

APRENDIZAJE ESTADÍSTICO PARA LA  
GESTIÓN DEL RIESGO DE VALIDACIÓN DE  
DOCUMENTOS TÉCNICOS

Autor:

Luis Daniel Chavarría Múnera

Trabajo de grado en modalidad de práctica profesional

Programa: Ingeniería Industrial

Presentado a:

Elkin Rodríguez Velásquez

Universidad Nacional de Colombia – Sede Medellín

Facultad de Minas

2019 – 2

# Contenido

1. Contexto empresarial .....	6
2. Áreas de la empresa y de desarrollo en la práctica profesional .....	8
3. Organigrama del área y funciones básicas .....	9
4. Selección del proyecto de enfoque para el trabajo de grado .....	11
5. Definición del proyecto seleccionado, estado inicial e indicadores .....	12
6. Metodología y desarrollo del proyecto.....	14
6.1. Entradas identificadas de la metodología y el aplicativo.....	14
6.2. Preguntas de investigación .....	14
6.3. Creación de base de datos para gestión documentos técnicos SIGNE – <i>Data wrangling</i> .....	15
6.4. Solución de BI integrando Tableau y R con registro histórico semanal.....	16
6.5. RPA para obtener de la información de validación de documentos técnicos .....	20
6.6. Metodología de gestión del riesgo de validación para documentos técnicos.....	21
6.7. Clasificación de comentarios mediante regresión logística multinomial penalizada con redes elásticas.....	22
6.8. Tablero de BI para la gestión del riesgo de validación y paso del modelo a operación .....	27
6.9. Resultados y conclusiones.....	30
6.10. Retroalimentaciones al programa de Ingeniería Industrial .....	31
7. Referencias .....	31

## Tabla de Figuras

Figura 1. Louis Renault con su primer vehículo. ....	6
Figura 2. Presencia de la alianza Renault-Mitsubishi-Nissan en el mundo.....	7
Figura 3. Información demográfica de los colaboradores en Renault-SOFASA.....	7
Figura 4. Proyectos vigentes en Renault-SOFASA .....	8
Figura 5. Misión y visión de la División de Ingeniería de la Alianza en Sistemas y Plataforma (DEA-SP)...	9
Figura 6. Organigrama de la DEA-SP Colombia .....	10
Figura 7. Funciones básicas, proveedores y clientes del puesto desarrollado. ....	11
Figura 8. Seguimiento manual de los documentos técnicos utilizando Excel. ....	13
Figura 9. Indicadores de estado de documentos técnicos automatizados. ....	17
Figura 10. Indicador de estado de documentos técnicos adaptado a la necesidad del usuario. ....	17
Figura 11. Tablero específico para planos en la DEA-SC. ....	18
Figura 12. Seguimiento temporal del indicador de documentos técnicos con trayectorias.....	19
Figura 13. Estado temporal de los documentos técnicos en la división mecánica-motor. ....	19
Figura 14. RPA extractor de archivos individuales para validación de planos.....	20
Figura 15. RPA consolidador de archivos. ....	21
Figura 16. Matriz de confusión del clasificador de libros de Julia Silge. ....	23
Figura 17. Veinte primeros valores de la matriz $\beta$ de dimensiones $(p = 347) \times (K = 657)$ . ....	25
Figura 18. Veinte primeras respuestas de la variable $y$ de dimensiones $p \times 1$ . ....	25
Figura 19. Multinomial deviance del ajuste vs log-lambda. ....	26
Figura 20. Palabras (coeficientes del modelo) y pesos asociados a la pertenencia a una categoría. ....	26
Figura 21. Matrices de confusión en el conjunto de prueba y del modelo sobreajustado. ....	27
Figura 22. Tablero principal para la gestión del riesgo de validación de planos. ....	28
Figura 23. Utilización del tablero de gestión de riesgo para el perímetro de ruedas, neumáticos y embellecedores. ....	29
Figura 24. Comentarios asociados a la selección realizada en el perímetro de ruedas y neumáticos. ....	29
Figura 25. Extracción de información directamente desde la visualización interactiva. ....	30

## Índice de Tablas

Tabla 1. Datos de origen – primeras 6 observaciones. ....	15
Tabla 2. Estructura de datos depurada para los desarrollos Renault-SOFASA. ....	16
Tabla 3. Estructura de datos consolidada de la primera fase de RPA. ....	21
Tabla 4. Porcentaje de riesgo consolidado por plano. ....	22
Tabla 5. Estructura de datos de comentarios asociados a los documentos técnicos.....	22
Tabla 6. Estructura de datos "tidy" para cada plano por palabra. ....	23

## **Agradecimientos**

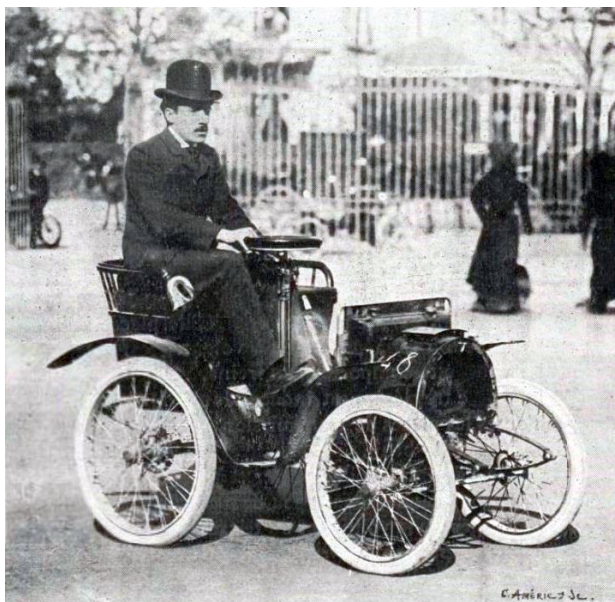
A todas esas maravillosas personas que tuve la oportunidad de conocer en Renault – SOFASA durante mi práctica, que no dudaron nunca en explicarme con la mayor pasión todo lo necesario para que esto fuera posible.

A mis docentes universitarios, por forjar mi personalidad a través de retos y enseñanzas con amor, sin ustedes no habría sido posible nada de esto.

A mi madre, Emilsen Múnera, por el constante apoyo a lo largo de mi carrera y su perseverancia inclusive cuando las circunstancias fueron muy complicadas.

## 1. Contexto empresarial

Renault es un fabricante francés de automóviles, vehículos comerciales y automóviles de carreras. Fundada en 1899 como Société Renault por Louis Renault y sus hermanos Marcel y Fernand. El primer vehículo de Renault fue el Renault Voiturette 1CV (Figura 1), que fue vendido a un amigo del padre de Louis después de darle un paseo de prueba el 24 de diciembre de 1898.



*Figura 1. Louis Renault con su primer vehículo.*

Actualmente, el gobierno francés posee el 15,7% de la empresa y Renault posee el 44,4% de Nissan, es propietaria además de Renault Samsung Motors y Dacia, formando entre los tres el “*Groupe Renault*”.

En Colombia, Renault-SOFASA es una empresa con 50 años en el mercado automotriz, dedicada al ensamble y comercialización de vehículos de la marca Renault. La Sociedad de Fabricación de Automotores S.A. fue fundada el 02 de Julio de 1969, a partir de una licitación del gobierno colombiano, donde se solicitaba la participación de una empresa extranjera con reconocimiento internacional. La planta se encuentra ubicada en Envigado, Antioquia, cuenta con un área total de 198.000 m<sup>2</sup> y una capacidad de 80.000 carros/año. El 100% de las acciones de Renault- SOFASA hacen parte de la Alianza (Renault - Nissan - Mitsubishi). En la figura 2 se presenta la distribución de la alianza en el mundo.

## Alliance Industrial plants



Figura 2. Presencia de la alianza Renault-Mitsubishi-Nissan en el mundo.

En la figura 2 se identifica una fuerte presencia global de la alianza en todo el globo. La posición estratégica de Colombia en América favorece que la marca pueda tomar de manera integrada decisiones de producción que permitan producir al menor costo posible y con la mejor calidad, además de vender los vehículos en varios países.

En la figura 3 se presentan algunos gráficos descriptivos de los colaboradores en Renault-SOFASA a octubre de 2019. El 66% de éstos son operarios en planta (*blue collar*), el 61% de los colaboradores tiene un contrato a término fijo y el 10% del total de empleados son mujeres. Aproximadamente el 60% de los empleados está en las edades de 30 a 45 años.

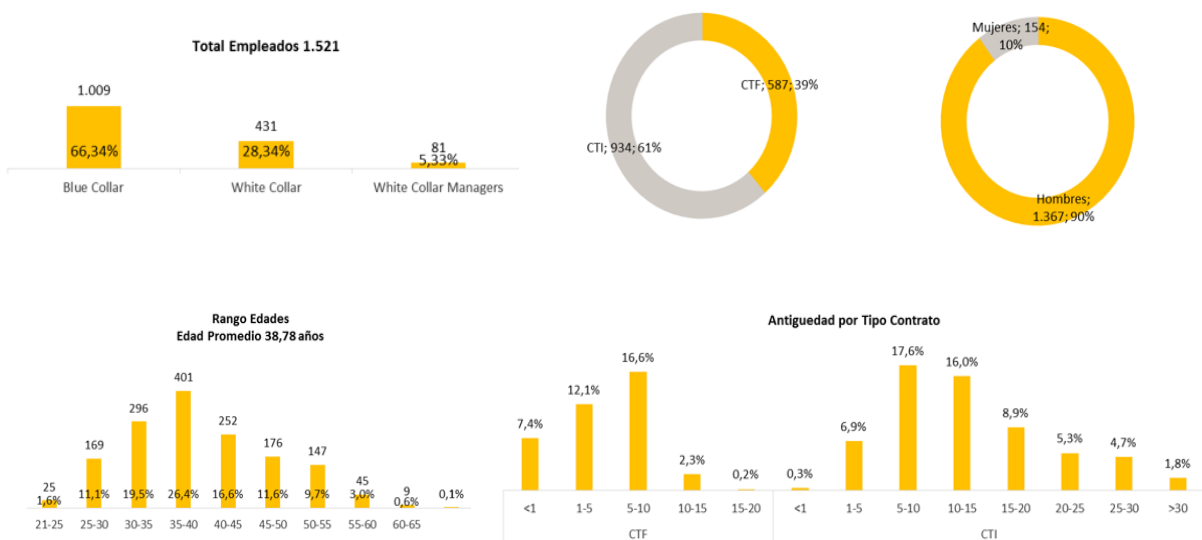


Figura 3. Información demográfica de los colaboradores en Renault-SOFASA

## PROYECTOS SOFASA

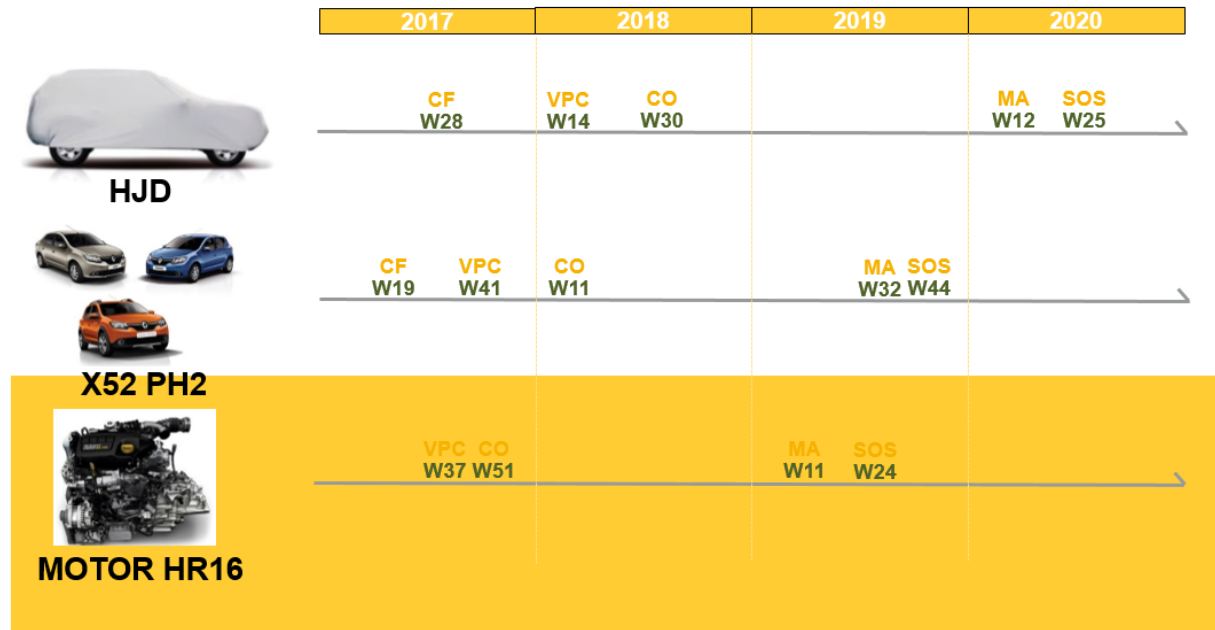


Figura 4. Proyectos vigentes en Renault-SOFASA

Actualmente se construye 1 nuevo vehículo proyecto en etapas tempranas, denominado HJD (Figura 4) con un nuevo motor llamado HR16, Los vehículos X52 PH2 (Logan 2020, Sandero 2020 y Stepway 2020) se lanzaron en diciembre del 2019. Con la introducción de nuevos retos en vehículos eléctricos y el boom de la industria 4.0 se espera que el número de vehículos proyecto incremente considerablemente.

