



Artista invitado

Fabio Andrés Arboleda Mendoza

Música, poder

De la serie *neWeed*

Acrílico sobre lienzo

20 x 20 cm

2020

Medellín



Analítica de datos aplicada al Plan Nacional de Desarrollo de 2018-2022, «Pacto por Colombia, pacto por la equidad»*

Enrique Arrieta Díaz (Colombia)**
Amanda Vargas Prieto (Colombia)**

Resumen

El Plan Nacional de Desarrollo (PND) 2018-2022 fue presentado durante la celebración de los sesenta años del Departamento Nacional de Planeación (DNP). Como el PND ha sido fuertemente criticado, nos interesa responder las siguientes preguntas: ¿cuáles son los temas más importantes del PND qué relación guardan con el discurso presidencial? La metodología utilizada se basa en algoritmos de analítica de datos, particularmente los que pertenecen al área de minería de texto. La analítica de datos aplicada a los documentos de políticas públicas ofrece una poderosa herramienta a la sociedad, la cual puede ser utilizada para comprender dichas políticas sin sesgos ideológicos. Este análisis encuentra que el tema más mencionado, entre los estudiados, es *territorio*, el menos mencionado es *paz*; los temas *equidad*, *legalidad* y *emprendimiento* no son tan relevantes en el texto como en el discurso presidencial, y la correlación entre estos términos no es relevante en el texto del PND. También se evidencia que los temas *salud*, *pobreza*, *infraestructura* y *seguridad* tienen poca trascendencia, así como las palabras *agro*, *cultura*, *infancia* y *energía*.

[215]

Palabras clave

Administración Pública; Políticas Públicas; Planeación; Analítica de Datos; Minería de Texto; Colombia.

* Artículo derivado del proyecto de investigación *Analítica de datos aplicada al Plan Nacional de Desarrollo de 2018-2022, «Pacto por Colombia, pacto por la equidad»*. Consejo Nacional de Planeación-CNP, Universidad del Magdalena, 2018-2022.

** Físico. Magíster en Física. Doctor en Física y Astronomía. Investigador asociado en Quantil. Profesor de la Universidad del Magdalena, Santa Marta, Colombia. Correo electrónico: earrieta@unimagdalena.edu.co - Orcid: 0000-0002-2408-3379 - Google Scholar: <https://scholar.google.com/citations?user=0wDQv8sAAAAJ&hl=es>

*** Administradora de Empresas. Magíster en Inteligencia Económica y Estrategias Competitivas. Doctora en Ciencias Económicas. Consejera Nacional de Planeación, representante del sector educativo y cultural. Profesora de la Universidad del Magdalena, Santa Marta, Colombia. Correo electrónico: avargasp@unimagdalena.edu.co - Orcid: 0000-0002-8508-6979 - Google Scholar: <https://scholar.google.com/citations?hl=es&user=-iKe3MgAAAAJ>

Fecha de recepción: agosto de 2020 • **Fecha de aprobación:** agosto de 2021

Cómo citar este artículo

Arrieta Díaz, Enrique y Vargas Prieto, Amanda. (2021). Analítica de datos aplicada al Plan Nacional de Desarrollo de 2018-2022, «Pacto por Colombia, pacto por la equidad». *Estudios Políticos* (Universidad de Antioquia), 62, pp. 215-239. <https://doi.org/10.17533/udea.espo.n62a09>

Data Analysis Applied to the National Development Plan for 2018-2022, «Pact for Colombia, Pact for Equity»

Abstract

The National Development Plan (PND) 2018-2022 was presented during the celebration of the sixtieth anniversary of the National Planning Department (DNP). The PND is strongly criticized by society on social media. For this reason, we are interested in finding what are the most important issues of the PND and how are they related to the president's speech. The methodology used is based on algorithms from data analysis, particularly those that belong to the text mining area. Analytics applied to public policy documents offer a powerful tool to society, which can be used to understand such policies without ideological biases. This analysis found that: the most mentioned topic, among those studied, is territory, the least mentioned is peace, the issues of equity, legality and entrepreneurship are not as relevant in the text as in the presidential speech, and the correlation between these terms is not relevant in the text of the PND. In addition, the topics health, poverty, infrastructure and security are not relevant, as well as the words agriculture, culture, childhood and energy.

[216]

Keywords

Public Administration; Public Politics; Planning; Data Analytics; Text Mining; Colombia.

Introducción

El Plan Nacional de Desarrollo (PND) es el documento que sirve de base y provee los lineamientos estratégicos de las políticas públicas formuladas por el presidente de la República de Colombia, Iván Duque Márquez, a través de su equipo de gobierno. Su elaboración, socialización, evaluación y seguimiento son responsabilidad directa del Departamento Nacional de Planeación (DNP) que se constituye como unidad rectora de planificación en el país, razón por la cual debe dirigir el proceso de formulación del PND 2018-2022 y coordinar la inclusión de los comentarios pertinentes por parte del Consejo Nacional de Planeación (CNP), el Consejo Nacional de Política Económica y Social (Conpes) y demás actores de la sociedad civil. El PND del presidente de la República se denomina: «Pacto por Colombia, pacto por la equidad».

El PND tiene tres componentes: la parte general o bases del PND, el plan de inversiones y las disposiciones instrumentales. La primera parte está dividida en dos: tres pactos estructurales y dieciséis pactos que representan las bases estructurales. Los pactos estructurales representan el pacto por la equidad: i) política social moderna centrada en la familia, eficiente, de calidad y conectada a mercados; ii) el pacto por el emprendimiento y la productividad, una economía dinámica, incluyente y sostenible que potencie todos nuestros talentos; iii) el pacto por la legalidad, justicia transparente y seguridad efectiva para que todos vivamos con libertad y en democracia. La segunda parte presenta el Plan Plurianual de Inversiones (PPI) 2019-2022, en la cual se estiman inversiones por COL\$1100 billones, de los cuales 66,6% corresponden a los recursos públicos y 33,4% a recursos privados. Los recursos públicos vienen del Presupuesto General de la Nación (PGN) (32,1%), recursos del Sistema General de Participaciones (SGP) (15,3%), recursos territoriales y de las empresas en el ámbito subnacional (11,7%), empresas estatales, industriales y comerciales del estado (EICE) (4,4%) y los recursos de regalías (3,1%) (DNP, 2018). El tercer componente «disposiciones instrumentales» está ausente en el documento (CNP, 2019).

[217]

Después de la publicación de las bases del PND, muchas han sido las críticas, en ese sentido, el Centro de Investigaciones para el Desarrollo (CID, 2019) de la Universidad Nacional de Colombia identificó «lo bueno, lo malo y lo feo» del Plan. Se pueden señalar dos preocupaciones principales sobre las propuestas: la primera se refiere al fortalecimiento territorial y la

continuación del modelo centralizador. Se advierte la debilidad del enfoque de capacidades, que no tiene propuestas para las falencias institucionales de los municipios y departamentos más vulnerables de Colombia, y las barreras normativas para capacitar personal como las causas de la falta de fortaleza institucional de las entidades territoriales. En otras palabras, se considera que la descentralización ha incidido en que las entidades territoriales sean consideradas como «gestoras de recursos» de inversión y esto no ha permitido que las Alcaldías y Gobernaciones tengan incidencia en las políticas que se ejecutan en sus territorios.

