporciona información valiosa para planificar y evaluar estrategias de gestión. También es crucial considerar la interacción con otros factores del mercado y el entorno económico.

Evaluar el rendimiento en la gestión de la demanda en Fabricato resulta beneficioso al identificar posibles deficiencias y aplicar intervenciones oportunas y adecuadas. En este proceso evaluativo, se sugiere emplear modelos predictivos basados en técnicas de Aprendizaje Automático para reconocer patrones de demanda y ajustar las estrategias de gestión en consecuencia.

### **3.4 ¿Qué es Machine Learning?**

El aprendizaje automático es un campo científico que forma parte de la inteligencia artificial. Implica permitir que los algoritmos identifiquen regularidades en conjuntos de datos. Estos datos pueden ser números, palabras, imágenes, estadísticas, entre otros. Cualquier información que se pueda almacenar digitalmente puede utilizarse como datos para el aprendizaje automático. Al reconocer patrones en estos datos, los algoritmos adquieren conocimiento y mejoran su desempeño en la realización de una tarea específica (Breiman, 2001).

### 3.5 ¿Cómo funciona el Machine Learning?

Hay cuatro fases clave en el desarrollo de un modelo de Aprendizaje Automático. Por lo general, es un Científico de Datos quien supervisa y dirige este proceso.

En primer lugar, se escoge y prepara un conjunto de datos de entrenamiento. Estos datos son empleados para enseñar al modelo de Aprendizaje Automático a resolver la tarea específica para la que fue diseñado. Los datos pueden estar etiquetados para indicar al modelo las características que debe reconocer o no estar etiquetados, en cuyo caso el modelo debe identificar y extraer las características importantes por sí mismo.

La siguiente etapa consiste en elegir un algoritmo adecuado para aplicar al conjunto de datos de entrenamiento. El tipo de algoritmo seleccionado dependerá del tipo y cantidad de datos disponibles y del problema particular que se desea abordar.

El tercer paso implica entrenar el algoritmo. Este proceso implica repetir iterativamente la ejecución del algoritmo con las variables pertinentes, comparando los resultados obtenidos con los esperados. Los “pesos” y el sesgo pueden ajustarse para mejorar la precisión del resultado.

Finalmente, la última fase abarca tanto la implementación como la optimización del modelo. Se aplica el modelo a nuevos conjuntos de datos, dependiendo siempre del problema específico que se esté tratando resolver.

Por ejemplo, en la administración de existencias de hilos se empleará un modelo de Aprendizaje Automático creado para anticipar la demanda venidera.

A pesar de sus beneficios, el machine learning enfrenta desafíos importantes relacionados con la interpretabilidad y la ética (Murphy, 2012). El futuro del aprendizaje automático incluye el desarrollo de algoritmos más robustos y adaptables (Chollet, 2018).

#### **4. Pregunta problema**

¿Una estrategia computacional basada en algoritmos de Machine Learning puede ayudar en la identificación temprana de la demanda de hilos en el proceso textil.

#### **5. Acercamiento a los datos**

Los datos empleados para este proyecto de mejora en la gestión de la demanda de hilos en Fabricato fueron recolectados internamente por la compañía. Estos datos se derivan de los registros internos, que incluyen detalles sobre la demanda de hilos, la manufactura textil, las adquisiciones de materias primas, las ordenes realizadas por los clientes y otros aspectos cruciales para la administración de la cadena de suministro y compras en Fabricato (Pedregosa).

#### **6. Descripción de variables**

**ANNO:** (Año): Variable de tipo entero que indica el año en el que se realizó la programación o pedido de las telas por el cliente.

**MES:** Variable de tipo entero que especifica el mes en el que se efectuó la programación o la programación de las telas.

**TA:** Variable de tipo entero que representa el código de la línea del producto.

**NOMBRE\_TELA:** Variable de tipo objeto (texto) que identifica el nombre de la y que fue asignadas por el departamento de nuevos productos.

**DITEPT:** Variable de tipo objeto (texto) que identifica el diseño que debe tener la tela el proceso de tejeduría.

**CANPED:** Variable de tipo entero que identifica la cantidad de metros que necesita el cliente para realizar su producción.

**TA\_PADRE:** Variable de tipo entero que identifica el peso de la tela y el tipo de mezcla intima para realizar la producción requerida de la referencia o base solicitada por el cliente.

**CLAVE1\_HIJO:** Variable de tipo objeto (texto) que identifica la posición en que el hilo debe tener en el proceso de tejeduría para realizar la producción requerida de la referencia o base solicitada por el cliente.

**CLAVE2\_HIJO:** Variable de tipo objeto (texto) que identifica el peso del hilo y los numero cabos que debe tener el hilo para realizar la producción requerida de la referencia o base solicitada por el cliente.

**CLAVE3\_HIJO:** Variable de tipo objeto (texto) que identifica la mezcla intima que tiene el hilo, es decir si es algodón 100%, si es poliéster algodón, entre otros mezclas, para realizar la producción requerida de la referencia solicitada por el cliente.

