



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

Rastreo de las Inversiones Públicas Nacionales y Subnacionales en Gestión del Riesgo de Desastres

Unidad de Científicos de Datos

Dirección de Economía Naranja y Desarrollo Digital

Julio, 2022



1. Descripción del proyecto

2. Metodología




3. Resultados

4. Conclusiones

1. Introducción



RASTREO DE LAS INVERSIONES PÚBLICAS NACIONALES Y SUBNACIONALES EN GESTIÓN DEL RIESGO DE DESASTRES

<h2>¿Qué?</h2>	<p>Implementar un modelo de rastreo automático de proyectos de inversión pública relacionados con gestión del riesgo de desastres (GRD)</p>	
<h2>¿Para qué?</h2>	<p>Apoyar el proceso de rastreo de proyectos de inversión relacionados con GRD reportados año a año en gesproy-SGR, SIIF y Ciclope</p>	
<h2>Resultados</h2>	<ol style="list-style-type: none"> 1. Se implementó un modelo SVM* para realizar predicciones de nuevos proyectos relacionados con GRD. 2. Se desarrollo una herramienta que permite cargar nuevas bases de datos para estimar proyectos relacionados con GRD. 	

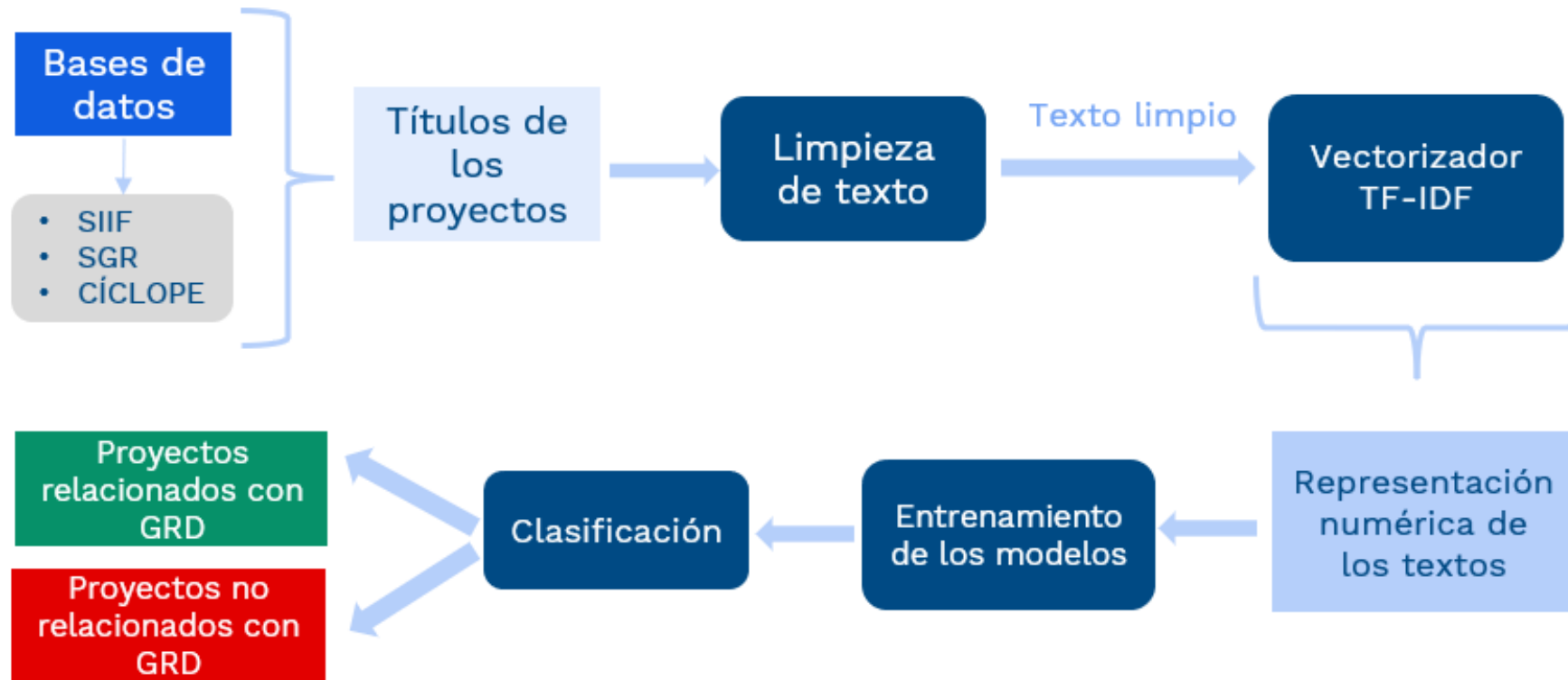
*SVM: Support Vector Machine con precisión promedio del 95%

3. Metodología



Metodología del proyecto

Figura 1: Metodología general para el rastreo automático de proyectos relacionados con GRD.