## 2. Áreas de la empresa y de desarrollo en la práctica profesional

SOFASA está conformada por diversas áreas: Departamento comercial, Ingeniería del producto, Ingeniería del proceso, Departamento de fabricación, entre otras. El desarrollo de la práctica profesional durante el segundo período del año 2019 se realizó en el Departamento de Ingeniería del producto, específicamente en la División de Ingeniería de la Alianza de Sistemas y plataforma (DEA-SP).

Esta área se encarga del diseño, especificación y adaptación de las piezas que conforman sistemas dentro del vehículo, como el aire acondicionado, el sistema de frenos, el montaje de rueda, etc.





*Figura 5. Misión y visión de la División de Ingeniería de la Alianza en Sistemas y Plataforma (DEA-SP)*

En la visión se tiene propuesto convertirse en una de las Unidades Elementales de Trabajo (UET) que lidere los proyectos en mejores prácticas en toda la alianza Renault-Nissan. Esto implica un cambio en la arquitectura de funcionamiento que haga uso de tecnologías fundamentadas en los datos para la correcta toma de decisiones.

### 3. Organigrama del área y funciones básicas

La DEA-SP es un equipo de 15 personas incluyendo el practicante con los objetivos misionales específicos que aparecen en la figura 5, el organigrama detallado se presenta en la figura 6. Las 7 personas que aparecen bajo la etiqueta STT están bajo un modelo de contratación diferente, en el cuál su empleador directo no es Renault-SOFASA, sino la Universidad EAFIT; sin embargo, desempeñan funciones cruciales dentro del área y son tratados como líderes de ingeniería (LI) en diferentes áreas del vehículo.

## ORGANIZACIÓN DE-S DE-P COLOMBIA | RTA COLOMBIA

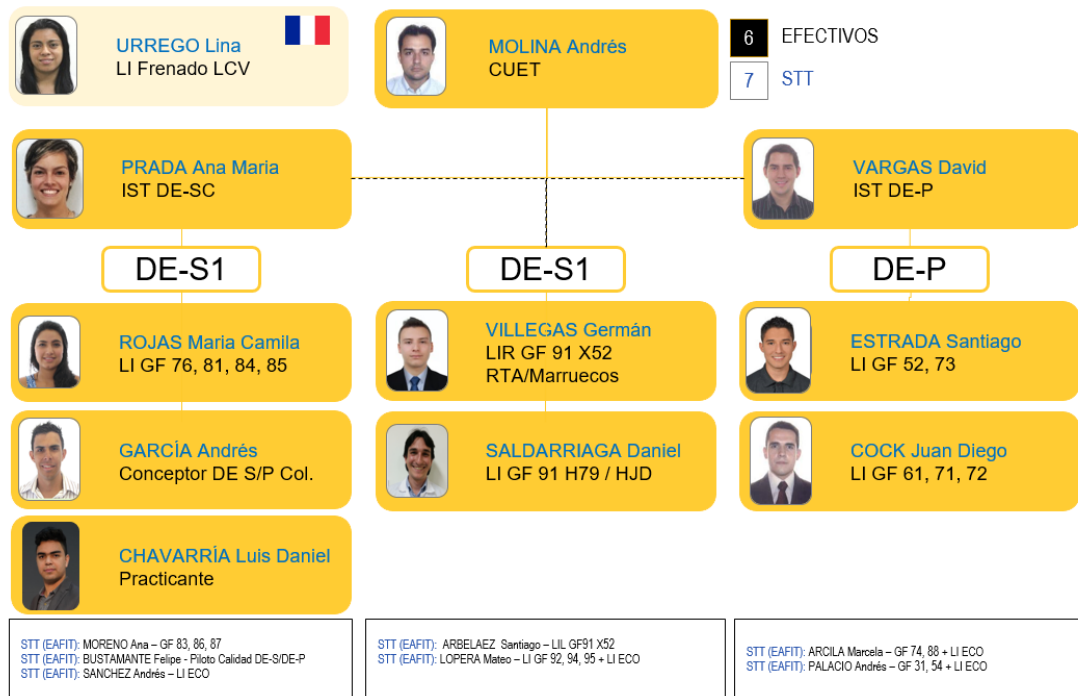


Figura 6. Organigrama de la DEA-SP Colombia

Dentro de la DEA-SP se designó al conceptor Andrés García, encargado de todos los temas asociados al modelamiento 3D de los sistemas y plataforma del vehículo como “padrino” de práctica; quien apoyó enormemente el desarrollo de este trabajo explicando conceptos, proponiendo preguntas de investigación y traduciendo problemas en necesidades y oportunidades abordables con los conocimientos obtenidos durante la carrera.

En la figura 7 se especifican todas las funciones básicas desarrolladas durante el cumplimiento de la práctica profesional, indicando los insumos utilizables, medios técnicos, competencias requeridas, proveedores, actividades, clientes e indicadores.


 <b>RENALT</b> Passion for life		<b>FICHA DE IDENTIDAD DE PROCESO PERSONAL</b>		Servicio : DEA-S/P Redactor: Luis Daniel Chavarría									
<b>Medios Técnicos</b> CATIA v5, NEWPDM, SIGNE, RStudio, Excel, Tableau, Office, UiPath, Share Points: DOC-AMERIQUE, DE-TD-RTA, Yammer UET			<b>Personal / Competencias</b> Formaciones: PROFIA, Todos los perímetros, 4Dx, GMF, QFS, O54, NPDM, RUET, 5s, PQB, Planning, RPA, Historia Renault Formaciones meter: Entrenamiento con concepor en CATIA, Entrenamiento con CUET y concepor en lectura de planos.										
<b>Proveedores</b> <ul style="list-style-type: none"> <li>Lis y PFEs RTA COL</li> <li>CUET</li> <li>ISTs proyectos</li> <li>ACT</li> </ul>	<b>Datos de entrada</b> <ul style="list-style-type: none"> <li>Extracciones de SIGNE</li> <li>Lista de desarrollos en los proyectos</li> <li>Documentos de Referencias que necesitan especificación</li> </ul>	<b>Actividades del proceso</b> <ul style="list-style-type: none"> <li>Validar y suministrar información de planos, 3D y listas de materiales para el programa de fomento a la industria automotriz (PROFIA) y otras áreas que lo requieran.</li> <li>Apoyar la gestión de envíos, montabilidades y pruebas para piezas y coordinar la salida de éstas de la línea (Rechazos) con los ingenieros de logística y el área de calidad para que no afecten el proceso de ensamble.</li> <li>Montaje de piezas en vehículos ya sea en taller o en planta para validar especificaciones técnicas, calidad, estética, entre otros; con el fin de generar alguna economía o mejora en el vehículo.</li> <li>Construir indicadores y tableros útiles para la mejora de procesos en el RTA que permitan tomar decisiones rápidas y ciertas utilizando herramientas de visualización y manipulación de datos.</li> <li>Subir planos a circuito de firmas de acuerdo a las solicitudes de los LI y el concepor para garantizar el cumplimiento de los valores.</li> <li>Desarrollar el proyecto de práctica a cabalidad y definir claramente la intensidad horaria de dedicación, documentando todos los pasos realizados en el trabajo final.</li> </ul>	<b>Datos de Salida</b> Tableros de indicadores y aplicaciones dinámicas Documentos especificados de PROFIA Fotografías de piezas montadas que ilustren el objetivo deseado Correos electrónicos a ACT para envíos y órdenes de salida Planos validados en circuito de firmas Documento final del trabajo de práctica	<b>Clientes</b> <ul style="list-style-type: none"> <li>Lis y PFEs RTA</li> <li>CUETs SOFASA</li> <li>ISTs Proyectos y VS</li> <li>ACT</li> <li>Nuevo practicante</li> </ul>									
<b>Documentos de referencia</b> 2D Funcionales, Yammer UET			<b>Indicadores</b> <table border="1"> <tr> <th>De dominio del proceso</th> <th>De resultados</th> </tr> <tr> <td>Puntualidad en tareas 4dX</td> <td>% Documentación proyectos (MDS, 2D, 3D)</td> </tr> <tr> <td>Cumplir hitos definidos por CUET</td> <td>% Riesgo validación 2D</td> </tr> <tr> <td>Aprobar requerimientos tutor de práctica</td> <td>% Cumplimiento PROFIA</td> </tr> </table>			De dominio del proceso	De resultados	Puntualidad en tareas 4dX	% Documentación proyectos (MDS, 2D, 3D)	Cumplir hitos definidos por CUET	% Riesgo validación 2D	Aprobar requerimientos tutor de práctica	% Cumplimiento PROFIA
De dominio del proceso	De resultados												
Puntualidad en tareas 4dX	% Documentación proyectos (MDS, 2D, 3D)												
Cumplir hitos definidos por CUET	% Riesgo validación 2D												
Aprobar requerimientos tutor de práctica	% Cumplimiento PROFIA												

Figura 7. Funciones básicas, proveedores y clientes del puesto desarrollado.

## 4. Selección del proyecto de enfoque para el trabajo de grado

Mediante lluvias de ideas en la DEA-SP e identificación de dificultades actuales del equipo de trabajo, se presentaron varias posibilidades para proyecto de grado, las cuales fueron evaluadas posteriormente con el tutor de práctica Elkin Rodríguez y se seleccionó la que mejor cumpliera con el rigor académico de la Universidad Nacional y permitiera entregar a la empresa un insumo útil que solucionara un problema real. A continuación, se especifica una lista de los 3 primeros proyectos presentados al tutor.

### ▪ Sistema de gestión de riesgos de validación para documentos técnicos

Diseñar desde cero un sistema de gestión del estado de validación de los documentos técnicos Renault, que permita además realizar una gestión integrada del riesgo de validación mediante la consulta y visualización histórica de los documentos técnicos más rechazados por los expertos en la casa matriz en Francia y sus causas, con el propósito de adoptar acciones que permitan anticipar la validación de los documentos técnicos.

### ▪ Aplicación de metodología de desarrollo de la maqueta numérica 3D

Aplicar la nueva metodología previamente implementada por la Ingeniería de Renault-SOFASA para el proceso de actualización de la maqueta numérica de los vehículos en su fase proyecto y vida serie mediante software RPA (*Robotic Process Automation*) conectando de manera eficiente los diferentes aplicativos de Renault durante el ciclo de vida del producto.

**Objetivos específicos:**

1. Utilizar el aplicativo de construcción de la definición técnica de los diferentes perímetros del vehículo para definir las entradas del robot de actualización de la maqueta numérica en dos proyectos de la empresa. **Entregable:** Archivos de entradas del robot (múltiples archivos de Excel).
2. Analizar los diferentes perímetros (sistemas del vehículo) construidos por el robot y revisar de la definición técnica de cada uno, realizando las correcciones requeridas (ubicación espacial de piezas fuera de posición, coeficientes de piezas mal especificados, interferencias entre piezas o contactos, entre otros) **Entregable:** Informes del estado de la definición técnica de cada perímetro.
3. Enviar el ensamble numérico a la GESC (Recurso humano subcontratado en India para tareas operativas) y hacer seguimiento del mismo de todos los sistemas del vehículo para asegurar la exhaustividad de la maqueta numérica oficial. **Entregable:** Tablero de seguimiento del GESC.

