UNIVERSITAT DE BARCELONA

MASTER EN BIG DATA & DATA SCIENCE

TRABAJO FINAL DE MÁSTER

ENTREGA 2. DESARROLLO

"MODELO PREDICTIVO DE DOWNTIME PARA MAQUINARIA DE COSTURA EN FOTL ES"

TUTORA DEL TFM:

Dolores Lorente Muñoz

INTEGRANTES GRUPO 2:

Luis Gamiño González
César Emilio García Ávalos
Andrey Ismagilov

29 DE JULIO DE 2024

ÍNDICE

1. Antecedentes

Fruit of the Loom® es una marca global que ofrece camisetas de colores, lana, ropa interior y prendas de vestir a consumidores de todas las edades. Actualmente, esta empresa es una de las principales exportadoras en El Salvador, según datos de CAMTEX (Cámara de la Industria Textil, Confección y Zonas Francas de El Salvador), hacia mercados de México y Estados Unidos. Para FOTL (Fruit of The Loom), existen dos factores clave de éxito: la calidad del producto y la eficiencia operativa. Como una empresa de producción masiva, FOTL exporta más de 37 millones de docenas de prendas de vestir al año desde El Salvador. Por lo tanto, los paros de máquina pueden resultar en pérdidas financieras considerables debido a la interrupción de la producción. En la industria textil, la detención de una máquina de costura puede detener toda la línea de producción, provocando una disminución directa en la capacidad productiva y, por ende, una destrucción de negocio.

Un paro de máquina puede afectar no solo a la empresa en sí, sino también a toda la cadena de suministro. Para FOTL, esto puede significar retrasos en la entrega de productos a minoristas o distribuidores, lo cual puede resultar en contratos incumplidos y pérdida de clientes.

2. Definición del objetivo principal del proyecto

Objetivo Principal

El objetivo principal de este proyecto es desarrollar un modelo de predicción del tiempo de paro de maquinaria de costura con el propósito de que el negocio asigne recurso (mecánicos, piezas para reemplazo, máquinas de respaldo, etc.) oportunamente.

Descripción del Proyecto

Este proyecto abordará la problemática de los paros no programados en máquinas de costura, que pueden causar pérdidas significativas en la producción y afectar la cadena de suministro.

- ✓ **Modelo de predicción de tiempo de paro**

Variable Objetivo (Y): "downtime_gross", es una variable cuantitativa continua, esta variable proporciona información sobre los minutos de paro generados por fallos en maquinaria de costura.

Beneficios Esperados

- ✓ **Optimización del Mantenimiento:** La predicción del tiempo de paro permite planificar el mantenimiento de manera más efectiva, evitando paros inesperados y programando intervenciones preventivas o correctivas en momentos que minimicen el impacto en la producción.
- ✓ **Reducción de Costos:** Al anticipar los tiempos de inactividad, las empresas pueden reducir los costos asociados con reparaciones de emergencia, paradas no planificadas y la pérdida de producción. Esto también puede ayudar a optimizar el uso de recursos y minimizar el costo de piezas de repuesto.
- ✓ **Aumento de la Eficiencia Operativa:** La capacidad de prever fallos permite a los equipos de operación y mantenimiento trabajar de manera más proactiva en lugar de reactiva, aumentando la eficiencia operativa y reduciendo el tiempo total de inactividad.
- ✓ **Mejora en la Gestión de Inventarios:** Con una predicción de tiempo de paro, se pueden gestionar mejor los inventarios de piezas de repuesto, asegurando que los componentes necesarios estén disponibles cuando se requieran, sin acumular exceso de stock.

Metodología

- ✓ **Recopilación y Preprocesamiento de Datos:** Recopilación de datos históricos de fallos y tiempos de inactividad de las máquinas de costura.
- ✓ **Limpieza y preprocesamiento de datos para garantizar su calidad y consistencia.**
- ✓ **Desarrollo de Modelos:** Selección de características relevantes y técnicas de ingeniería de características.
- ✓ **Entrenamiento y validación de los modelos**
- ✓ **Evaluación y Optimización** Evaluación del desempeño de los modelos mediante métricas estándar. Optimización de hiperparámetros para mejorar la precisión y robustez de los modelos.
- ✓ **Implementación y Prueba:** Implementación de los modelos en un entorno de prueba para evaluar su efectividad en condiciones reales.
- ✓ **Feedback y ajustes basados en los resultados obtenidos.**
- ✓ **Documentación y Presentación:** Documentación completa del proceso, resultados y conclusiones.

- ✓ **Presentación de los hallazgos y recomendaciones a los departamentos interesados.**

3. Punto de partida

Según datos del año 2024, el paro de máquina en la compañía ha sido de más de 31,000 horas, de las cuales el top 3 de operaciones que generan mayor impacto son:

1. Hem Sleeve
2. Hem Bottom
3. Hem Around

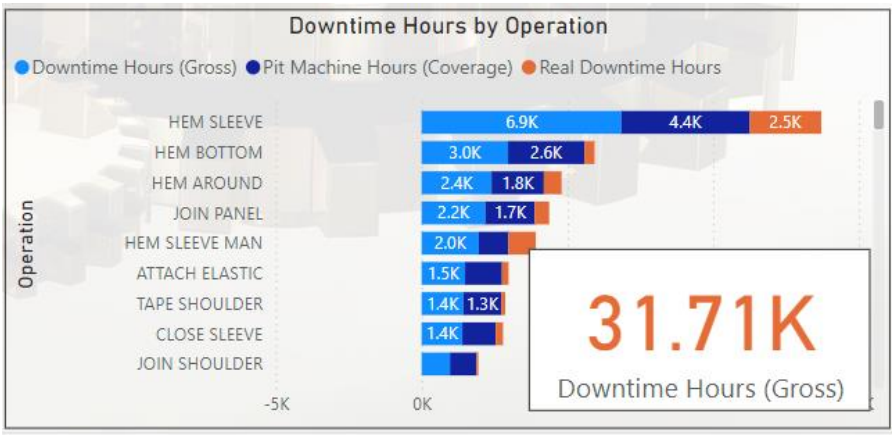


Imagen 1.0 *detalle de horas de paro por operación, autoría propia.*

Estas 31,000 horas (gross) tienen un impacto directo en la capacidad productiva afectando al negocio.