Por otra parte, es preocupante que el PND deje de lado el cierre de brechas —utilizado en el plan del gobierno de Juan Manuel Santos—, «una noción de desarrollo territorial que no responde a un enfoque de brechas regionales que deben ser compensadas, ni a un canal indirecto de política social focalizada espacialmente, sino que busca promover el canal directo del crecimiento (Baghwati, 1988) para fomentar mercados e impulsar nodos de desarrollo tanto urbanos como de zonas intermedias y rurales» (DNP, 2018, p. 910). Este nuevo plan propone, al contrario, el camino de desarrollo económico a través de la lógica de mercado, insistiendo en procesos de emprendimiento. Esto va en contravía de lo señalado en el Acuerdo de paz entre el Gobierno Colombiano y las Fuerzas Armadas Revolucionarias de Colombia-Ejército del Pueblo (FARC-EP), pues abandona la idea de disminuir las desigualdades y reconocer el conflicto colombiano (CID, 2019).

[218]

De acuerdo con estas preocupaciones, se propone como pregunta de investigación: ¿cuáles son los temas más importantes del Plan Nacional de Desarrollo (PND) y cuál es su relación con el discurso del presidente? La respuesta se encuentra con base en algoritmos de analítica de datos, particularmente los que pertenecen al área de minería de texto. Los algoritmos de la minería de texto transforman textos no estructurados en un formato estructurado, del cual se identifican patrones significativos y nuevos conocimientos que permiten explorar y descubrir relaciones ocultas dentro de sus datos no estructurados. El procesamiento del lenguaje escrito o del lenguaje natural (Chowdhury, 2005) utiliza métodos de diversas disciplinas como la informática, la inteligencia artificial y la lingüística para analizar textos extrayendo la información relevante de los documentos escritos. La minería de texto permite seleccionar las características significativas de un documento escrito para contribuir al máximo en la creación de modelos de análisis predictivo, seleccionar un subconjunto de características para

mejorar la precisión de una tarea de clasificación y seleccionar y categorizar entidades específicas en el texto, como nombres o ubicaciones. La analítica de datos aplicada a los documentos de políticas públicas, entre otros documentos gubernamentales, ofrece una poderosa herramienta a la sociedad, la cual puede ser utilizada para comprender dichas políticas sin sesgos ideológicos. De igual manera, la herramienta es de suma utilidad para los gobernantes, ya que a través de ella pueden verificar si sus discursos públicos son coherentes con los documentos legales que los describen, ya que son los documentos los que realmente determinan la hoja de ruta para el desarrollo del país.

1. Metodología de la investigación

La minería de datos es una ciencia que cada vez acapara más atención en la sociedad (Han, Pei y Kamber, 2011; Vargas y Yepes, 2018; Vargas y Arrieta, 2021). Esto se debe a su poder de reconocimiento de patrones y a su alto poder de predicción con una certidumbre medible. En las ciencias naturales, la minería de datos se volvió una herramienta fundamental de análisis, ya que por medio de esta las máquinas pueden reconocer patrones y correlaciones con mucha más facilidad que los humanos y en tiempos mucho más cortos.

[219]

El campo de la física de las altas energías —física de partículas elementales— es pionero en el campo de minería de datos. Experimentos como los que se llevan a cabo en la Organización Europea para la Investigación Nuclear (CERN) —laboratorio de física de partículas más grande del mundo— o en el Laboratorio Nacional Fermi (Fermilab) —laboratorio de física y acelerador de partículas más importante del continente americano— producen decenas de *petabytes*¹ de información cada año, los cuales nunca podrían ser analizados si solo se utilizara el capital humano de dichos experimentos. Científicos de estos laboratorios, apoyados en científicos del campo de la informática, han desarrollado técnicas, métodos y algoritmos para examinar con gran detalle y rigurosidad la gran colección de datos de los que disponen, y así poder analizarlos en tiempos cortos y con resultados y predicciones que superan ampliamente los que pueden alcanzar los humanos. Gracias a la minería de datos han sido posibles descubrimientos, analizando grandes cantidades de datos complejos, como los del bosón de Higgs, el *quark top* o las ondas gravitacionales.

¹ Equivalente a 10^{15} (un trillón de) bytes de información digital.

Dentro del campo de la minería de datos se encuentra la minería de texto (Guernsey, 2003, october 16), la cual se vale de diversos algoritmos de analítica de datos utilizando herramientas computacionales y matemáticas aplicadas para estudiar el lenguaje natural escrito y poder extraer información con mayor eficiencia y precisión que un humano. El objetivo principal de la minería de texto es resumir el contenido de un documento o colección de documentos, de forma que los seres humanos obtengan la mayor cantidad de información posible sin necesidad de leer el contenido escrito, aplicando técnicas de minería de datos-texto como *machine learning*, vectorización y mapeo de vecinos cercanos (Chirici et al., 2016), para obtener la información más relevante. Estos métodos no buscan escribir resúmenes que simplemente acortan el contenido de un escrito, resaltando sus aspectos más importantes, sino mostrar los temas más relevantes de un escrito, cómo se definen estos temas, sus correlaciones más importantes con otros temas, el sentimiento con el que se escribe el documento y otras estadísticas como palabras frecuentes y vecinos cercanos.

[220]

Los temas que se extraen de un documento pueden ser definidos por el analista a través del análisis supervisado (Way, Scargle, Ali y Srivastava, 2012) o pueden ser establecidos por los algoritmos en el denominado análisis no supervisado (Hebb, 1949). Ambas técnicas son útiles en la minería de texto y su aplicabilidad depende del objetivo del análisis. Incluso se pueden usar las dos técnicas combinadas, como se hizo en el presente estudio. Los temas, en algunos casos, se explican por sí mismos —una sola palabra los define—, por ejemplo, *equidad* o *paz*; en otros casos se requiere de algún complemento adicional que explique su significado, por ejemplo, *territorio* o *ambiental*. En este último caso, los algoritmos de minería de texto aquí desarrollados buscan un conjunto de términos que forman lo que se denomina *tema*. Siguiendo con el ejemplo anterior, el tema *ambiental*, se complementa con palabras relacionadas con la raíz *ambiente*, *ambientales*, *ambientalmente* y *ambientes*. De esta manera, se hace uso de la técnica de análisis supervisado, es decir, los autores definen los temas. En otros casos, como los de los temas *infraestructura* y *competitividad*, son definidos por el análisis no supervisado, el cual los identifica con base en su capacidad para descubrir similitudes y diferencias en la información.

La técnica de minería de texto aplicada al texto del PND se lleva a cabo en tres etapas: en primer lugar, el texto del PND se digitaliza y luego se le aplican algunos filtros para hacer más sencillo su análisis. En este proceso se descartan las referencias bibliográficas, los textos de tablas y gráficas, y las *stop words* —palabras que no aportan al contexto de un documento como artículos y preposiciones, entre otras—. Una vez se han aplicado estos filtros, la segunda etapa consiste en vectorizar el texto resultante con el fin de encontrar la frecuencia de aparición de cada palabra y los vecinos más cercanos. Un vecino cercano es una palabra —o conjunto de palabras— que aparece con frecuencia en la misma oración donde aparece uno de los temas estudiados. A manera de ejemplo: *Estos proyectos son demandados por entidades territoriales, esquemas asociativos locales, o son priorizados por el Gobierno Nacional.* El término *territorio* (territoriales), tiene como vecinos cercanos las palabras *proyectos, demandados, entidades, esquemas, asociativos y locales*. La cantidad de vecinos cercanos es definida por los investigadores. En este caso, se buscaron vecinos cercanos hasta la cuarta posición, es decir, palabras que estén a máximo cuatro palabras de distancia del término analizado dentro de una misma oración.

El concepto de vecinos cercanos permite producir grupos de palabras que aparecen juntas en el contexto del documento con cierta frecuencia. La cantidad de elementos en estos grupos es definida por los investigadores. La frecuencia de aparición para que estos grupos se formen y se tengan en cuenta es un parámetro libre en el análisis. En el presente análisis se han limitado el grupo de elementos a dos o tres. Estos elementos son parejas —bigrama— o tríos —trigrama— de palabras que aparecen juntas con una frecuencia dada. Al bigrama y al trigrama más frecuentes se les asigna un puntaje de 1 —esto no quiere decir que ambos tengan la misma frecuencia, ya que cada escala es independiente—. El resto de bigramas y trigramas tienen un puntaje que representa una frecuencia relativa con respecto al más frecuente de su categoría. Si un bigrama tiene un puntaje asignado de 0.5 quiere decir que aparece en el documento con la mitad de la frecuencia con la que aparece el bigrama más frecuente.

Finalmente, dentro del marco del análisis de los temas más importantes del texto del PND se produce una nube de palabras que le permite al lector comparar la frecuencia con la que estos aparecen en el documento. Una nube

[221]

de palabras es una figura donde cada palabra mostrada tiene un tamaño que es proporcional a su frecuencia de aparición en el documento. De esta manera se pueden comparar visualmente las frecuencias de aparición de palabras de interés. Con el fin de mostrar resultados relevantes para el lector, las nubes de palabras omiten aquellas que no aportan al contexto del análisis, pero que no pertenecen al conjunto de *stop words*. Dado que esta es una herramienta visual, se evita saturar la imagen con exceso de palabras, por lo que se establece un límite inferior para descartar palabras cuya frecuencia de aparición sea baja.

Cada uno de los temas establecidos o encontrados por los análisis supervisados y no supervisados tiene un conjunto de vecinos cercanos que le proporcionan un contexto. En un tema como la educación dentro del PND, sus vecinos cercanos permiten presentarle al lector el contexto dentro del cual se menciona el tema. Por ejemplo, si el tema *educación* se menciona frecuentemente con la palabra *calidad* o con la palabra *superior*, estas le dan al lector una noción sobre lo que el texto del PND quiere resaltar o enfatizar con respecto a la educación. De esta manera, los temas analizados presentan un conjunto de diez palabras cercanas —vecinos cercanos— que ilustran los contextos alrededor de los cuales se presentan dichos temas. Las frecuencias de aparición de estos vecinos cercanos disminuyen rápidamente, de manera que en la mayoría de los temas no vale la pena presentar más de diez vecinos cercanos. Estos vecinos cercanos se presentan junto con su frecuencia de aparición relativa al tema central que los agrupa. Por ejemplo, si un vecino cercano aparece con una frecuencia de 0.2 quiere decir que de cada cinco veces que su tema central se menciona el vecino cercano aparece una vez.

Los temas centrales del análisis pueden ser vecinos cercanos entre sí. La frecuencia de aparición en estos casos establece una correlación entre los temas que se puede cuantificar. Con el fin de presentar de manera sencilla dicha cuantificación, este análisis asigna un puntaje a cada correlación encontrada entre los temas centrales. A la correlación más fuerte se le asigna un puntaje de 1. Un puntaje de 0 implica que en el documento dos temas no tienen correlación. Cualquier valor intermedio del puntaje da a entender una correlación relativa entre los temas. Por ejemplo, si el tema 1 tiene una correlación de 0.5 con el tema 2 y el tema 3 tiene una correlación de 0.1 con el tema 4, esos puntajes quieren decir

[222]

que por cada vez que los temas 3 y 4 se correlacionan en el documento, los temas 1 y 2 se correlacionan cinco veces. Un ejemplo de correlación entre los temas *educación* e *investigación* se da cuando en una misma oración se mencionan ambos de manera conexas: *La investigación es un pilar fundamental en la educación superior*.

2. Resultados

El primer resultado del análisis se aprecia en la figura 1, una nube de palabras que presenta de manera visual los términos más frecuentes en el PND. De esta manera, se observa que el término más frecuente en el PND es *territorio*. El término más pequeño que se lee en la nube de palabras es *propiedad*, y su frecuencia de aparición es 5% de la frecuencia encontrada para *territorio*. Términos con una frecuencia menor a la de *propiedad* no se incluyen en la nube de palabras.

Figura 1. Nube con las palabras más frecuentes en el Plan Nacional de Desarrollo 2108-2022.



[223]

Fuente: elaboración propia a partir de DNP (2018).

Los términos que aparecen con un asterisco (*) representan varias raíces de una misma palabra: plurales, géneros o adverbios. Por ejemplo, el tema *territorio* está compuesto de *territorio*, *territorial*, *territorios*, *territoriales* y *territorialidad*. Algunos términos como *pública* o *público* se presentan por separado, ya que con frecuencia se considera que sus contextos son diferentes. El género de la palabra en el ejemplo distingue entre diferentes conceptos: el femenino se utiliza junto con palabras como *educación* o *deuda*, mientras que el masculino se utiliza junto con palabras como *sector* o *servicio*.

Teniendo en cuenta el contenido de la nube de palabras de la figura 1 y el discurso del presidente de la República durante el evento de celebración de los sesenta años del DNP el 14 de noviembre de 2018 (DNP Colombia, 14 noviembre de 2018), se decidió analizar los treinta temas que aparecen en la figura 2, dada su importancia tanto en el discurso presidencial como en el PND.

Figura 2. Temas seleccionados del Plan Nacional de Desarrollo 2108-2022.