- **Diseño de un esquema de validación de la encuesta de calidad QFS**

El departamento de calidad de la compañía genera un reporte utilizando la encuesta de retroalimentación de la calidad (QFS) entregada a los clientes, sin embargo, estos indicadores no incluyen un modelamiento estricto que considere la variabilidad natural de ésta; por lo que en ocasiones se podrían estar entregando resultados estadísticamente iguales en un intervalo de confianza que no está siendo calculado a pesar de que las medias son diferentes. El proyecto consistiría en crear una metodología robusta de la encuesta de calidad que permita identificar si el indicador efectivamente bajó o subió una cantidad estadísticamente significativa.

## 5. Definición del proyecto seleccionado, estado inicial e indicadores

Luego de reunirse con el asesor de práctica, se identifica que el proyecto de diseño de una metodología para la gestión del estado de los documentos técnicos que involucre la gestión del riesgo de validación es un reto académico y de implementación apropiado para la práctica, que además contribuirá enormemente al desarrollo del flujo de trabajo del departamento de ingeniería.

En Renault existen 3 documentos técnicos clave para cada pieza:

- **MDS:** Describe la lista de materiales completa de la pieza, además de sus composiciones y características químicas, incluyendo la información de los proveedores y la disponibilidad.
- **2D o Plano:** Describe la geometría de la pieza en un esquema bidimensional, incluyendo cotas técnicas y especificación de reglamentaciones necesarias para cada país bajo estándares internacionales y Renault.
- **3D:** Especifica el esquema tridimensional de la pieza y permite su incorporación en el *Digital Mockup* (DMU) o maqueta numérica del vehículo completo, lo cual permite llegar a validaciones previas al montaje por parte del equipo de arquitectura, que se encarga de evitar contactos entre las partes y superposiciones que harían el vehículo inviable.

Cada uno de estos documentos técnicos debe ser enviado a la casa matriz en Francia donde un *pool* de expertos valida técnicamente cada uno de éstos, añade remarcas, acepta o rechaza los documentos;

dependiendo de la aprobación de este conjunto de expertos en un área específica del vehículo, los líderes de ingeniería pueden incorporar o no una modificación en un vehículo Renault que pueda ser vendido.

Actualmente, el seguimiento de los documentos técnicos es realizado manualmente por el conceptor de cada división utilizando el sistema de información de documentos técnicos SIGNE y Excel (Figura 8); esto es un proceso dispendioso y mecánico donde se debe abrir manualmente cada documento en el sistema de información y ver si ya está validado o no por los expertos; posteriormente se marca en una hoja de Excel si el documento está “OK” o “NOK”.

Con base en la información provista por los conceptores, esta validación manual puede tomar hasta 15 minutos por referencia, con más de 500 piezas que requieren seguimiento, es decir:

$$15 \frac{\text{min}}{\text{pieza}} * 500 \text{ pieza} = 7.500 \text{ min} = 125 \text{ h} \quad (\text{Eq. 1})$$

Es decir, que en promedio para actualizar únicamente la base de datos del indicador de documentos técnicos sin crear el reporte en sí se deben contratar al menos 3 personas que trabajen 48 horas semanales para dar cumplimiento a este requerimiento en horas-hombre en una semana.

Dada la excesiva manualidad de este proceso y a que el conceptor cuenta con funciones de mayor importancia y valor agregado en los equipos como el modelamiento 3D para el diseño y modificación de piezas, sólo ciertas referencias (20 máximo) tienen el seguimiento al día y están en el radar del equipo. Usualmente los proyectos se retrasan debido al desconocimiento de partes sin la definición técnica completa, que en ocasiones son diseñadas desde cero por el equipo (pieza local) o que son traídas de otra planta Renault y podrían requerir alguna modificación *Carry-Over Carry-Across* (pieza COCA) para poder ser instaladas en el vehículo.

REFERENCIA	COCA	INDICE PIEZA	TIPOLOGIA PIEZA	PLANO	INDICE PLANO	VERSION PLANO	STATUS PLANO NEWPDM	STATUS PLANO SIGNE	STATUS 3D NEWPDM	STATUS 3D SIGNE	STATUS MDS SIGNE	STATUS PIEZA SIGNE	COMENTARIOS	COMENTARIOS AG
164467923R	Orange							NOK	NOK	NOK	NOK		Revisar MDS, va a evolucionar por cambio de malla	2D S40.1
164464012R	Rouge							NOK	NOK	NOK	NOK		Revisar MDS, va a evolucionar por cambio de malla	2D S40.1
164463166R	Orange							NOK	NOK	NOK	NOK		Revisar MDS, va a evolucionar por cambio de malla	2D S40.1
175067342R	Orange							NOK	NOK	NOK	NOK		Revisar MDS, no cambia	2D S40.1
175062815R	Orange							NOK	NOK	NOK	NOK		Revisar MDS, va a evolucionar por cambio de malla	2D S40.1
172022790R	Orange							NOK	NOK	NOK	NOK			REVISAR QUE 3D SEA COHERENTE CON PLANO, S42
172025249R	Orange							NOK	NOK	NOK	NOK			PLANOS LISTO PARA CS
164468120R	Orange							NOK	NOK	NOK	NOK		Revisar MDS, va a evolucionar por cambio de malla	2D S40.1
172026454R	Orange							NOK	NOK	NOK	NOK			2D LISTO PARA CS
172023154R	Orange							NOK	NOK	NOK	NOK			REVISAR QUE 3D SEA COHERENTE CON PLANO, S42
172025115R	Orange							NOK	NOK	NOK	NOK			PLANOS LISTO PARA CS
172217545R	Orange							NOK	NOK	NOK	NOK			2D LISTO PARA CS

Figura 8. Seguimiento manual de los documentos técnicos utilizando Excel.

Adicionalmente, el sistema actual no involucra nada de la información de qué tan probable es que estas validaciones sean rechazadas por los *pools* de expertos en Francia. La empresa cuenta con un sistema de información denominado NPDM para los planos y los 3D donde está consignada la información de validación; en promedio, realizar un análisis de riesgo manualmente para una pieza toma alrededor de 60 minutos para un ingeniero entrenado en el tema teniendo en cuenta los comentarios o remarcas anteriores del historial de validación de la pieza en todos los proyectos anteriores. Es decir:

$$60 \frac{\text{min}}{\text{pieza}} * 500 \text{ pieza} = 30.000 \text{ min} = 500 \text{ h} \quad (\text{Eq. 2})$$

Esto significa que tener un indicador actualizado que involucre el seguimiento de los documentos técnicos y la probabilidad de que estos sean rechazados, añadiendo comentarios sobre los riesgos de éstos bajo la metodología actual tomaría 750 horas por semana de personal altamente capacitado, es decir un equipo de al menos 16 ingenieros dedicados exclusivamente al seguimiento del estado de los documentos técnicos del proyecto y su nivel de riesgo.

Con salarios de entrada aproximados de COP 4'000.000 mensuales luego de las deducciones y reglamentaciones laborales, el costo aproximado promedio de mantener esta información actualizada sería de COP 64'000.000 por mes.

## 6. Metodología y desarrollo del proyecto

### 6.1. Entradas identificadas de la metodología y el aplicativo

El proyecto tiene como objetivo construir un indicador automatizado que indique el estado de los documentos técnicos y además su probabilidad de riesgo de ser rechazado, se identifica que éste debe tomar las siguientes entradas:

- Extracción semanal del sistema de información SIGNE con el estado de validación de todas las piezas en el vehículo en MDS, 2D y 3D o conexión directa a la base de datos.
- Información estructurada del historial de validación de todas las piezas de todos los proyectos (vehículos) en la división de ingeniería de América *Renault Technique Amerique* (RTA).
- Listado de la referencia de las piezas desarrolladas en el proyecto al que se le realizará seguimiento, con su tipología asociada: pieza local, COCA, PLDA (Pieza de Larga distancia de aprovisionamiento) o POI (Pieza desarrollada por un proveedor externo dentro de la fábrica, pero con definición técnica Renault).

### 6.2. Preguntas de investigación

Se plantean las siguientes preguntas claves que pretende solucionar el proyecto:

- ¿Cuáles son los planos o funciones del vehículo con mayor probabilidad de ser rechazados en los circuitos de firmas?
- ¿Cuál es el estado actual de validación de los documentos técnicos de los vehículos proyecto?
- ¿Cuál es el perímetro con mayor riesgo de validación? ¿Quién tiene la mayor responsabilidad en un proyecto?
- ¿Cómo obtener la información para responder las preguntas si no se pueden generar extracciones del sistema de gestión de planos NPDM?
- ¿Cómo agregar valor a la clasificación de riesgo que permita tomar decisiones particulares con respecto a las estrategias de control con los proveedores o el proceso?

### 6.3. Creación de base de datos para gestión documentos técnicos SIGNE – *Data wrangling*

Renault cuenta con un sistema propio para la gestión de la Ingeniería producto del vehículo denominado SIGNE. De éste es posible generar extracciones mensuales de la información técnica en diferentes formatos estructurados. Luego de limpiar los registros y seleccionar las variables generadas por éste se obtiene la siguiente información de interés en una estructura de datos tabular.

*Tabla 1. Datos de origen – primeras 6 observaciones.*

ELTDEC_numero	ELTDEC_designation	PIE_numero	PIE_designation	PIE_gest_MDS	PIEIND_typologie	DOCTECH_numero	DOCTECH_type
11	SUPERSTRUCTURE	763568061R	GOUTTIERE LAT AR D	Oui	Plan et Num 3D obligatoires	2D6031454R	Plan
11	SUPERSTRUCTURE	763568061R	GOUTTIERE LAT AR D	Oui	Plan et Num 3D obligatoires	713008490 / 1.00	MDS
11	SUPERSTRUCTURE	763568061R	GOUTTIERE LAT AR D	Oui	Plan et Num 3D obligatoires	3DRP115652	3D
11	SUPERSTRUCTURE	763571615R	GOUTTIERE LAT AR G	Oui	Plan et Num 3D obligatoires	2D6031454R	Plan
11	SUPERSTRUCTURE	763571615R	GOUTTIERE LAT AR G	Oui	Plan et Num 3D obligatoires	713008497 / 1.00	MDS
11	SUPERSTRUCTURE	763571615R	GOUTTIERE LAT AR G	Oui	Plan et Num 3D obligatoires	3DRP115752	3D

En la tabla 1 se presenta un ejemplo de los datos de origen, donde se evidencian registros para 2 piezas pertenecientes a la superestructura del vehículo. La columna *PIE\_gest\_MDS* indica que todas éstas requieren el documento técnico de los materiales “MDS”; adicionalmente la columna *PIEIND\_typologie* indica que se requieren también los documentos técnicos “2D” y “3D”. En la columna *DOCTECH\_numero* se encuentra el número de documento asociado, si está vacío, es porque el documento no está validado todavía (no ha sido aprobado en la casa matriz).

Con esta información y utilizando reglas de decisión simples, se cambia la columna *DOCTECH\_numero* a una estructura longitudinal y se crean las columnas “2D-ok”, “3D-ok” y “MDS-ok”, que permiten diferenciar si el documento está al día o no para la pieza asociada. Posteriormente se añaden las demás variables de interés utilizando el código de la pieza como campo clave, lo que conduce a la estructura de datos presentada a continuación.



*Tabla 2. Estructura de datos depurada para los desarrollos Renault-SOFASA.*

[illegible]

Nótese que en la tabla 2 se encuentran las 3 variables de estado del documento técnico, con valores posibles de “OK” si el documento existe, “NOT OK” si el documento falta y “NO APLICA” en caso de no necesitarlo. Adicionalmente se han cruzado otras fuentes de la compañía para añadir información de interés, como la dirección a la que pertenece la pieza, el perímetro del vehículo, entre otros.