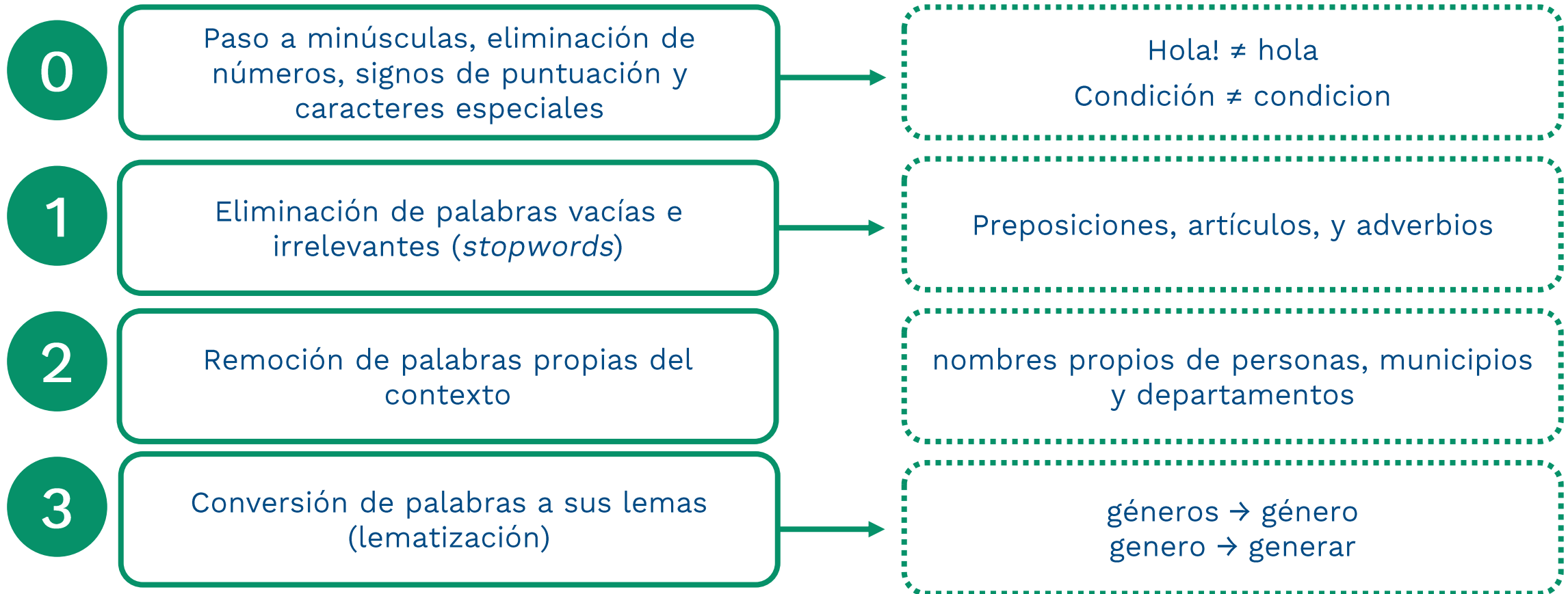
Fuente: elaboración propia.

*GRD Gestión de Riesgo de Desastres



Limpieza de texto

Minúsculas, remover caracteres no deseados, suprimir palabras irrelevantes y otros



Vectorización de texto TF-IDF

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$w_{i,j}$: importancia de la palabra i en el nombre del proyecto j

$tf_{i,j}$: número de ocurrencias de la palabra i en el nombre del proyecto j

df_i : número de nombres de proyectos que contienen la palabra i

N : número proyectos

Figura 2: Representación a modo de ejemplo de la matriz TF de una vectorización TF-IDF

		Títulos en la base de datos				
		T1	T2	T3	...	TN
Vocabulario	desarrollo	1		3		
	sostenible		2	1		1
	estudio	1		1		
	ambiental			2		1
	gestion	1			2	
	desastre		1	2		
	riesgo	1				2
	...					
	plan		1		1	

Fuente: elaboración propia.

