La salud es de todos

Minsalud

PAPELES EN SALUD
Edición No. 19
Marzo de 2019
Bogotá D.C.

Minería de texto para el análisis de los Planes Territoriales de Salud

PENSEMOS LA SALUD: EVIDENCIA, ANÁLISIS Y DECISIÓN

La salud es de todos Minsalud

JUAN PABLO URIBE RESTREPO
Ministro de Salud y Protección Social

IVÁN DARÍO GONZÁLEZ ORTIZ
Viceministro de Salud Pública y Prestación de Servicios

DIANA ISABEL CÁRDENAS GAMBOA
Viceministra de Protección Social

GERARDO BURGOS BERNAL
Secretario General

WILSON FERNANDO MELO VELANDIA
Jefe Oficina Asesora de Planeación y Estudios Sectoriales

Minería de texto para el análisis de los Planes Territoriales de Salud

Documento elaborado por

DAVID EDUARDO GÓMEZ LIZARAZÚ
Oficina Asesora de Planeación y Estudios Sectoriales – Grupo de Estudios Sectoriales y de Evaluación de Políticas Públicas

JAIRO AUGUSTO NUÑEZ MENDEZ
Oficina Asesora de Planeación y Estudios Sectoriales – Grupo de Estudios Sectoriales y de Evaluación de Políticas Públicas, 2018

OFICINA ASESORA DE PLANEACIÓN Y ESTUDIOS SECTORIALES
GRUPO DE ESTUDIOS SECTORIALES Y DE EVALUACIÓN DE POLÍTICA PÚBLICA
GRUPO DE COMUNICACIONES

ISSN: 2500-8366 (En línea)
Documento de trabajo No: 19

Contenido

RESUMEN

Mediante la implementación de la metodología de Minería de Texto se analizan los Planes Territoriales de Salud de Departamentos, Distritos y Municipios con un doble objetivo: primero, se utilizan herramientas descriptivas para resumir la información contenida en los Planes Territoriales de Salud, se realizan nubes de palabras, árboles de asociación y análisis de asociaciones. En segundo lugar se utilizan modelos inferenciales para analizar la relación existente entre los instrumentos de planeación del orden nacional (Plan Nacional de Desarrollo - PND y Plan Decenal de Salud Pública - PDSP) y los instrumentos de planeación territorial. De esta manera las dos principales conclusiones del trabajo son: primero, la metodología de Minería de Texto permite capturar las diferencias de enfoque del conjunto de instrumentos de planeación (tanto nacional como local). Segundo, utilizando modelos binarios y de temas se puede afirmar que la planeación territorial en salud pública es congruente, desde el punto de vista conceptual, con el Plan Decenal de Salud Pública.

Palabras clave: Instrumentos planeación salud pública Colombia, minería de texto, planes territoriales de salud, minería de texto, plan decenal de salud pública, modelo binario Bayes Ingenuo, Modelo de Temas.

Códigos JEL: H75, I1, P46

Introducción[1]

El documento que a continuación se presenta muestra los resultados de la implementación de la metodología de Minería de Texto para analizar los instrumentos de planeación territorial en salud, especialmente los Planes Territoriales de Salud de los Departamentos, Distritos y Municipios construidos bajo la estrategia PASE a la equidad (periodo 2016-2019), el Plan Nacional de Desarrollo - PND y el Plan Decenal de Salud Pública - PNSP. Siguiendo a (Pasquali, 2016) la minería de texto se puede definir como un conjunto de algoritmos que buscan descubrir una estructura semántica en un conjunto de textos. Este conjunto de algoritmos se construye a partir de herramientas estadísticas como la Minería de Datos, que a partir del análisis de grandes volúmenes de información extrae patrones que permiten realizar análisis y pronósticos, lingüística aplicada, entre otras. La minería de texto busca extraer conocimiento de un texto o un conjunto de estos sin que un ser humano tenga que leerlos directamente.