**CLAVE4\_HIJO:** Variable de tipo objeto (texto) que identifica el sentido de la torsión del hilo, es decir si es Z o S para realizar la producción requerida de la referencia solicitada por el cliente.

**CLAVE5\_HIJO:** Variable de tipo objeto (texto) que identifica si el hilo debe tener el proceso de parafinado o no en el momento de ser envuelto para realizar la producción requerida de la referencia o base solicitada por el cliente.

**KGS:** Variable de tipo objeto (texto) que identifica el tipo de color con que se requiere el hilo para realizar la producción requerida de la referencia o base solicitada por el cliente.

**YYARWA:** Variable de tipo entero que identifica la cantidad de kilos necesarios para realizar la producción requerida de la referencia o base solicitada por el cliente.

## 7. Posibles aplicaciones.

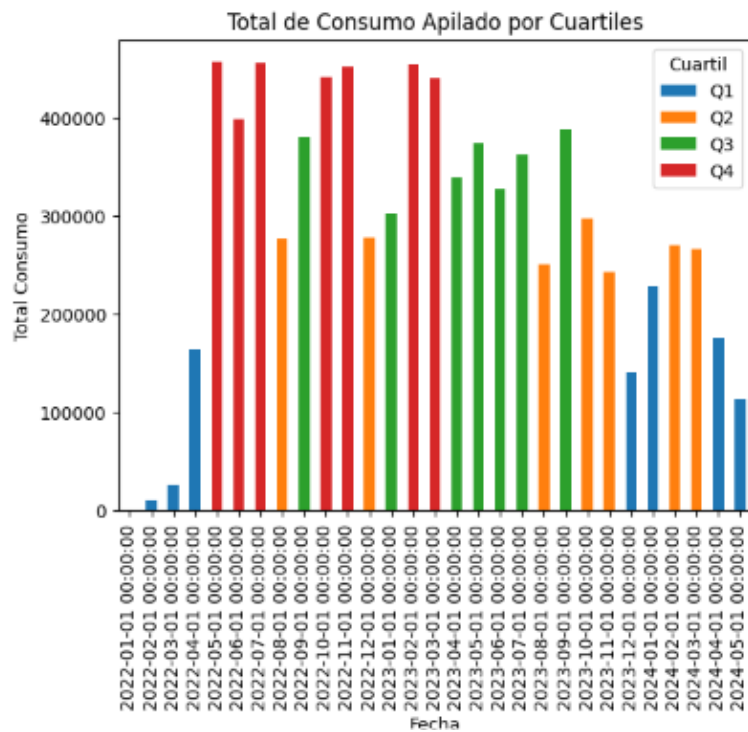
El plan para implementar un modelo de regresión lineal de múltiples entradas y una salida ofrece diversas oportunidades estratégicas a Fabricato en la gestión de la demanda de hilos. Este modelo permitirá optimizar los inventarios al prever con precisión las necesidades de materia prima, reduciendo el exceso y garantizando disponibilidad. Además, facilitará la planificación de la producción al anticipar la demanda según pedidos y tendencias del mercado, mejorando así la eficiencia operativa. En cuanto a la gestión de compras, posibilitará tomar decisiones más fundamentadas al determinar cuándo y cuánto adquirir hilos, minimizando los costos asociados con compras urgentes o innecesarias. El análisis de tendencias emergentes contribuirá a ajustar proactivamente las estrategias comerciales y productivas, mientras que la predicción de costos permitirá una planificación financiera más precisa. Introducido como parte integral de un proceso constante de mejora, el modelo perfeccionará continuamente sus algoritmos y ajustará variables en función de los resultados y comentarios recibidos, asegurando así una

personalización efectiva del servicio y una automatización eficiente en las decisiones empresariales. Estas aplicaciones resaltan cómo este modelo puede fortalecer la competitividad y eficiencia de Fabricato en el sector textil mediante valiosas perspectivas predictivas.

El uso de redes neuronales profundas ha demostrado ser efectivo en una variedad de aplicaciones complejas (Goodfellow, 2016). El algoritmo K-Nearest Neighbors (KNN) es simple pero potente, especialmente en problemas de clasificación (Bishop, 2006). El plan para implementar un modelo de regresión lineal de múltiples entradas y una salida ofrece diversas oportunidades estratégicas a Fabricato en la gestión de la demanda de hilos (Hastie, 2009).

## **8. Aproximaciones con gráficos.**

Con el desarrollo del trabajo, se espera generar gráficos que permita la toma de decisiones sobre el consumo de hilos de hilos por cuartiles, en la figura 1 del artículo Total Consumo por cuartiles encontramos el comportamiento en porcentajes consumo



*Figura 1 Total de consumo aplicado por cuartiles*

## 9. Objetivos

### 9.1. Objetivo general.

Implementar un algoritmo computacional para el análisis de datos y pronósticos precisos, para que ayude a la empresa Fabricato a mejorar la gestión de la demanda de hilos utilizando estrategias de machine learning.

