

Technological incident classification model from a machine learning approach in insurance services

Paola Andrea Gómez-Jaramillo^a, Favián González-Echavarría^b & Jorge Iván Pérez-Rave^c

^a Universidad de Antioquia, Medellín, Colombia: paolaa.gomez@udea.edu.co

^b Universidad de Antioquia, Departamento de Ingeniería Industrial, Medellín, Colombia: favian.gonzalez@udea.edu.co

^c IDINNOV research group, IDINNOV S.A.S, Medellín, Colombia: investigacion@idininnov.com

Received: December 13th, 2021. Received in revised form: May 20th, 2022. Accepted: May 26th, 2022.

Abstract

Managing technological incidents in insurance companies requires a correct and timely assignment of these to the problem-solving teams. Classifying these incidents by humans demands time and knowledge and is frequently executed erroneously. This paper addresses this classification problem from a machine learning approach. The performance of five supervised learning methods (logistic regression, classification trees, random forest, discriminant linear analysis, support vector machines) is compared in three scenarios of inclusion of predictors: structured, texts, and both. The use of unstructured variables considerably improves the accuracy of the models (e.g., Random Forest, validation sample: 0.709 using structured data; 0.881 using text data). Moreover, considering the practical implications of the human correct classification rate (66%) vs. machine (88%, Random Forest, SVM, or linear regression), the machine favors resource-saving in the organization. This article is a successful case of machine learning in the insurance industry.

Keywords: classification of incidents; technology incidents; insurance; machine learning.

Modelo de clasificación de incidentes tecnológicos desde un enfoque de aprendizaje automático en servicios de seguros

Resumen

La gestión de incidentes tecnológicos en las compañías de seguros requiere una asignación correcta y oportuna de estos a los equipos de resolución de problemas. La clasificación de tales incidentes por humanos demanda tiempo y conocimiento y, con frecuencia, se ejecuta de manera errónea. Este documento aborda este problema de clasificación desde un enfoque de aprendizaje automático. Se compara el desempeño de cinco métodos de aprendizaje supervisado (regresión logística, árboles de clasificación, bosque aleatorio, análisis lineal discriminante y máquinas de vectores de apoyo) en tres escenarios de inclusión de predictores: estructurado, textos y ambos. El uso de variables no estructuradas mejora considerablemente la exactitud de los modelos (ej., Random Forest, muestra de validación: 0,709 con datos estructurados; 0,881 con datos de texto). Además, considerando las implicaciones prácticas de la tasa de clasificación humana correcta (66%) frente a la máquina (88%, Random Forest, SVM o regresión logística), la máquina favorece el ahorro de recursos en la organización. Este artículo es un caso exitoso del aprendizaje automático en la industria de seguros.

Palabras clave: clasificación de incidentes; incidentes tecnológicos; seguros; aprendizaje automático.

1. Introducción

El incremento en el uso de la tecnología en los procesos organizativos ha venido generando un crecimiento en la demanda de los servicios de las áreas de infraestructura

tecnológica (TI), específicamente los relacionados con la gestión de incidentes. De acuerdo con la Biblioteca de Infraestructura de Tecnología de la Información (ITIL), tal proceso busca resolver, rápida y eficazmente los problemas que causen interrupción en el servicio [1].

How to cite: Gómez-Jaramillo, P.A., González-Echavarría, F. and Pérez-Rave, J.I., Modelo de clasificación de incidentes tecnológicos desde un enfoque de aprendizaje automático en servicios de seguros., DYNA, 89(221), pp. 161-167, April - June, 2022.

Uno de los tópicos de investigación con mayor popularidad al respecto es la clasificación de incidentes, la cual busca asignarlos de forma correcta a los equipos técnicos encargados de dar solución a cada incidente [2]. La clasificación adecuada de un incidente es esencial para dirigirlo rápidamente al equipo correcto y, así, evitar reprocesos por devoluciones y minimizar su tiempo de solución [1]. De hecho, una mala clasificación genera ineficiencias en el recurso humano, aumenta el tiempo de resolución y disminuye la satisfacción del usuario final [1].

A pesar de la importancia de dicho proceso de clasificación, su desempeño en las organizaciones amerita intervenciones de mejora, debido, entre otros aspectos, a la alta dependencia de actividades humanas que derivan en sesgos interpretativos del agente clasificador [2]. Ello puede corroborarse en estudios recientes, en ámbitos como los de servicios de tecnología [3] y aviación [4].

Uno de los principales recursos hoy en día para reducir la subjetividad humana en la clasificación de incidentes corresponde a los métodos de aprendizaje automático (machine learning, en inglés). A pesar de ello, su desempeño no debe asumirse desde un enfoque universalista, ya que no existe un único método que siempre conduzca al mejor resultado en todo tipo de contexto. Tal desempeño depende de variedad de factores, entre ellos: el conjunto y la naturaleza de los datos, las tareas de preprocesamiento (ej., en la descripción narrativa del incidente) y los parámetros del algoritmo [1]. De ahí la sugerencia de estudios como [4,5] de entender las particularidades de cada escenario y realizar una comparativa entre varios métodos y tipos de características (predictores), en busca de la alternativa más adecuada a cada escenario de estudio.

Esta necesidad es aún mayor, al considerar el escenario de compañías de seguros, cuya supervivencia y éxito son altamente sensibles a riesgos reputacionales derivados de la insatisfacción del usuario, a causa de los reprocesos y el incremento de tiempos/costos que traen consigo la clasificación de incidentes tecnológicos [7]. Adicionalmente, la literatura científica poco ha discutido sobre el abordaje de este problema en escenarios de seguros, desde un enfoque de aprendizaje automático. La mayoría de las aplicaciones al respecto se han realizado en ámbitos de servicios de compañías de aviación [4], tecnología [6] y salud [7]. Este hecho limita la comprensión acerca de qué método presenta mejores niveles de desempeño en el ámbito de pólizas de seguros. Asimismo, se hace necesario validar si en tal escenario las variables de texto, alusivas a la descripción del incidente tecnológico, son suficientes para entrenar/validar modelos altamente precisos, como lo reportó [5]; o, por el contrario, valdría la pena combinarlas con datos estructurados, como sucedió en [2].

Con miras a una mejor claridad del tema, este artículo analiza datos de una compañía de seguros en Colombia desde un enfoque de aprendizaje automático para dar respuestas a dos interrogantes:

P1 ¿Qué método, entre árboles de clasificación, máquinas de vectores de soporte, random forest, regresión logística y análisis lineal discriminante, presenta las mejores tasas de clasificación de incidentes para la suscripción de pólizas de seguros?

P2 ¿Qué implicaciones prácticas pueden atribuírsele al modelo con el mejor desempeño en el escenario de seguros observado, considerando la clasificación actual del agente

humano?

Mediante el abordaje de las preguntas expuestas, este artículo contribuye con lo siguiente:

- Provee información sobre las particularidades del escenario de clasificación de incidentes en el ámbito de seguros, desde un enfoque de aprendizaje automático. Por tanto, sirve de guía para investigaciones futuras al respecto, en lo que concierne al método de aprendizaje y la naturaleza de los datos que resultaron más útiles en tal contexto.
- Asiste la toma de decisiones en el ámbito estudiado, respecto a cómo mejorar el desempeño del proceso de clasificación de incidentes para la suscripción de pólizas de seguros, a través de métodos intensivos en computación. De hecho, en el escenario observado, cerca del 34% de los incidentes tecnológicos reportados por sus expedidores de pólizas son clasificados erróneamente por el agente humano, lo cual deriva en ineficiencias internas e insatisfacción del usuario final, debido a reprocesos y largos tiempos de espera (cerca de 12 días por cada incidente mal clasificado).

El artículo cuenta con las siguientes secciones: La sección 1 expuso la introducción, en la cual se justificó el estudio. La sección 2 aporta la revisión de literatura llevada a cabo. La sección 3 describe los materiales y métodos orientados al entrenamiento de cinco modelos de aprendizaje automático, así como a la validación y comparación con el agente humano. La sección 4 aporta los resultados obtenidos para las dos preguntas expuestas, considerando tres escenarios de análisis (datos estructurados, textos y ambos). La sección 5 provee las conclusiones y el trabajo futuro.

2. Revisión de literatura

Esta sección se apoyó en lineamientos de alfabetización informacional, sin ser alcance del presente manuscrito realizar una revisión sistemática. Inicialmente, la búsqueda se realizó por medio de las bases de datos Scopus y Web de la Ciencia, empleando el término “aprendizaje automático” en conjunto con expresiones alusivas a dominios de incidentes tecnológicos bajo mesas de ayuda. Tales consultas arrojaron pocos trabajos al respecto (ej., en Scopus nueve documentos fueron reportados usando TITLE ("machine learning") AND TITLE ("help desk" OR "service desk systems" OR "IT incidents" OR "IT service delivery" OR "ticket classification" OR "incident ticket" OR "IT Tickets" OR "IT Service"). Esto llevó a extender la consulta a la sección de referencias citadas por tales trabajos, así como a explorar otras fuentes por medio de Google Académico. A medida que se fueron recuperando documentos al respecto, se verificaba el título/resumen y luego el texto completo, con el fin de corroborar el tema y dominio de acción, así como el hecho de que correspondieran a trabajos empíricos. Finalmente, considerando la afinidad (desde la óptica de los autores) entre los propósitos del presente estudio y los documentos localizados, catorce trabajos fueron tomados como base para el desarrollo de la presente sección. A continuación, se describen los principales elementos de soporte.

La Tabla 1 muestra la preponderancia de los servicios de infraestructura de TI para la clasificación de incidentes usando un enfoque de aprendizaje automático, con una participación en 12 (86%) de los 14 trabajos de referencia.

Tabla 1.

Dominio y objetivo principal de los estudios revisados

Dominios	Objetivos	Estudios
Servicio de infraestructura TI.	Automatizar la tarea de clasificación de incidentes, a través del aprendizaje supervisado, con el fin de mejorar el desempeño del sistema de gestión de incidentes tecnológicos.	[1-3,5,6,8-14]
Servicio de salud.	Automatizar la tarea de clasificación de condiciones de salud e historia clínica de pacientes con VIH para indicar qué tratamiento se ajusta mejor a sus necesidades.	[7]
Mesas de ayuda en la industria de aviación	Desarrollar un prototipo de un sistema que identifica automáticamente la información relativa a incidentes, a partir de textos disponibles en la web, con el fin de mejorar el proceso de análisis de riesgos de una compañía de aviación.	[4]

Fuente: Los autores

Tabla 2.

Atributo de entrada

Atributo	Estudio
Descripción del incidente	[1-4,6-8,11-14]
Título (o asunto)	[1,5,9,11,12]
Equipo de solución	[6,11]
Texto de resolución	[2,10]
Todos los atributos (incluido el equipo de solución, la fecha, la criticidad, la prioridad y el texto de solución o cierre)	[10]

Fuente: Los autores

Asimismo, emerge la necesidad de comprender el tema en otros dominios, encontrándose, en esta oportunidad, experiencias en ámbitos de salud y aviación.

Por otro lado, la Tabla 2 provee una síntesis de los atributos que los trabajos revisados emplearon para entrenar los modelos de clasificación de incidentes.

La Tabla 2 muestra que el 79% de los estudios ha utilizado como atributo de entrada la descripción (o cuerpo del incidente). En segundo lugar, se encuentra el título (o asunto) del incidente, con participación en el 36% de los estudios revisados. Otros atributos, como el equipo de solución, la fecha, la criticidad, la prioridad y el texto de solución o cierre, son utilizados en menor medida por estudios que se interesan no solo por la clasificación del incidente sino también en la predicción de fallas de equipos. Cabe resaltar que solo uno de los estudios revisados utilizó todos los atributos del incidente para la clasificación automática.

En general, los mejores resultados reportados por los trabajos previos han sido atribuidos al uso del título (o asunto) y la descripción del incidente. Es decir, datos en formato texto, los cuales son complejos de abordar dada la diversidad de agentes, conocimientos/capacidades, estilos de escritura y conocimientos previos del productor del texto. De ahí la importancia de considerar recursos de minería de textos dentro de los modelos de aprendizaje automático [4].

Tabla 3.

Variable por predecir

Variable	Estudio
Tipo de problema	[2-4,6-8,10,12-14]
Área de solución	[1,9,11]
Nombre servidor	[5,8]
Prioridad	[11]

Fuente: Los autores

En la Tabla 3 se describe la variable objeto de clasificación, reportada por los trabajos revisados. En dicha tabla se observa que la variable respuesta de mayor interés por los trabajos de referencia fue el tipo de problema (ej., apagado inesperado, reinicio, hardware defectuoso), con una participación del 71%, seguido del área de solución (ej., equipo de bases de datos, administradores de red y servidores), con un 21%. La variable respuesta identificada como servidor que presentó fallos representa el 14%.

Con respecto a las etapas empleadas para abordar la clasificación de incidentes, los trabajos de referencia suelen coincidir en el siguiente derrotero: recolección de datos, preprocesamiento de los datos, construcción del vector de características, selección del método de clasificación, entrenamiento del modelo, clasificación y validación comparativa entre métodos seleccionados. Al considerar datos de textos, es fundamental emplear tareas como “tokenización”, remoción de palabras vacías (ej., stopwords), de expresiones poco relevantes (ej., con uso de la frecuencia de términos – frecuencia inversa de documentos TF-IDF) y, en casos, de operaciones de lematización y/o reducción de expresiones a su raíz [7]. Es así como [9] señalan que un modelo de aprendizaje automático adecuado para la clasificación del servicio de incidentes depende de los datos de entrenamiento, del preprocesamiento del texto, de la vectorización de características, del algoritmo de aprendizaje automático y de los parámetros del algoritmo.

Con respecto a la fase de construcción del vector de características, [6] señalan que, para implementar los modelos de clasificación, los datos que se han de utilizar en la fase de entrenamiento deben ser convertidos a representación vectorial. Para la construcción del vector de características, [12] sugiere aplicar la ponderación booleana o la frecuencia de palabras.

Para entrenar los modelos de clasificación, autores como [6] recomiendan dividir los datos objeto de estudio, así: 80% para entrenar y 20% para la fase de validación. Sin embargo, [2] sugiere conservar 70% para entrenar y 30% para validar. Más aún, [7] indicaron que el 90% debía ser para entrenamiento y el 10% para validar. Con relación a la validación de los modelos, además del uso del 10% - 30% de los datos para tal fin, se encuentra otra alternativa, alusiva a la validación cruzada (véase, por ejemplo, [4 y 9])

En la Tabla 4 se presentan los métodos de clasificación

empleados por los trabajos previos.

En la Tabla 4 se muestra que los métodos de aprendizaje automático más utilizados para la clasificación de incidentes fueron Naive Bayes y Máquinas de Soporte Vectorial, ambos con un 50% de participación en los estudios de referencia. Seguido, están regresión logística, K vecinos más cercanos y máquinas de aumento de gradiente (GBM), cada uno con uso

Tabla 4.

Métodos de aprendizaje automático empleados

Métodos	Estudios
Naive Bayes	[1,3,4,6,7,11,13]
Máquinas de soporte vectorial (SMV)	[1,3,7,9,10,11,13]
Regresión logística	[2,3,8]
K vecinos más cercanos (KNN)	[10,11,13]
Máquinas de aumento de gradiente (GBM)	[2,3,8]
Árboles de clasificación	[1,11]
Adaboost	[8]
Redes neuronales	[14]
FastText - Facebook AI Research	[12]
CRF (Campo aleatorio condicional)	[5]

Fuente: Los autores

en el 21% de los estudios. Árboles de clasificación participa en el 14% y otros métodos de menor uso, son random forest, adaboost, redes neuronales, CRF y FastTxt con un 7% cada uno.

3. Métodos

Con el fin de cubrir un amplio espectro de los métodos en mención, el presente estudio toma en consideración representantes de los grupos de alta (SVM), media (regresión logística) y baja participación (árboles de clasificación) en los trabajos previos sobre clasificación de incidentes tecnológicos (véase Tabla 4). Adicionalmente, para complementar la comparativa por llevar a cabo, se incluyen análisis lineal discriminante y random forest, los cuales han resultado útiles en otras experiencias de los autores (ej., precios inmobiliarios). Así, el presente trabajo considera un total de cinco métodos de aprendizaje automático, cuyos desempeños serán comparados en tres tipologías de escenarios: solo atributos estructurados, solo atributos de texto, y ambos tipos de atributos.

A continuación, se resumen los pasos seguidos para preparar y desplegar la comparativa entre los cinco métodos en consideración:

3.1 Reconocimiento del proceso

Esta etapa abarca un primer acercamiento con los analistas y directivos involucrados en el proceso de gestión de incidentes. Además, profundiza en el procedimiento actual de clasificación de incidentes desde la óptica del analista de seguros generales, con lo que también se contemplan posibles criterios que los llevan a tomar determinadas decisiones de clasificación.

3.2 Recolección y preparación de datos

A partir del software utilizado en la gestión de incidentes en la compañía aseguradora, se extrae un informe que corresponde a los incidentes relacionados con las fallas en los aplicativos. El rango de observación comprende nueve meses (desde julio de 2019 hasta abril de 2020). Durante este periodo de tiempo, se asignaron 2816 incidentes a los grupos solucionadores “0” y “1”, lo que se configura como la variable respuesta.

Adicionalmente, se construye el diccionario de variables que se configuran como relevantes de acuerdo con la literatura y la profundización en el contexto de la compañía aseguradora; estas han de ser la base para el diseño de los modelos de clasificación, a través de los cinco métodos de aprendizaje supervisado seleccionados. En la presente fase también se realizan actividades como homologación de datos y creación de variables adicionales, a partir de las existentes. Luego, se procede con la partición de la muestra, conservando el 70% de los datos para entrenamiento y el 30% restante para validación.

3.3 Entrenamiento y validación de los modelos

A través del paquete Caret [15,16], en R, se programan funciones específicas para que permitan entrenar la clasificación de los incidentes de interés, a través de los cinco métodos en comparación (SVM, random forest, regresión logística, árboles de clasificación y análisis lineal discriminante). Asimismo, se procede con el análisis de parámetros de refinamiento (ej., poda de los árboles de clasificación), hasta dejar a punto los modelos, lo cual es guiado por la observación del desempeño de cada uno en la muestra de entrenamiento y el uso de métricas de la matriz de confusión.

La validación de los modelos se ejecuta considerando principalmente la “exactitud”, derivada de los resultados obtenidos para el 30% de las instancias separadas para ello. Dicha métrica es obtenida a partir de los datos de la matriz de confusión, así: suma de la diagonal (verdaderos positivos y verdaderos negativos), dividida entre el total de casos.