La estructura de este documento es la siguiente: en un primer momento se abordan los principales aspectos teóricos de la metodología Minería de Texto, en segundo lugar se presentan los textos a analizar. En las secciones tres y cuatro se muestran los resultados de la implementación de la metodología; el capítulo tres presenta un análisis descriptivo de los textos usando conteos de palabras, nubes de palabras, árboles y coeficientes de asociación, principalmente. En la sección cuarta se presentan dos modelos estadísticos aplicados a los instrumentos de planeación en salud, el modelo Bayes Ingenuo (Naive Bayes) y los Modelos de Temas (Topic Models). La quinta sección presenta las conclusiones, donde el principal resultado obtenido se relaciona con la con la coherencia entre los instrumentos de planeación en salud territoriales y nacionales y, en mayor medida, entre el PDSP y los PTS distritales y departamentales.

[1] El presente documento fue construído en el marco de la Evaluación del Plan Decenal de Salud Pública llevada a cabo por el Grupo de Estudios Sectoriales y de Evaluación de Políticas Públicas.

Metodología y datos:

Welbers, Van Atteveldt, & Kenneth (2017) presentan un estado del arte de la Minería de Texto donde se destacan dos procesos que integran la implementación estándar de esta herramienta (Ilustración 1); el primero relacionado con la preparación de los datos, donde los textos a analizar se convierten a lenguaje matricial, y un segundo proceso, donde se realiza el análisis de información utilizando modelos supervisados y no supervisados. El objetivo de la preparación de los datos es transformar los textos a analizar en un conjunto de información que pueda ser procesada.

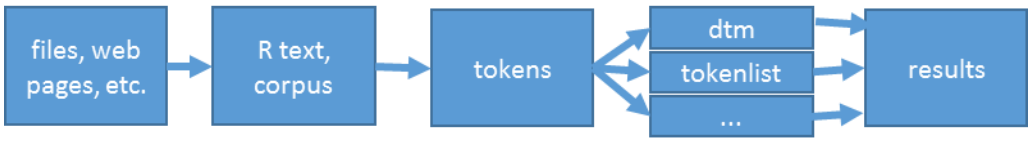
Este primer paso de la metodología incluyó la importación de textos, la cual se realizó usando los paquetes *pdftools* y *textreadr* del programa de software libre R Project. Siguiendo la metodología se importan los textos en *Word* y *PDF*, por ser los formatos utilizados por las administraciones locales y nacionales. Una vez realizada la importación de los textos se procede a depurar los datos, esto es, remover todo el léxico que no aporta información al análisis. Se usa particularmente el paquete *tm* (Ford, 2016) de *R Project*, para eliminar artículos, signos de puntuación, guiones, entre otros, del texto a analizar.

Como parte del objetivo es tener textos que sean comparables, se convierten todos los caracteres a minúsculas y se realiza un proceso para utilizar las raíces de las palabras en el análisis, proceso conocido como de *stemming* (Statistics University of Michigan, s.f.); de esta manera las palabras *plan, planes, planeación, planificación*, quedan reducidas en el análisis a la raíz *plan*, se considera que esta manera de proceder permite extraer del texto la esencia del concepto al que se hace referencia.

Ilustración 1. Proceso para el análisis de texto

Operación	Paquetes R
Preparación de los datos	
Importar texto	*Readtext*
Operación de *strings*	*Stringi*
Pre procesamiento	*Quanteda*
Matriz de términos por documento (DTM)	*Quanteda*
Filtrar y ponderar	*Quanteda*
Análisis	
Diccionario	*quanteda*

Aprendizaje de máquinas supervisado	*quanteda*
Aprendizaje de máquinas no supervisado	*topicmodels*
Estadísticas de texto	*quanteda*
Tópicos avanzados	
Avanzado NPL	*spacyr*
Posición de las palabras y sintaxis	*corpustool*



Fuente: (Welbers, Van Atteveldt, & Kenneth, 2017)

Después de realizado el proceso de depuración y estandarización del contenido de la base de datos, el siguiente paso en la metodología consiste en llevar los textos analizados a forma matricial. Primero se implementa un proceso conocido como *tokenization*, el cual consiste en identificar las unidades de análisis que permitan la representación matricial de los textos; en este caso la unidad de análisis son las palabras, como se puede observar en la Ilustración 2, donde se sigue a Pasquali (2016).