## 9.2 Objetivos específicos.

- Recolectar y procesar una base de datos de alrededor de 6,335 registros, asegurando que la información sea de calidad y consistente para su posterior análisis.
- Utilizar modelos predictivos para analizar las tendencias de las cantidades de hilos necesarias para la producción y compra en los próximos meses, teniendo en cuenta las variaciones del mercado y los tiempos de entrega de la materia prima.
- Evaluar la eficacia y precisión del software desarrollado a través de métricas de rendimiento, como la reducción de costos por excedentes en inventario, la mejora en la planificación de compras y la optimización de la cadena de suministro de hilos en Fabricato.

## 10. Desarrollo e implementación del aprendizaje

Para aplicar el modelo de regresión lineal con múltiples entradas en Fabricato, se llevará a cabo un procedimiento que implica la preparación de datos históricos internos, la elección y capacitación del modelo, la validación a través de métricas como el error cuadrático medio, la integración en el sistema de gestión empresarial y ajustes continuos basados en retroalimentación y nueva información.

**FECHA:** Variable de tipo entero que indica el año y el mes en el que se realizó la programación o pedido de las telas por el cliente.

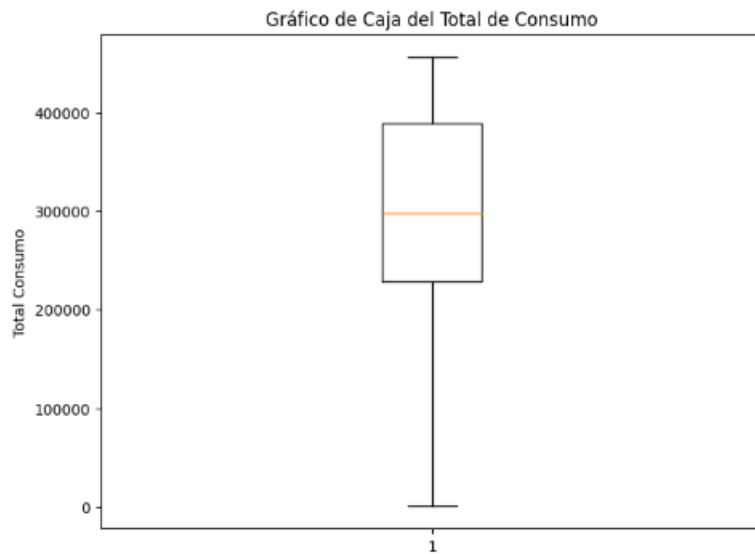
**HILOS:** Variable de tipo objeto (texto) que identifica la materia prima que es necesaria para cumplir con los requerimientos solicitados por el cliente en su programación inicial.

**CANTIDAD:** Variable de tipo entero que identifica la cantidad de kilos necesarios para realizar la producción requerida de la referencia o base solicitada por el cliente.

Se indagó por el comportamiento del consumo del hilo en los años 2022, 2023 y 2024, con el fin de analizar cuál es el comportamiento del valor medio en estos años, esto a través de un gráfico de bigotes. Este proceso se realizó usando la herramienta Colab de Google, especializada para machine Learning a través del código de la figura 2, en esta se evidencia que su consumo promedio es de 300.000 kilos mensuales. En la figura 3 se evidencia que en el año 2022 y 2023 el promedio de consumo es más alto que el promedio global, en estos años el consumo es 330.000 kilos en el año 2022 y 350.000 kilos en el año 2023, es decir que el consumo más representativo a tener en cuenta es el año 2023 debido a su alto consumo.

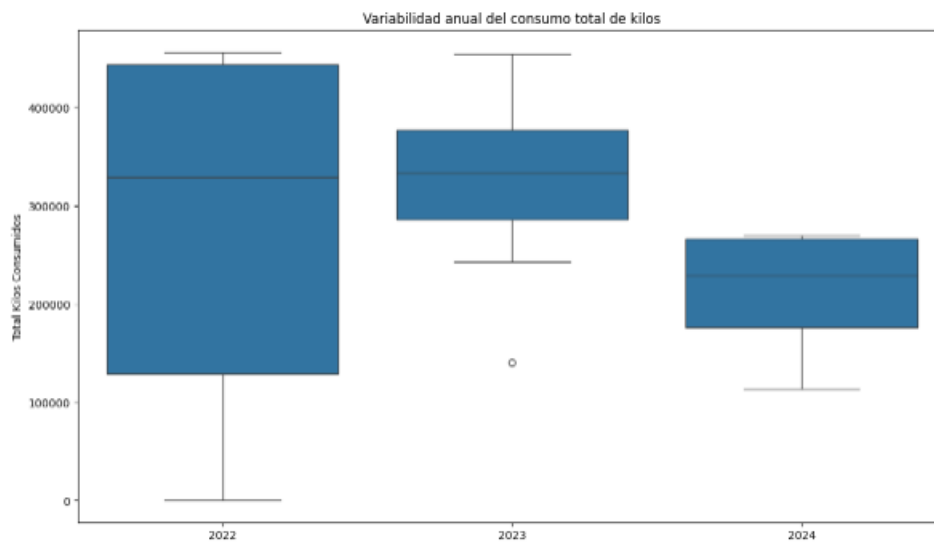
El código utilizado para esta gráfica fue el siguiente:

```
# Gráfico de Caja del Total de Consumo
plt.figure(figsize=(8, 6))
plt.boxplot(data['Total consumo'])
plt.title('Gráfico de Caja del Total de Consumo')
plt.ylabel('Total Consumo')
plt.show()
```



