

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

DETECCIÓN Y RASTREO DE PROYECTOS DE INVERSIÓN EN ALINEACIÓN CON LOS OBJETIVOS DE DESARROLLO SOSTENIBLE (ODS)

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Dirección de Seguimiento y Evaluación de Políticas Públicas

Sector

Inversión y finanzas públicas

Lenguaje

R

Fuente de datos

Sistema Unificado de Inversiones y Finanzas Públicas (SUIFP)
Agencia Presidencial de Cooperación Internacional (APC)

Presentación

Los Objetivos de Desarrollo Sostenible (ODS) surgen como respuesta a los problemas primordiales de las naciones y buscan una mejora en la calidad de vida de sus habitantes. Si bien las metas asociadas a cada ODS contemplan unos indicadores para hacer seguimiento, verificar el avance y cumplimiento, es necesario contar con indicadores financieros complementarios que permitan tener análisis más completos sobre cómo la inversión de recursos públicos se orienta hacia la consecución de cada ODS. Para ello, la UCD y la Dirección de Seguimiento y Evaluación de Políticas Públicas del DNP desarrollaron el presente proyecto, que estima la asignación de recursos de inversión destinados a cada ODS mediante el análisis y clasificación automática de 112 mil proyectos de inversión financiados con recursos del Sistema General de Regalías (SGR), del Presupuesto General de la Nación (PGN), de entidades territoriales y de cooperación internacional.

The Sustainable Development Goals (SDG) emerge as a response to the primary problems of nations and seek an improvement in the quality of life of their inhabitants. Although each SDG includes indicators for monitoring, verifying progress and compliance, it is necessary to have complementary financial indicators that allow for a more complete analysis of how the investment of public resources is oriented towards the achievement of each SDG. To this end, the UCD and the Directorate of Public Policy Monitoring and Evaluation (DSEPP) of the DNP developed the present project, which estimates the allocation of investment resources for each ODS through the automatic analysis and classification of 112 thousand investment projects financed with resources from the General System of Royalties, the General Budget of the Nation, territorial entities and international cooperation.

Objetivo general

Analizar la inversión pública realizada en alineación con los Objetivos de Desarrollo Sostenible (ODS) a través de la aplicación de analítica de datos.

Objetivos específicos

1. Consolidar datos sobre los proyectos de inversión y los ODS, mediante la realización de consultas en bases de datos públicas y del DNP con información estructurada o no estructurada.
2. Etiquetar los proyectos de inversión como relacionados o no relacionados con cada uno de los ODS a través de la aplicación de técnicas de análisis de texto y procesamiento de lenguaje natural.
3. Presentar los resultados de manera organizada, con enfoque territorial y con enfoque sectorial a través del desarrollo de una herramienta de visualización.



El futuro es de todos

DNP
Departamento
Nacional de Planeación

Antecedentes

En 2018, el DNP realizó el Reporte Nacional Voluntario de Colombia en ODS con la coordinación de la Comisión ODS de la Dirección de Seguimiento y Evaluación de Políticas Públicas (DSEPP). Con miras a la realización de dicho documento, la comisión ODS identificó la necesidad de definir indicadores de inversión de presupuesto en alineación con cada ODS, motivo por el cual se propuso a la Unidad de Científicos de Datos el proyecto “Clasificación de proyectos de inversión para la estimación del presupuesto estatal en función del cumplimiento de los Objetivos de Desarrollo Sostenible”. Este proyecto permitió identificar que los recursos destinados a cumplir los ODS disminuyeron entre 2015 y 2017 en un 11% y que existe una mayor alineación entre algunas fuentes de financiación y algunos ODS específicos, como los recursos del PGN que tienen mayor participación en temas de pobreza, salud y educación (ODS 1, 3 y 4), los recursos del SGR que se enfocan más en temas de infraestructura y construcción de ciudades sostenibles (ODS 9 y 11) y los de cooperación internacional que se caracterizan por la inversión en proyectos de construcción de paz y acción por el clima (ODS 16 y 13), entre otros resultados que pueden consultarse en el Reporte Nacional Voluntario.

Finalizado dicho proyecto, se manifestó un gran interés en extender el alcance del proyecto para generar nuevos indicadores de inversión de recursos que consideraran la correlación existente entre los distintos ODS, incluyendo además un enfoque territorial y sectorial que permitiera realizar análisis con un mayor nivel de profundidad.