El dinero perdido por las plantas de FOTL ES por pago de Off Std por causa de máquina mala, es **para el año 2024 un costo de $27,000 en horas no productivas por fallo de máquinas de costura.**

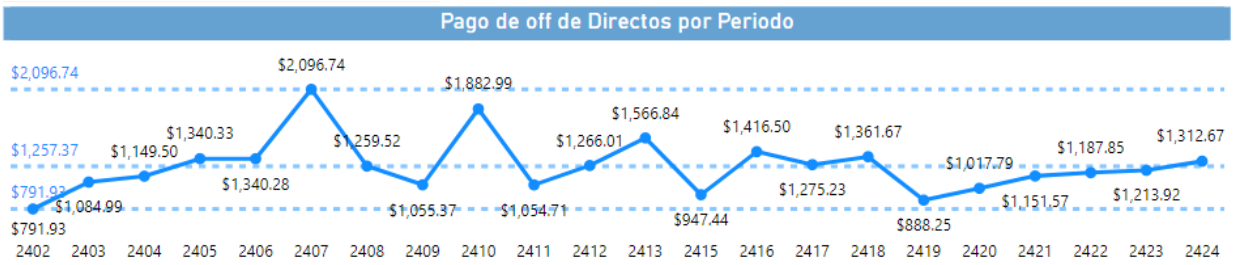


Imagen 2.0 *Tendencia de paros por Off Std. por fallo de maquinaria*

Para entender mejor el problema y el desempeño del equipo de mantenimiento, los indicadores

relevantes a considerar son el MTBF (Mean Time Between Failures) y el MTTR (Mean Time To Repair). A continuación, se proporciona una explicación detallada de estos indicadores y cómo se aplican a la situación dada:

MTBF (Mean Time Between Failures): Es el tiempo promedio que transcurre entre una falla y la siguiente en un sistema o máquina. Un alto MTBF indica que la máquina es confiable y no falla con frecuencia. Es una medida de la confiabilidad del equipo.

MTTR (Mean Time To Repair): Es el tiempo promedio que se tarda en reparar una máquina después de que ha fallado. Un bajo MTTR indica que las reparaciones se realizan rápidamente, minimizando el tiempo de inactividad. Es una medida de la eficiencia del equipo de mantenimiento.

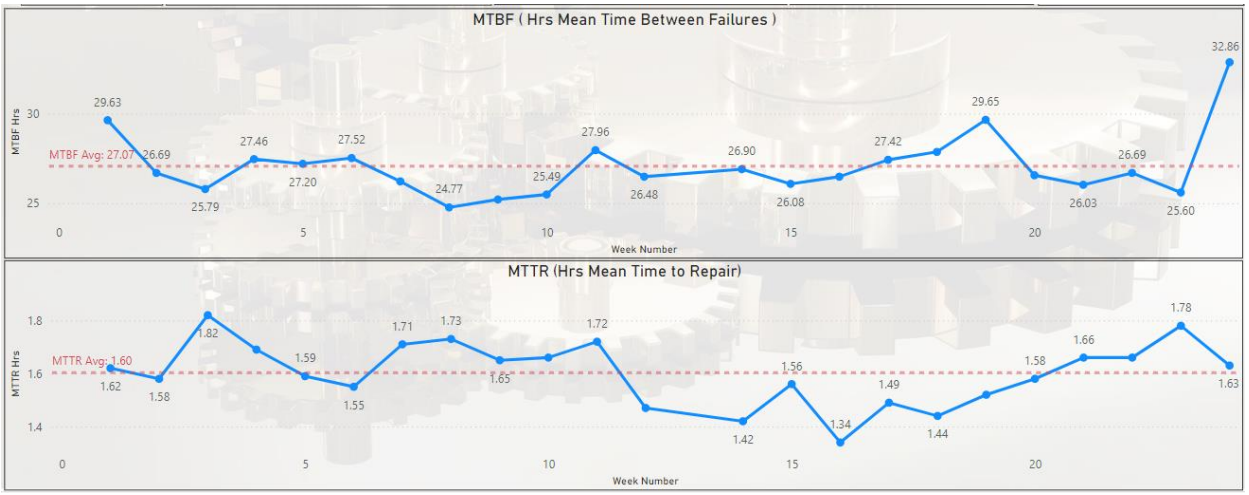


Imagen 3.0 *Indicadores de Mean Time Between Failures y Mean Time to Repair*

En el caso particular de FOTL la frecuencia de fallas (MTBF) y la eficiencia de las reparaciones (MTTR) impactan directamente la productividad. Con un MTBF de 27.07 horas, las interrupciones son frecuentes, lo que puede afectar la producción si no se gestionan adecuadamente, Un MTTR de 1.6 horas sugiere que el equipo de mantenimiento es bastante eficiente en términos de tiempo de reparación.

La empresa cuenta con datos detallados sobre el historial de fallas de los equipos utilizados. Estos datos se registran manualmente en Órdenes de Trabajo (WO) y luego se ingresan al sistema de Gestión de Activos Empresariales (EAM). Las bases de datos que contienen esta información están alojadas en servidores ORACLE.

4. Selección del dataset

4.1. Descripción del dataset

Se cuenta con un registro de aproximadamente 150,000 órdenes de trabajo ingresadas en el sistema EAM, el dataset cuenta con las siguientes variables:

Variable	Descripción	Tipo de variable
Equipment	Identificador único de la máquina de costura.	Variable cualitativa nominal
Operation	Operación de costura para la cuál es utilizada la maquinaria.	Variable cualitativa nominal
Location	Planta de costura a la cual pertenece.	Variable cualitativa nominal
WO type	Tipo de work order realizada.	Variable cualitativa nominal
Model Base	Modelo base de la máquina.	Variable cualitativa nominal
Problem Code	Código del problema de maquinaria reportado.	Variable cualitativa nominal
Problem Code Descripion	Descripción del código de problema reportado.	Variable cualitativa nominal
Failure Code	Código de falla con la cual se reporta el problema de la máquina.	Variable cualitativa nominal
Failure Code Description	Descripción del código de falla.	Variable cualitativa nominal
Cause Code	Código de la causa que ha generado la falla de maquinaria.	Variable cualitativa nominal
Cause Code Description	Descripción de la causa que ha generado la falla de maquinaria.	Variable cualitativa nominal
Action Code	Código de acción necesaria para solucionar el problema de máquina.	Variable cualitativa nominal
Action Code Description	Descripción de la acción necesaria para resolver el problema de maquinaria.	Variable cualitativa nominal
Mechanic Name	Nombre del mecánico que solucionó el problema.	Variable cualitativa nominal
Day	Fecha de reporte de la falla.	Variable cualitativa ordinal
Work Order Number	Número único de orden de trabajo.	Variable cualitativa nominal
Downtime Hours (Gross)	Horas de paro crudas.	Variable cuantitativa continua
Pit Machine Hours	Tiempo de cobertura de máquinas de respaldo.	Variable cuantitativa continua
Real Downtime Hours	Tiempo improductivo.	Variable cuantitativa continua
Parts Cost	Costo de las partes utilizadas para la reparación.	Variable cuantitativa continua
Repair hours	Horas invertidas en la reparación de la maquinaria.	Variable cuantitativa continua
Waiting hours	Horas de espera.	Variable cuantitativa continua

5. Limpieza del dataset

La limpieza de datos es un paso esencial en el proceso de análisis y modelado predictivo. Este apartado describe los pasos seguidos para preparar el conjunto de datos, asegurando que los datos sean consistentes, completos y libres de errores.

5.1. Identificación y Manejo de Valores Faltantes

Se verificó la presencia de valores faltantes en todas las columnas utilizando la función *sapply* en R. No se encontraron valores faltantes, lo cual permitió continuar con el análisis sin la necesidad de imputación o eliminación de datos incompletos, se concluye que existe fiabilidad suficiente para continuar con el análisis.

5.2. Conversión de Formatos y Limpieza de Datos Numéricos

- Se eliminó el símbolo de dólar y las comas de la columna parts_cost, convirtiéndola a tipo numérico.
- Se convirtió la columna date al formato de fecha adecuado y se ordenó el DataFrame por esta columna con el propósito de calcular la variable “days_between_failures”.

5.3. Eliminación de valores atípicos

Se identificaron datos atípicos donde las horas de downtime_gross son negativas, los datos negativos no tienen sentido pues la variable indica el tiempo de paro de la maquinaria textil, estos datos atípicos representan un 2.68% de un dataset conformado por 149,066 instancias por lo que, se decidió eliminar los datos con esta característica.

5.4. Eliminación de datos irrelevantes

Se identificaron 8 variables que no eran relevantes para el análisis y modelado predictivo. Estas columnas fueron eliminadas para simplificar el conjunto de datos y enfocarse únicamente en las variables necesarias. Las columnas eliminadas fueron:

- ✓ problem_code
- ✓ failure_code
- ✓ cause_code
- ✓ action_code
- ✓ pit_coverage
- ✓ real_downtime
- ✓ waiting_hours
- ✓ repair_hours

5.5. Justificación de la Eliminación

- **problem_code, failure_code, cause_code, action_code:** Estas columnas contienen códigos categóricos que duplican la información ya presente en las columnas descriptivas problem_desc, failure_desc, cause_desc, y action_desc. Al mantener solo las descripciones, se evita la redundancia y se facilita la interpretación de los datos.
- **pit_coverage:** Esta columna no aporta información relevante para el análisis específico del tiempo de inactividad y la frecuencia de fallos (esta variable es calculada como la diferencia entre real_downtime y downtime_gross).
- **real_downtime:** Se decidió utilizar downtime_gross como la métrica principal de tiempo de inactividad. La columna real_downtime se consideró redundante y potencialmente confusa.
- **repair_hours:** Las horas de reparación se registran una vez ha finalizado el mantenimiento, es decir, si el propósito es realizar una predicción, esta es una variable con la cual no se dispone al inicio del mantenimiento.
- **waiting_hours:** Esta columna representa el tiempo de espera de las asociadas de costura mientras el mecánico se acerca a la unidad de producción para resolver el

problema, que no es directamente relevante para el análisis del tiempo de paro de máquina. Su eliminación ayuda a centrar el análisis en las variables más importantes.

5.6. Corrección de errores estructurales

Se corrigió la inconsistencia en los nombres de los mecánicos, unificando las diferentes variantes del nombre de un mismo mecánico.

6. Análisis de variables

- **Operation:** La variable operación hace referencia a la operación de costura realizada con la maquinaria, existen 25 operaciones diferentes; para el análisis se calculó la frecuencia relativa de los fallos de maquinaria para cada operación obteniendo que, el top 3 de operaciones con más fallos es, Hem Sleeve con un 28.2%, Hem Bottom con un 11.7% y Join Panel con un 5.7%.

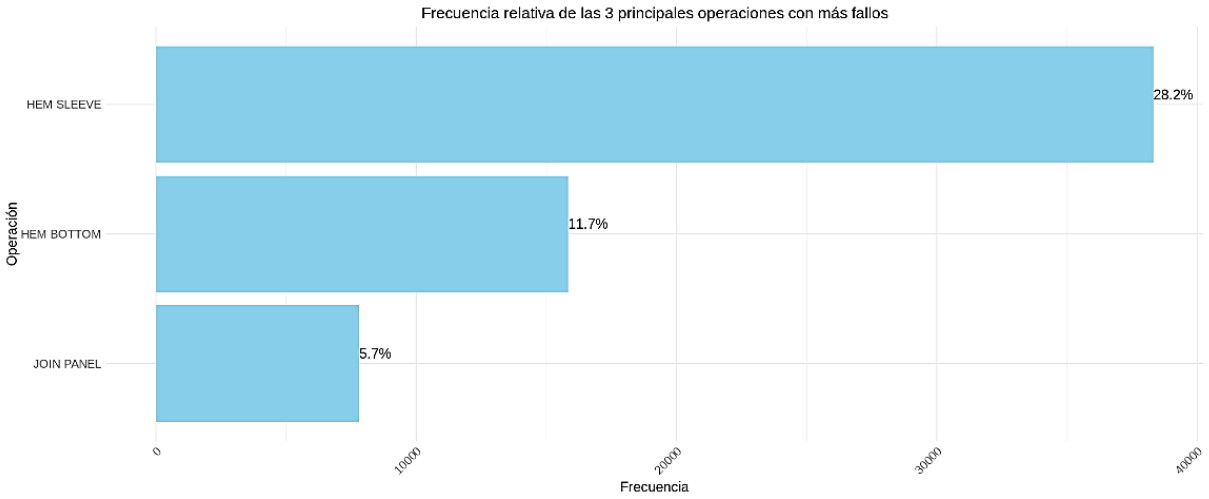


Imagen 4.0 *Frecuencia relativa de operaciones con fallos de maquinaria*

- **problem_desc:** La variable problem_desc hace referencia a la descripción del problema de maquinaria de costura, son 3 problemas reportados en ordenes de trabajo, para su análisis, se calculó la frecuencia relativa de los problemas reportados teniendo que Cose con defecto es el principal problema reportado con 69.94%, No cose con un 23.5% y Otros con un 6.55%.