*Figura 2 Gráfica de Bigotes sobre el consumo total en kilos*

```
# Boxplot para la variabilidad del consumo total de kilos por años
data_subset['año'] = data_subset['fecha'].dt.year
plt.figure(figsize=(14, 8))
sns.boxplot(x='año', y='Total_consumo', data=data_subset)
plt.title('Variabilidad anual del consumo total de kilos')
plt.xlabel('Año')
plt.ylabel('Total Kilos Consumidos')
plt.show()
```



*Figura 3 Gráfica de Bigotes sobre el consumo total en kilos por año*

La Figura 4 “Gráfica de líneas del consumo de hilos por meses” nos permite evaluar el comportamiento de la necesidad de la materia prima e identificar meses con mayor demanda de producción, adicionalmente nos permite analizar que las necesidades de hilo a partir del segundo semestre del año 2023 empiezan a tener un declive bastante considerable.

Este también es un código de Machine Learning realizado en la herramienta de Google Colab.

El código con el que se realizó el análisis fue el siguiente:

```
# Gráfico de líneas para el consumo total de kilos por fecha
plt.figure(figsize=(14, 8))
plt.plot(data_subset['fecha'], data_subset['total_kilos_consumidos'], marker='o')
plt.title('Tendencia del consumo total de kilos por fecha')
plt.xlabel('Fecha')
plt.ylabel('Total Kilos Consumidos')
plt.grid(True)
plt.show()
```

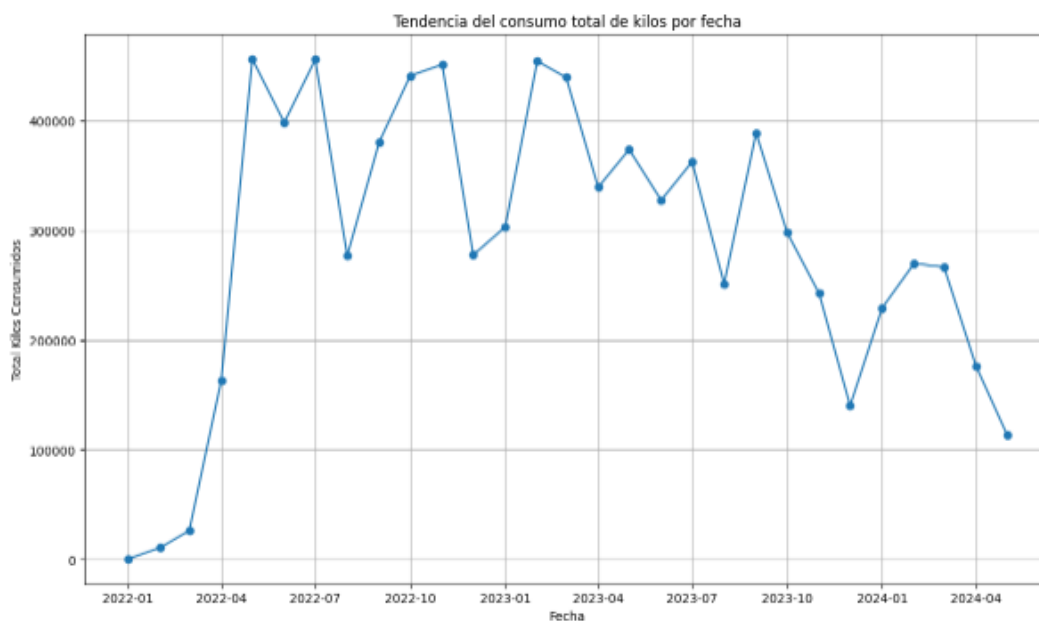


Figura 4 Gráfica de líneas del consumo de hilos por meses.

La siguiente gráfica se realizó para identificar si el consumo de la materia prima presenta un promedio estable entre los años 2022 y 2024, esto con el fin de tener una cantidad de stop de hilo en inventario, pero de acuerdo al comportamiento de la línea amarilla (Promedio móvil Figura 5), no es posible determinar un valor mínimo de inventario ya que los valores son muy cambiantes, el cual también pudimos analizar por medio de algoritmos de Machine Learning en la herramienta Colab.