Entrenamiento de modelos de clasificación

Tres etapas para desarrollar

Probar con diferentes tamaños del vocabulario de la vectorización TF-IDF

- Cada nombre de proyecto se representa por un vector numérico de alta dimensión, donde cada dimensión esta asociada a un *unigrama* o a un *bigrama*.

Dividir las matrices generadas con la vectorización TF-IDF en entrenamiento y validación.

- Grupo de entrenamiento 75%
- Grupo de validación 25%

Ajustar el modelo de clasificación con validación cruzada sobre grupo de entrenamiento y validar sobre grupo de validación

- Métricas utilizadas para evaluar el rendimiento del modelo
 - Precisión
 - Recall
 - F1-Score



4. Resultados



Entrenamiento del modelo

Tabla 1: Distribución de los datos para el entrenamiento del modelo

	Porcentaje de la base de datos	Clase positiva	Clase negativa	Total
Entrenamiento	75%	1.933	17.276	19.209
Validación	25%	645	5.759	6.404

Fuente: elaboración propia.

Tabla 2: Matriz de confusión del modelo sobre los datos de validación

	Precisión	Recall	F1-score	Muestras
Clase negativa	0,99	0,99	0,99	5.759
Clase positiva	0,90	0,93	0,91	645

Fuente: elaboración propia.


Herramienta de Rastreo

Rastreo de inversión Nacional para Gestión del Riesgo de Desastres

Esta herramienta permite apoyar el rastreo de las inversiones públicas en gestión del riesgo de desastres a través de la información reportada en el orden nacional y subnacional proveniente de fuentes como el [Sistema Integrado de Información Financiera \(SIIF\)](#), (SGR) y del Sistema de Información de Cooperación (CÍCLOPE), perm de las inversiones para la Política Nacional de Gestión del Riesgo de