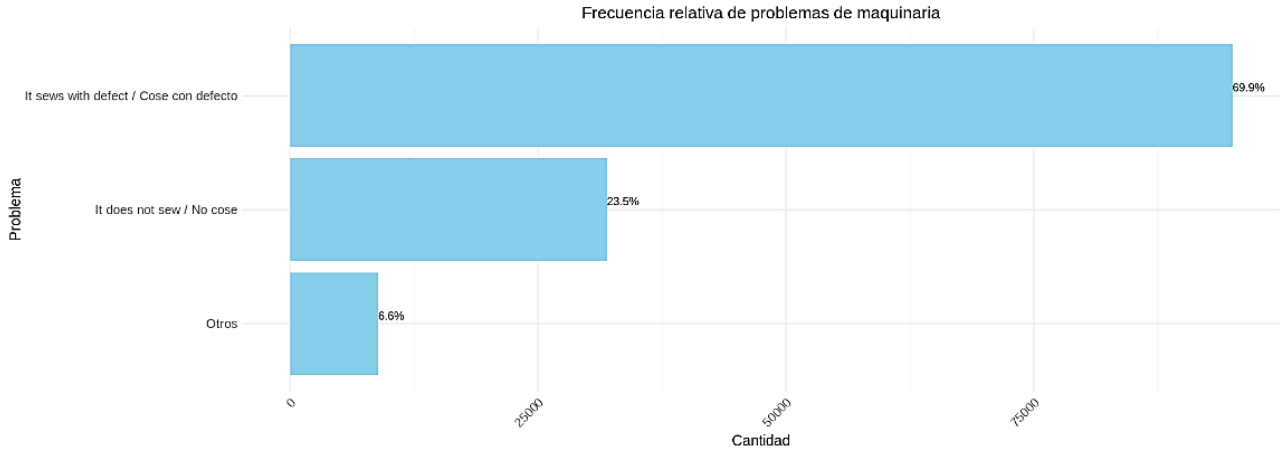


Imagen 5.0 *Frecuencia relativa de problemas de maquinaria*

- **failure_desc:** Esta variable describe la falla detectada en la maquinaria de costura, se han reportado 4 fallas diferentes en ordenes de trabajo, se realizó un análisis de la frecuencia relativa obteniendo que la falla del cabezal dejando de coser es el más común, representando casi la mitad de todos los fallos registrados con un 47.1%. Es una falla crítica que detiene completamente el proceso de costura, luego se tiene la falla del cabezal cosiendo con defecto con un 27.1%, Otros con un 15.7% y la falla del dispositivo que corta desperdicio con un 10.1%.

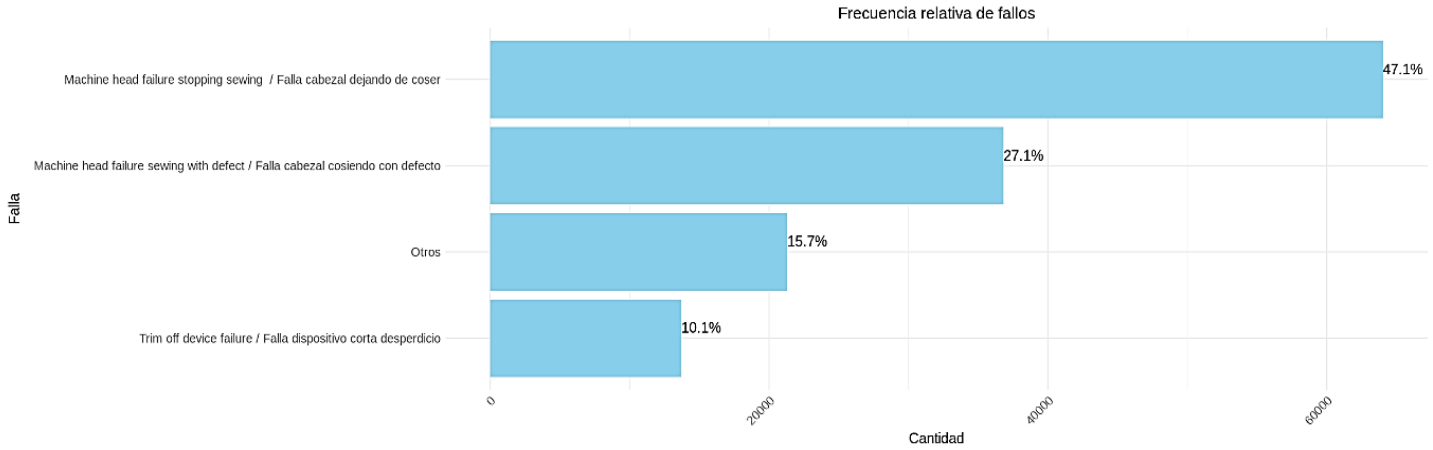


Imagen 6.0 *Frecuencia relativa de fallos de maquinaria*

- **cause_desc:** esta variable hace referencia a la descripción de la causa del fallo de maquinaria de costura, se han reportado 29 causas diferentes, para el análisis se calculó la frecuencia relativa de cada una de las fallas y se tiene que en el top 3, el enlazador de cabezal desajustado representa un 37.5% de las causas, la cuchilla de corta desperdicio en mal estado representa un 9.56% y la Guarda de aguja de cabezal desajustada representa un 8.34%

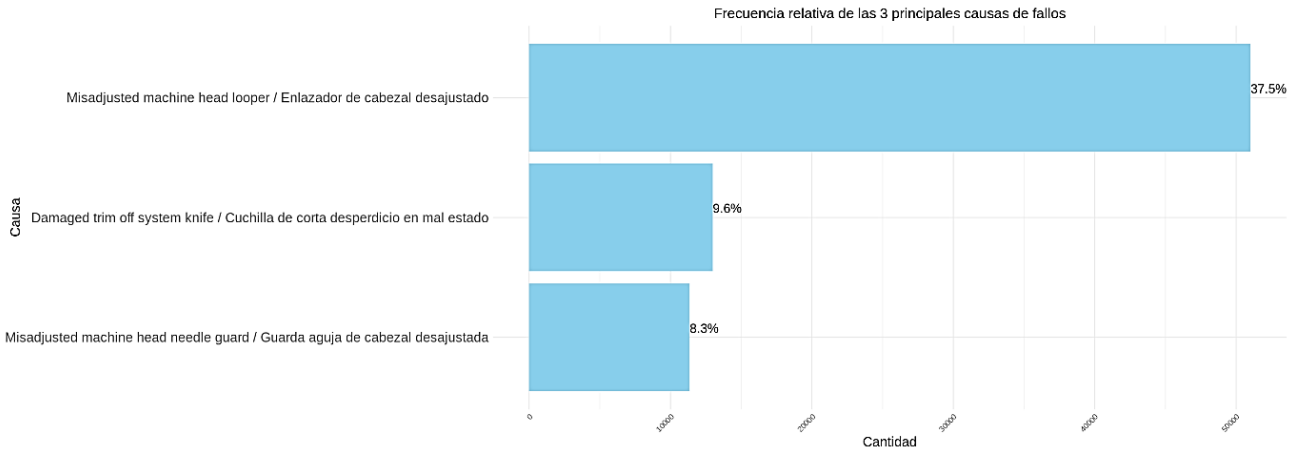


Imagen 7.0 *Frecuencia relativa de causa de fallos de maquinaria*

- **action_desc:** Se han reportado 29 acciones correctivas en ordenes de trabajo, la acción correctiva más frecuente es ajustar el enlazador de cabezal, que representa el 33.06% del total de acciones correctivas. Con una frecuencia relativa del 8.40%, ajustar la guarda de la aguja de cabezal es la segunda acción más común. Ajustar los dientes de cabezal es la tercera acción más frecuente, con una frecuencia relativa del 7.14%.