Si bien las reglas de construcción de la base de datos son simples, se necesitan conocimientos específicos en bases de datos y tecnología de información para la coordinación con los departamentos de TI para obtener los registros correctos, realizar la limpieza de los más de 60.000 registros por semana en el vehículo proyecto, procesos de validación, mantenimiento y escalabilidad. Todas estas particularidades, añadidas al desconocimiento generalizado de soluciones fundamentadas en código para este tipo de problemas prevenían que la empresa pudiera saber el estado actual de los vehículos en desarrollo sin tener que dedicar el recurso humano expuesto anteriormente. (Lee & Shin, 2020) tratan las causas de algunos desafíos similares.

Con sólo la construcción de la base de datos utilizando esquemas automatizados para las 509 referencias en desarrollo se llega a ahorros sustanciales en tiempo, dinero y oportunidad (Eq. 1 y 2) en la solución de inconvenientes en el proyecto.

#### 6.4. Solución de BI integrando Tableau y R con registro histórico semanal

Dentro de los objetivos del proyecto de seguimiento al estado de los documentos técnicos está la presentación ágil e interactiva de resultados, de manera que sea posible llegar a decisiones en la operación a partir de tableros de control tácticos, exportar subconjuntos de información rápidamente, disponibilidad en dispositivos móviles y despliegue web. La compañía de BI Tableau cumplió con todas estas características; por lo que se desarrolló un aplicativo interactivo en esta plataforma que funcionara como tablero de control con frecuencia de actualización semanal (figura 9).



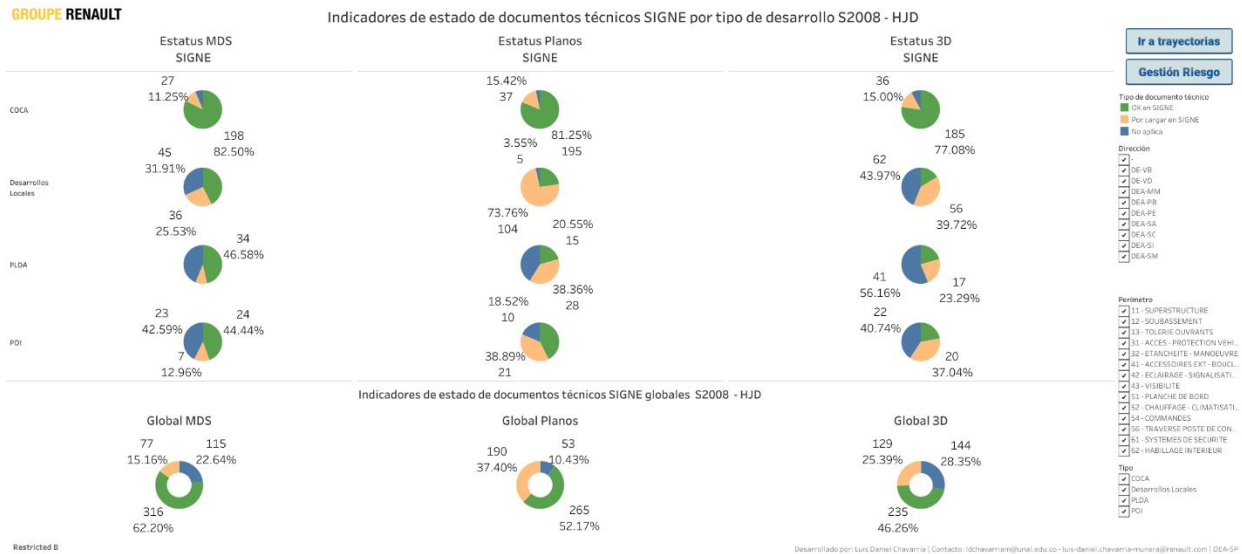


Figura 9. Indicadores de estado de documentos técnicos automatizados.

En la figura 9 se encuentran 3 columnas clave, una para cada tipo de documento técnico. En las 4 primeras filas de la parte superior, se realiza una separación por tipología de pieza y en la fila inferior se muestra un indicador global para cada tipo de documento técnico incluyendo todas las piezas. Adicionalmente, en la parte superior derecha se cuenta con filtros para las direcciones, la parte del vehículo (perímetro) y tipología de pieza, que cambian el tablero de forma dinámica a la medida de lo que desea ver el usuario.

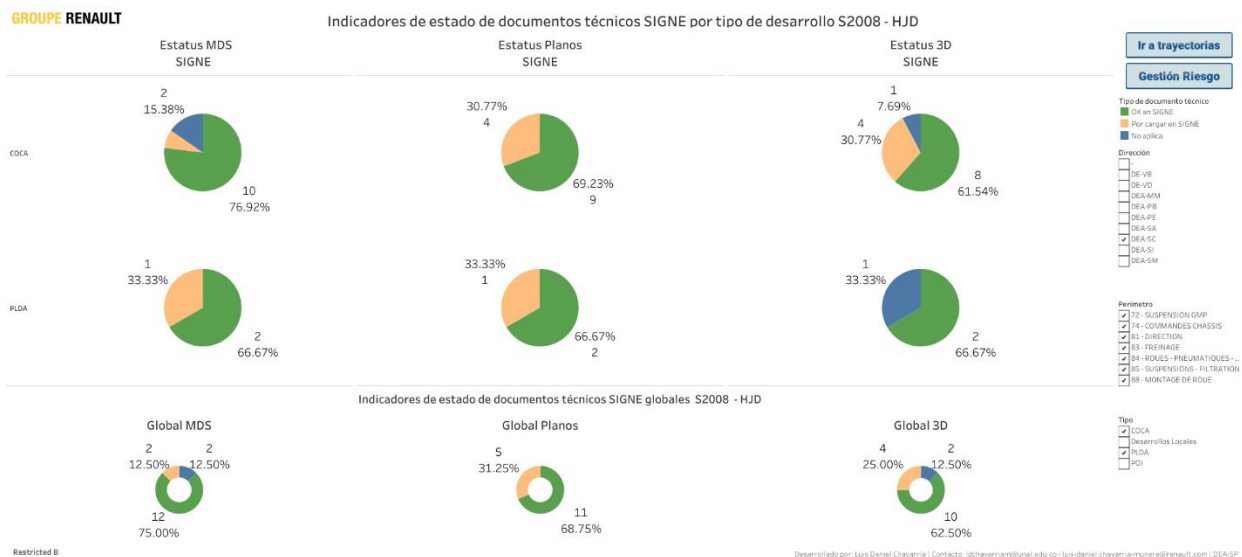


Figura 10. Indicador de estado de documentos técnicos adaptado a la necesidad del usuario.

En la figura 10 se presenta uno de estos casos particulares, donde se ha seleccionado la división de ingeniería de Sistemas y Chasis (DEA-SC) en sus desarrollos de tipología COCA y PLDA. Adicionalmente, al pasar el ratón encima de éstos se presenta información descriptiva de la selección. Se programó una interacción adicional al hacer clic en cualquiera de los indicadores, de manera tal que al hacerlo se lleve al usuario a saber exactamente cuáles son las piezas faltantes asociadas a la necesidad del usuario.

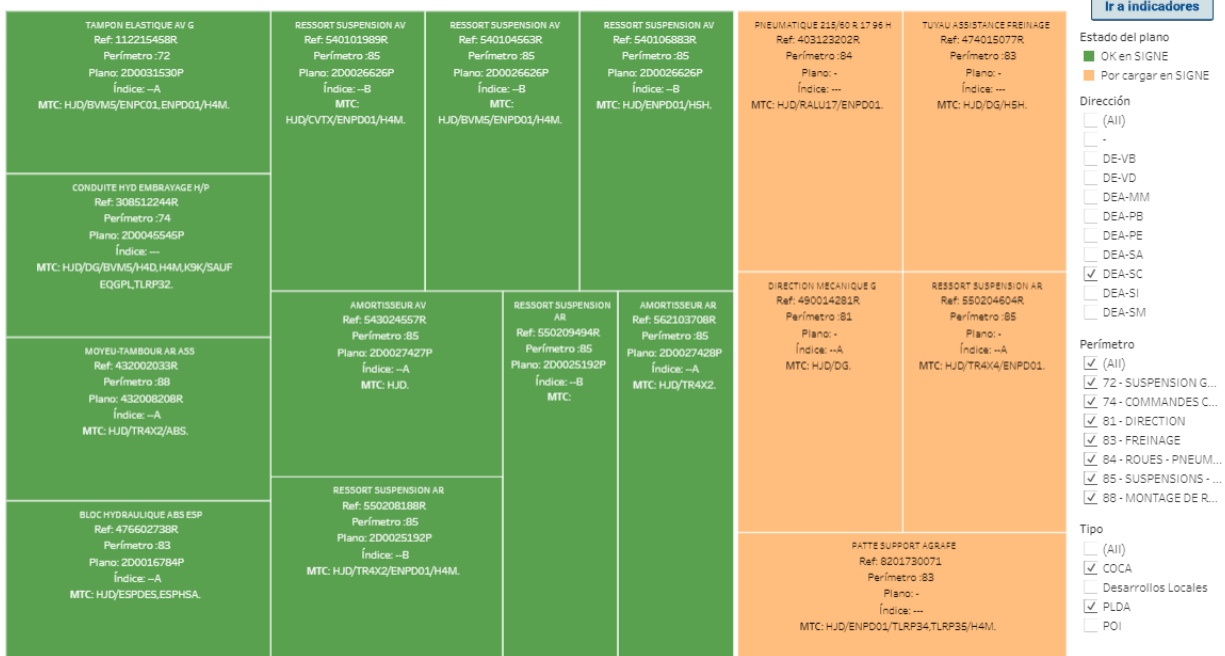


Figura 11. Tablero específico para planos en la DEA-SC.

En la figura 11 se presenta el resultado una de estas interacciones, donde al hacer clic en uno de los indicadores de la columna Planos de la figura 10 se llega al detalle de cuáles son las piezas que tienen planos faltantes. Se evidencia la congruencia entre ambos tableros: el indicador global muestra que hay 9 planos validados y 5 faltantes, en éste se presentan cuáles son los planos faltantes, quién es el responsable de gestionarlos (perímetro), a qué vehículo y en qué lugar específico pertenece (MTC), cuántas veces se ha intentado validar (índice) y su nombre y referencia. Esto permite integrar la visión estratégica (gerencia del proyecto), táctica (coordinadores de UET) y operativa (líderes de ingeniería) mediante la desagregación visual de la información actualizada.

Adicionalmente, se generó un esquema de seguimiento temporal del estado de los documentos técnicos (figura 12) accesible mediante el botón “Ir a trayectorias” en la figura 10, utilizando versiones históricas donde se presenta una trayectoria lineal representada en área gris desde el inicio del seguimiento hasta la fecha esperada del *Part Trial 2* (ABPT2), donde se espera que todos los documentos técnicos del vehículo proyecto estén 100% validados.

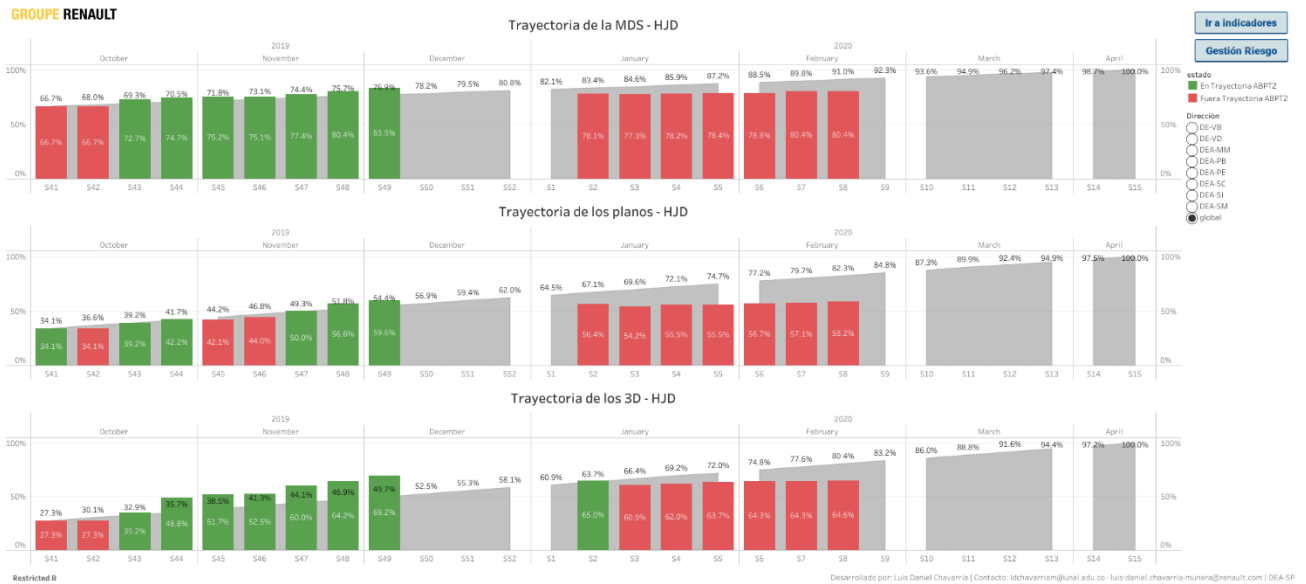


Figura 12. Seguimiento temporal del indicador de documentos técnicos con trayectorias.