Metodología

La metodología utilizada para la realización de este proyecto se puede dividir en 5 grandes fases: (1) la consolidación y preparación de los insumos, (2) el preprocesamiento de los textos, (3) el cálculo de una medida de similitud entre textos de ODS y textos de proyectos de inversión, (4) el ajuste de modelos de mixturas para etiquetar los proyectos como alineados o no con cada uno de los ODS y (5) el desarrollo de un aplicativo de visualización. Cabe mencionar que la metodología propuesta inicialmente consistía en utilizar una lista de sintagmas clave (palabras y bigramas) por ODS, la cual se obtuvo del proyecto descrito en los antecedentes, para etiquetar los proyectos mediante reglas duras (si algún sintagma clave del i-ésimo ODS se encuentra en el texto del proyecto, entonces el proyecto está alineado con el i-ésimo ODS) y así desarrollar modelos de aprendizaje de máquina que predijeran, a partir del texto de los proyectos, la probabilidad de que el proyecto estuviera alineado con cada ODS. Sin embargo, el desempeño de estos modelos no fue muy satisfactorio, lo que llevó a adoptar la metodología alternativa que se detalla en este documento.

Consolidación y preparación de los insumos

Para llevar a cabo el presente proyecto, fue necesario consolidar información de DNP junto con información no estructurada de fuentes externas. A nivel de proyectos de inversión, se realizaron consultas sobre tres tablas de SQL asociadas a la base de datos de Mapa Inversiones (pry.IdentificacionProyecto, pry.FuentesFinanciacion y pry.Regionalizacion), obteniendo para cada proyecto las variables presentadas en la tabla 1. En estas consultas se delimitó la información a los registros provenientes de SUIFP PGN y SUIFP SGR, que corresponden al presupuesto aprobado (y no al presupuesto ejecutado) para los proyectos de inversión, sean estos financiados con recursos del Presupuesto General de la Nación (PGN), del Sistema General de Regalías (SGR), del Sistema General de Participaciones (SGP), de las entidades territoriales, de empresas públicas o con recursos privados.



El futuro es de todos

DNP
Departamento
Nacional de Planeación

Variable	Significado	Tablas asociadas
Bpin	Identificador del proyecto (Llave).	pry.IdentificacionProyecto pry.FuentesFinanciacion pry.Regionalizacion
NombreOrigen	Sistema de información del cual se obtiene el registro.	pry.IdentificacionProyecto pry.FuentesFinanciacion pry.Regionalizacion
NombreProyecto	Título del proyecto de inversión.	pry.IdentificacionProyecto
ProblemaCentral	Problema que se busca resolver con el proyecto de inversión.	pry.IdentificacionProyecto
Descripcion	Descripción del proyecto de inversión.	pry.IdentificacionProyecto
ObjetivoGeneral	Objetivo general del proyecto de inversión.	pry.IdentificacionProyecto
AnioInicio	Año de inicio del proyecto.	pry.IdentificacionProyecto
AnioFinal	Año de finalización (real o estimada) del proyecto.	pry.IdentificacionProyecto
SectorGasto	Sector asociado al proyecto de inversión.	pry.IdentificacionProyecto
EsRegionalizable	Variable que permite identificar si el proyecto es regionalizable ("S" = Sí / "N" = No) o si se financió con recursos de entidades territoriales ("T").	pry.IdentificacionProyecto
ValorInicial	Presupuesto aprobado para el proyecto de inversión (se interpreta como "aprobado" solo para la información que proviene de SUIFP PGN o de SUIFP SGR, pues si la información proviene de otros sistemas de información, puede corresponder al presupuesto ejecutado).	pry.FuentesFinanciacion pry.Regionalizacion
PeriodoAnio	Vigencia (año) en la cual se aprueba una cantidad dada de recursos (ValorInicial).	pry.FuentesFinanciacion pry.Regionalizacion
CodigoDepartamento	Código DIVIPOLA del departamento para el cual se aprobaron los recursos.	pry.Regionalizacion
Departamento	Nombre del departamento para el cual se aprobaron los recursos.	pry.Regionalizacion

Tabla 1: Variables consultadas para los proyectos de inversión.

Con respecto a este punto, es importante aclarar que la base de datos contiene también una variable "Fuente" que especifica la fuente de los recursos (diferente de "NombreOrigen", que indica de qué sistema de información proviene el registro), sin embargo, se prefirió omitirla y clasificar los registros con base en las variables "NombreOrigen" y "EsRegionalizable", dando lugar a las categorías "PGN" (para todos los proyectos registrados en SUIFP PGN), "SGR" (para los proyectos registrados en SUIFP SGR identificados con "EsRegionalizable = S") y "Territorial" (para los proyectos registrados en SUIFP SGR identificados con "EsRegionalizable = T")¹. Esta clasificación se puede interpretar como "recursos del PGN", "recursos del SGR" y "recursos de entidades territoriales" (incluyendo privadas), respectivamente, pues estas fuentes de financiación y las categorías propuestas se encuentran altamente correlacionadas, a pesar de que los proyectos con cofinanciación del PGN (1521) se registran en SUIFP SGR, al igual que unos pocos proyectos que aparecen con financiación de entidades territoriales (118), del SGR (74) y con recursos privados (800) se registran en SUIFP PGN.

¹ Nótese que ningún proyecto registrado en SUIFP SGR se encuentra registrado como "no regionalizable".



Otra observación que debe realizarse sobre la información es respecto a la variable “ValorInicial”: esta variable permite conocer el presupuesto aprobado por región (en este caso, departamento) y por vigencia (año) si se consulta en la tabla “Regionalizacion”, sin embargo, es necesario consultar la tabla “FuentesFinanciacion” para conocer el presupuesto aprobado para los proyectos no regionalizables. Así pues, la información de “FuentesFinanciacion” se utiliza para los análisis que se hacen por vigencia y por sector, mientras que la información de “Regionalizacion” solo se utiliza para el análisis territorial. En todo caso, se encuentra que la información disponible en ambas tablas es bastante consistente (es decir, coincide el valor total de un proyecto regionalizable en las dos tablas) desde 2012, cuando pasa de un 21% de consistencia (2011) a un 65% de consistencia, llegando a alcanzar un 93% de consistencia en 2018. Como última salvedad con respecto a esta variable, debe considerarse que el valor aprobado para una vigencia específica puede ser negativo, pues una reprogramación del presupuesto puede llevar a registros de este tipo, que tienen un fin meramente contable. Sin embargo, para evitar confusiones por parte de los usuarios, estas cifras se presentan como ceros en el aplicativo, a pesar de que sí se consideran en los agregados por sector, región y vigencia.

Obtenidos los insumos de SUIFP para los proyectos de inversión, se pasó a consolidar la información de la Agencia Presidencial de Cooperación (APC) para los proyectos de cooperación internacional. En este caso, se tomó la base con corte a 2018 que fue facilitada por la comisión ODS del DNP y el único tratamiento que se hizo fue totalizar los aportes registrados en la base de datos, independientemente del actor que realice el aporte, obteniendo un valor total destinado a cada proyecto de cooperación internacional. Finalmente, para enriquecer los textos sobre ODS que se utilizaron en el proyecto descrito en los antecedentes, se agregaron las cartillas de “Por qué es importante” cada ODS, así como su correspondiente introducción y sus datos destacables, todos ellos disponibles en la página web de las Naciones Unidas², complementando así las metas e indicadores que se usaron previamente como insumo principal.

Preprocesamiento de los textos

Habiendo obtenido los datos necesarios, se consolidaron los textos asociados a cada proyecto y los asociados a cada ODS. Para cada proyecto de inversión, se creó una cadena de texto única que contenía el nombre del proyecto, el problema central, su descripción y su objetivo general; para cada proyecto de cooperación internacional, se unieron los textos de “Nombre de la intervención” y “Objetivo general”; y para cada ODS se realizó el mismo procedimiento, pero con los textos descritos en la sección de “Consolidación y preparación de los insumos” (párrafo anterior).

La limpieza de las cadenas de texto obtenidas se realizó con R y consistió en la transformación del texto a minúsculas y en la remoción de números, signos de puntuación y demás caracteres que no fueran letras; también se removieron conectores, preposiciones y palabras que no agregan significado al texto (*stopwords*). Para este proyecto no se realizó la remoción de tildes, pues la inclusión de palabras como “género” (que sin tilde se relacionaría con “generar”) era relevante en este análisis, particularmente para ODS como el 5 (igualdad de género). El resultado de este tratamiento es lo que se denominará en adelante como “texto limpio”.

Cálculo de medidas de similitud

Habiendo obtenido las cadenas de texto limpio, se realizó una vectorización (es decir, una representación numérica) de los textos de cada ODS. Para ello, se utilizó el modelo de bolsa de palabras (*bag of words*), que consiste en construir una matriz con palabras en las columnas y ODS en las filas que representa el número de veces que aparece cada palabra en el texto de cada ODS. Por ejemplo, si el texto del ODS 1 fuera únicamente

² Página web de las Naciones Unidas: <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>



“acabar pobreza extrema”, las columnas “acabar”, “pobreza” y “extrema” tendrían un valor de 1 en la primera fila de la matriz, mientras las demás columnas tendrían un valor de 0. Esta representación numérica se ilustra con mayor claridad en la figura 1. Luego, se ponderó la matriz obtenida utilizando TF-IDF (*Term frequency – Inverse document frequency*), una técnica muy conocida en análisis de texto que permite dar una mayor relevancia a las palabras que son más características de cada unidad de texto (en este caso, de cada ODS o de cada proyecto). Cabe mencionar que esta técnica es bastante sencilla en comparación con otras como doc2vec y BERT, con las cuales podría tenerse una metodología más cercana al estado del arte, pero esta fue suficiente para realizar los análisis requeridos, como se ilustrará más adelante.

	este	es	un	texto	otro	último
este es un texto	1	1	1	1	0	0
este texto es otro texto	1	1	0	2	1	0
un último texto	0	0	1	1	0	1

Figura 1: Representación numérica de textos con el modelo de bolsa de palabras.

Habiendo realizado la vectorización de los textos de ODS, se procedió a vectorizar los textos de los proyectos (tanto de inversión como de cooperación internacional) con base en el diccionario de ODS (palabras presentes en los textos de ODS), pues era esencial que los vectores finales fueran comparables, para lo que se requería que estuvieran en el mismo espacio vectorial. En términos prácticos, la realización de este procedimiento sería equivalente a eliminar de los textos de proyectos las palabras que no estuvieran en los textos de los ODS, para luego vectorizarlos fácilmente como nuevas filas en la misma matriz.

Posteriormente, para cada proyecto se procedió a calcular la similitud coseno entre el vector “texto del proyecto” y cada uno de los vectores “texto de ODS”, tanto con la ponderación TF-IDF como sin ella. Esta medida de similitud se escogió porque ha sido ampliamente utilizada en análisis de texto y se calcula con un bajo costo computacional. Se determinó luego que ambas medidas (con y sin TF-IDF) eran importantes para el análisis ya que los resultados podrían ser mejores con uno o con otro enfoque, pero dada la necesidad de contar con una única medida de similitud, se decidió agregarlas calculando el promedio entre ambas, considerando además que las dos medidas toman valores entre 0 y 1, pues no existen frecuencias negativas en la matriz construida con el modelo de bolsa de palabras. Se obtuvieron así 16 medidas de similitud para cada proyecto (una con respecto a cada ODS) que permiten ordenar los proyectos del más alineado al menos alineado, para cada ODS.

Ajuste de modelos de mixturas

Habiendo calculado la medida de similitud, se procedió a estimar un umbral de clasificación para cada ODS a partir del cual los proyectos se pudieran considerar en alineación con un ODS dado. Para ello, se partió del supuesto de que el cálculo de una medida de similitud, en un caso ideal, permitiría discriminar perfectamente entre dos conjuntos: el conjunto de proyectos que no están alineados con determinado ODS y el conjunto de proyectos que sí lo están; pues el conjunto de no alineados tendría similitudes bajas (por ejemplo, entre 0 y 0,4) que no se superpondrían con las de los proyectos alineados, que serían altas (por ejemplo, entre 0,6 y 1). Este caso ideal se ilustra en la figura 2.

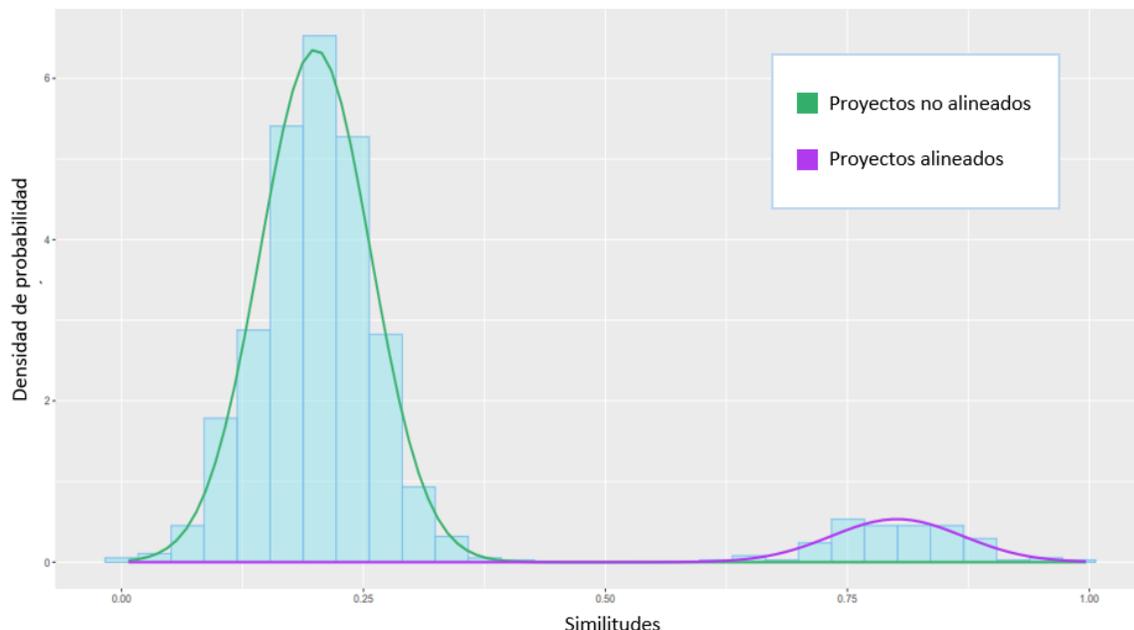


Figura 2: Distribución de proyectos alineados y no alineados con un ODS dado en un escenario ideal.

Bajo este supuesto, se puede pensar entonces en un modelo de probabilidad para cada ODS (en el ejemplo que sigue se formulará para el ODS 1) como:

Sea X_i la similitud coseno calculada entre el texto del proyecto i y el texto del ODS 1, con $i = 1, 2, \dots, n$, donde n corresponde al número de proyectos que conforman la muestra. Se define entonces la variable aleatoria X (similitud coseno) con función de distribución dada por $F(x) = \int_0^x f(t) dt$, donde $x \in [0,1]$. Aunque $f(x)$ es desconocida, es posible tener una idea de su forma realizando el histograma de las observaciones X_i , como se muestra en la figura 3 (izquierda).

Existen distribuciones de probabilidad que podrían modelar una distribución similar a la que se ilustra, como la distribución beta o la distribución gamma (aunque esta última se define en el rango $[0, \infty)$ y no en $[0,1]$, como la beta), pero el objetivo del modelo es identificar la distribución de similitudes de los proyectos no alineados con el ODS 1 (llámese $f_1(x)$) y la de los proyectos alineados con el ODS 1 (llámese $f_2(x)$), con la dificultad de que las observaciones correspondientes a ambos grupos se encuentran mezcladas en el mismo conjunto de los X_i . A partir de esto, se puede definir la variable latente (no observada) Z , que permite identificar si una observación X_i proviene de la distribución $f_1(x)$ o de $f_2(x)$, que es precisamente la variable que se quiere estimar, es decir, Z corresponde a la etiqueta asociada a un proyecto que permite saber si está alineado o no con el ODS 1.

Con las anteriores consideraciones, puede pensarse entonces que X sigue una distribución de mixtura, definida matemáticamente por:

$$f(x) = \eta_1 f_1(x) + \eta_2 f_2(x),$$



donde los parámetros η_1 y η_2 se conocen como los pesos de la mixtura, que cumplen con las propiedades:

$$\eta_k \geq 0, \quad \eta_1 + \eta_2 = 1,$$

y que en este contexto se pueden interpretar como los porcentajes de proyectos no alineados y alineados (respectivamente) con el ODS 1.

El modelo descrito se conoce como “modelo de mixturas” y existe una gran cantidad de literatura al respecto de su formulación y de la estimación de sus parámetros a partir de los datos, detalle que excede el alcance de este informe. En todo caso, habiendo formulado este modelo, la información de interés para el análisis corresponde a estimar η_1 , η_2 , $f_1(x)$ y $f_2(x)$ a partir de los datos, ejercicio para el cual es muy común suponer que las funciones de densidad de probabilidad $f_1(x)$ y $f_2(x)$ corresponden a una misma distribución paramétrica. En el marco de este proyecto, se supuso una distribución normal para cada uno de los componentes de la mixtura, dada la facilidad que brinda esta distribución para realizar la estimación de parámetros, aunque se reconoce las limitaciones que puede tener puesto que la distribución no es simétrica y puesto que está definida en el rango $(-\infty, \infty)$, mientras que los valores de X_i se encuentran en el rango $[0, 1]$. Con este supuesto, el modelo propuesto se puede expresar como:

$$f(x) = \eta_1 \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu_1}{\sigma_1} \right)^2} + \eta_2 \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu_2}{\sigma_2} \right)^2}$$

Definido este modelo bajo el supuesto de normalidad, se utilizó el algoritmo EM (esperanza-maximización) para estimar los parámetros del modelo (η_1 , η_2 , μ_1 , μ_2 , σ_1 y σ_2) mediante máxima verosimilitud. Los resultados de este procedimiento se ilustran en la figura 3 (derecha), donde se dibujan $\eta_1 f_1(x)$ (verde) y $\eta_2 f_2(x)$ (púrpura) por separado para poder comparar con el caso ideal que se presentó en la figura 2.

Habiendo resuelto el modelo, el último paso consistió en definir un umbral de similitud a partir del cual un proyecto se clasifica como “alineado con el ODS 1”. Con base en el modelo, se decidió que este umbral debía ser una similitud tal que la probabilidad de que un proyecto no estuviera alineado con el ODS 1 fuera igual a la probabilidad de que lo estuviera, es decir, debía ser igual la probabilidad de que la observación proviniera de una u otra de las distribuciones componentes (probabilidades a posteriori, en el algoritmo EM), lo que gráficamente equivale a encontrar la intersección de las funciones presentadas en la figura 3 (derecha). Matemáticamente, esto es igual que encontrar el máximo valor de $x \in [0, 1]$ que satisface:

$$\eta_1 \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu_1}{\sigma_1} \right)^2} = \eta_2 \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu_2}{\sigma_2} \right)^2} = 0,5$$

valor que se aproximó utilizando métodos numéricos. Así se completó la clasificación de los proyectos en el ODS 1 (en el ejemplo dado) y en cada uno de los 16 ODS analizados, lo que hace posible identificar proyectos alineados con más de un ODS, dando solución a una de las principales limitaciones que tuvo el proyecto presentado en los antecedentes.



Desarrollo del aplicativo

Finalizada la clasificación de los proyectos, se desarrolló un aplicativo que permite a un usuario visualizar los resultados de manera general, con los filtros de interés y con distintas desagregaciones. En el aplicativo, el usuario escoge el ODS que desea analizar y puede modificar el umbral de clasificación correspondiente (por defecto se presenta el que resulta del modelo de mixturas), algo que permite mayor flexibilidad si se desean aumentar o disminuir las tasas de falsos negativos y falsos positivos. Definido el umbral, el usuario puede filtrar los proyectos de inversión por fuente de recursos, sector (para el caso de “SGR” y “Territorial”) y vigencia, así como puede filtrar los proyectos de cooperación internacional por año de inicio o por estado del proyecto. Habiendo escogido los filtros, la aplicación muestra al usuario una tabla con los proyectos organizados de mayor a menor similitud con respecto al ODS escogido, junto con el valor aprobado para el proyecto cada una de las vigencias escogidas. De manera más agregada, la aplicación reporta (según la información disponible) el monto total de los proyectos clasificados (con los filtros y el umbral escogidos), un mapa de calor con los montos de inversión por departamento (enfoque territorial), una serie de tiempo con los recursos por vigencia, un gráfico de barras con los recursos por sector (enfoque sectorial) y un gráfico de torta comparando la inversión programada en alineación con el ODS contra el total de recursos aprobados.

Resultados

El presente proyecto permitió clasificar exitosamente los 112 mil proyectos utilizados en el análisis, calculando para cada proyecto una medida de similitud con respecto a cada ODS y su correspondiente umbral de clasificación, resultado que se ilustra en la figura 3.

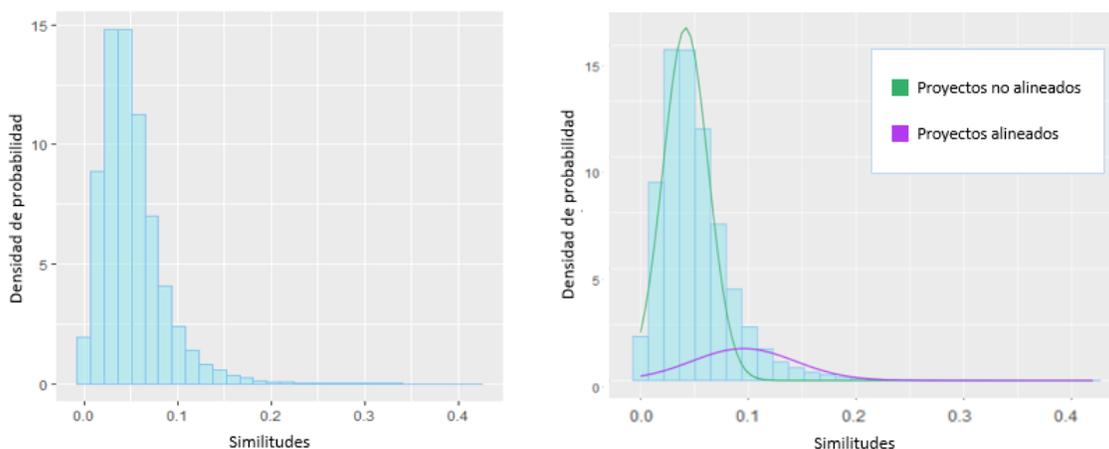


Figura 3: Distribución de la similitud coseno entre los textos de los proyectos de inversión y el texto del ODS 1 (izquierda) y distribuciones componentes de la mixtura de normales, estimadas con el algoritmo EM.