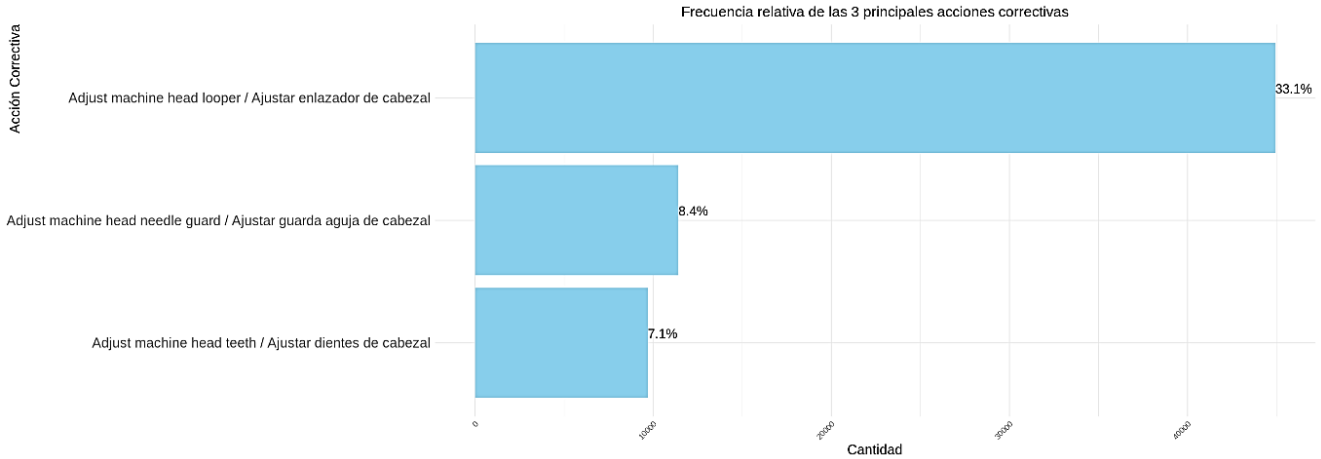


Imagen 8.0 *Frecuencia relativa de acciones correctivas implementadas*

- **parts_cost:** Esta variable hace referencia al costo de las partes utilizadas para ejecutar las acciones correctivas en la maquinaria, esta variable puede tomar el valor de cero si el reemplazo de piezas se realiza con piezas en existencia de otras máquinas, la media de costo de partes es de $12.11, la mediana es $0, la desviación estándar es 41.74, el rango es $2,812.49, estos datos indican que se tiene una alta variación, hay costos de piezas altos en comparación con otros.

6.1. Análisis de variable objetivo

La variable objetivo a predecir es downtime_gross, este es el tiempo de paro de maquinaria de costura registrado en ordenes de trabajo por el departamento de mecánica, la variable downtime_gross muestra una distribución con tiempos de inactividad que varían ampliamente, desde 0.00 hasta 835.00 minutos. La mediana es 50.00 minutos, mientras que la media es más alta, 87.97 minutos, indicando que los datos están sesgados hacia la derecha por algunos valores atípicos. El primer cuartil es 25.00 minutos y el tercer cuartil es 110.00 minutos, sugiriendo que el 75% de los registros tienen tiempos de inactividad menores o iguales a 110.00 minutos.

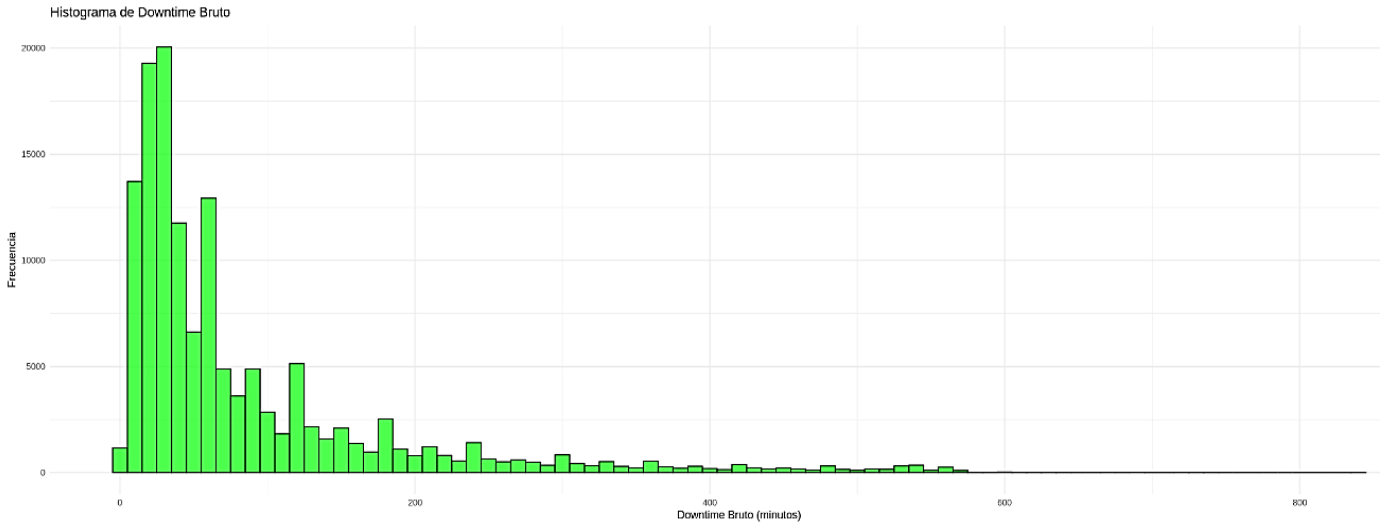


Imagen 9.0 *Histograma de downtime_gross*

Se realizó un gráfico de caja de la variable downtime_gross, se visualiza una alta frecuencia de outliers, por este motivo se decidió analizar si estos datos atípicos siguen un patrón temporal, es decir, se analizó si hay meses donde la cantidad de datos atípicos incrementa.