[224]



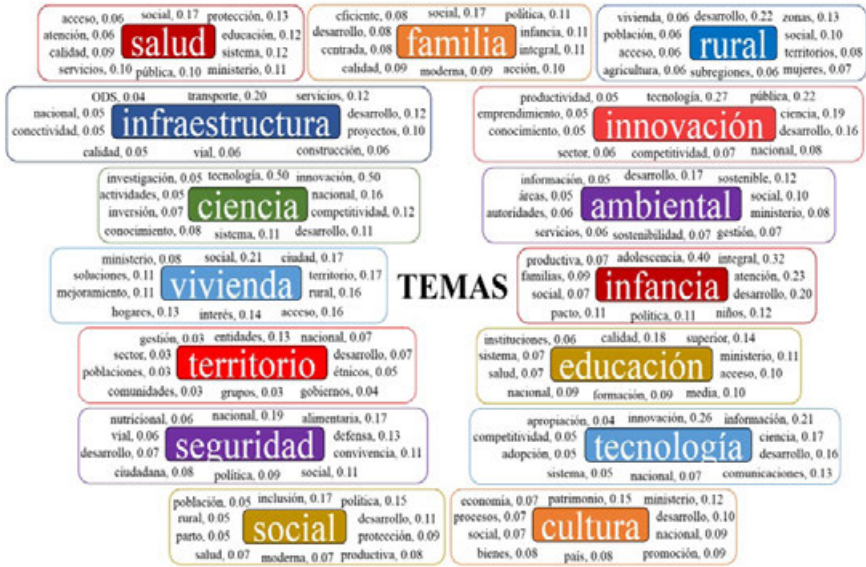
Fuente: elaboración propia a partir de DNP (2018).

Los verbos *implementar* y *fortalecer* se definen como temas con el fin de analizar qué campos son los más afectados por estas dos acciones. Después de estudiar con cuidado el PND, se observa que la frecuencia con la que cada uno de estos temas es mencionado es proporcional a su transversalidad en el PND y, por lo tanto, la frecuencia es proporcional a la importancia. Una vez más, el tamaño de los términos en la figura 2 es proporcional a su frecuencia de aparición en el PND. De aquí se concluye que el tema más recurrente en el PND es *territorio* y el menos frecuente es *paz*, con una frecuencia de tan solo 6% del tema *territorio*. Los temas *social* e *implementar* tienen una frecuencia aproximada de 60% respecto del tema más frecuente, ocupando el segundo y tercer lugar, respectivamente. Los temas *fortalecer* y *economía* tienen una frecuencia aproximada de 50% de la frecuencia máxima, ocupando el quinto y sexto lugar, respectivamente. Los temas centrales del PND, según la ecuación $equidad = legalidad + emprendimiento$ tienen cada uno una frecuencia inferior a 20% —*equidad*, 13,9%, *legalidad*, 12,5% y *emprendimiento*, 19,5%—. Los temas *salud*, *pobreza*, *infraestructura* y *seguridad* tienen una frecuencia cercana a un cuarto de la máxima, mientras que los temas *agro*, *cultura*, *infancia* y *energía* tienen una frecuencia cercana a un quinto de la máxima. Los temas *investigación* (12,2%), *ciencia* (12,1%) y *justicia* (11,6%) ocupan los últimos lugares de importancia en el PND y solo superan al tema *paz*.

[225]

La vectorización del PND permite encontrar los vecinos más cercanos a cada término en el documento. Con base en este concepto, se encontraron los diez vecinos más cercanos a los términos que definen cada uno de los treinta temas (véanse figura 3 y figura 4). Esta cercanía se mide por medio de la frecuencia de aparición de un vecino cercano junto con su término (tema). Esta frecuencia se presenta como un número de 0 a 1, al lado derecho de cada vecino cercano. Los resultados del análisis de vecinos cercanos se presentan en estas dos figuras y no en una sola para facilitar su lectura. Los datos de cada tema en las dos figuras deben leerse de manera independiente. Los datos relacionados con cada tema se delimitan con una línea del mismo color al del tema. En estas dos figuras la ubicación de un tema particular en la gráfica no es relevante y la aparición de un tema en una figura o en la otra es aleatoria.

Figura 3. Vecinos cercanos al primer grupo de quince temas del Plan Nacional de Desarrollo 2108-2022.



Fuente: elaboración propia a partir de DNP (2018).

[226]

Figura 4. Vecinos cercanos al segundo grupo de quince temas del Plan Nacional de Desarrollo 2108-2022.



Fuente: elaboración propia a partir de DNP (2018).

En las figuras 3 y 4 se observa que las palabras con una correlación más fuerte a su tema son *tecnología* e *innovación* en el tema *ciencia*, con una frecuencia de 50%. Esto quiere decir que de cada 10 veces que se trata el tema de *ciencia*, 5 veces se mencionan las palabras *tecnología* e *innovación*. La segunda correlación más fuerte se da entre la palabra *adolescencia* y el tema *infancia*, con una frecuencia de 40%. La tercera correlación más fuerte se da entre la palabra *integral* y el tema *infancia*, con una frecuencia de 32%. Por otro lado, la correlación más alta del tema *fortalecer* se da con la palabra *capacidades*, con una frecuencia de 10%. Esa relación se resalta ya que en su categoría (tema-vecino cercano) es la más baja.

Es importante leer con atención los vecinos cercanos de cada tema, ya que estos definen en qué tipo de contextos aparecen los treinta temas estudiados. En el tema *pobreza* se entiende que el PND plantea formas de reducir (reducción) o superar (superación) la pobreza extrema. En el tema *educación* se ve cómo se acompaña este tema con conceptos como *calidad* y *acceso*. En el tema *salud* también se habla de *calidad* junto con *atención* y *protección*. En el tema *paz* se habla de *estabilización* y de *víctimas*. En el tema *emprendimiento* se habla de *acceso* e *inclusión*. En el tema *legalidad*, que incluye también el concepto *ilegalidad*, se menciona el concepto *marco* y se habla de *grupos ilegales*. En el tema *justicia* se prioriza el *derecho* a su *acceso*. El foco del tema *competitividad* es el concepto *regional*, mientras que el foco del tema *economía* es *economía naranja* con base en el crecimiento sostenible. El concepto *equidad* se relaciona con una *política social moderna* que ofrezca *oportunidades de desarrollo*. El tema *seguridad* tiene dos facetas: la *alimentaria* y la que tiene que ver con el concepto *seguridad ciudadana*. El tema *tecnología* hace énfasis en *comunicaciones* e *información*. El tema *agro* hace alusión a actividades *productivas*. En el PND se habla de *fortalecer: capacidades, estrategias y programas*; así como se habla de *implementar: estrategias, proyectos y programas*. El concepto *objetivos de desarrollo sostenible (ODS)* es vecino cercano de *infraestructura*, *justicia* y *paz*. ODS es una de las palabras clave que aparecen en la figura 1, con una frecuencia de aparición que representa 10,5% de la frecuencia de aparición del término *territorio* (véase [anexo 1. Tabla 1. Matriz de correlación](#)).