Se utilizan dos formas de representación matricial de los textos, donde por un lado se presentan las unidades de análisis como un conteo, y por otro lado, la representación se puede hacer de manera booleana donde los textos se representan como matrices de unos y ceros. Dada esta forma de expresar matricialmente los textos, es posible construir ponderaciones a las palabras usadas teniendo en cuenta la frecuencia de la aparición de las mismas; particularmente está disponible la ponderación conocida como Frecuencia inversa, la cual consiste en dar mayor ponderación a las palabras que aparecen con alta frecuencia en un texto y que, adicionalmente, tienen mayor frecuencia relativa en un documento particular. De esta manera, se considera que una palabra que aparece mucho y está asociada a pocos o un solo texto, da mucha información del contenido de este.

Ilustración 2. Metodología Minería de Texto. Preparación de datos y representación matricial

- Preparación de datos. *Tokenization*

Document	Text
1	There is no cure for curiosity
2	Curiosity killed the cat
3	My dog ate my lunch

Document	Vector
1	['There', 'is', 'no', 'cure', 'for', 'curiosity']
2	['Curiosity', 'killed', 'the', 'cat']
3	['My', 'dog', 'ate', 'my', 'lunch']

- Representación vectorial

ate	for	no	is	there	dog	cat	lunch	cure	curiosity	the	my	killed
1	1	1	1	1	1	1	1	1	2	1	2	1

- Boolean vectors representation

	ate	for	no	is	there	dog	cat	lunch	cure	curiosity	the	my	killed
1	0	1	1	1	1	0	0	0	1	1	0	0	0
2	0	0	0	0	0	0	1	0	0	1	1	0	1
3	1	0	0	0	0	1	0	1	0	0	0	1	0

- Terms Frequency – Inverse Document Frequency

N: número de documentos.
D: universo de documentos
$\{d \in D : t \in d\}$: número de documentos donde *t* aparece

$$IDF(t, D) = log \frac{N}{\{d \in D : t \in d\}}$$

t: término específico (ngram, token, palabra) en *d*
d: documento específico que pertenece a *D.*

$$TFIDF(t, d, D) = tf(t, d) idf(t, D)$$

Se pondera con mayor valor los términos que aparecen asociados a pocos documentos; términos no comunes dan más información

Fuente: (Pasquali, 2016)

Con la base de datos depurada se realizan dos análisis: en primer lugar un análisis descriptivo donde, usando conteos de palabras y coeficiente de asociación entre ellas se describe el contenido de los textos analizados. En segundo lugar, se utilizan dos modelos estadísticos, uno supervisado y otro no supervisado para analizar en profundidad la estructura de los textos; los modelos supervisados se denominan de esta forma porque, a partir de un subconjunto de los textos, denominado conjunto de entrenamiento, son calibrados para dar cuenta de la estructura de estos. Particularmente se usa el modelo de Bayes Ingenuo (Naive Bayes), el cual es un modelo de clasificación binaria que se utiliza para determinar la semejanza entre los textos analizados. En cuanto los modelos no supervisados, se utiliza el modelo de temas (Topic