El código con el que se realizó el análisis fue el siguiente:

```
# Gráfico de Línea con Promedio Móvil del Total de Consumo
data['Promedio Móvil'] = data['Total consumo'].rolling(window=3).mean()
plt.figure(figsize=(12, 6))
plt.plot(data['fecha'], data['Total consumo'], label='Total Consumo')
plt.plot(data['fecha'], data['Promedio Móvil'], label='Promedio Móvil (3 meses)', linestyle='--')
plt.title('Total de Consumo con Promedio Móvil')
plt.xlabel('Fecha')
plt.ylabel('Total Consumo')
plt.legend()
plt.grid(True)
plt.show()
```

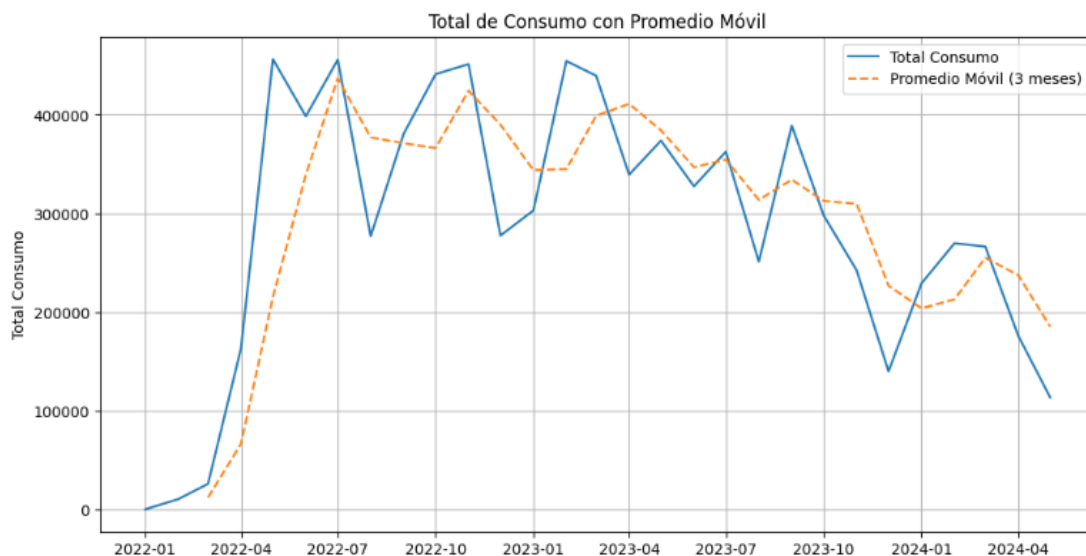


Figura 5 Gráfica de total de consumo con promedio movil.

Se realiza un gráfico de dispersión (Figura 6) sobre el dataset para evidenciar que tan dispersos se encuentran los datos, adicionalmente se ajusta el grafico de dispersión con la línea de regresión, identificando que los datos menos dispersos se encuentran entre 350.000 kilos y 250.000 kilos, tal como se muestra en la Figura 7, el cual también pudimos analizar por medio de algoritmos de Machine Learning en la herramienta Colab (Ng, 2015).

El código que se utilizó para este análisis fue el siguiente

```
# Gráfico de Dispersión del Total de Consumo
plt.figure(figsize=(12, 6))
plt.scatter(data['fecha'], data['Total consumo'])
plt.title('Dispersión del Total de Consumo a lo Largo del Tiempo')
plt.xlabel('Fecha')
plt.ylabel('Total Consumo')
plt.grid(True)
plt.show()
```

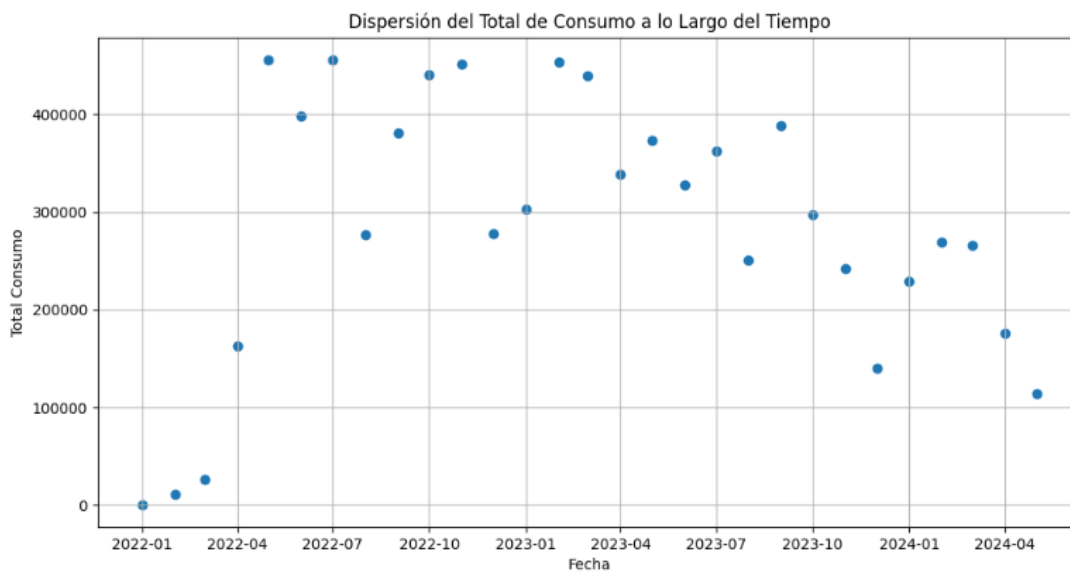


Figura 6 Gráfica de dispersión del total de consumo a lo largo del tiempo.

El código que se utilizó para este análisis fue el siguiente

```
#Gráfico de Dispersión del Total de Consumo con Línea de Regresión
plt.figure(figsize=(12, 6))
sns.regplot(x='fecha_num', y='Total consumo', data=data, line_kws={'color': 'yellow'})
plt.xticks(ticks=data['fecha_num'], labels=data['fecha'].dt.strftime('%Y-%m-%d'), rotation=45)
plt.title('Dispersión del Total de Consumo a lo Largo del Tiempo con Línea de Regresión')
plt.xlabel('Fecha')
plt.ylabel('Total Consumo')
plt.grid(True)
plt.show()
```

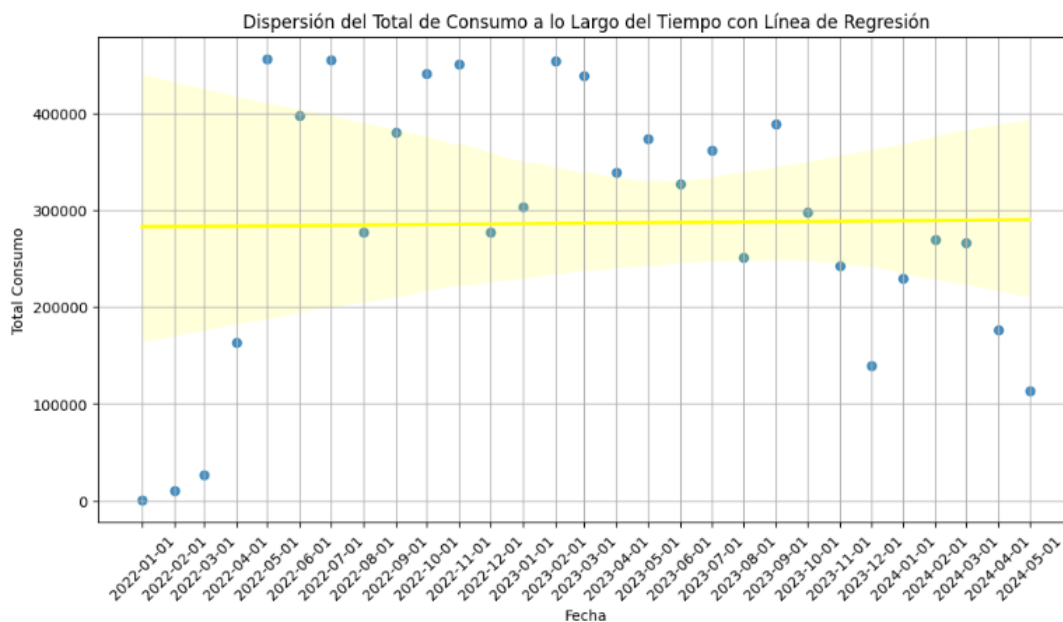


Figura 7 Gráfica de dispersión del total del tiempo con línea de regresión.

En nuestro enfoque de regresión, empleamos los datos de nuestro conjunto de datos y los modificamos para poder analizarlos numéricamente considerando la fecha, la cantidad de hilos requeridos y los títulos de los hilos. Gracias a esta estructuración de la información, se logra desarrollar los modelos predictivo KNNR, MLPRegressor, DecisionTreeRegressor, RandomForestRegressor y SVR , para así realizar un comparativo de confiabilidad y definir cuál sería el modelo más indicado para lograr anticipar el consumo de hilo en los próximos 4 meses (Vapnik, 1995).

```

# Lista de modelos a usar
models = {
    'KNNR': KNNR(),
    'MLPRegressor': MLPRegressor(max_iter=1000),
    'DecisionTreeRegressor': DecisionTreeRegressor(),
    'RandomForestRegressor': RandomForestRegressor(),
    'SVR': SVR()
}

```

Figura 8 Código de implementación de los modelos aplicados.