Unidad de Científicos de Datos - UCD

© 2022



El futuro es de todos DNP Departamento Nacional de Planeación

Rastreo de Inversión Nacional para Gestión del Riesgo de Desastres

Desarrollado por:
Unidad de Científicos de Datos - UCD

ucd@dnpp.gov.co

© 2022

Rastreo Automático

Cargar archivo

Drag and drop file here
Limit 200MB per file • XLSX

Browse files

resultado_siif.xlsx
0.6MB

Rastreo de inversión Nacional para Gestión del Riesgo de Desastres

Resultados del rastreo

codBPIN	nomProy
nan	ATENCIÓN DE DESASTRES Y EMERGENCIAS EN EL TERRITORIO NACIONAL - FONDO NACIONAL DE GESTIÓN DEL RIESGO DE DESASTRES
nan	ATENCIÓN DE DESASTRES Y EMERGENCIAS EN EL TERRITORIO NACIONAL - FONDO NACIONAL DE GESTIÓN DEL RIESGO DE DESASTRES
nan	ATENCIÓN DE DESASTRES Y EMERGENCIAS EN EL TERRITORIO NACIONAL - FONDO NACIONAL DE GESTIÓN DEL RIESGO DE DESASTRES
nan	ATENCIÓN DE DESASTRES Y EMERGENCIAS EN EL TERRITORIO NACIONAL - FONDO NACIONAL DE GESTIÓN DEL RIESGO DE DESASTRES
nan	FONDO DE MITIGACIÓN DE EMERGENCIAS - FOME
nan	FONDO DE MITIGACIÓN DE EMERGENCIAS - FOME
nan	FONDO DE MITIGACIÓN DE EMERGENCIAS - FOME
nan	FONDO DE MITIGACIÓN DE EMERGENCIAS - FOME
nan	FONDO DE MITIGACIÓN DE EMERGENCIAS - FOME
nan	FONDO DE MITIGACIÓN DE EMERGENCIAS - FOME
nan	FONDO DE MITIGACIÓN DE EMERGENCIAS - FOME
2021011000018	IMPLEMENTACION DE ACCIONES DE MITIGACION DEL RIESGO EN ZONAS DEGRADADAS EN LAS VEREDAS CASA GRANDE Y CANADA EN EL
2021011000198	IMPLEMENTACION DE ACCIONES CORRECTIVAS PARA LA REDUCCION Y MITIGACION DEL RIESGO POR MOVIMIENTOS EN MASA E INUNDA
2021011000267	IMPLEMENTACION DE UN PROCESO DE RESTAURACION ECOLOGICA COMO MEDIDA DE ADAPTACION AL CAMBIO CLIMATICO EN CUNDIN
2018011001058	FORTALECIMIENTO DE LA GESTIÓN DEL RIESGO DE DESASTRES EN LA ZONA DE AMENAZA VOLCÁNICA ALTA-ZAVA DEL VOLCÁN GALER
2021011000149	IMPLEMENTACION DE MEDIDAS ESTRUCTURALES Y NO ESTRUCTURALES PARA EL CONTROL DE LA EROSION E INUNDACION EN LA CORRI
2018011000599	FORTALECIMIENTO DE LOS CUERPOS DE BOMBEROS DE COLOMBIA - NACIONAL
2021011000239	MANTENIMIENTO LIMPIEZA, RECTIFICACION DE CAUCE Y PROTECCION DE TALUD DEL K0+000 AL K1 + 300 DEL CANAL DE AGUAS LLUVIAS
2021011000212	IMPLEMENTACION DE MEDIDAS DE ADAPTACION A LA VARIABILIDAD Y CAMBIO CLIMATICO EN JURISDICCION DE CORPOGUAVIO CUNDIN
2018011000472	ASISTENCIA PARA FORTALECER TÉCNICAMENTE A LAS ENTIDADES TERRITORIALES EN LA IMPLEMENTACIÓN DE LOS COMPONENTES DEL SI
2020011000069	FORTALECIMIENTO FINANCIERO DE LA POLÍTICA NACIONAL DE GESTIÓN DEL RIESGO DE DESASTRES EN EL TERRITORIO NACIONAL

Descargar tabla

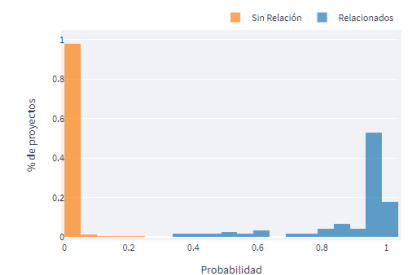
|| Proyectos Rastreados * ||

125/4492

↑ 4367 sin relación

* Los resultados tienen una precisión promedio del 95%

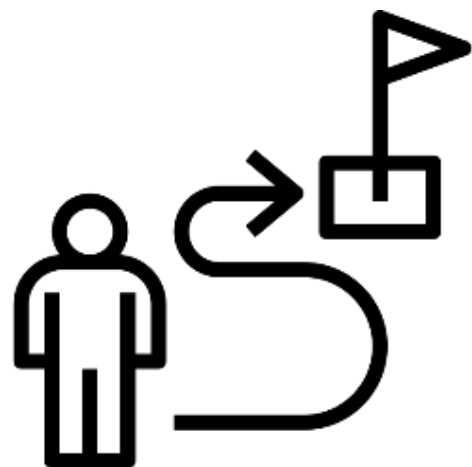
Distribución de Probabilidades



5. Conclusiones y recomendaciones



Conclusiones

**1**

El modelo de rastreo presentó un acierto de clasificación del 98% sobre el conjunto de datos de validación, con una tasa de falsos positivos del 10% y una tasa de falsos negativos del 1%

2

Se identificó que proyectos rastreados para GRD por parte de la Subdirección de Cambio Climático y Gestión del Riesgo de Desastres, no cumplían con la exclusión por filtros negativos dentro de la metodología que se usa actualmente para el rastreo. Por este motivo el clasificador automático mejora el rastreo y reduce la probabilidad de cometer errores debido a rechazo o aceptación por palabras clave.

3

Existen proyectos dentro que en su nombre no contiene ninguna palabra que tenga relación con GRD, pero que fueron rastreados por el conocimiento del experto temático. Este tipo de observaciones son difíciles de modelar por el rastreador automático por lo que puede ocasionar malas clasificaciones para futuras observaciones.

4

Se recomienda que siempre se haga una validación los resultados que entrega la herramienta de rastreo, pues, aunque la confiabilidad es alta, existe una probabilidad del 2% de cometer errores de rastreo, especialmente por las condiciones mencionadas en el numeral 3.



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación