



DEPARTAMENTO NACIONAL DE PLANEACIÓN



DEPARTAMENTO
NACIONAL DE PLANEACIÓN

ANALÍTICA PREDICTIVA DE PROYECTOS DE INVERSIÓN FINANCIADOS CON RECURSOS SGR – ETAPA 2



UNIDAD DE CIENTÍFICOS DE DATOS

Dirección de Economía Naranja y Desarrollo
Digital

febrero, 2023

1. Introducción

2. Descripción del proyecto




3. Metodología

4. Resultados

5. Conclusiones

1. Introducción

ANALÍTICA PREDICTIVA DE PROYECTOS DE INVERSIÓN FINANCIADOS CON RECURSOS SGR – ETAPA 2

¿Qué?	<p>Este proyecto busca continuar el proyecto de analítica desarrollado en el año 2021 buscando predecir si un proyecto con fondos de regalías va a presentar dificultades a futuro.</p>	
¿Para qué?	<p>Este proyecto busca redefinir y eliminar algunas variables involucradas en el modelo del año 2021 con el fin de mejorar la predicción del modelo.</p>	
Logros	<ol style="list-style-type: none">1. Se redefinieron y eliminaron las variables sugeridas2. Se mejoró la predicción del modelo a partir de la redefinición y eliminación de las variables.	

2. Descripción del proyecto

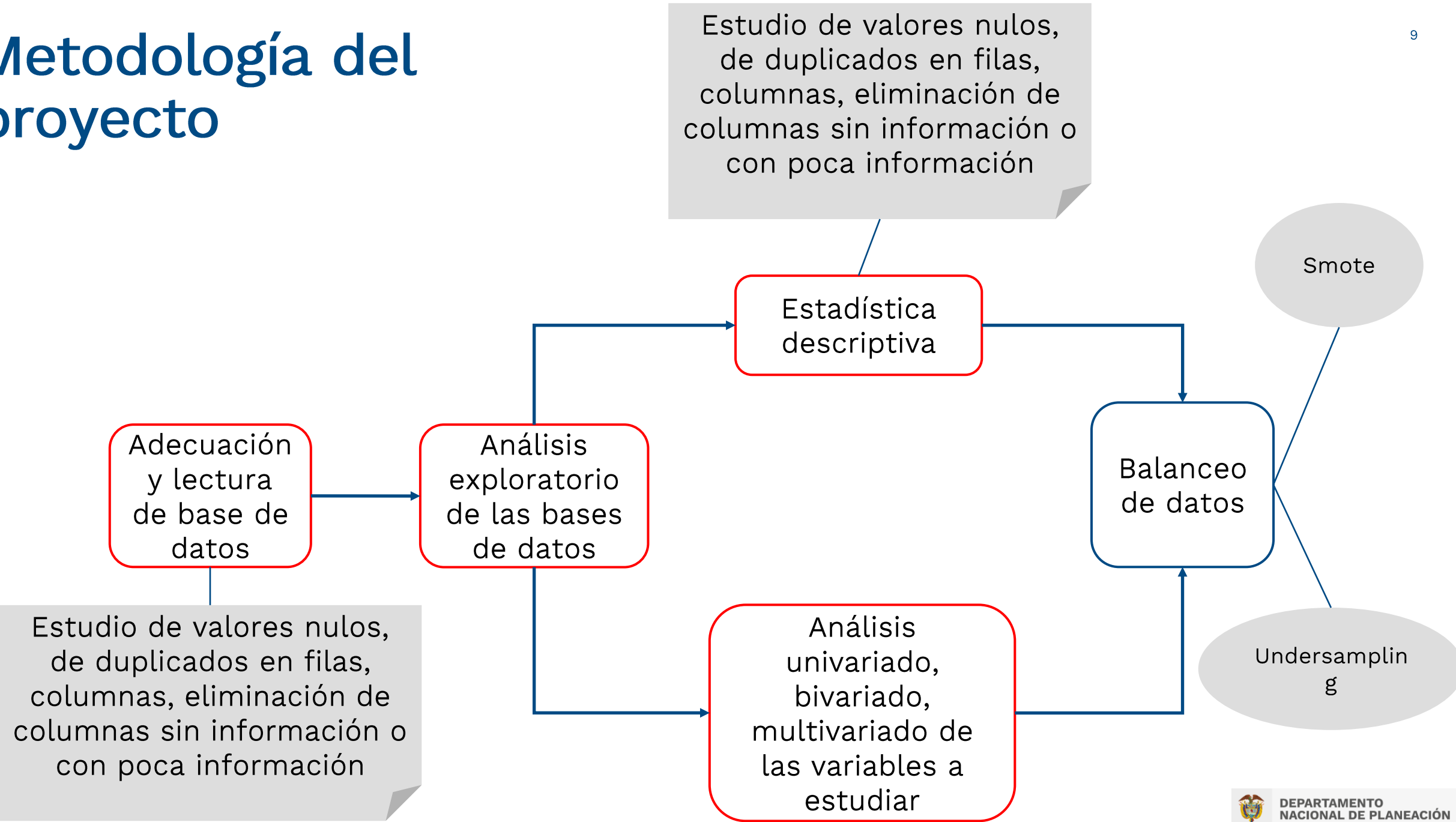
Descripción del proyecto

La motivación de este proyecto es lograr atender a tiempo los proyectos con fondos de regalías que puedan tener dificultades a futuro.

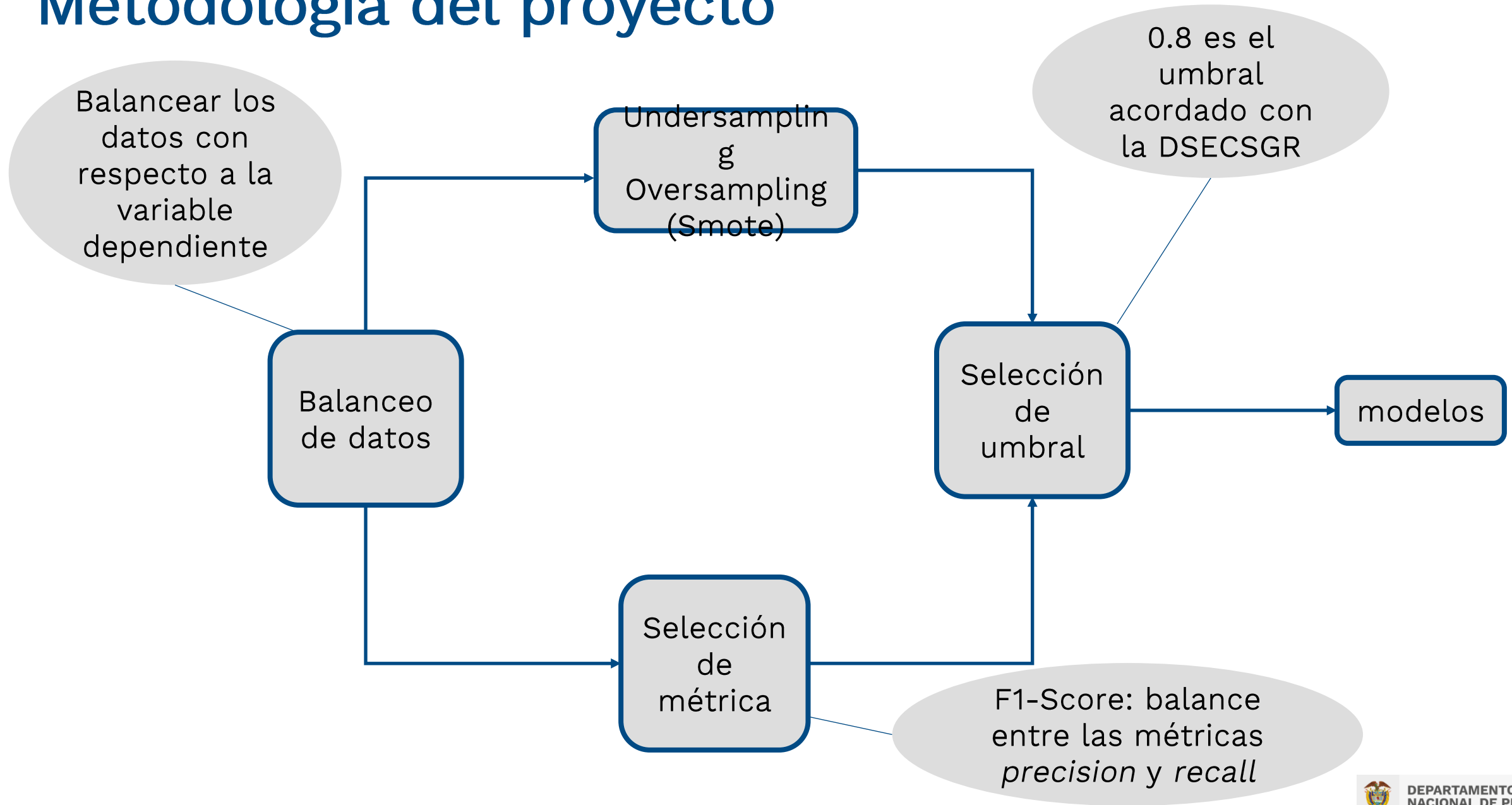
Para resolver este problema la UCD usó el proyecto del año 2021 con el fin de redefinir y eliminar variables para mejorar la métrica del modelo desarrollado. Una vez definidas las variables a usar se construye un modelo de aprendizaje automático para resolver el problema.

3. Metodología

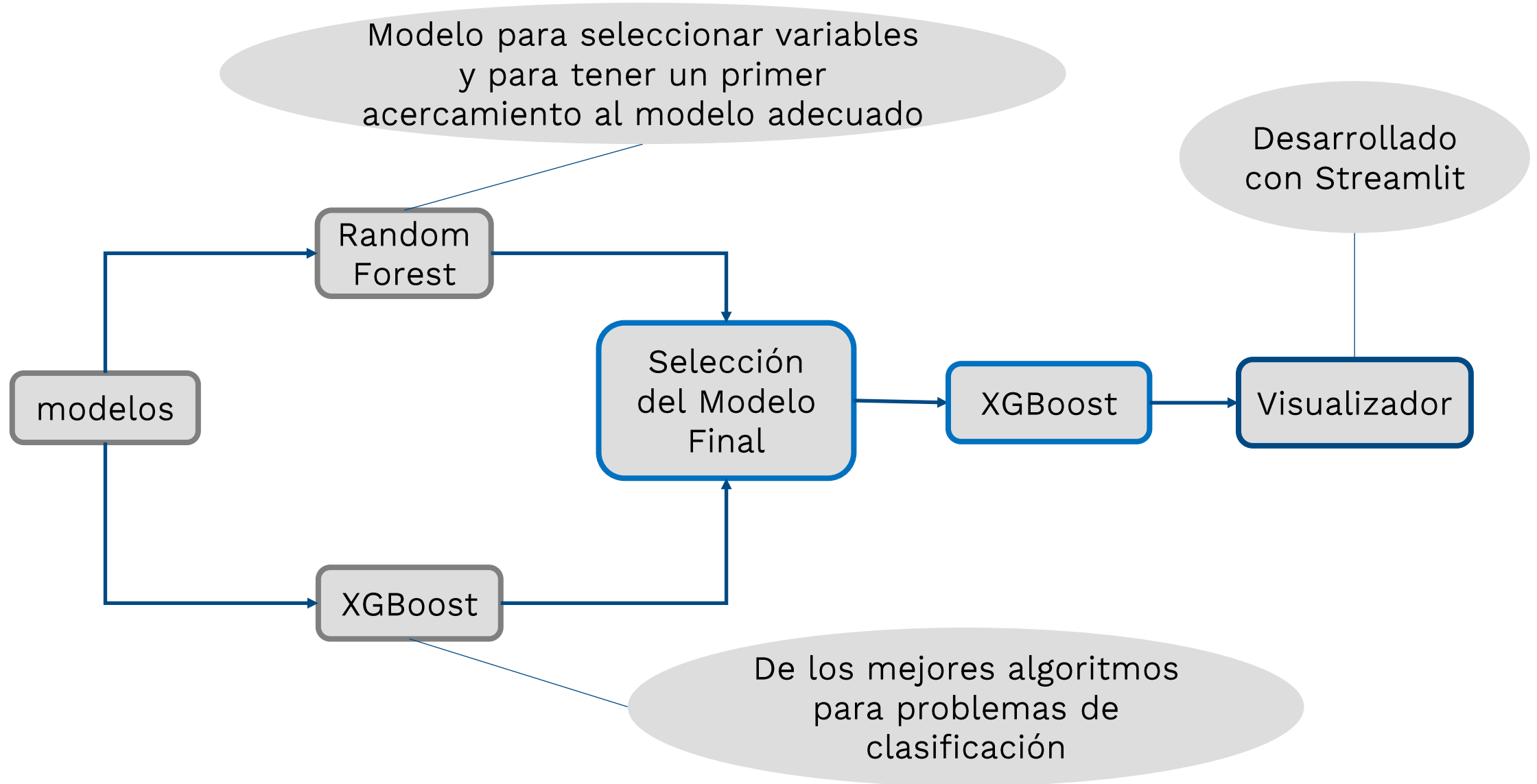
Metodología del proyecto



Metodología del proyecto



Metodología del proyecto



4. Resultados

Resultados

Los resultados se realizaron con respecto a la métrica *f1-score*. La métrica *f1-score* se usa para problemas desbalanceados como en este caso. Varía entre 0 y 1, donde un valor cercano a 1 indica que es altamente probable que la predicción sea correcta.

Variables dependiente	Valor f1-score
Cambios significativos en tiempos de ejecución iniciales	0.94
Cambios significativos en el valor monetario original de la inversión	0.84
Cambios significativos en el valor monetario programado	0.85

- La selección del umbral se realizó de acuerdo con el balanceo de los datos, la selección de la métrica y el modelo escogido.

5. Conclusiones y recomendaciones



Conclusiones y recomendaciones

1

Con respecto al índice IGPR se observó que tiene distintas estandarizaciones y no se cuenta con suficiente información histórica del mismo. Se recomienda el estudio de variables adicionales que puedan influenciar en dicho índice.

2

Se redefinieron las variables propuestas por la DSECSGR: *longitud inicial del proyecto*; *retraso inicio proyecto*; y *demora en el inicio del proyecto*, siguiendo los lineamientos de la DSECSGR . Estas definiciones mejoraron la precisión de la estimación.

3

No se incluyeron las variables encontradas en los distintos modelos debido a que de la discusión con la DSECSGR no se consideró apropiado incluirlas. Se recomienda su estudio para futuros modelos.

4

Se logró balancear los datos con respecto a cada variable dependiente con el fin de evitar sesgos en el modelo. Este balanceo de datos también mejoró la precisión en la clasificación.

Conclusiones y recomendaciones

5

Se implementaron distintos modelos y se escogió el XGBoost como el modelo final ya que arroja los mejores resultados con una probabilidad superior al 84%.

6

Se desarrolló una herramienta de visualización que contiene el modelo desarrollado y que la DSECSGR podrá usar.





DEPARTAMENTO NACIONAL DE PLANEACIÓN