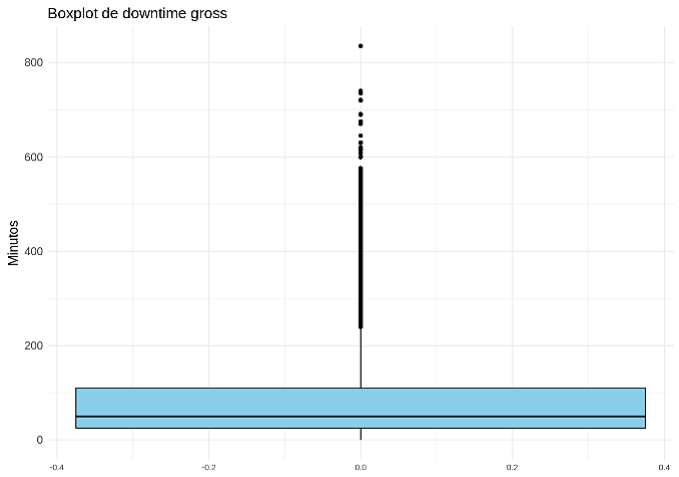


Imagen 10.0 *Boxplot de downtime_gross*

Se identifica que los datos atípicos en cuanto a tiempo elevado de paro el primer semestre del año, a partir de julio estos outliers se ven reducidos, esto puede ser debido a que el mantenimiento preventivo de los equipos se programa prácticamente a mitad de año.

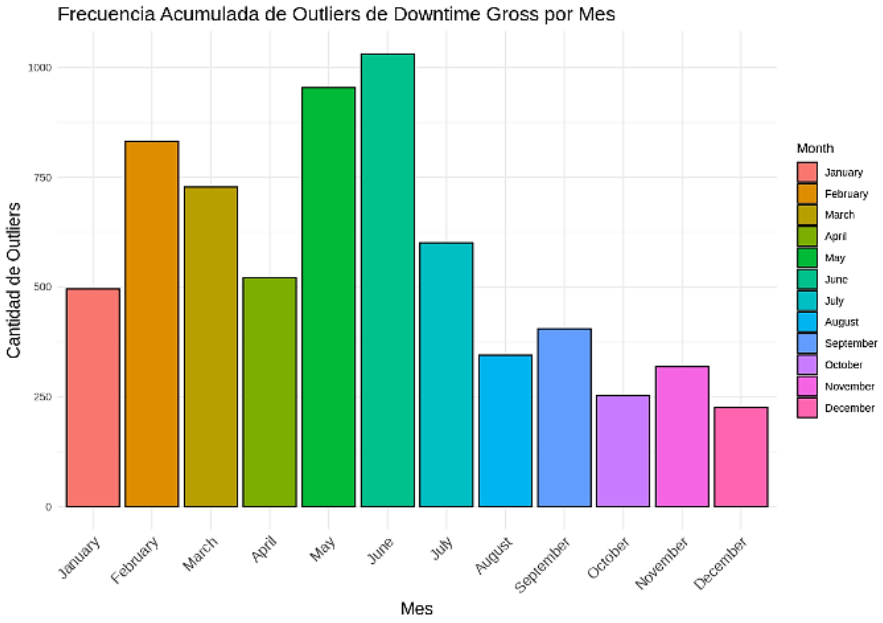


Imagen 11.0 *Análisis de distribución de outliers por mes*

En el análisis realizado sobre el tiempo improductivo de las operaciones, se calculó la proporción de tiempo improductivo atribuible a cada operación. Los resultados muestran que la maquinaria utilizada para las operaciones de Hem Sleeve es la que genera un mayor tiempo de inactividad en el negocio. Específicamente, se encontró que el 33.6% del tiempo total de paro se atribuye a esta operación (esto tiene sentido pues, la principal causa de fallo es el cabezal de la maquinaria, esto corresponde específicamente al equipo utilizado para la operación de Hem Sleeve).

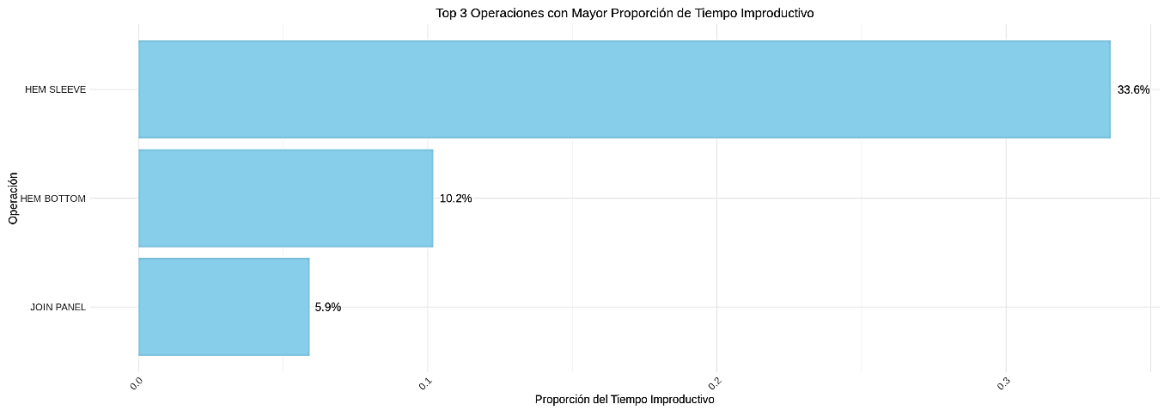


Imagen 12.0 *Proporción de tiempo improductivo por operación*

7. Realización de modelos predictivos de downtime_gross

7.1. División del conjunto de datos

Para asegurar la robustez y la generalización de los modelos predictivos desarrollados, el conjunto de datos se dividió en tres subconjuntos: entrenamiento, validación y prueba. La división se realizó de la siguiente manera:

- ✓ **Reproducibilidad:** Se estableció una semilla aleatoria para garantizar la reproducibilidad de los resultados. Esto permite que los experimentos puedan ser replicados con los mismos resultados.
- ✓ **Proporciones de División:** Se definieron las proporciones para cada subconjunto de datos: Entrenamiento: 70% (95,100 instancias), Validación: 20% (27,171 instancias) y Prueba: 10% (13,587 instancias).

7.2. Realización de modelos GLM

Un modelo lineal generalizado (GLM) es una extensión del modelo lineal clásico; por ello, antes de definir un modelo lineal generalizado conviene recordar las tres condiciones que debe verificar un modelo lineal:

1. Los errores se distribuyen normalmente.
2. La varianza es constante.
3. Las variables independientes están relacionadas con la variable dependiente de manera lineal.

De manera analítica se tiene que, dada una muestra (Y_i, X_{i1},..., X_{ip}) con i=1,...,n, la relación entre las observaciones Yi y las variables independientes se expresa como:

$$Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} + \varepsilon_i \qquad \text{con i=1,...,n.}$$

o equivalentemente

$$\mu_i = E(Y_i) = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip}$$

Suponiendo que el error verifica

Para la evaluación del rendimiento de los modelos se han calculado 3 métricas.