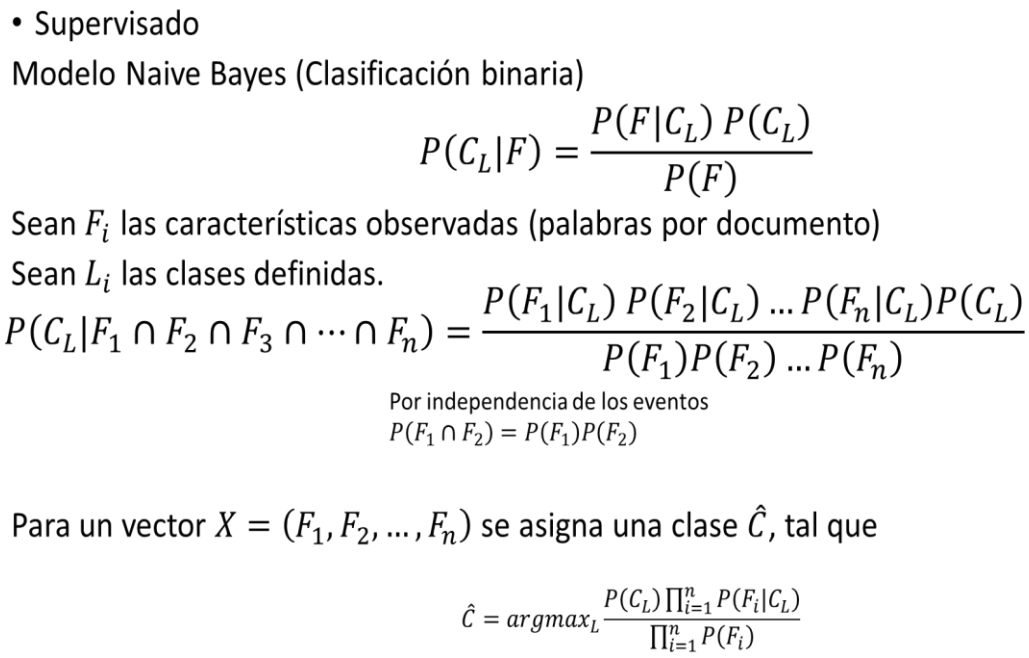
Model) el cual permite entender la estructura de temas en un documento, a partir de un conjunto de palabras.

La Ilustración 3 muestra la lógica del modelo de Bayes Ingenuo. Se parte del Teorema de Bayes definiendo el conjunto de categorías a clasificar C, a través de los niveles L= (1,2). Se hicieron varios ejercicios pero, en general, se utilizó como contraste el PND en el capítulo de salud[2] y el PDSP. Las características a observar fueron las palabras por documento F. Utilizando el supuesto de independencia de eventos se llega a la probabilidad posterior.

$$P(C_L|F_1,\dots,F_n) = \frac{P(C_L)\prod_{i=1}^{n} P(F_i|C_L)}{\prod_{i=1}^{n} P(F_i)}$$

Donde el argumento maximizador $\hat{C}$ de la ecuación anterior, permite ubicar los textos en alguna de las dos categorías planteadas, para un vector observado de características (palabras) $X = (F_1, F_2, \dots, F_n)$ por documento.

Ilustración 3. Lógica del modelo Bayes Ingenuo.



Fuente: (Statistics University of Michigan, s.f.)

[2] El componente de salud del Plan Nacional de Desarrollo 2014-2018: Todos por un nuevo país, está incluido en el Capítulo IV, Movilidad Social, Objetivo 2.

En cuanto a los modelos no supervisados se utiliza el Modelo de Temas (Topic Model). El Topic Model permite determinar los principales temas que se abordan en un texto y de esta manera comparar las temáticas en dos o más de ellos. El *Topic Model* es un modelo bayesiano que permite responder preguntas tales como: ¿qué se discute en el documento X? ¿Si estoy interesado en el tema Z cuál documento debería leer primero?

La Ilustración 4 muestra la lógica que se sigue en el Topic Model. La intuición del modelo es la siguiente: la variable observable son las palabras que aparecen en un texto y su distribución empírica observada (W en el círculo sombreado). Esta distribución empírica de palabras define el tema (que viene siendo una distribución específica de palabras) y la distribución de temas define el documento. Cada tema tiene unas palabras más probables de estar asociadas a un tema particular. Para cada una de estas distribuciones se asocia una función de distribución y se estima usando información a priori o por muestreo de Gibbs (Pasquali, 2016).