Un resultado de interés con las medidas calculadas se muestra en la tabla 2, que presenta los cinco proyectos cuyos textos tuvieron una mayor medida de similitud con respecto al del ODS 1, es decir, los cinco proyectos más alineados con el ODS 1 (fin de la pobreza).



El futuro es de todos

DNP
Departamento
Nacional de Planeación

BPIN	Nombre	Similitud ODS 1
2019502260011	Mejoramiento de viviendas como estrategia de erradicación de las privaciones multidimensionales de familias en condición de pobreza extrema y pobreza monetaria en la zona rural y urbana del municipio de Cumaral.	0.4195615
2015005810019	Apoyo para la superación de la pobreza extrema a familias acompañadas por la ANSPE en el departamento de Arauca.	0.4042324
2011011000131	Implementación de unidos red de protección social para la superación de la pobreza extrema – nacional.	0.3989361
2018684060106	Fortalecimiento, sensibilización y capacitación sobre inclusión laboral y social de la población vulnerable con perspectiva de género en el municipio de Lebrija.	0.3905803
2018058370050	Implementación del programa “Más Familias Fuera de la Pobreza Extrema” en el municipio de Turbo.	0.3897449

Tabla 2: Proyectos con mayor alineación con respecto al ODS 1.

Finalmente, en la figura 4 se muestran dos capturas de pantalla del aplicativo desarrollado.

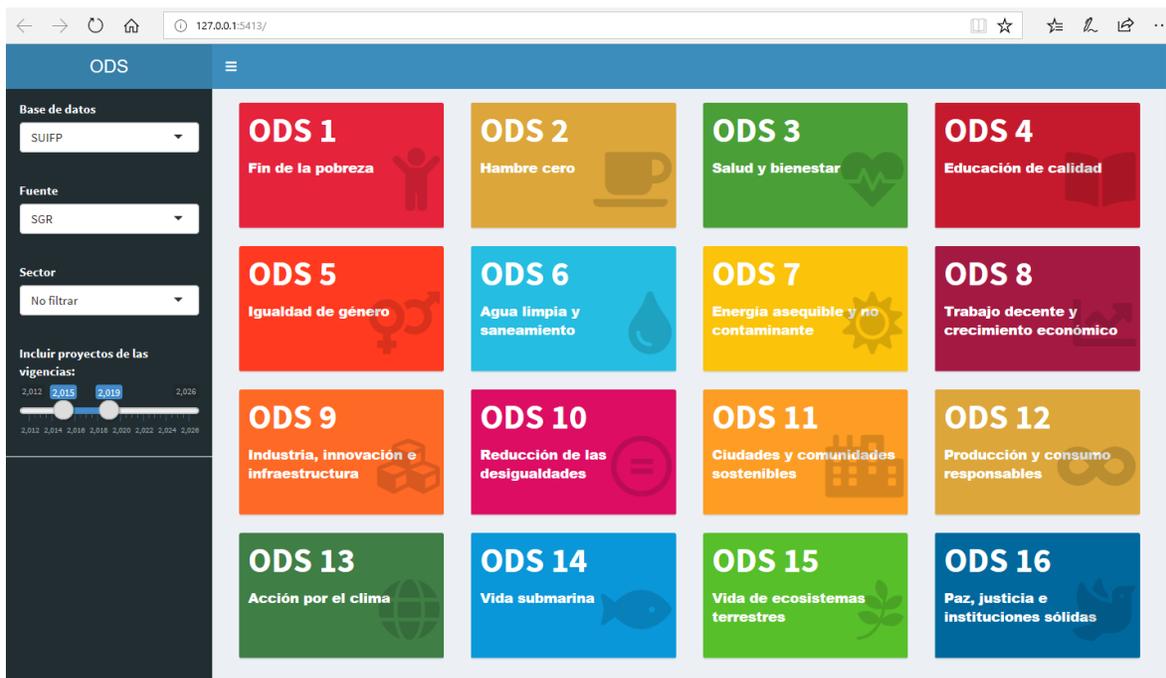


Figura 3: Vista inicial de la interfaz de usuario.