- **Raíz del Error Cuadrático Medio (RMSE):** El RMSE es la raíz cuadrada del MSE. Proporciona una medida del error en las mismas unidades que la variable de salida, lo que facilita su interpretación.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- **Error Absoluto Medio (MAE):** El MAE mide el promedio de los errores absolutos, es decir, la diferencia absoluta entre los valores predichos y los valores reales. Al igual que el RMSE, el MAE está en las mismas unidades que la variable de salida.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Coeficiente de Determinación (R^2):** mide la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Un valor de R^2 cercano a 1 indica que el modelo explica bien la variabilidad de la variable de salida.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Se han desarrollado en total 6 GLM, a continuación, se presenta cada uno de los modelos realizados:

1. El modelo 1 es un modelo GLM, la formula utilizada es la siguiente downtime_gross ~ operation + base_model + problem_desc + failure_desc + cause_desc + action_desc + mechanic + parts_cost donde se identifica que el costo de las partes es una variable estadísticamente significativa para explicar la variación en el downtime_gross, (Pvalue < 0.05). El primer modelo muestra un ajuste débil a los datos:

```
Test RMSE:  92.9449
Test MAE:  61.42112
Test R-squared:  0.1623169
```

2. El modelo 2 es un modelo GLM, se calcularon las siguientes variables con el propósito de explicar más variación en la variable downtime_gross; “Days_Betwen_Failures”, “Cumulative_Failure_Count” y “Cumulative_Maintenance_Cost”, obteniendo que las variables son estadísitcamente significativas para explicar la variación en el modelo.

```
Days_Between_Failures          < 2e-16
Cumulative_Failure_Count      0.041475
Cumulative_Maintenance_Cost   0.019561
```

Las métricas para el modelo 2 son las siguientes:

```
Test RMSE:  92.88477
Test MAE:  61.39467
Test R-squared:  0.1634004
```

3. Para el modelo 3 se realizó un GLM, se solicitó al departamento de mecánica la obsolescencia de la maquinaria para determinar si es una variable que puede explicar el tiempo de paro obteniendo que es estadísticamente significativa (Pvalue = 3.42e-13). Las métricas para este modelo son las siguientes:

```
Test RMSE:  92.81877
Test MAE:  61.36279
Test R-squared:  0.164589
```

4. El cuarto modelo incorpora la variable “mechanic antiquity”, se refiere a la cantidad de años de experiencia del trabajador, se obtuvo que esta variable no es estadísticamente significativa para explicar la variación de downtime_gross (Pvalue<0.05).

```
Test RMSE:  92.81877
Test MAE:  61.36279
Test R-squared:  0.164589
```

5. El modelo 5 incorpora la variable “changed_pieces”, hace referencia a la cantidad de piezas reemplazadas durante el mantenimiento correctivo, obteniendo que es una variable estadísticamente significativa (Pvalue = 2e-16), las métricas de este modelo mejoran considerablemente con la incorporación de esta variable:

```
Test RMSE:  66.48923
Test MAE:  39.69079
Test R-squared:  0.5713219
```

6. En el modelo 6 se agrega una interacción entre las variables mechanic*chaged_pieces, esta interacción no es estadísticamente significativa para explicar la variación de downtime_gross.

```
Test RMSE:  66.48923
Test MAE:  39.69079
Test R-squared:  0.5713219
```

7.3 Eliminación Forward

Se utilizó eliminación forward con el propósito de hacer una selección de modelos y variables para mejorar el rendimiento y la interpretabilidad del modelo.
Al realizar la eliminación forward se obtuvo lo siguiente:

```
Step:  AIC=800006.1
downtime_gross ~ changed_pcs + mechanic + operation + cause_desc +
    problem_desc + Age + base_model + action_desc + failure_desc +
    Days_Between_Failures + Cumulative_Maintenance_Cost + parts_cost

                          Df Sum of Sq       RSS    AIC
<none>                                426363553 800006
+ Cumulative_Failure_Count  1    3350.3 426360203 800007
```

Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
<I<chr>>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
	NA	NA	95099	989850205	879713.1
+ changed_pcs	-1	500060217.90	95098	489789987	812805.0
+ mechanic	-85	55991828.28	95013	433798158	801430.1
+ operation	-24	3676760.47	94989	430121398	800668.6
+ cause_desc	-28	1086040.06	94961	429035358	800484.2
+ problem_desc	-2	707543.63	94959	428327814	800331.2
+ Age	-1	396302.71	94958	427931512	800245.2
+ base_model	-21	542075.50	94937	427389436	800166.6
+ action_desc	-28	582600.09	94909	426806836	800092.9
+ failure_desc	-3	226750.11	94906	426580086	800048.4
+ Days_Between_Failures	-1	114443.02	94905	426465643	800024.9
+ Cumulative_Maintenance_Cost	-1	84764.35	94904	426380879	800008.0
+ parts_cost	-1	17325.58	94903	426363553	800006.1

Se obtiene un modelo número 7 utilizando la fórmula reducida, las métricas de dicho modelo son las siguiente:

```
Test RMSE:  66.48391
Test MAE:  39.68947
Test R-squared:  0.5713905
```