Ilustración 4. Metodología del Topic Model

• Modelo: Latent Dirichlet Allocation

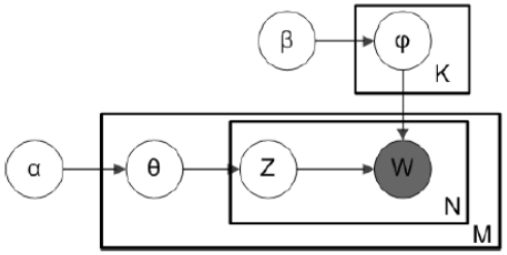


Figure 2.1: Latent Dirichlet Allocations represented in plate notation

K: Número de temas
N: Número de palabras en el documento
M: Número de documentos para analizar
α*: Dirichlet prior* parámetro de concentración de la distribución del tema por documento
β*:* Parámetro de la distribución de palabras por tema
φ*(k):* Distribución de palabras por tema *k*
θ*(i):* Distribución de temas por documento *i*
w(i,j): j-ésima palabra en el *i-ésimo* documento
z(i,j): es la asignación de tema para *w(i,j)*
ϕ y θ ~ Dirichlet

Fuente: (Pasquali, 2016)

Instrumentos de planeación en salud

Dada la estructura orgánica del Estado colombiano, en donde operan autoridades civiles a diferente niveles de gobierno, con competencias diferenciadas y fuentes y usos de recursos también diferenciados por nivel, se puede hablar de dos conjuntos de instrumentos de planeación en salud, unos territoriales, los cuales incluyen las administraciones departamentales, distritales y municipales y otros de carácter nacional donde se incluye el PND Salud y el PDSP.

Plan Nacional de Desarrollo: Salud

El PND es el principal instrumento de planeación con el que cuentan los gobiernos para cumplir con las promesas de campaña y establecer los objetivos, metas, lineamientos y estrategias de mediano y largo plazo (Congreso de la República de Colombia, 1994). En el año 2015 el Departamento Nacional de Planeación (DNP) expide el Plan Nacional de Desarrollo 2014-2018: Todos por un nuevo país. Este documento contiene diez capítulos en donde quedan enmarcadas las líneas de acción de cada uno de los sectores para el cuatrienio, las metas y los objetivos de política.

El componente de salud se reconoce al interior del Capítulo VI Movilidad Social, éste capítulo presenta 7 objetivos donde el *Objetivo 2 es: "Mejorar las condiciones de salud de la población colombiana y propiciar el goce efectivo del derecho a la salud, en condiciones de calidad, eficiencia, equidad y sostenibilidad."* El PND trae cuatro objetivos específicos para el sector en el cuatrienio: 1) aumentar el acceso efectivo a los servicios y mejorar la calidad de la atención; 2) mejorar las condiciones de salud de la población y reducir las brechas de resultados en salud; 3) recuperar la confianza y la legitimidad del sistema y 4) asegurar la sostenibilidad financiera del sistema de salud en condiciones de eficiencia.

El objetivo específico 1 se descompone en nueve estrategias, que incluyen: Consolidar la cobertura universal y unificar la operación del aseguramiento, generar incentivos para el mejoramiento de la calidad entre los actores del sistema, la Política de Atención Integral en Salud, incentivar la inversión pública hospitalaria en condiciones de eficiencia, desarrollar esquemas alternativos de operación de hospitales públicos, avanzar en el desarrollo de la política de talento humano en salud, mejorar la capacidad de diagnóstico de los laboratorios de salud pública a nivel nacional y territorial, implementar la política nacional de sangre e implementar el sistema indígena de salud propia e

intercultural. La Ilustración 5 muestra las metas incluidas en este objetivo específico.

Ilustración 5. Metas objetivo 1 PND Todos por un nuevo país

Metas

Meta intermedia	Línea base 2013	Meta a 2018
Percepción de acceso a los servicios de salud	46 %	50 %

Producto	Línea base 2013	Meta a 2018
Porcentaje de población afiliada al sistema de salud	96 %	97 %
Porcentaje de personas entre 18 a 25 años afiliadas al sistema de salud	95 %	99 %
Porcentaje de puntos de atención en IPS públicas con servicios de telemedicina en zonas apartadas o con problemas de oferta	34,5 %	43,1 %
Minutos de espera para la atención en consulta de urgencias para el paciente clasificado como Triage II	32,6	20
Días para la asignación de cita en la consulta con médico general y odontólogo general, respecto a la fecha para la que se solicita	3,92	3
Oportunidad en el inicio del tratamiento de leucemia en niños menores de 18 años (días)	12	5
Porcentaje de avance en la implementación del modelo de atención integral en salud para zonas con población dispersa	29,2 %	100 %
Guías de práctica clínica gestionadas con herramientas de implementación elaboradas	0	30
Porcentaje de casos de VIH detectados tempranamente	56,5 %	70 %
Hospitales públicos que adoptaron alguna de las medidas expedidas para mejorar su operación	0	955
Proyectos de infraestructura física o de dotación de las empresas sociales del Estado cofinanciados	0	37

Fuente: Tomado de (Plan Nacional de Desarrollo., 2015)

El objetivo especifico 2 se descompone en 8 estrategias: Implementar territorialmente el PDSP 2012-2021; generar hábitos de vida saludable y mitigar la pérdida de años de vida saludable por condiciones no transmisibles; prevenir y controlar las enfermedades transmisibles, endemoepidémicas, desatendidas, emergentes y re emergentes; promover la convivencia social y la salud mental; mejorar las condiciones nutricionales de la población colombiana; asegurar los derechos sexuales y reproductivos; atender integralmente en salud al adulto mayor y promover el envejecimiento activo y mentalmente saludable y mejorar la operación del programa ampliado de inmunizaciones.

El objetivo especifico 3 se descompone en cinco estrategias: Acercar la inspección, vigilancia y control al ciudadano; fortalecer la institucionalidad para la administración de los recursos del SGSSS; simplificar procesos; consolidar

el sistema integral de información de la protección social (SISPRO) y promover la transparencia, participación ciudadana y rendición de cuentas.

El objetivo especifico 4 se descompone en ocho estrategias: establecer medidas financieras para el saneamiento de pasivos; obtener nuevas fuentes de recursos; generar estabilización financiera y fortalecimiento patrimonial; consolidar la regulación del mercado farmaceútico; disminuir costos de transacción; revisar el mecanismo de redistribución del riesgo, restricciones de financiación y definir el mecanismo técnico participativo de exclusión de beneficios en salud.

Plan Decenal de Salud Pública

Mediante la Resolución 1841 de 2013 se adopta el PDSP para todo el territorio nacional. El PDSP es "un pacto social y un mandato ciudadano que define la actuación articulada entre actores y sectores públicos, privados y comunitarios para crear condiciones que garanticen el bienestar integral y la calidad de vida en Colombia" (Ministerio de Salud y Protección Social, 2013). Es un plan con perspectiva de mediano y largo plazo planteado para un decenio. Presenta tres objetivos centrales: 1) alcanzar mayor equidad en salud; 2) mejorar las condiciones de vida y salud de la población y 3) cero tolerancia con la morbilidad, la mortalidad y la discapacidad evitable.

Para alcanzar estos objetivos plantea el trabajo en 8 dimensiones prioritarias y dos transversales. Las dimensiones prioritarias son: salud ambiental, vida saludable y condiciones no transmisibles, convivencia social y salud mental, seguridad alimentaria y nutricional, derechos y sexualidad, vida saludable libre de enfermedades transmisibles, salud pública en emergencias y desastres. Las dos dimensiones transversales son: Gestión diferencial de poblaciones vulnerables y gestión para el fortalecimiento institucional y de los servicios de salud.

El PDSP se plantea a partir de siete enfoques conceptuales: enfoque de derechos, enfoque de determinantes sociales de la salud, enfoque diferencial, enfoque de ciclo de vida, enfoque de género, enfoque étnico y el enfoque poblacional. Se resalta como fundamental en este plan la intervención articulada de cada uno de los sectores relacionados con salud pública y el diseño de políticas que busquen la afectación positiva de los determinantes sociales de la salud.