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

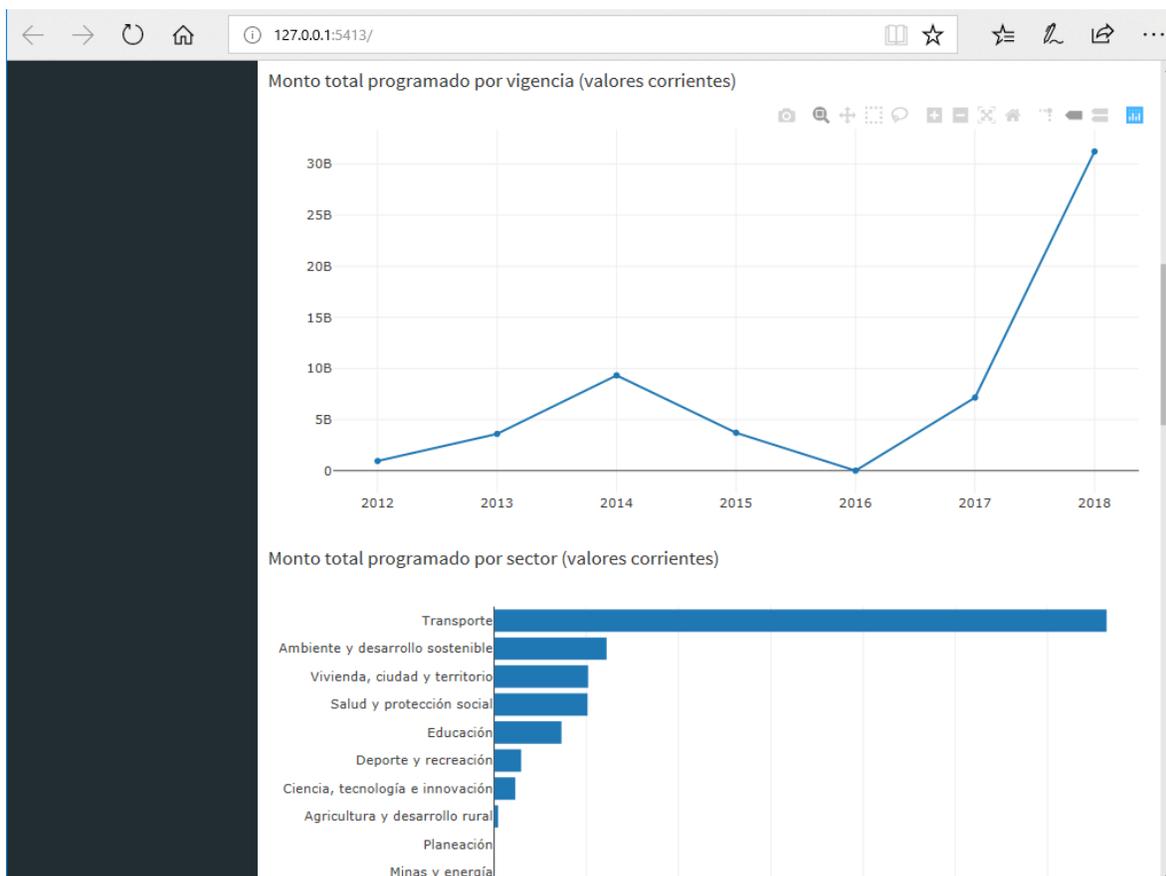


Figura 4: Vista de gráficos de inversión por año y por sector en la interfaz de usuario.

Conclusiones y recomendaciones

1. El presente proyecto permitió consolidar insumos valiosos para el análisis de inversión pública y de recursos de cooperación internacional alineados con los ODS, insumos que pueden ser utilizados para otros contextos en que la minería de texto represente una alternativa viable para facilitar la obtención de información sobre inversión pública o sobre ODS.
2. El uso de técnicas de análisis de texto, procesamiento de lenguaje natural y modelos estadísticos representa una buena alternativa para analizar la inversión pública en alineación con cada ODS, pues la metodología propuesta permitió realizar una clasificación de los proyectos de inversión en cada uno de los ODS de manera independiente, considerando que los proyectos de inversión pueden estar alineados con varios ODS.
3. El aplicativo desarrollado logra resumir los resultados de manera organizada, presentando la información de mayor valor para los tomadores de decisiones utilizando un enfoque territorial y un enfoque sectorial.