El concepto *desarrollo* es transversal a la mayoría de los treinta temas estudiados (23), ocupando los primeros lugares de importancia entre los vecinos cercanos. No en vano es el segundo término más popular de la figura 1. Sin embargo, resulta curioso que ese concepto no venga acompañado más

[227]

de cerca del concepto *inversión*, pues solo aparece entre los vecinos cercanos del tema *ciencia*. Si se analiza *inversión* como un tema independiente, se aprecia una relación no muy alta con *ciencia*, *tecnología*, *innovación* (CTI) entre sus diez vecinos más cercanos. Dentro de este mismo grupo también se encuentran *privada* y *extranjera*. No se ve *inversión* como vecino cercano frecuente de *salud* o *pobreza*.

Anteriormente se omitió la discusión de temas como vecinos cercanos de otros temas. Es interesante revisar si los treinta temas son vecinos cercanos entre sí, por lo que se construyó la matriz que se presenta en el [anexo 1, matriz de correlación](#).

La forma de leer esta matriz es escoger un tema en la fila superior y mirar los puntajes que aparecen dentro de esa columna, frente a cada tema de los que aparecen en la primera columna del lado izquierdo. La relación más fuerte ocurre entre los temas *territorio* y *desarrollo*, por eso se le ha asignado un puntaje de 1. La fuerte relación ocurre porque el tema *innovación* es el vecino cercano que con más frecuencia aparece junto con el tema *tecnología* dentro del conjunto de treinta temas, por esta razón se le asigna un puntaje de 1. La escala de puntajes va de 0 a 1 y los puntajes intermedios son proporcionales a la frecuencia de aparición de los temas en la primera fila con respecto a cada uno de los temas en la primera columna.

[228]

Las correlaciones más fuertes de cada tema son:² *territorio-fortalecer*, *social-territorio* (productividad), *implementar-fortalecer*, *productividad-social* (territorio), *fortalecer-productividad*, *economía-social* (productividad), *rural-social* (productividad), *educación-salud*, *ambiente-social* (economía), *innovación-ciencia*, *empresa-productividad* (territorio), *salud-social*, *pobreza-rural*, *seguridad-social* (legalidad), *infraestructura-economía*, *cultura-economía*, *energía-productividad* (territorio), *infancia-social* (familia), *agro-productividad* (territorio), *emprendimiento-productividad* (territorio), *familia-social* (infancia), *competitividad-territorio* (productividad), *vivienda-social* (rural), *equidad-social* (productividad), *legalidad-economía*, *investigación-innovación*, *ciencia-innovación*, *justicia-legalidad* y *paz-justicia*. Es interesante notar la baja correlación entre *investigación-ciencia*,

² Cuando el vecino más cercano es: social, se presenta entre paréntesis el segundo vecino más cercano. También se presenta un vecino cercano entre paréntesis si hay dos vecinos cercanos con el mismo puntaje en el primer puesto.

agro-ambiente, pobreza-equidad y paz-cualquier tema entre las correlaciones bajas más notorias.

Retomando el eje central del PND, la ecuación que define *equidad*, se debe resaltar que la correlación entre *equidad* y *legalidad* tiene un puntaje de apenas 0.14 (0.07); mientras que la correlación entre *equidad* y *emprendimiento* es tan solo 0.04 y la correlación entre *legalidad* y *emprendimiento* es de 0.09 (0.04). Estos resultados no respaldan la definición que se da en el PND sobre equidad, ya que los tres temas no van de la mano con una frecuencia relevante. De la misma manera, las pocas correlaciones entre equidad y la mayoría de los otros veintinueve temas dejan ver que, en realidad, la equidad no hace parte de las conversaciones que se llevan a cabo sobre la mayoría de los otros temas importantes.

2.1 Análisis del tema *Educación* en el Plan Nacional de Desarrollo

El tema *educación* se considera aquí como uno de los más importantes dentro del marco de referencia de un país en vías de desarrollo, y por eso se le dedica una sección especial. Una población educada conoce sus debilidades y enfoca sus esfuerzos colectivos para fortalecerlas. Un electorado educado es capaz de elegir las propuestas adecuadas para el desarrollo de su sociedad, distinguiéndolas de aquellas que presentan planes vacíos, desviados de la realidad o irrealizables. En este mismo sentido, la Misión de Sabios Colombia 2019, referente al tema de desarrollo social con equidad, afirmó que la sociedad colombiana «requiere generar y transformar conocimiento en ciencias humanas, sociales y educación útil para la sociedad y el país, puesto que crea herramientas fundamentales para el desarrollo humano» (Ministerio de Ciencia, Tecnología e Innovación, 2020, p. 25). Por esta razón, este grupo de intelectuales colombianos recomendó al Estado colombiano garantizar una educación de calidad como un medio para la solución de necesidades presentes en sectores clave de la sociedad como el político, el social, el económico y el cultural. Asimismo, en su diagnóstico del PND, los sabios entienden que a través de la educación y la ciencia se pueden encontrar respuestas a problemáticas ya identificadas por el Estado colombiano.

[229]

Tal vez la recomendación más significativa de la Misión de Sabios Colombia 2019 tiene que ver en el sector de la educación. Ellos consideran «como eje fundamental la creación de modelos etno-educativos, modelos

educativos flexibles y modelos educativos con enfoque de ruralidad que mejoren las experiencias pedagógicas y a su vez contribuyan a cerrar la brecha frente a la inequidad y la inclusión social» (Ministerio de Ciencia, Tecnología e Innovación, 2020, p. 51). Dado que el enfoque del PND es precisamente la equidad, resulta interesante estudiar en detalle la pertinencia que tiene la educación en este, sobre todo, como herramienta contra la inequidad.

Con respecto al análisis del sector educativo y cultural del CNP, se revisaron todos los pactos (19) para encontrar aquellos en los que se trataba el tema. En los pactos estructurales se encontró únicamente en el pacto por la equidad, específicamente en los puntos B, C y G. En los pactos transversales, el VII hace propuestas sobre ciencia, tecnología e innovación. A continuación, se presenta el análisis de los dos pactos, teniendo en cuenta únicamente la parte denominada «estrategias» del PND. Este análisis se realiza usando los criterios de la ficha de recolección de insumos trabajada por el CNP: aclaración, modificación, inserción y eliminación. La figura 5 presenta los vecinos cercanos al tema *educación* del puesto 11 al 30, para complementar los primeros 10 que se presentaron en la figura 3.

[230]

Figura 5. Vecinos cercanos al tema *educación* del Plan Nacional de Desarrollo 2108-2022.



Fuente: elaboración propia a partir de DNP (2018).