Cuando el indicador de la figura 12 se encuentra en verde, es porque está en la trayectoria lineal propuesta, en otro caso es rojo. Se realiza un seguimiento semanal que además se puede desagregar por divisiones para ver las contribuciones individuales de cada división al proyecto e identificar dónde hay retrasos utilizando el filtro en la parte superior derecha.

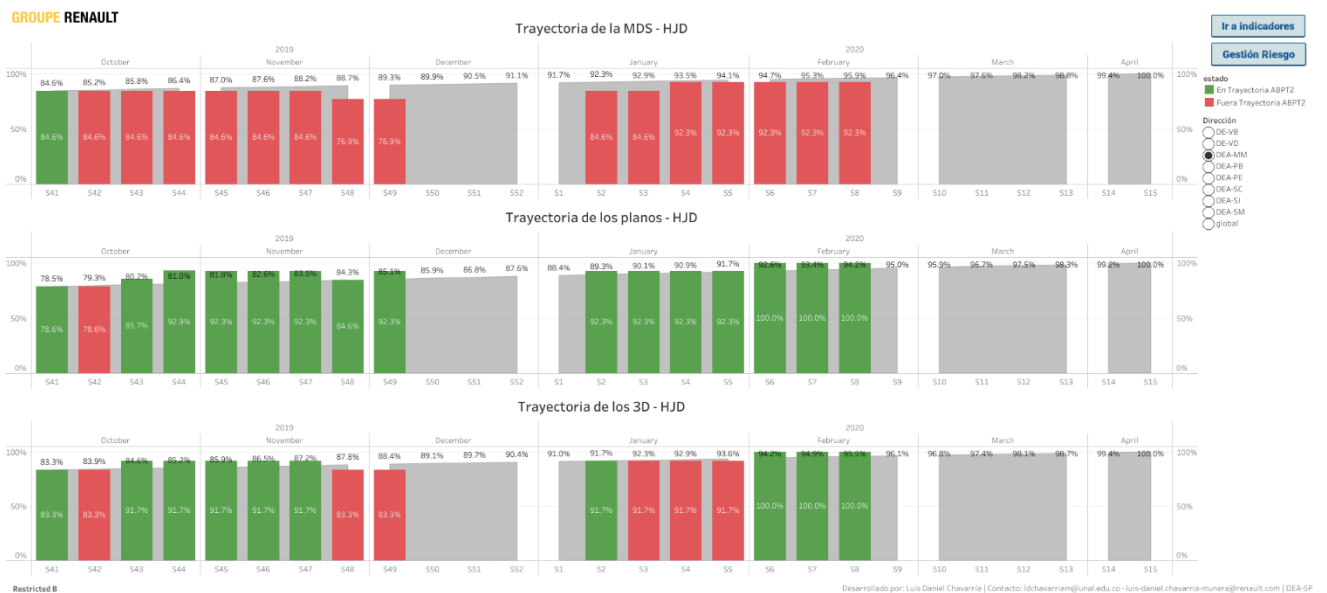


Figura 13. Estado temporal de los documentos técnicos en la división mecánica-motor.

En la figura 13 se presenta la dinámica temporal de la división mecánica-motor (DEA-MM), donde se identifica que han validado todos los documentos técnicos requeridos en planos y 3D antes del hito propuesto, sin embargo, todavía les falta realizar algunas especificaciones adicionales en los materiales (MDS).

## 6.5. RPA para obtener de la información de validación de documentos técnicos

Dado que no es posible generar extracciones escaladas desde el sistema de gestión de vida del producto “NPDM”, se construyó un robot virtual, más conocidos como RPA de (*Robotic Process Automation*) (Hofmann, Samp, & Urbach, 2019) cuyo trabajo fue descargar y consolidar la información para la gestión del riesgo y los comentarios de los validadores de documentos técnicos.

Este robot funciona mediante un flujo de “clics” en objetos HTML específicos y decisiones que permiten reaccionar a diferentes situaciones que se presentan en la página, por ejemplo, cuando un documento técnico nunca ha tenido un plano, en cuyo caso no habría información disponible para descargar. En la figura 14 se presenta una de las rutinas para descargar la información de validación de los planos utilizando el software UiPath (UiPath, 2016).

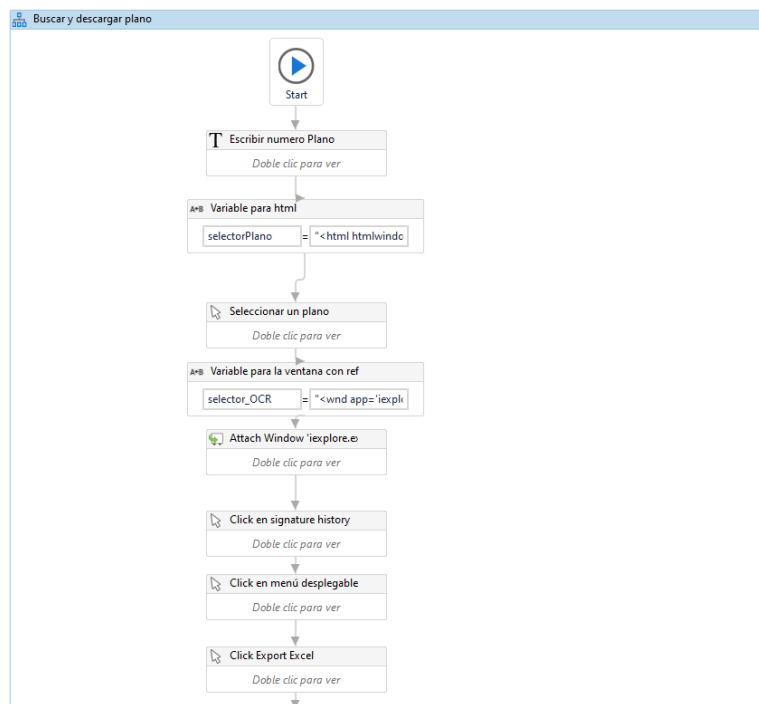


Figura 14. RPA extractor de archivos individuales para validación de planos.

Por restricciones de confidencialidad, el funcionamiento del aplicativo de gestión del producto de Renault no puede ser publicado ni tampoco el robot operando sobre éste. Este aplicativo está construido sobre HTML y funciona como una página web, como pasos generales visibles en la figura 14:

- El robot escribe el número de plano en una barra de búsqueda, el cuál toma de una lista de Excel.
- Selecciona el plano correcto una vez la búsqueda ha finalizado
- Utilizando la tecnología OCR (*Optical Character Recognition*) de UiPath, que sirve para el reconocimiento de caracteres, el programa toma algunas decisiones de flujo dependiendo de los contenidos de la página web.
- Se accede al historial de validación o “signature history”

- Se exportan los resultados a Excel y se guardan en una carpeta del equipo

Posteriormente, se consolidan los archivos de Excel utilizando otro RPA, cuyo flujo base en UiPath se presenta en la figura 15.

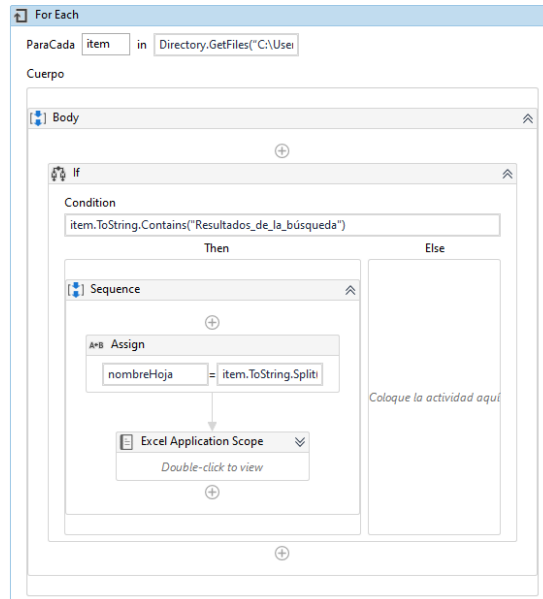


Figura 15. RPA consolidador de archivos.

Para cada ítem en la carpeta donde están guardados las salidas del robot extractor, se recorren las filas y se realizan operaciones de unión entre las filas, ya que todos los archivos cuentan con los mismos encabezados.

## 6.6. Metodología de gestión del riesgo de validación para documentos técnicos

Con la información obtenida y consolidada mediante RPA, el propósito es construir una base de datos útil que permita, para cada plano, calcular el número de veces que ha ido a circuito de firmas y de éstas, cuántas veces ha sido rechazado. Para ello, se realizan las siguientes transformaciones:

- Seleccionar únicamente los planos con estado de madurez “Validated” o “Refused”, ya que son los estados de interés para el cálculo del riesgo.
- Se eliminan los planos cuya información vive en el aplicativo anterior de gestión de vida del proyecto, “GDG”, ya que desde la implementación del nuevo sistema “NPDM” se empezó a llevar registro de la información de interés. Un ejemplo de la estructura de datos resultante se presenta en la tabla 3, donde se identifican 3 diferentes planos con su estado de madurez.

Tabla 3. Estructura de datos consolidada de la primera fase de RPA.

Name / Id	Index	Maturity	Designation	Family Project
8200309441	--F	Validated	PLAQUETTE FIX ARRET PORTE ASS	L90

763572736R	--M	Refused	GOUTTIERE LAT AR G	B52
763572736R	--M	Validated	GOUTTIERE LAT AR G	B52
763572736R	--N	Validated	GOUTTIERE LAT AR G	B52
801192401R	--C	Validated	RENFORT CHARNIERE SUP PIED AV G ASS	B52

- Se agrupan los planos de acuerdo a la columna **Name / Id**, de ésta manera, los cálculos realizados ocurren dentro de cada plano, en la tabla 3, se cuenta con 3 grupos únicos. Se calcula la proporción de documentos rechazados sobre el total del grupo, es decir, un plano con riesgo cero nunca ha sido rechazado y uno con riesgo 1 ha sido rechazado todas las veces que fue enviado a circuito de firmas. La estructura de datos resultante es la siguiente para estos 3 planos.

Tabla 4. Porcentaje de riesgo consolidado por plano.

Name / Id	Designation	total	riesgo
763572736R	GOUTTIERE LAT AR G	3	0,333333
801192401R	RENFORT CHARNIERE SUP PIED AV G ASS	1	0
8200309441	PLAQUETTE FIX ARRET PORTE ASS	1	0

Con estas transformaciones realizadas, la información base está lista para ser incorporada al aplicativo de seguimiento de documentos técnicos mediante la pestaña “Gestión del riesgo” presente en la figura 9.

## 6.7. Clasificación de comentarios mediante regresión logística multinomial penalizada con redes elásticas

Luego la segunda fase de *web-scrapping* y consolidación de información con RPA, se obtiene la siguiente estructura de datos para todos los planos en Renault América.

Tabla 5. Estructura de datos de comentarios asociados a los documentos técnicos.