	Modelo	Confiabilidad
0	DecisionTreeRegressor	100.000000
1	KNNR	28.750439
2	MLPRegressor	-31.625119
3	RandomForestRegressor	83.152157
4	SVR	-12.321823

---

Figura 9 Resultados de confiabilidad de los modelos.

El resultado obtenido por medio de este análisis permite identificar que los modelos más confiables para lograr realizar la proyección serian, DecisionTreeRegressor con un 100% de confiabilidad y RandomForestRegressor con un 83.15% de confiabilidad

Gracias a este análisis se toma la decisión de aplicar los dos modelos anteriormente informados, dando como resultado Figura 10 Proyección de los modelos DecisionTreeRegressor y en la Figura 11 Proyección de los modelos RandomForestRegressor



datos se lean correctamente sin importar el formato (por ejemplo, comas, punto y coma, tabulaciones).

Una vez que se ha cargado el archivo, se procede con la limpieza y transformación de los datos. En primer lugar, se extraen y formatean los valores de año (ANNO) y mes (MES), convirtiéndolos a un formato numérico para mantener la coherencia. Luego, se fusiona la información del año y mes para crear una columna de fecha (fecha), lo que facilita el análisis temporal. Además, se cambian los nombres de varias columnas para que sean más descriptivos y coincidan con el contenido de los datos.

Las columnas TITULO, MEZCLA, MT, CARACTERISTICA y COLOR se combinan en una sola columna (HILO) separada por guiones. Esta combinación permite identificar de forma única cada conjunto de características del producto, simplificando así el análisis y la agregación de datos. Se eliminan varias columnas que no son necesarias para el análisis posterior a fin de reducir el tamaño de los datos y mejorar la eficiencia en el procesamiento.

Los datos se agrupan según la columna de fecha y una nueva columna llamada HILO, donde se suman los valores de KILOS. Esta agrupación es crucial para obtener una visión general de la cantidad de producto por combinación de características y por período de tiempo. Se construye una tabla pivote donde las fechas funcionan como índices y las combinaciones de características (HILO) como columnas, mostrando los valores de KILOS en el contenido de la tabla. Este formato pivote facilita el análisis de tendencias y patrones en los datos. Posteriormente, los datos transformados y reorganizados se guardan en un archivo Excel.

## **12. Modelo de toma de decisiones**

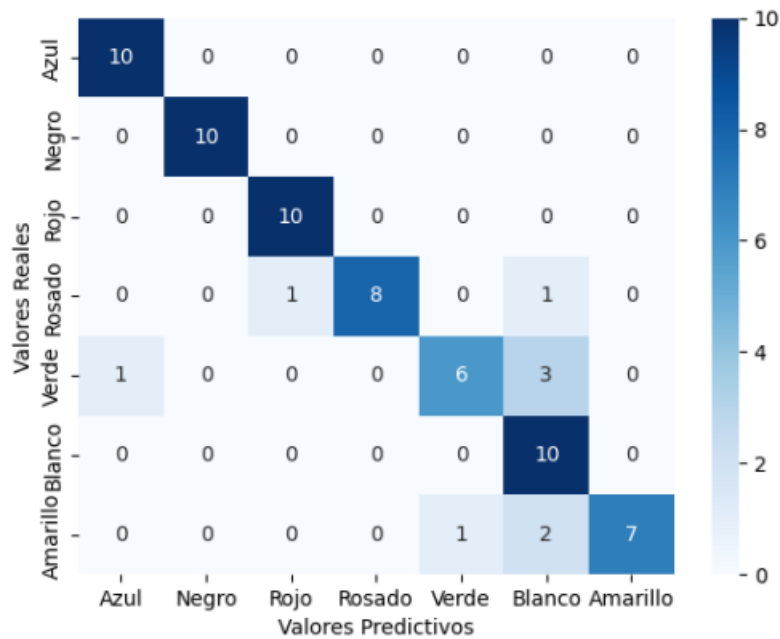
### **12.1 Análisis de datos**

El análisis de los datos se llevó a cabo utilizando dos modelos de regresión, DecisionTreeRegressor y RandomForestRegressor, con el fin de ofrecer una representación visual y estadística sobre la cantidad de hilos que se requerirá adquirir en los próximos 4 meses. Los gráficos y estadísticas permiten observar y comprender las tendencias, fluctuaciones y posibles patrones que puedan surgir a lo largo del tiempo. Además, para plasmar adecuadamente la información obtenida, se emplearon métodos visuales que facilitan la identificación de correlaciones, picos de actividad y otros aspectos fundamentales para comprender mejor el comportamiento en cuanto a las colocaciones de pedidos, las necesidades de materias primas y las compras realizadas.