La Fig. 1 resume el diseño de los modelos en consideración. Allí se muestran los siete predictores por incluir en el diseño de los modelos, seis de los cuales son de naturaleza estructurada y, el restante (“descripción”), es tipo texto. Estos predictores permiten examinar tres escenarios de entrenamiento/validación (solo variables estructuradas, solo texto, y ambos tipos de variables). [17] explica que para realizar la clasificación a partir de texto se suele comenzar con el análisis textual (o caracterización), proceso en el que

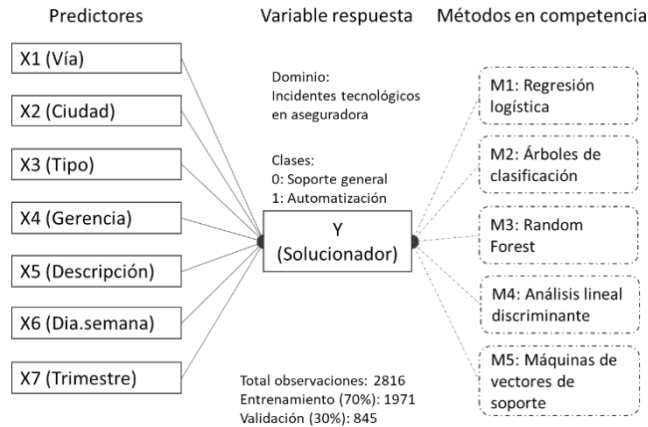


Figura 1. Resumen del diseño de los modelos y su comparación
Fuente: Los autores

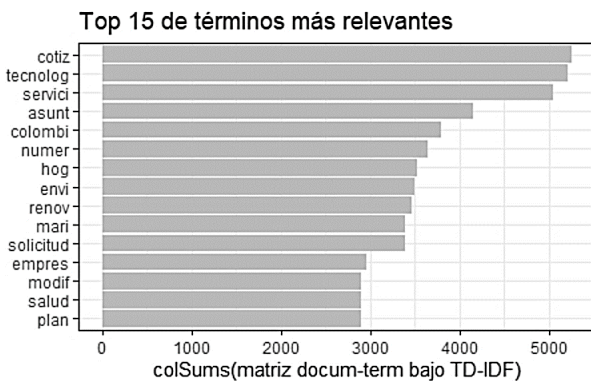


Figura 2. Top 10 de expresiones (previa limpieza de datos) más relevantes de la descripción del incidente
Fuente: Los autores

el documento se divide en características (palabras o frases) y se le aplican una serie de métodos como la normalización y la eliminación de palabras vacías, con el fin de generar el conjunto de características.

Respecto a la variable de texto, se optó por extraer los vectores de las 50 expresiones más relevantes (suma de TF-IDF) en las descripciones del incidente, previas operaciones de normalización a minúsculas, eliminación de números y caracteres especiales, tokenización, remoción de palabras no semánticas (stopwords), stemming (raíz de las palabras) y construcción de la matriz.

A modo de evidencia, en la Fig. 2 se presenta el diagrama de barras de las 15 expresiones más relevantes extraídas de la descripción del incidente, previa elaboración de la matriz de documentos (filas)-términos (columnas) con TF-IDF. De este modo, cada vector representa una variable predictora por examinar usando TF-IDF.

Con los modelos objeto de estudio, enmarcados en cinco métodos de aprendizaje automático, se busca clasificar el equipo de trabajo más adecuado para dar solución al incidente, el cual presenta categorías “1” (automatización) y “0” (soporte general).

4. Resultados

En la Tabla 5 se presentan los resultados consolidados de la exactitud de los modelos construidos bajo los cinco métodos en competencia, tanto para la muestra de entrenamiento como de validación, considerando los tres escenarios de análisis (atributos estructurados, textos y ambos).

En la Tabla 5 se observa que bajo el escenario 1 (solo variables estructuradas), el método que arroja el modelo con el mejor desempeño en los datos de entrenamiento y validación es random forest, con una clasificación correcta del 67.9% y 70.9% de las instancias, respectivamente.

Asimismo, en el escenario solo con variables de texto, se obtiene lo siguiente: Random forest presenta el mejor desempeño en la muestra de entrenamiento (99.5% de exactitud), lo cual refleja sobreajuste a los datos, dado que en la muestra de validación dicha exactitud se reduce a 88.1%. Sin embargo, este valor aún forma parte de los resultados más favorables en tal conjunto de datos de validación, compartiendo la supremacía con regresión logística (88.1%) y SVM (88.4%).

Ahora bien, al combinar los predictores en mención (escenario 3, véase Tabla 5), random forest sigue presentando la mayor exactitud en las instancias de entrenamiento (88.7%), pero a costa de sobreajuste que luego deriva en 77.5% de clasificaciones correctas en la muestra de validación. Sin embargo, sorprende que los árboles de clasificación, los cuales no consideran los beneficios del ensamble (como sí ocurre en random forest), muestran estabilidad en el desempeño (entrenamiento: 84.9%; validación: 85%). Además, dicho modelo (bajo árboles de clasificación) arroja el mejor desempeño en este último conjunto de datos. Asimismo, el desempeño en ambas muestras por parte de regresión logística y SVM son cercanos al arrojado por árboles de clasificación (84% y 83.2%, respectivamente).