Name/Id	Maturity	Designation	Assignee	Approval Status	Comment
2D0017696P	Refused	BARRE ANTIDEVERS AV	a036324 bw60262 ay15765	Reject Approve None	- Couple de serrage des BBAD à 62 Nm sinon réduire épaisseur peinture des pattes forgées à 80µm maxi ald 160µm - Merci de demander au fournisseur de mettre les informations protection pour les brides du ce projet dans la note R&D07-1-I et de nous faire parv

El objetivo del modelo es obtener una clasificación útil de los rechazos que permita tomar acciones de producción con respecto a los proveedores, ingenieros encargados, planes de control, entre otros. Dentro de la organización se formularon 3 categorías de motivo de rechazo:

- Técnico:** El plano fue rechazado por fallas en el rigor técnico, por ejemplo: reglamentación faltante, tolerancias inadecuadas, geometría incorrecta, entre otros.
- Forma:** El plano fue rechazado por aspectos de normalización, ausencia o exceso de simbología, cajetines o sellos fuera de posición, entre otros.

- **Procedimiento:** El plano fue rechazado por estar en la versión (índice incorrecto), por número de reglamentaciones de seguridad o de ley mal contadas, firmantes citados incorrectos, entre otros.

Dado que esta clasificación no es una variable a priori que exista en la base de datos, se realizó un muestreo aleatorio estratificado entre los perímetros (diferentes áreas del vehículo), garantizando un tamaño de muestra mínimo de 15 observaciones en cada uno, con un número mayor para las categorías con mayor número de partes. (Lohr, 2019) Con base en esta clasificación, el motivo de rechazo del plano en la tabla anterior sería técnico ya que habla de una especificación incorrecta para el “*Couple de serrage des BBAD*”.

Se tuvieron en cuenta 30 perímetros diferentes del vehículo, para garantizar al menos 15 observaciones en cada uno se tendrían que tomar 450, sin embargo, hay algunas categorías con mayor número de planos, por lo que la muestra tomada fue de  $n = 473$ , con la que se realizó manualmente la clasificación con el propósito de construir un conjunto de entrenamiento que permita entrenar un algoritmo que realice el trabajo con base en las palabras utilizadas por los validadores.

Gran parte de la inspiración de afrontar este reto es la científica de datos Julia Silge, autora del texto “*Text Mining with R: A Tidy Approach*”, (Silge & Robinson, 2016) quien hace no mucho escribió un artículo titulado “*text classification with tidy data principles*” (Silge, 2018) donde se toman dos libros: “orgullo y prejuicio” y “la guerra de los mundos”. Julia implementó un sólido modelo de clasificación binomial que, con base en las palabras de los párrafos, asocia el párrafo a uno de los libros, obteniendo la siguiente matriz de confusión.

##	Truth		
## Prediction	Pride and Prejudice	The War of the Worlds	
## Pride and Prejudice	2508	180	
## The War of the Worlds	122	1194	

Figura 16. Matriz de confusión del clasificador de libros de Julia Silge.

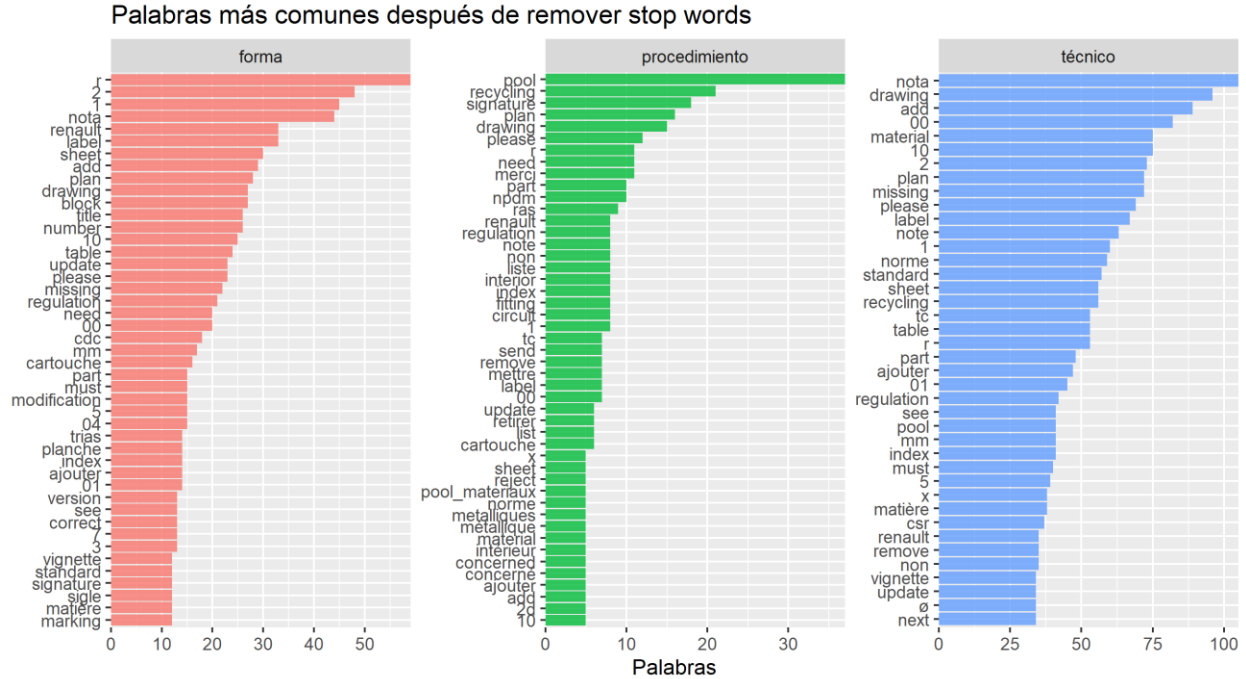
El problema de los motivos de rechazo en Renault es multinomial en lugar de binomial, ya que existen 3 categorías posibles para predecir, gran parte de las matemáticas asociadas cambian y el problema adquiere una complejidad mayor, que es el valor agregado que presenta esta sección del trabajo.

El objetivo inicial consiste en asociar las palabras individuales a las causas de rechazo. Luego de algunas transformaciones de datos, se llega a la siguiente estructura:

Tabla 6. Estructura de datos “tidy” para cada plano por palabra.

rowname	Name/Id	ELTDEC_numero	PIE_designation	causa_rechazo	word
2366	767421670R	11	RENFORT INF AR COTE CAISSE D	forma	to
2366	767421670R	11	RENFORT INF AR COTE CAISSE D	forma	be
2366	767421670R	11	RENFORT INF AR COTE CAISSE D	forma	corrected
2366	767421670R	11	RENFORT INF AR COTE CAISSE D	forma	in
2366	767421670R	11	RENFORT INF AR COTE CAISSE D	forma	drawing
2091	625124642R	11	TRAVERSE SUP FACE AV PARTIE CTL	procedimiento	validation
2091	625124642R	11	TRAVERSE SUP FACE AV PARTIE CTL	procedimiento	le
2091	625124642R	11	TRAVERSE SUP FACE AV PARTIE CTL	procedimiento	plan
2091	625124642R	11	TRAVERSE SUP FACE AV PARTIE CTL	procedimiento	sera
2091	625124642R	11	TRAVERSE SUP FACE AV PARTIE CTL	procedimiento	en

Si se cuentan las palabras más comunes para las categorías y se remueven las conjunciones (*stop words*) en inglés, francés, español y portugués, no es complejo obtener la siguiente figura.



Con estas palabras ya sería posible intentar “predecir” cuál es la categoría de un comentario con un modelo rudimentario que tenga en cuenta el porcentaje de palabras más repetidas en una categoría, pero ocurre una dificultad y es que varias palabras como “nota” que se repiten en todas las categorías y que en realidad no son un buen predictor de ésta. El objetivo es construir un modelo de clasificación con variables  $x_1, \dots, x_n$  que representan las palabras asociadas y predecir la probabilidad asociada de cada palabra a la categoría.

Este problema ha sido previamente solucionado para el caso general, (Yuan, Ho, & Lin, 2012) (Friedman, Hastie, & Tibshirani, 2010), quienes crearon el paquete *glmnet* (Friedman, Hastie, & Tibshirani, 2009) en múltiples software estadísticos utilizando rutinas en FORTRAN para resolverlo.

Se utilizará una variante híbrida de regularización denominada *Elastic-net*, que combina las fortalezas de dos métodos de penalización conocidos: LASSO (*Least Absolute Shrinkage and Selection Operator*) y Ridge regression. el cual se presenta adaptado a la situación presentada con 3 categorías.

Sea  $y = \Pr(G = k|X = x)$  la variable respuesta con 3 niveles  $G = \{1, 2, 3\}$  (1 – procedimiento, 2 – técnico, 3 – forma), modelamos la siguiente función de regresión logística multinomial:

$$\Pr(G = k|X = x) = \frac{e^{\beta_{0k} + \beta_k^T x}}{\sum_{\ell=1}^K e^{\beta_{0\ell} + \beta_\ell^T x}}. \quad (\text{Eq. 3})$$

Sea  $Y$  una matriz indicadora de las variables respuestas  $N \times K$  con elementos  $y_{il} = I(g_i = l)$ , entonces la función de log-verosimilitud que resuelve el paquete *glmnet* es:



$$\ell(\{\beta_{0k}, \beta_k\}_1^K) = - \left[ \frac{1}{N} \sum_{i=1}^N \left( \sum_{k=1}^K y_{ik} (\beta_{0k} + x_i^T \beta_k) - \log \left( \sum_{k=1}^K e^{\beta_{0k} + x_i^T \beta_k} \right) \right) \right] + \lambda \left[ (1 - \alpha) \|\beta\|_F^2 / 2 + \alpha \sum_{j=1}^p \|\beta_j\|_q \right] \quad (\text{Eq. 4})$$

En la ecuación 4,  $\beta$  es una matriz  $p \times K$  de coeficientes.  $\beta_k$  Se refiere a la  $k$ -ésima columna (de salida en la categoría  $k$ ) y  $\beta_j$  a la  $j$ -ésima fila (vector de  $K$  coeficientes para la variable  $j$ ). El último término  $\|\beta_j\|_q$  dónde existen dos posibilidades para  $q$ :  $q \in \{1, 2\}$ . Cuando  $q = 1$  ocurre una penalización LASSO en cada uno de los parámetros; cuando  $q = 2$  ocurre una penalización grupal en todos los coeficientes  $K$  para un grupo de variables, haciéndolas todas cero o no cero de manera conjunta.

Luego del planteamiento formal del problema, el siguiente reto es llevar la información al formato requerido por el modelo de regresión. Inicialmente se divide la muestra en dos conjuntos, uno de entrenamiento que corresponde al 75% de la información y el 25% restante, que permanecerá desconocido por el modelo para probar sus resultados de clasificación. Se realizan algunas transformaciones y se construye la matriz  $\beta$  de coeficientes, que tiene dimensiones  $347 \times 657$  ( $p \times K$ ) y se presenta una versión acotada de ésta en la figura 17.

	1	2	add	block	drawing	functional	index	number	part	renault	supplier	title	00	01	510	efforts	inutile	malgré	non
1001	1	1	2	2	1	1	1	1	1	2	1	2	0	0	0	0	0	0	0
1009	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
1012	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
102	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1029	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
103	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1031	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1033	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1042	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1053	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1058	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1060	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1064	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1067	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1068	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1070	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1071	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1085	0	0	0	0	1	0	0	0	0	0	0	0	2	2	2	1	1	1	2
1086	0	0	1	0	1	0	0	0	0	0	0	0	2	2	2	1	1	1	2
1087	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Figura 17. Veinte primeros valores de la matriz  $\beta$  de dimensiones ( $p = 347$ )  $\times$  ( $K = 657$ ).

Nótese que en esta matriz hay un índice único para cada comentario en las filas  $p$  y las palabras están en las columnas  $K$ , es decir que los valores corresponden al número de veces que la palabra  $K$  aparece en el comentario  $p$ . Por otra parte, la variable respuesta  $y$  es un vector con las categorías asociadas a cada comentario  $p$ , como se muestra a continuación transpuesto.

1 2 3 1 3 1 2 1 1 1 3 1 2 2 2 2 2 2 2 2

Figura 18. Veinte primeras respuestas de la variable  $y$  de dimensiones  $p \times 1$ .