El foco del tema de la educación en el PND es el acceso a una educación superior y media de calidad. De una manera menos recurrente se aprecia la aparición de la educación básica-preescolar. Esto se podría justificar porque en el plan la educación se reorienta hacia la inserción al mundo productivo (CID, 2019). En este aspecto, las propuestas entran en contradicción con la Ley de Educación (Ley 115 de 1994), pues los principios proclamados por dicha ley son el pluralismo ideológico, la universalidad de saberes y la capacidad crítica, y en este sentido el plan los reduce al Sistema Nacional de Cualificaciones (DNP, 2018). En el análisis de minería de texto se encontró que la formación de talento humano por medio de programas para generar oportunidades también desempeña un papel significativo en este tema. La inclusión social, tan recurrente en el tema *social*, hace énfasis en el acceso a educación y salud de calidad, y por eso se da la alta correlación entre estos dos temas.

Se evidencia también que los temas *cobertura* e *inclusión* son clave en el desarrollo de los ODS. De hecho, la iniciativa de la economía naranja se encuentra en las metas planteadas en los ODS, por esto se resalta la educación de calidad, el trabajo decente y crecimiento económico, y el de industria, innovación e infraestructura. Brilla por su ausencia en los vecinos cercanos la palabra *inversión*, la cual no aparece aquí ni en el conjunto de vecinos cercanos definidos al cortar las palabras con una frecuencia superior a 10%. Aunque *educación superior* casi que tiene 15% de la atención en el tema *educación* en el PND, también se nota la ausencia de la palabra *universidad* en todo el conjunto definido de vecinos cercanos. Tal vez la sorpresa más grande en el tema *educación* es la ausencia de correlaciones significativas entre *educación* y *ciencia*, *investigación* e *innovación*, como se deduce del [anexo 1, matriz de correlación](#). Esta falta de correlaciones sugiere que en el PND no hay una conexión entre estos tres temas que deberían estar altamente relacionados en un país en vías de desarrollo. Esto se podría explicar por el enfoque de la propuesta de la economía naranja de instrumentalizar la intervención del capital humano en el Sistema Nacional de Cualificaciones, pues esto

[231]

Deja de lado el papel fundamental de las instituciones de educación superior, no solo en la provisión de la formación sino en la posibilidad de establecer estrategias de emprendimiento incubados desde los centros de formación. En este sentido, la propuesta iniciaría con problemas de fallas de coordinación entre la demanda y el sistema educativo, pero también deja de lado el concepto del ODS industria, innovación e infraestructura, fundamentado en la tradicional definición de la triple

hélice: gobierno, empresa y universidad, requisito necesario para la conformación de las Áreas de Desarrollo Naranja - ADN (CID, 2019, p. 108).

2.2 Análisis de los grupos de palabras

Continuando con la búsqueda de correlaciones significativas, se presentan dos productos que resultan del ejercicio de vecinos cercanos: bigramas y trigramas. Los veinte bigramas y trigramas más frecuentes se presentan en la tabla 2.

Tabla 2. Bigramas y trigramas más frecuentes del Plan Nacional de Desarrollo 2108-2022.

Bigramas	Pun.	Trigramas	Pun.
gobierno, nacional	1,00	ciencia, tecnología, innovación	1,00
grupos, étnicos	0,74	niños, niñas, adolescentes	0,82
entidades, territoriales	0,74	meta, cuatrienio, 2022	0,78
inclusión, social	0,60	inclusión, social, productiva	0,77
política, pública	0,52	tecnologías, información, comunicaciones	0,75
meta, cuatrienio	0,52	comercio, industria, turismo	0,72
cambio, climático	0,51	política, social, moderna	0,65
economía, naranja	0,44	departamento, nacional, planeación	0,56
servicios, públicos	0,44	plan, nacional, desarrollo	0,54
desarrollo, sostenible	0,41	salud, protección, social	0,54
atención, integral	0,41	ambiente, desarrollo, sostenible	0,51
ordenamiento, territorial	0,41	agua, potable, saneamiento	0,49
ciencia, tecnología	0,41	agricultura, desarrollo, rural	0,47
bienes, servicios	0,40	atención, integral, infancia	0,44
plan, nacional	0,39	potable, saneamiento, básico	0,42
política, social	0,38	2017, meta, cuatrenio	0,41
asistencia, técnica	0,38	entidades, orden, nacional	0,39
infancia, adolescencia	0,37	nivel, nacional, territorial	0,38
toma, decisiones	0,36	territorios, gobiernos, poblaciones	0,33
tecnología, innovación	0,36	integral, infancia, adolescencia	0,33
*Puntaje (Pun) de 0 (menos frecuente) a 1 (más frecuente).			

Fuente: elaboración propia a partir de DNP (2018).

Vale la pena resaltar que entre los bigramas más frecuentes se encuentra *economía naranja*, concepto que frecuentemente aparece como motor de desarrollo en el discurso del presidente de la República. Asimismo, la importancia de los ODS se ve reflejada en la presencia de *desarrollo sostenible* en la tabla 2. De igual forma, la preocupación por el cambio climático es evidente en el PND. Alrededor de este asunto se encuentran vecinos cercanos como *adaptación, mitigación, riesgo, gestión y desastres*. Un tema cercano conceptualmente al cambio climático es *ambiente*, este tema también tiene presencia en la tabla 2, donde se ve acompañado de *desarrollo sostenible*. El conjunto *ciencia, tecnología, innovación* (CTI) es el trigramma más frecuente. Este resultado contrasta con el hecho de que la *ciencia* ocupa el puesto 28 e *innovación* el puesto 13 entre los treinta temas estudiados. *Equidad, legalidad y emprendimiento* no aparecen en la tabla 2, lo que corrobora que estos conceptos no tienen altas correlaciones con otras palabras en el PND.

2.3 Discurso presidencial versus Plan Nacional de Desarrollo

Finalmente, se hace una comparación entre los temas que el presidente de la República resaltó como de suma importancia en su discurso y lo que se encontró en el PND. Esta comparación permite identificar algunos hechos estilizados que facilitarían la reflexión por parte de los lectores.

[233]

- *Ingresos altos*. Este bigrama no aparece en el PND; sin embargo, aparecen bigramas como *ingresos dignos, ingresos bajos, ingreso familia, aumentar ingresos* y *generación ingresos*, con puntajes relativos entre 0.14 y 0.04.
- *Crecimiento clase media*. Las combinaciones posibles de dos o más de estas tres palabras tienen una participación muy pobre en el PND, con puntajes inferiores a 0.05.
- *La informalidad vista como una gran amenaza a la sostenibilidad del régimen pensional, al sistema de salud y al sistema fiscal*. La palabra *informalidad* tiene una frecuencia de aparición de 26% de la frecuencia de *propiedad*, que es el término menos frecuente presentado en la figura 1. Asimismo, los conceptos *sistema de salud* y *sistema de seguridad social* tienen unos puntajes en la escala de bigramas (trigramas) inferiores a 0.01.
- *Micro, pequeña y mediana empresa*. Los bigramas y trigramas que aparecen con las palabras mencionadas por el presidente tienen puntajes

inferiores a 0.01; por otro lado, estas tres palabras no aparecen entre los veinte primeros vecinos cercanos del tema *empresa*.