Tabla 5.

Exactitud de los modelos bajo la comparativa en estudio.

Escenario 1: Atributos estructurados		
Métodos	Entrenamiento	Validación
Regresión logística	0.641	0.657
Árboles de clasificación	0.614	0.611
Random forest	0.679	0.709
LDA	0.640	0.656
SVM	0.625	0.662
Escenario 2: Atributos tipo texto (vectores considerando TF-IDF; las 50 palabras más relevantes en la descripción del incidente, previa limpieza de datos)		
Métodos	Entrenamiento	Validación
Regresión logística	0.868	0.881
Árboles de clasificación	0.851	0.856
Random forest	0.995	0.881
LDA	0.856	0.873
SVM	0.874	0.884
Escenario 3: Ambos tipos de atributos		
Métodos	Entrenamiento	Validación
Regresión logística	0.825	0.840
Árboles de clasificación	0.849	0.850
Random forest	0.887	0.775
LDA	0.798	0.815
SVM	0.822	0.832

Fuente: Los autores

Otro aspecto por considerar es que los resultados arrojados por la combinación de variables estructuradas y de texto en un mismo modelo, no están mostrando un incremento en la exactitud de los modelos. De hecho, en el escenario 2 (solo datos de texto), ninguno de los métodos proporcionó un modelo que superase el 88.1% que produjo random forest o regresión logística, y menos el 88.4% arrojado por SVM. En otras palabras, la complejidad y el consumo de recursos que implica el hecho de combinar variables estructuradas y de texto, no están siendo justificados por los resultados obtenidos.

Por consiguiente, siendo conscientes de la mejoría que produce el escenario de predictores de texto (véase escenario 2, Tabla 5) y tomando en cuenta la exactitud en la muestra de entrenamiento y en la de validación, cualquiera de los modelos de random forest, SVM o regresión logística resultan útiles para el uso en la organización observada, con tasas de acierto del 88%.

Para examinar las implicaciones prácticas al respecto, se tomó una muestra adicional de 50 incidentes ya cerrados (conociendo el equipo solucionador correcto). Luego, se indagó en la compañía la trazabilidad de cada incidente, con el fin de identificar si este había sido transferido una o más veces a otros equipos antes de ser solucionado por el equipo correcto. Asimismo, se extrajo la fecha de creación y cierre de cada incidente, para calcular los días que estuvo en tránsito. Esto permitió dar cuenta del promedio de días transcurridos desde la primera clasificación que recibió el incidente reportado a la mesa de ayuda.

Con estas estimaciones, se recurrió a una representación del tiempo estimado de resolución de un incidente típico desde que este se clasifica la primera vez. Dicho indicador considera (a) la exactitud de la clasificación (humano vs modelo elegido), usando la nueva muestra (50 instancias), y (b) dos tipologías de tiempos de resolución del incidente: cuando este se clasifica correcta o incorrectamente en el primer contacto. Así, la representación del indicador de interés se plasma en la ec. 1.

$$t_{it} = \tilde{t}_2 - \rho_1(\tilde{t}_2 - \tilde{t}_1) \quad (1)$$

Siendo:

t_{it} = tiempo estimado de resolución de un incidente típico desde que este se clasifica la primera vez.

ρ_1 = probabilidad de clasificar correctamente cuando se reporta el incidente (a la primera ocasión).

\tilde{t}_1 = tiempo medio de resolución de un incidente cuando éste es correctamente clasificado al primer contacto con él.

\tilde{t}_2 = tiempo medio de resolución de un incidente cuando éste es clasificado incorrectamente al primer contacto con él (considera el tiempo de tránsito por reclasificaciones hasta que el incidente llega al equipo correcto de trabajo y se le brinda solución).

En la Tabla 6 se plasman los resultados de la ecuación 1 y sus componentes.

Tabla 6. Comparativo humano vs máquina con relación al desempeño práctico de la clasificación de incidentes

	t_1	t_2	p_1	t_{it}
Humano	6.27 días	11.88 días	66%	8.18 días
Modelo			88%	6.94 días

Fuente: Los autores

La Tabla 6 muestra que al llevar a cabo la clasificación de incidentes mediante alguno de los modelos de aprendizaje automático (Random Forest, SVM o regresión logística, cada uno con 88% de exactitud, Tabla 5, validación), el ahorro en tiempo es de cerca de 30 horas de consumo/uso de recursos (ej., equipo humano, tecnología, espacios físicos) por incidente (1.23 días; 29.6 horas) en comparación con la tasa de exactitud del humano. Ahora bien, una determinada mesa de ayuda en la organización observada suele recibir en promedio 30 incidentes diarios en época de baja demanda y 200 en temporada de alta demanda. Por tanto, la clasificación mediante alguno de los modelos sugeridos (Random Forest, SVM o regresión logística) representaría, en cada mesa de ayuda, un ahorro diario de entre 888 horas y 5920 horas de consumo/uso de recursos. De ahí que los 22 puntos porcentuales de más exactitud por parte de la máquina (88%) en comparación con el humano (66%), constituyen una importante implicación práctica para la organización observada.