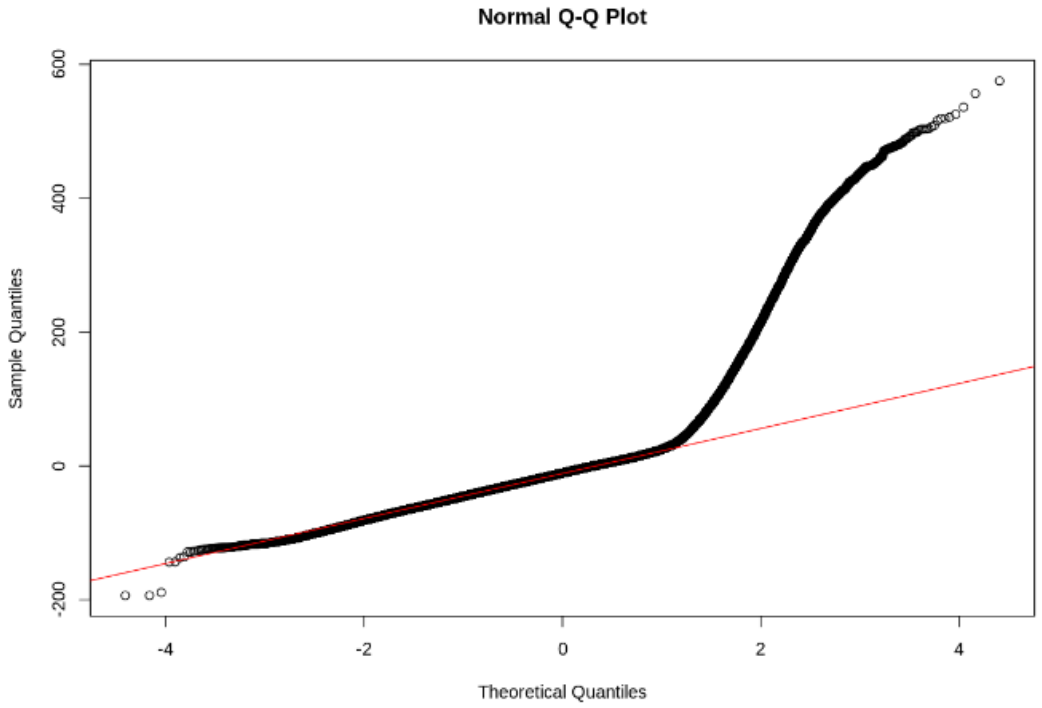
7.3. Summary de modelos

Modelo	RMSE	MAE	R-squared	Tiempo de ejecución
1	92.9449	61.42112	16.23%	7.692935 secs
2	92.88477	61.39467	16.34%	8.559741 secs
3	92.81877	61.36279	16.46%	8.689681 secs
4	92.81877	61.36279	16.46%	8.038722 secs
5	66.48923	39.69079	57.13%	8.290193 secs
6	66.65351	39.35493	56.44%	16.57506 secs
7	66.48391	39.68947	57.14%	7.829709 secs

El modelo escogido para validación es el número 7, en comparación con el modelo 5 tiene métricas de rendimiento similares (tienen el error más bajo y el mejor coeficiente de determinación) pero el tiempo de computación es más corto.

7.4. Análisis de residuos

Se realizó un gráfico para evaluar la normalidad de los residuos del modelo número 7, se visualiza que los residuos no son normales, se identifican colas cortas en el gráfico.



Este modelo no es adecuado para la predicción del downtime, incumple el supuesto de normalidad de los residuos, por lo tanto, se procede a realiza un modelo menos sensible a distribuciones no normales, en el siguiente apartado se presenta la realización de un modelo Random Forest.

7.5. Modelo Random Forest

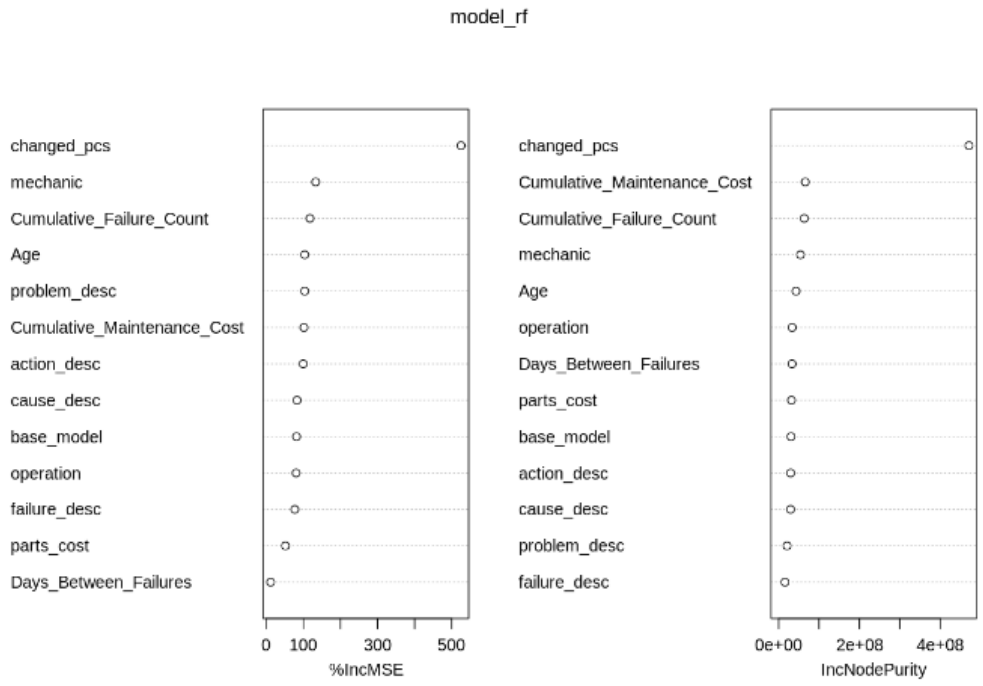
Se realizó un modelo Random Forest para predecir el tiempo de paro de maquinaria de costura, los resultados son de acuerdo con lo siguiente:

```
Time difference of 1.537768 hours

Call:
 randomForest(formula = downtime_gross ~ . - equipment - downtime_gross -      date - antiquity, data = train_data4, importance = TRUE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 4

          Mean of squared residuals: 3455.089
                    % Var explained: 66.81
```

El modelo explica un 66.81% de la variación del downtime, en el conjunto de prueba el modelo obtuvo el siguiente resultado:

```
Test RMSE:  58.69398
Test MAE:  34.05731
Test R-squared:  0.6659467
```

No se identifica una diferencia considerable entre el resultado para los datos de entrenamiento o de prueba por lo que, no existe riesgo de sobreajuste del modelo. Se relalizó un gráfico de importancia de características, la variable más importante es “changed_pieces”.



La desventaja de este modelo es el tiempo requerido de cómputo, el modelo requirió 1.54 horas para entrenarse.

8. Recomendaciones

- ✓ Se identifica una oportunidad de mejora en cuanto al registro de la información, actualmente las ordenes de trabajo se escriben en papel y luego se digitan en el sistema EAM, esto puede dar lugar a errores estructurales.

- ✓ Cuando se reportan ordenes de trabajo, estas no contienen información que describa el diagnóstico de la máquina en cuanto a su seteo por ejemplo, RPM, ajuste de cuchillas, estado de la plancha, ajuste de barra de agujas, por hacer mención de algunos, estas podrían ser variables que expliquen el tiempo de paro de una maquinaria.

- ✓ El análisis de frecuencia de outliers de downtime indica que estos incrementan en el primer semestre del año y disminuyen en el segundo semestre, esto puede estar relacionado al hecho de que los mantenimientos preventivos se ejecutan cada 6 meses, esta frecuencia puede resultar insuficiente para mantener las máquinas operando adecuadamente.

- ✓ Se identifica que la maquinaria utilizada para Hem Sleeve (VC2700 y VC1700) tiene mayor cantidad de outliers de downtime, los mantenimientos preventivos deberían estar enfocados a esta operación de costura.

- ✓ Es importante que la empresa cuente con niveles adecuados de inventario de partes para reemplazo de maquinaria, el reemplazo de partes es una variable estadísticamente significativa para explicar el tiempo de paro de los equipos.

- ✓ Se identifica que la mayor de cantidad de fallos está relacionada a problemas en el cabezal de la maquinaria de Hem Sleeve, los mantenimientos preventivos deberían estar enfocados en esa parte específica de la máquina.

- ✓ Se recomienda medir el nivel de conocimiento de los mecánicos, esto sería de utilidad para mantener programas de capacitación y como una potencial variable para explicar la variación en los tiempos de paro.