### **12.2 Implementación en contextos reales**

En situaciones reales, se llevan a cabo investigaciones sobre la precisión de un sistema mediante el análisis de una matriz de confusión multiclase con siete categorías de colores: azul, negro, rojo, rosado, verde, blanco y amarillo. Se recopilieron un total de 70 observaciones, distribuidas equitativamente con 10 observaciones por cada color. En esta matriz, las filas representan los valores reales mientras que las columnas muestran los valores predichos. Los resultados revelaron altos niveles de precisión para negro y blanco, con el 100% de las observaciones correctamente identificadas. No obstante, se detectaron errores en la clasificación en otras categorías como la confusión entre rojo y

rosado, verde con azul y blanco, así como amarillo con blanco y verde (Osejos Rocha, 2022) .



### 13. Resultados adicionales

Como resultado adicional, se creó un modelo de regresión para examinar datos históricos sobre la colocación de pedidos y la demanda de materias primas en un período que abarca los últimos 24 meses. Esto facilita la identificación rápida de los materiales más necesarios para los próximos 4 meses, garantizando que los niveles de existencias de hilos y materias primas en la empresa sean los adecuados. El propósito de este análisis fue comprender mejor la situación presente y proyectar posibles escenarios futuros.

## 14. Conclusiones

La aplicación de un enfoque analítico preciso de datos y pronósticos tiene el potencial de anticipar con certeza las cantidades exactas de hilos necesarias, lo cual mejorará significativamente la planificación en términos de compras y producción. Esto conducirá a una notable reducción del exceso de inventario y los costos relacionados con la gestión de materiales no requeridos (Zhang, 2019).

El proceso de digitalización e integración de algoritmos de aprendizaje automático jugará un papel crucial en la optimización tanto de los procesos como en la toma de decisiones dentro del ámbito del suministro de hilos en Fabricato. La recopilación detallada y análisis exhaustivo de datos, combinado con estrategias avanzadas para clasificarlos, contribuirán significativamente a mejorar la eficacia en las entregas de pedidos.

La precisión en la planificación y adquisición adecuada de hilos permitirá a Fabricato satisfacer completamente las demandas planteadas por sus clientes, lo que se traducirá en una mejora general tanto en términos operativos como económicos al reducir los costos asociados al excedente de inventario. Este método fortalecerá aún más la posición competitiva del fabricante textil dentro del mercado.

La recolección y procesamiento sistemático continuo basado en una base sólida y coherente revela ser un pilar fundamental para llevar a cabo modelos predictivos y

análisis detallados sobre las tendencias emergentes dentro de la gestión logística relacionada con el suministro textil.

La evaluación del desempeño del software creado, mediante el uso de métricas particulares, demostrará su eficacia y exactitud en la mejora de la cadena de suministro de hilos. Esto se traducirá en una disminución de los costos por exceso de inventario, lo que a su vez mejorará la planificación de compras y la optimización general del proceso.

## 15. Referencias

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. Obtenido de <https://link.springer.com/book/10.1007/978-0-387-45528-0>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32. Obtenido de <https://doi.org/10.1023/A:1010933404324>
- Cerosetenta. (2023). *[Suicidio indígena] La otra cara de los suicidios en el Vaupés*. Bogota.
- Chollet, F. (2018). *Deep Learning with Python*. Manning Publications. Obtenido de <https://www.manning.com/books/deep-learning-with-python>
- Goodfellow, I. B.-C. (2016). *Deep Learning*. MIT Press. Obtenido de <https://www.deeplearningbook.org/>
- Hastie, T. T.-F. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. Obtenido de <https://link.springer.com/book/10.1007/978-0-387-84858-7>
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press. Obtenido de <https://mitpress.mit.edu/books/machine-learning-1>
- Ng, A. Y. (2015). *Machine Learning Yearning*. Obtenido de Machine Learning Yearning: <http://www.mlyearning.org>
- Osejos Rocha, S. M. (2022). *Desarrollo de un sistema para detectar los colores en tejidos con hilo utilizando visión artificial*. Ibarra: Universidad Técnica del Norte. Obtenido de <https://repositorio.utn.edu.ec/handle/123456789/12958>
- Pedregosa, F. V. (s.f.). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. Obtenido de <http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
- Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65, 386-408. Obtenido de <https://doi.org/10.1037/h0042519>
- Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 85-117. Obtenido de <https://doi.org/10.1016/j.neunet.2014.09.003>
- Srivastava, N. H. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929-1958. Obtenido de <http://jmlr.org/papers/v15/srivastava14a.html>
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer. Obtenido de <https://link.springer.com/book/10.1007/978-1-4757-2440-0>
- Witten, I. H. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann. Obtenido de <https://www.elsevier.com/books/data-mining/witten/978-0-12-804291-5>
- Zhang, Z. (2019). Machine Learning and Deep Learning: From Basics to Practice. *Journal of Applied Mathematics and Computation*, 361, 187-208. Obtenido de <https://doi.org/10.1016/j.amc.2019.04.033>
- Zhou, Z. H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC. Obtenido de <https://www.crcpress.com/Ensemble-Methods-Foundations-and-Algorithms/Zhou/p/book/9781439830031>