- *Generación de empleo y mejor remuneración pensando en la productividad*. Las palabras *desempleo*, *empleos* y *remuneración* tienen cada una frecuencia de aparición muy cercanas a 25% de la frecuencia de *propiedad*.
- *Cerrar brechas*. Este bigrama tiene un puntaje inferior a 0.01 y tiene como vecinos cercanos *talento humano* y *movilidad social*.
- *Aumento de la demanda de energía*. El bigrama *eficiencia energética* tiene un puntaje inferior a 0.01. El tema *energía* tiene un norte definido, ya que como vecinos cercanos se encuentran *consolidación productiva*, *mercados*, *renovables convencionales*, *fuentes alternativas* y *promover competitividad*.
- *Cultivos ilícitos*. El bigrama tiene un puntaje menor a 0.01. El concepto *ilícito* aparece con más frecuencia asociado a temas de minería ilegal, aunque esta frecuencia es comparativamente baja con otros temas tratados aquí.
- *Transparencia para la generación de confianza*. La palabra *transparencia* tiene una frecuencia de aproximadamente 66% de *propiedad*. Como vecinos cercanos están *acceso a información*. La palabra *confianza* tiene una frecuencia de aparición muy cercana a 25% de la frecuencia de *propiedad*. No tiene vecinos cercanos relevantes.
- *Trámites innecesarios*. El bigrama no tiene importancia en el PND. La palabra *trámites* tiene una frecuencia levemente inferior a *propiedad* y sus vecinos cercanos más relevantes son *solicitudes* y *racionalización*.
- *Corrupción*. Esta palabra tiene una frecuencia de aproximadamente 66% de *propiedad*. *Anticorrupción* y *corruptos* tienen frecuencias mucho menores. Como vecino cercano tiene *lucha*.
- *Conflictos e impunidad*. El tema *conflicto* tiene una frecuencia muy parecida a la de *propiedad*, mientras que la palabra *impunidad* tiene una frecuencia de 10% de esta. *Conflicto* tiene como vecinos cercanos *armado*, *resolución*, *socioambientales* y *víctimas*. El tema *víctimas* tiene una frecuencia muy similar a la del tema *legalidad*, con vecinos cercanos como *mujeres*, *unidad*, *reparación*, *población*, *atención*, *política* y *conflicto armado*.
- *Reinserción, reincorporación y reintegración*. El conjunto de estas tres palabras tiene una frecuencia de aproximadamente 75% de *propiedad*. Como vecino cercano tiene *víctimas*.

[234]

- *Proyectos productivos*. El bigrama tiene una frecuencia inferior a 0.01.
- *Contrabando y piratería*. La frecuencia combinada de estas dos palabras es muy similar a 10% de *propiedad*.
- *Impuestos e inflación*. La frecuencia combinada de estas dos palabras es muy similar a 25% de *propiedad*, con una contribución mayor *impuestos*.
- *Campo y campesinos*. La frecuencia del concepto *campesino* es muy similar a 10% de *propiedad*. No es sencillo distinguir la palabra *campo* en este contexto de cuando se usa en otros; sin embargo, la palabra *campo*, sin distinción, tiene una frecuencia baja, similar a 50% de *propiedad*.
- *Gratuidad*. La palabra es irrelevante en el PND.
- *Exportaciones*. La frecuencia del tema *exportaciones* es aproximadamente la mitad de *propiedad*. Como vecinos cercanos tiene *agropecuarias*, *fomentar* y *atracción de inversiones*.

Este análisis de los términos utilizados en el documento de las bases del PND permite visualizar las propuestas realizadas por el gobierno de Iván Duque Márquez y validar las políticas públicas que impulsarían el crecimiento económico más allá del desarrollo social. Se identifica en el plan, como políticas directrices, la economía naranja, la regulación y aceleración de los proyectos de 4G, el impulso a la agroindustria y el mayor gasto público en educación y temas a cargo del Departamento de Prosperidad Social (DNP, 2018, pp. 1061-1062).

[235]

Conclusiones

El primer resultado del análisis de los temas más significativos, tanto del PND como del discurso del presidente, permiten concluir que el tema más mencionado es *territorio* y el menos mencionado es *paz*. En el actual PND los temas centrales, según el Gobierno nacional, son *equidad*, *legalidad* y *emprendimiento*. Este análisis muestra una diferencia notable entre la importancia de estos. El tema *emprendimiento* tiene una frecuencia de aparición de 19,5%, mientras que *equidad* y *legalidad* tienen frecuencias de aparición de 13,9% y 12,5%, respectivamente. El análisis de correlación entre los temas *equidad* y *legalidad* tiene un puntaje de apenas 0.14, mientras que la correlación entre *equidad* y *emprendimiento* es tan solo 0.04, y la correlación entre *legalidad* y *emprendimiento* es de 0.09. La diferencia de aproximadamente 50% entre las frecuencias de aparición de *emprendimiento* y los otros dos temas, junto con las bajas

correlaciones entre los tres temas, alrededor de 10%, permite concluir que los tres temas realmente no son el núcleo del PND, como lo argumenta el Gobierno nacional. De la misma manera, las pocas correlaciones entre *equidad* y la mayoría de los otros veintinueve temas dejan ver que, en realidad, la equidad no hace parte de las conversaciones que se llevan a cabo sobre la mayoría de los otros temas importantes del PDN.

Los temas *salud, pobreza, infraestructura y seguridad* tienen frecuencias de aparición bajas —un cuarto de la máxima—, algo similar a lo que ocurre con los temas *agro, cultura, infancia y energía* —un quinto de la máxima—. Estos temas con un cuarto y un quinto de frecuencia de aparición en comparación con el tema más relevante. Debería tener mayor importancia dentro del PND, teniendo en cuenta los ODS con los que se ha comprometido Colombia.

[236] Es interesante notar que el tema *paz* se encuentra en último lugar de importancia en el PND, acompañado de cerca, en los últimos lugares, de otros temas como *investigación, ciencia y justicia*. Estas posiciones no son coherentes con la importancia que le ha dado la sociedad colombiana a la paz y a la justicia en la última década. Asimismo, *investigación y ciencia* no tienen la importancia en el PDN que sí le da la Misión de Sabios Colombia 2019 en sus recomendaciones.

Con respecto al análisis realizado con los vecinos cercanos, se puede concluir que el concepto *desarrollo* es transversal a la mayoría de los treinta temas estudiados; sin embargo, resulta curioso que ese concepto no venga acompañado más de cerca del concepto *inversión*, pues solo aparece entre los vecinos cercanos del tema *ciencia*. Si se analiza *inversión* como un tema independiente, tampoco se aprecia una relación significativa con *ciencia, tecnología, innovación* (CTI) entre sus diez vecinos más cercanos. Dentro de este mismo grupo de términos también se encuentran *privada* y *extranjera*. Tampoco se observa el tema *inversión* como vecino cercano frecuente de *salud o pobreza*, lo que va en contravía con los ODS.