Planes Territoriales de Salud

El artículo 6 de la Resolución 1536 de 2015 define el Plan Territorial de Salud - PTS como el "instrumento estratégico e indicativo de política pública en salud, que permite a las entidades territoriales contribuir con el logro de las metas estratégicas del PDSP" (Ministerio de Salud y Protección Social, 2015). Es competencia de los Departamentos, Distritos y Municipios "Diseñar, gestionar y

articular en el Plan Territorial de Salud, según los procesos de concertación con los actores sectoriales, intersectoriales, transectoriales y comunitarios, los compromisos intersectoriales, que permitan actuar sobre las inequidades en salud y determinantes sociales" (Ministerio de Salud y Protección Social, 2013).

Si bien la Resolución 1536 de 2015 establece en su Título III los contenidos del Plan Territorial de Salud, las Entidades Territoriales organizan de manera dispar los contenidos incluidos. La extensión de los PTS es por esta razón también altamente variable, se observan algunos de no más de 30 hojas mientras en otros la extensión supera las 300 hojas.

Tabla 1. Planes Territoriales disponibles para el análisis.

	PTS	
Departamentos y Distritos	Deptos	Municipios
Antioquia	1	7
Atlántico	0	0
Cundinamarca	1	4
Valle	1	40
Santander	0	1
Caquetá	1	1
Meta	0	2
Tolima	0	3
Huila	0	3
Norte de Santander	1	3
Nariño	1	1
Risaralda	0	1
Córdoba	0	1
Boyacá	1	119
Magdalena	1	1
Casanare	1	1
Arauca	0	1
Caldas	1	27
Guajira	0	1
Vaupés	1	0
Amazonas	1	0
Chocó	0	0
Cauca	0	1
Bolívar	0	0
Guanía	0	0
Cesar	0	0
Guaviare	0	0
Putumayo	1	0
Quindío	1	0

San Andrés y Providencia	0	0
Sucre	0	0
Vichada	1	2
Bogotá	1	0
Santa Marta	0	0
Cartagena	1	0
Buenaventura	0	0
Barranquilla	1	0
Total	18	220
	Total PTS	238

Fuente: Elaboración propia.

La Tabla 1 presenta el conjunto de PTS disponibles para el análisis. En total se tienen 238 PTS disponibles para ser analizados, 18 corresponden a departamentos y distritos y 220 corresponden a municipios.

Resultados

A continuación se presentan los resultados de la implementación de la metodología de Minería de Texto para los instrumentos de planeación en salud.

En primer lugar se presenta información descriptiva de los diferentes instrumentos usando nubes de palabras, coeficientes de asociación, árboles de asociación y conteos de palabras, donde el objetivo es responder la pregunta ¿de qué hablan los instrumentos de planeación en salud?

En segundo lugar se presenta el resultado de implementar dos modelos inferenciales uno supervisado y otro no supervisado para responder la pregunta ¿hay relación entre el PDSP y los PTS departamentales?, se realizan otros ejercicios usando PTS municipales con conclusiones similares y que serán brevemente mencionados.

Para realizar este análisis se dividen los instrumentos de planeación en 4 grupos: en primer lugar el PDSP, en segundo lugar el PND Salud, en tercer lugar los PTS departamentales y distritales disponibles y, por último, los PTS de las cuatro ciudades más grandes del país (Bogotá, Medellín, Cali, Barranquilla) que representan alrededor del 30% de la población total colombiana.