5. Conclusiones

Este artículo demuestra que la utilización de modelos de aprendizaje automático para la clasificación de incidentes tecnológicos en una compañía aseguradora es una alternativa prometedora, considerando la comprensión de un nuevo dominio y uso, así como sus implicaciones prácticas.

Con relación al método con mejor exactitud, el estudio encuentra supremacía de Random Forest en el escenario del mero uso de predictores estructurados. Asimismo, aunque dicho método también resulta prometedor en el escenario de predictores tipo texto (0.881), los resultados de los modelos diseñados bajo los demás métodos no son notablemente diferentes, oscilando entre 0.85 (árbol de clasificación y 0.884 SVM). Esto da flexibilidad a la elección del método en cuanto a exactitud y, además, posibilita tomar en cuenta otras variables como criterios de elección, entre ellas, eficiencia computacional de las alternativas (ej., tiempo de entrenamiento).

En referencia al escenario de observación, el uso de variables no estructuradas representa una considerable mejoría de la capacidad de predicción de los modelos, pasando Random Forest, en la muestra de validación, de 0.709 en datos estructurados a 0.881 en el escenario de datos de texto. De igual forma ocurre con los demás métodos, siendo superiores los resultados en el escenario de datos de texto en más de 20 puntos porcentuales, en comparación con el de datos estructurados. Así, las 50 de expresiones más relevantes (bajo suma de TF-IDF) en la descripción narrativa del incidente (ej., “cotiz”, “tecnolog”, “servici”, “asunt”), representan predictores esenciales para mejorar la exactitud de los modelos. Este último hallazgo es consistente con la importancia que [3] le atribuye a la descripción del incidente para favorecer el desempeño de los modelos de clasificación.

Al considerar las implicaciones prácticas de la tasa de clasificación correcta humano (66%) vs máquina (88%, Random Forest, SVM o regresión logística, véase Tabla 5, escenario 2), la máquina representa una mejoría importante

desde el punto de vista práctico, posibilitando ahorros anuales de 10656 horas de consumo de recursos humanos, físicos y tecnológicos en condiciones de temporada baja y de 71040 horas en temporada de alta demanda.

Este artículo constituye un caso de éxito del aprendizaje automático en el sector de seguros. Por consiguiente, sirve de base para la preparación de la organización observada y de otras organizaciones, en cuanto al afrontamiento del paradigma de la industria 4.0, centrado en el uso de tecnologías emergentes (ej., Big Data, inteligencia artificial, robótica) y en su integración con las personas para transformar el entorno de trabajo.

Si bien este estudio examina la mayoría de los datos disponibles sobre el incidente enviado vía correo electrónico, como lo es la descripción narrativa, las imágenes de pantallazos de errores alusivos al incidente no fueron consideradas. El procesamiento de dichas imágenes, en futuros trabajos, podría generar nuevo conocimiento al respecto y favorecer el desempeño de los modelos aquí entrenados/validados. De igual forma, otros futuros estudios podrían abordar también datos de voz, generados al momento de realizar el reporte del incidente a través de llamada telefónica.