Otro resultado significativo es el análisis del tema *educación* en el PND. En este se encontró la ausencia del tema *inversión* en el conjunto de los vecinos cercanos, así como la baja presencia de la palabra *universidad*. El resultado más importante fue la ausencia de correlaciones significativas entre *educación y ciencia, investigación e innovación*.

Finalmente, se observa que los temas que el presidente de la República resaltó en su discurso como claves en su PND no son todos necesariamente los más importantes en el texto del PND. Los temas *cultura, vivienda, familia, infancia, energía y empresa*, a los que el presidente de la República les dio suma importancia, aparecen dentro de los treinta temas investigados después del puesto 18, con excepción del tema *empresa*, que aparece en el puesto 14. La importancia del tema *víctimas* es comparable con los temas analizados menos trascendentales. El resto de los temas mencionados en la comparación entre el discurso presidencial y el PND están muy por debajo en importancia al compararlos con los 26 temas estudiados.

Por último, es oportuno aclarar que el análisis descrito aquí es solo un resumen de toda la información que se puede extraer de las diferentes tablas y figuras, las que a su vez son un resumen de toda la información analizada. Toda la información analizada no se presenta en su totalidad, sino solo sus aspectos más relevantes. La lectura rigurosa de las ayudas visuales aquí presentadas proporciona un entendimiento más profundo del PND del que se ha descrito en todas las secciones del artículo. El análisis cualitativo de estos resultados permite concluir cuáles son los temas de mayor relevancia para las políticas públicas del gobierno actual. También permite identificar su posicionamiento con respecto a temas fundamentales del desarrollo de Colombia, como lo es la educación, entre otros temas.

[237]

Este tipo de análisis de texto permite comparar la información de los PND con los discursos políticos y es una herramienta de suma importancia para la creación y aprobación de futuros planes de desarrollo nacionales o territoriales, pues permite a la academia y a la opinión pública analizar más a fondo estos documentos que muchas veces no son consultados o contruidos con la sociedad civil.

Invitamos a los científicos de datos, físicos y matemáticos, a continuar con este esfuerzo por realizar un trabajo interdisciplinar que incluya expertos en políticas públicas, en gobierno, en diplomacia, en economía, pero también a entidades como el Consejo Nacional de Planeación y a los Consejeros Territoriales de Planeación del país a sumarse en estos estudios rigurosos que nos ofrece la academia y que pueden ser aprovechados en la mejora de la planeación participativa de Colombia.

Referencias bibliográficas

1. Baghwati, Jagdish N. (1988). Poverty and Public Policy. *World Development*, 16 (5), pp. 539-555. [https://doi.org/10.1016/0305-750X\(88\)90184-2](https://doi.org/10.1016/0305-750X(88)90184-2)

2. Centro de Investigaciones para el Desarrollo (CID). (2019). Invitados, vips y excluidos en el gran pacto para Colombia. Análisis crítico del plan nacional de desarrollo 2018-2022 «Pacto por Colombia, pacto por la equidad». *Investigaciones y Productos CID*, 28. <http://www.fce.unal.edu.co/media/files/CentroEditorial/documentos/investigacionesCID/documentos-CID-28.pdf>

3. Colombia. Congreso de la República. Ley 115. (8 de febrero de 1994). Por la cual se expide la ley general de educación. https://www.mineducacion.gov.co/1621/articles-85906_archivo_pdf.pdf

4. Consejo Nacional de Planeación (CNP). (2019). Concepto a las bases del Plan Nacional de Desarrollo 2018-2022. «Pacto por Colombia, pacto por la equidad». <https://colaboracion.dnp.gov.co/CDT/Prensa/PlanNacionaldeDesarrolloconceptoCNP%202018-2022.pdf>

5. Chirici, Gherardo et al. (2016). A Meta-Analysis and Review of the Literature on the k-Nearest Neighbors Technique for Forestry Applications that Use Remotely Sensed Data. *Remote Sensing of Environment*, 176, pp. 282-294. <https://doi.org/10.1016/j.rse.2016.02.001>

[238]

6. Chowdhury, Gobinda G. (2005). Natural Language Processing. *Annual Review of Information Science and Technology*, 37 (1), pp. 51-89. <https://doi.org/10.1002/aris.1440370103>

7. Departamento Nacional de Planeación (DNP). (2018). Bases del Plan Nacional de Desarrollo 2018-2022. «Pacto por Colombia, pacto por la equidad». Bogotá, D. C.: DNP. <https://colaboracion.dnp.gov.co/CDT/Prensa/BasesPND2018-2022n.pdf>

8. DNP Colombia. (14 noviembre de 2018). DNP 60 años - Entrega bases Plan Nacional de Desarrollo [Video]. *Youtube*. https://www.youtube.com/watch?v=42_yCaHa620&ab_channel=DNPColombia

9. Guernsey, Lisa. (2003, october 16). Digging For Nuggets Of Wisdom. *The New York Times*. <https://www.nytimes.com/2003/10/16/technology/digging-for-nuggets-of-wisdom.html>

10. Han, Jiawei; Pei, Jian & Kamber, Micheline. (2011). *Data Mining: Concepts and Techniques*. Massachusetts: Elsevier

11. Hebb, D. O. (1949). *The Organization of Behavior. A Neuropsychological Theory*. New York: John Wiley & Sons.

12. Ministerio de Ciencia, Tecnología e Innovación. (2020). *Colombia hacia una sociedad del conocimiento. Reflexiones y propuestas. Volumen 1*. Bogotá, D. C.: Ministerio de Ciencia, Tecnología e Innovación. https://minciencias.gov.co/sites/default/files/upload/paginas/ebook-_colombia_hacia_una_sociedad_del_conocimiento.pdf

13. Vargas-Prieto, A., & Arrieta-Díaz, E. (2021). Risks of organizational growth to the identity of cooperatives. *Clio América*, 15 (29). <https://revistas.unimagdalena.edu.co/index.php/clioamerica/article/view/4278>

14. Vargas-Prieto, Amanda y Yepes Lugo, Cristian. (2018). Aportes teóricos y metodológicos sobre el emprendimiento agrícola. En: Vargas Prieto, Amanda y Sánchez Álvarez, César (eds.). *Desarrollo, estructuras económicas, políticas públicas y gestión. Reflexión interdisciplinar* (pp. 131-160). Bogotá, D. C. y México, D. F.: De la Salle.

15. Way, Michael J.; Scargle, Jeffrey D.; Ali, Kamal M. & Srivastava, Ashok N. (Eds.). (2012). *Advances in Machine Learning and Data Mining for Astronomy*. Boca Raton: CRC. <https://doi.org/10.1201/b11822>