Con los requisitos de entrada se procede a realizar la estimación del modelo haciendo uso del paquete *glmnet* en R (R Core Team, 2019). Se obtiene la siguiente gráfica de deviance (no confundir con la desviación) multinomial con respecto a log-lambda, deviance es el negativo de dos veces la log-verosimilitud maximizada, entre menor la deviance, mejor el ajuste del modelo. (James, Witten, Hastie, & Tibshirani, 2013)

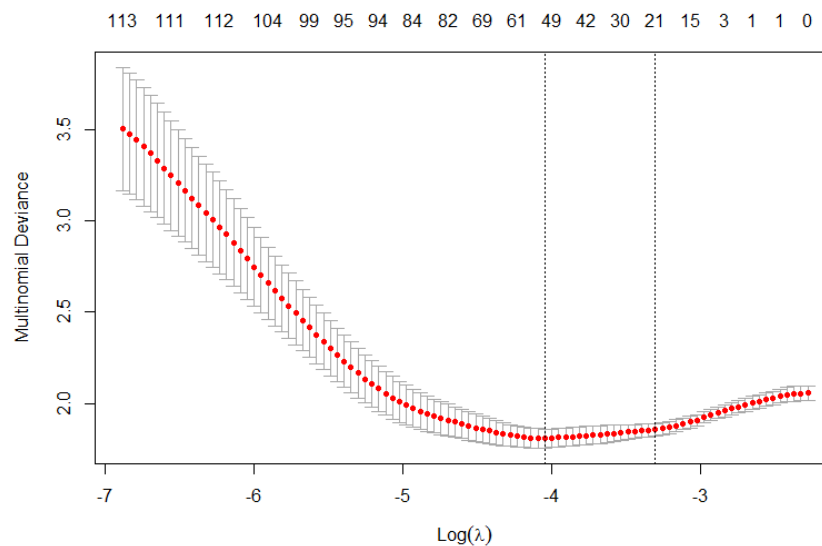


Figura 19. Multinomial deviance del ajuste vs log-lambda.

En este caso, se selecciona el modelo para el que se minimiza la *deviance* y se tiene la menor variación en *deviance*, en este caso, se escoge el modelo con  $\lambda = 0.02649253$ , tal que  $\log(\lambda) = -3.165724$ . Con estos parámetros, los coeficientes reales asociados a las categorías serían:

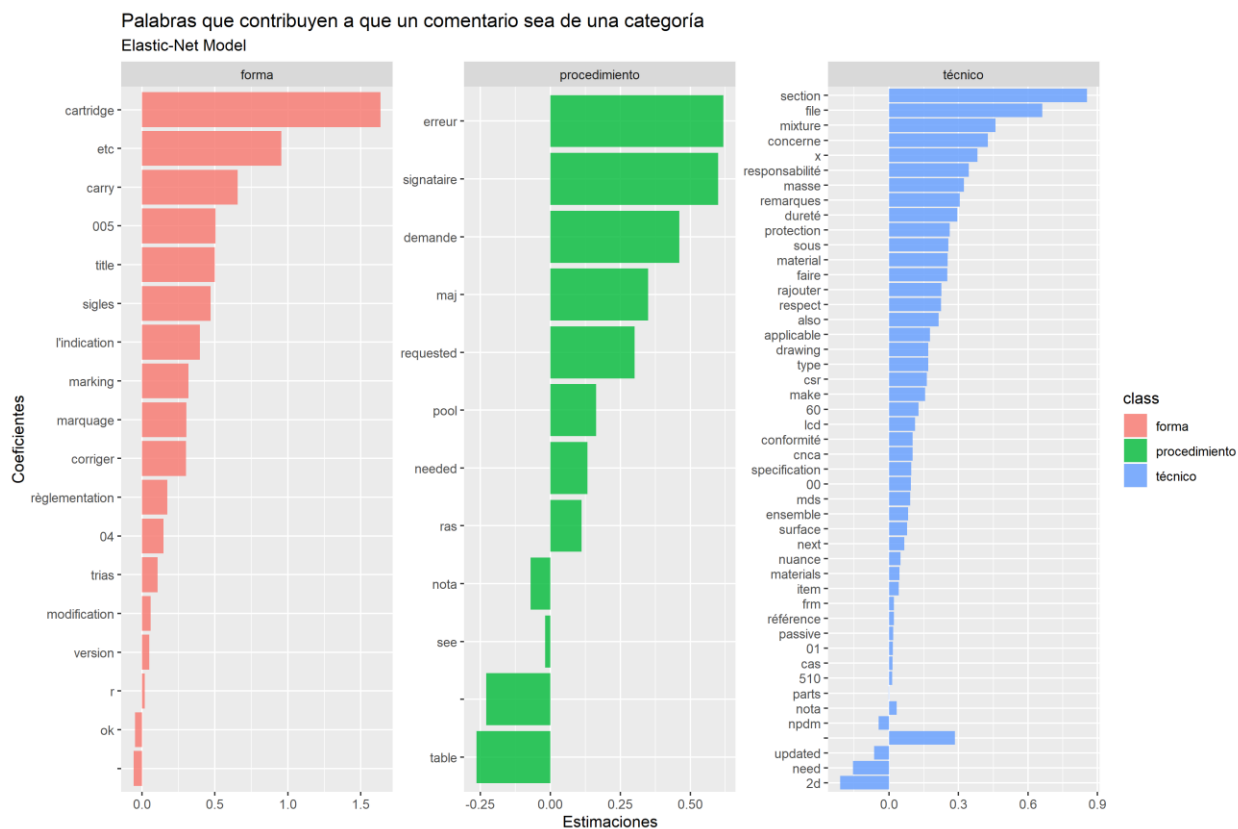


Figura 20. Palabras (coeficientes del modelo) y pesos asociados a la pertenencia a una categoría.

Más allá de funcionar como coeficientes para la predicción de las categorías, la figura anterior permite hacer inferencias sobre cuáles son los motivos de rechazo en cada categoría. Un ingeniero o conceptor con experiencia en la organización sabría llegar a conclusiones como: “Si la palabra con mayor peso es *Cartridge*, es porque en muchas ocasiones los comentarios de forma se deben a que los cajetines del plano están mal ubicados”, y de la misma manera para las demás categorías. Por ende, estos coeficientes son un insumo real que permite tomar planes de acción para prevenir el rechazo de los planos.

Posteriormente se aplican los coeficientes a los comentarios reales, tanto en el conjunto de prueba como con un modelo sobre ajustado utilizando toda la información de prueba, obteniendo los siguientes resultados:

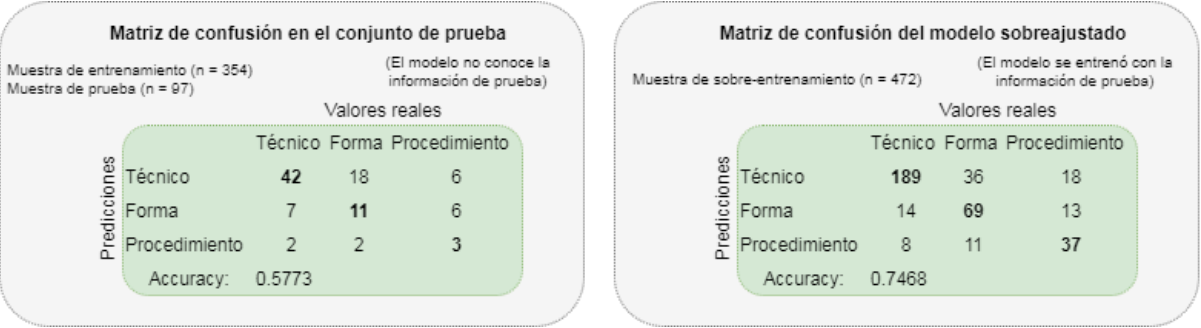
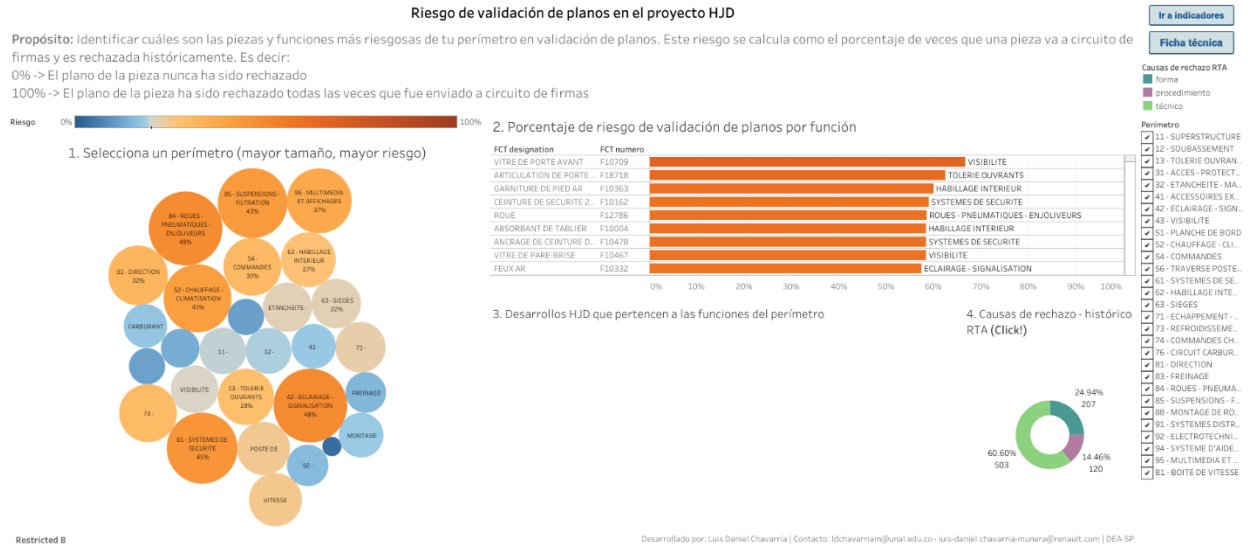


Figura 21. Matrices de confusión en el conjunto de prueba y del modelo sobreajustado.

Este último modelo es el que utiliza con el propósito de clasificar los más de 22.000 comentarios en diferentes categorías en el RTA dado a que cuenta con un mayor tamaño de muestra.

6.8. Tablero de BI para la gestión del riesgo de validación y paso del modelo a operación

Con los resultados del modelo de clasificación y la estimación del riesgo para cada una de las piezas realizada, se utiliza como entrada las referencias en desarrollo del vehículo proyecto en curso con el propósito de ver únicamente información relevante a éste. A continuación, el tablero de control para la gestión del riesgo de validación de planos.



**Figura 22. Tablero principal para la gestión del riesgo de validación de planos.**

El tablero de validación del riesgo de planos cuenta con 5 momentos,

1. Se presenta el porcentaje de riesgo de validación asociado al tamaño de los círculos y al color, donde azul representa un menor riesgo y rojo-naranja un riesgo mayor; el cambio de tonalidad se da alrededor del 20%.
2. Se presentan los porcentajes de riesgo de validación para las funciones, es decir, una subdivisión más específica que el perímetro donde se llega al propósito de la pieza.
3. Se presentan los desarrollos actuales pertenecientes a la selección realizada.
4. Se presenta el resultado de la clasificación del modelo de *Machine Learning* con respecto a las causas de rechazo.
5. Al hacer clic en las causas de rechazo, se lleva al usuario a los comentarios de los validadores que ocasionaron los rechazos.