Referencias

- [1] Al-Hawari, F. and Barhamb, H., A machine learning based help desk system for IT service management, *Journal of King Saud University - Computer and Information Sciences*, 33(6), pp. 702-718, 2019. DOI: <https://doi.org/10.1016/j.jksuci.2019.04.001>.
- [2] Giurgiu, I., Wiesmann, D., Bogojeska, J., Lanyi, D., Stark, G., Wallace, R., Pereira, M.M and Hidalgo, A.A., On the adoption and impact of predictive analytics for server incident reduction, *IBM Journal of Research and Development*, 61(1), pp. 9-98, 2017. DOI: <https://doi.org/10.1147/JRD.2016.2631400>.
- [3] Paramesh, S. and Shreedhara, K., Automated IT service desk systems using machine learning techniques, in *Data analytics and learning*, Singapore, Springer, [online]. 2019, pp. 331-346. Available at: https://link.springer.com/chapter/10.1007/978-981-13-2514-4_28.
- [4] Sulaman, S., Weyns, K. and Höst, M., Identification of IT incidents for improved risk analysis by using machine learning, *Euromicro Conference on Software Engineering and Advanced Applications*, [online]. vol. 41, 2015, pp. 1-5. Available at: <https://ieeexplore.ieee.org/document/7302476>.
- [5] Jan, E., Ayachitula, N., Ni, J. and Zhang, Z., A statistical machine learning approach for ticket mining in IT service delivery, in: 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), [online]. 2013, pp. 541-546. Available at: <https://ieeexplore.ieee.org/document/6573030>.
- [6] Son, G., Hazlewood, V. and Peterson, G., On automating XSEDE user ticket classification, *National Institute for Computational Sciences University of Tennessee*, [online]. 14, pp. 1-7, 2014. Available at: <https://dl.acm.org/doi/abs/10.1145/2616498.2616549>.
- [7] Bogojeska, J., Stöckel, D., Zazzi, M. and Kaiser, R., History-alignment models for bias-aware prediction of virological response to HIV combination therapy, *Journal of Machine Learning Research*, [online]. 22, pp. 118-126, 2012. Available at: <https://proceedings.mlr.press/v22/bogojeska12.html>.
- [8] Maksai, A., Bogojeska, J. and Wiesmann, D., Hierarchical incident ticket classification with minimal supervision, *IEEE International Conference on Data Mining*, 2014, pp. 923-928. DOI: <https://doi.org/10.1109/ICDM.2014.81>.
- [9] Akbar, M. and Jianglei, H., Vertical domain text classification: towards understanding IT tickets using deep Neural Networks, in *The Thirty-Second AAAI Conference on Artificial Intelligence*, [online]. 2018. Available at: <https://ojs.aaai.org/index.php/AAAI/article/view/11375>.
- [10] Silva, S., Pereira, R. and Ribeiro, R., Machine learning in incident categorization automation, *Iberian Conference on Information Systems and Technologies [CISTI]*, [online]. vol. 13, 2018. Available at: <https://ieeexplore.ieee.org/document/8399244>.
- [11] Altintas, M. and Cuneyd, A., Machine learning based ticket classification in issue tracking systems, [online]. 2014. Available at: <https://docplayer.net/1775410-Machine-learning-based-ticket-classification-in-issue-tracking-systems.html>.
- [12] Kallis, R., Di Sorbo, A., Canfora, G. and Panichella, S., Ticket tagger machine learning driven, *International Conference on Software Maintenance and Evolution [ICSME]*, [online]. pp. 1-4, 2019. Available at: <https://ieeexplore.ieee.org/document/8918993>.
- [13] Gore, R., Daillo, S., Padilla, J. and Ezell, B., Assessing cyber-incidents using machine learning, *International Journal of Information and Computer Security [IJICS]*, 10(4), pp. 341-360, 2018. DOI: <https://doi.org/10.5555/3292793.3292795>.
- [14] Li, H. and Zhan, Z., Machine learning methodology for enhancing automated process in IT incident management, in: *Symposium on Network Computing and Applications*, [online]. 2012. Available at: <https://ieeexplore.ieee.org/document/6299094>.
- [15] <https://cran.r-project.org/>, <https://cran.r-project.org/>, 9 10 2021. [Online]. Available at: <https://cran.r-project.org/web/packages/caret/caret.pdf>.
- [16] cran.r-project.org, Package 'caret', 9 10 2021. [Online]. Available at: <https://cran.r-project.org/web/packages/caret/caret.pdf>.
- [17] Diao, Y., Jamjoom, H. and Loewenstern, D., Rule-based problem classification in IT service management, in: *2009 IEEE International Conference on Cloud Computing*, New York, 2009. DOI: <https://doi.org/10.1109/CLOUD.2009.80>.

P. Gómez-Jaramillo, es Ing. Industrial de la Universidad de Antioquia, Colombia. MSc en Ingeniería de la Universidad de Antioquia, Colombia. ORCID: 0000-0002-8745-0850

F. González-Echavarría, es Ing. Industrial de la Universidad de Antioquia, Colombia. MSc. en Economía, PhD (c) en Dirección de Empresas, de la Universidad de Valencia, España. Es profesor del Departamento de Ingeniería Industrial, de la Universidad de Antioquia, Medellín, Colombia. ORCID: /0000-0002-1540-9859

J.I. Pérez-Rave, es Ing. Industrial de la Universidad de Antioquia, Colombia. Con Esp. En Estadística y en Ingeniería de Sistemas de la Universidad Nacional de Colombia. MSc. en Ingeniería de Sistemas de la Universidad Nacional de Colombia, y en Análisis y Visualización de Datos Masivos de la UNIR, España. Es PhD. en Ingeniería – Sistemas e Informática de la Universidad Nacional de Colombia, y PhD (c) en Dirección de Empresas, de la Universidad de Valencia, España. Es director del Grupo de investigación IDINNOV, IDINNOV S.A.S. ORCID: 0000-0003-1166-5545