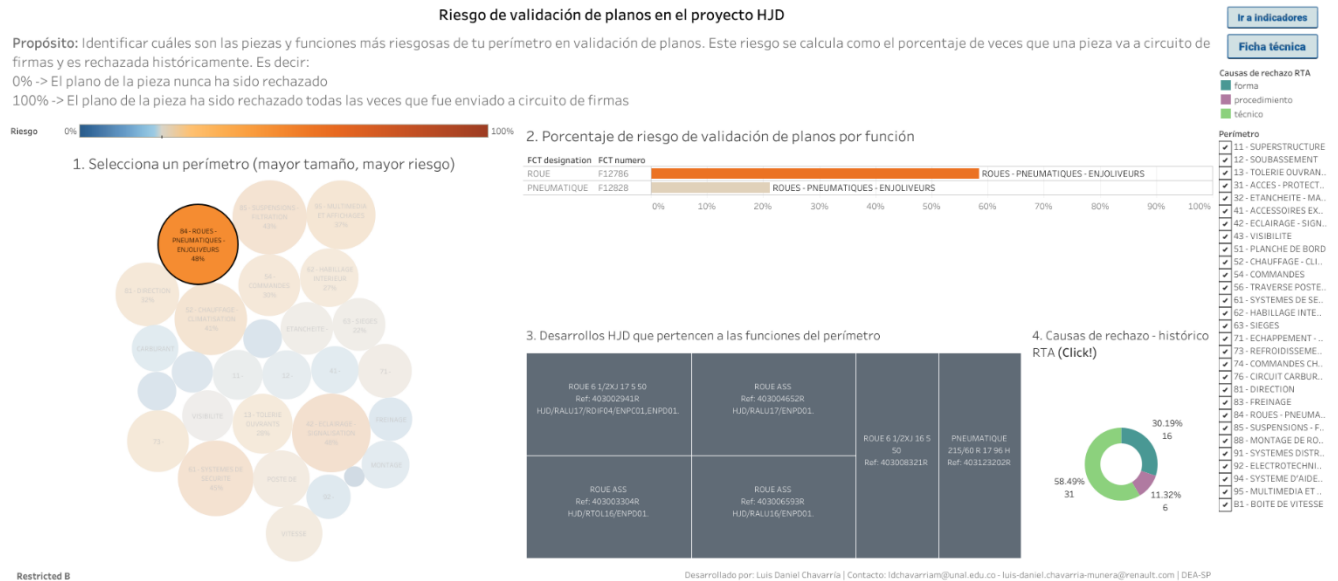


Figura 23. Utilización del tablero de gestión de riesgo para el perímetro de ruedas, neumáticos y embellecedores.

En la figura 23 se selecciona el perímetro 84, donde se encuentran las ruedas y neumáticos del vehículo, en el apartado 2, el tablero muestra que las ruedas son considerablemente más riesgosas en validación que los neumáticos, con un 59% de probabilidad de ser rechazadas al enviarlas a circuito de firmas, con base en los registros históricos. En el apartado 3 se presentan las piezas en desarrollo asociadas a este perímetro y el algoritmo de clasificación concluye que en el apartado 4 que el 58.49% de las veces, los rechazos de planos en este perímetro ocurren por motivos técnicos; al hacer clic en este momento, se despliega el siguiente tablero con los comentarios.

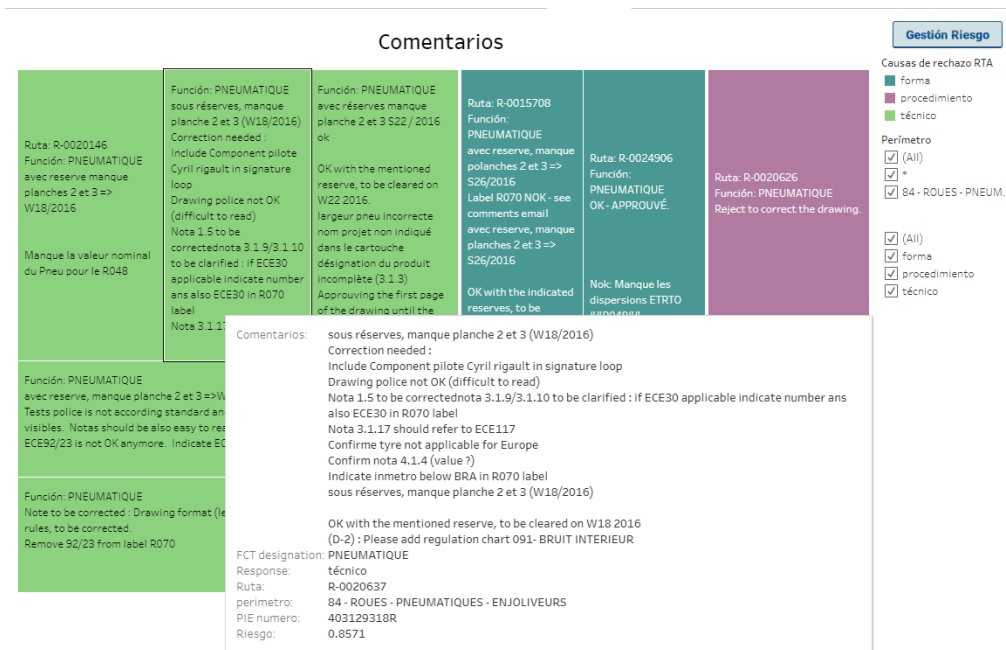


Figura 24. Comentarios asociados a la selección realizada en el perímetro de ruedas y neumáticos.

Al seleccionar los comentarios de rechazo para neumáticos, el tablero presenta la información asociada a éstos, como se ve en la figura 24, al pasar el ratón sobre cada uno de éstos se puede leer con mayor detalle los comentarios, además, es posible seleccionar un conjunto particular de elementos en los tableros y generar extracciones en formato .csv para importarlos rápidamente para un análisis más detallado, como se muestra en la figura 25.

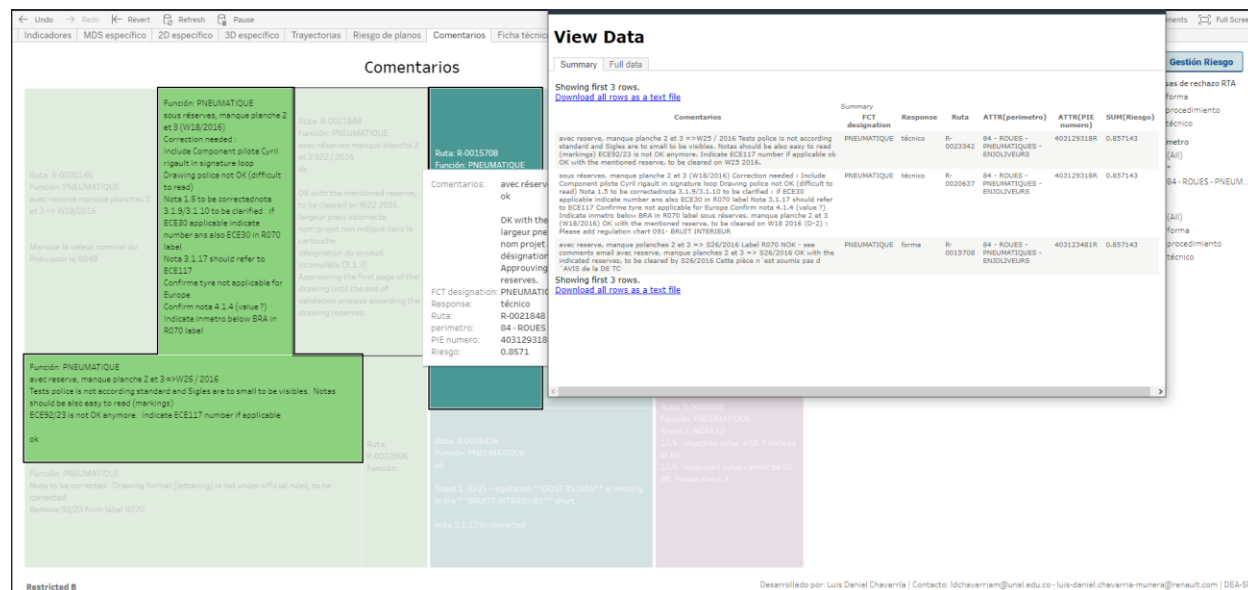


Figura 25. Extracción de información directamente desde la visualización interactiva.

De esta manera, es posible construir planes de acción, por ejemplo, para prevenir a un proveedor de una función riesgosa, retroalimentándolo con información de rechazos anteriores en algún tipo particular de pieza y permitiéndole aprender de experiencias pasadas de otros proveedores.

## 6.9. Resultados y conclusiones

Con respecto a los indicadores presentados en la ecuación 2 y 3, donde se llegó a un costo estimado de COP \$64.000.000 por mes, la solución propuesta presenta un ahorro del 100% de éste, con beneficios adicionales no cuantificados en agilidad en la toma de decisiones y medidas de prevención que trae esta nueva herramienta.

En ejercicios de validación, se identificó que es posible utilizar la herramienta de gestión del riesgo para crear planes de acción estratificados para cada proveedor dependiendo de los motivos de rechazo; en el 2016 Renault América empleó más de 2 millones de dólares en formación de los proveedores con el propósito de tecnificar sus procesos y disminuir el porcentaje de rechazos en los planos.

Esto dejó de hacerse debido a la ineffectividad de un programa de educación generalizado, esta herramienta permite establecer formaciones diferenciales enfocados en las necesidades de cada sector, como ejemplo, las medidas de control para un fabricante de vidrios son totalmente distintas a las que requiere uno de neumáticos.

Se presentó el proyecto ante la junta directiva de Ingeniería en Renault América, liderada por el directivo Antonio Fleischmann, donde se aprobó la implementación de la herramienta en todo el RTA mediante la integración al software “IPM”, para la gestión de los vehículos proyecto. El impacto generado llevó a que se desarrollara documentación adicional, como manuales de actualización y videos de capacitación en la creación de tableros para futuros proyectos. A marzo de 2020, la herramienta está vigente con seguimiento semanal del estado de los documentos técnicos utilizando metodologías ágiles en reuniones de 15 minutos para cada división del vehículo.

Existen múltiples maneras de afrontar grandes preguntas dentro de las empresas; sin embargo, en muchas ocasiones se ignora el potencial que se tiene para dar respuesta a éstas desde los datos actuales de la organización. Salir de la zona de confort, explorar nuevas soluciones y bajar a la realidad empresarial terminología técnica como “*Machine Learning*” “*Artificial Intelligence*” ó “*Deep learning*” en soluciones de impacto, interpretables y funcionales, debe ser el enfoque de cualquier profesional que desee trabajar desde el ámbito de los datos.

## 6.10. Retroalimentaciones al programa de Ingeniería Industrial

- Nuestro programa de Ingeniería Industrial tiene un fuerte componente estadístico, con al menos 6 asignaturas fundamentada en ésta (Estadísticas 1 a 3, Control estadístico de la calidad, Simulación de sistemas e Investigación de operaciones 2) y algunas basadas en la programación. Sería de gran valor crear una línea de especialización clara y tangible que complemente esta rama y le entregue a la sociedad científicos de datos con competencias industriales, de programación y de toma de decisiones que puedan llevar a cabo proyectos de analítica de impacto.
- Presentarse en entrevistas, venderse como profesional y tener la confianza propia para afrontar el mundo laboral son habilidades que no aprendí muy bien en la Universidad Nacional. Nuestro programa académico está al nivel de universidades internacionales prestigiosas, sin embargo, es mi percepción que profesionales de universidades privadas tienen mejores posibilidades porque su Universidad se preocupó más por conseguir alianzas estratégicas con el sector laboral, involucrando inclusive a la empresa en lo académico. Expertos de SOFASA por ejemplo dan clases formales en diseño en la Universidad EAFIT y los llevan a la práctica desde el inicio.

## 7. Referencias

- Friedman, J., Hastie, T., & Tibshirani, R. (2009). glmnet: Lasso and elastic-net regularized generalized linear models. *R Package Version*.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v033.i01>
- Hofmann, P., Samp, C., & Urbach, N. (2019). Robotic process automation. *Electronic Markets*. <https://doi.org/10.1007/s12525-019-00365-8>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. <https://doi.org/10.1007/978-1-4614-7138-7>

- Lee, I., & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2), 157–170. <https://doi.org/10.1016/j.bushor.2019.10.005>
- Lohr, S. L. (2019). Stratified Sampling. In *Sampling* (pp. 73–116). <https://doi.org/10.1201/9780429296284-3>
- R Core Team. (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. Retrieved from <https://www.r-project.org/>
- Silge, J. (2018). *Text classification with tidy data principles*. Retrieved from <https://juliasilge.com/blog/tidy-text-classification/>
- Silge, J., & Robinson, D. (2016). Text Mining with R. *The Journal of Open Source Software*. <https://doi.org/10.21105/joss.00037>
- UiPath. (2016). UiPath Robotics Process Automation. Retrieved from 2016 website: <https://www.uipath.com/rpa/robotic-process-automation>
- Yuan, G. X., Ho, C. H., & Lin, C. J. (2012). An improved GLMNET for l1-regularized logistic regression. *Journal of Machine Learning Research*.