Conteos de palabras y nube de palabras

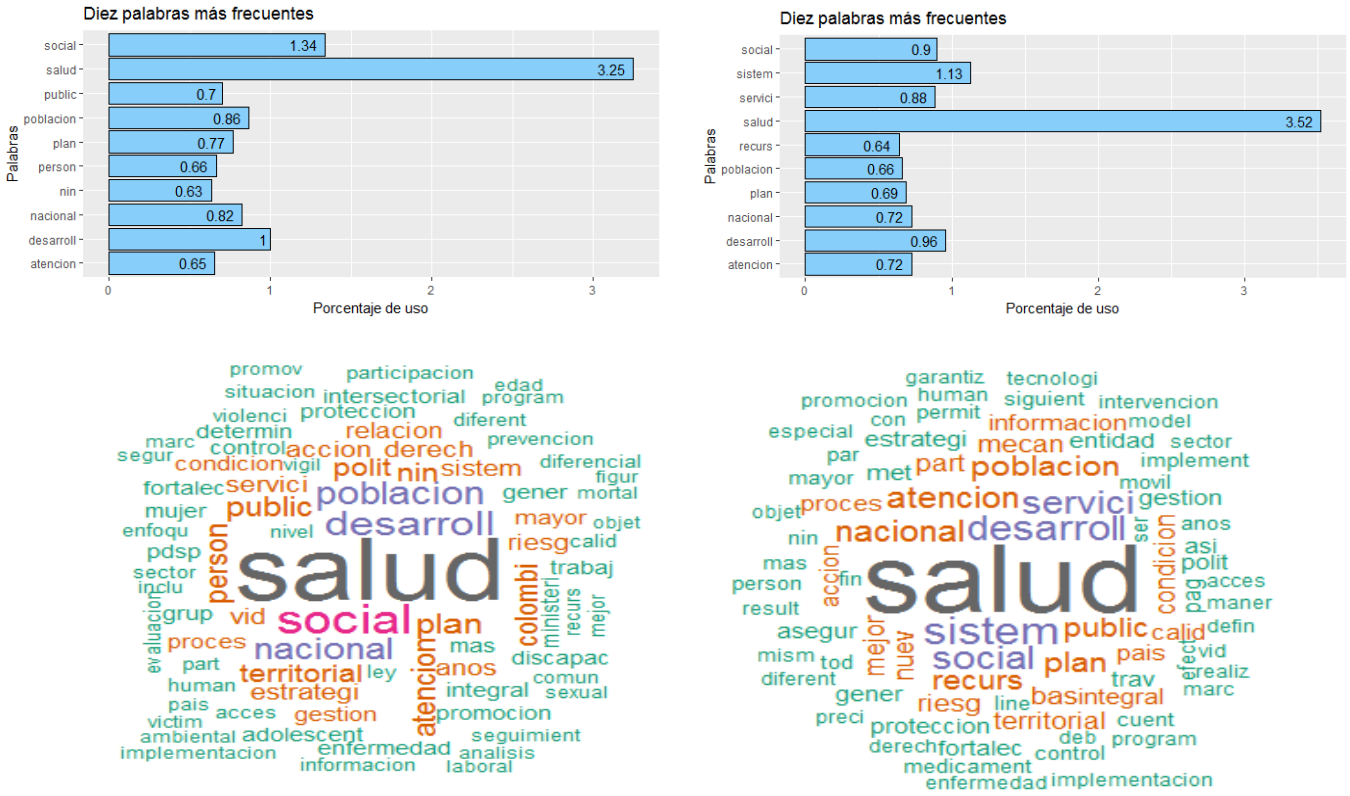
Los conteos de palabras y las nubes de palabras son la primera manera de acercarse al contenido de los instrumentos de planeación en salud.

No es de extrañar que las palabras más mencionadas en los instrumentos de planeación del nivel nacional en salud sean salud, nacional, desarrollo, atención, social, entre otros. Pero se observan diferencias interesantes, en primer lugar, el PND Salud incluye la palabra recursos mientras que en el PDSP no tiene la misma importancia; esto es coherente por tanto, el objetivo específico 4 del PND Salud está relacionado con la sostenibilidad financiera del sistema. Por su parte en PDSP aparece la raíz nin[3], que no aparece tan frecuente en el PND, lo que habla también de un enfoque de población en el PDSP.

[3] En el proceso de identificar las raíces de las palabras nin hace referencia a los conceptos asociados a niño, niña, niñez.

Las nubes de palabras dan información del conteo de palabras de una manera más gráfica. Se observa que salud es la palabra más utilizada en los dos instrumentos de planeación. En el PDSP aparecen palabras como determin, intersectorial, nin, condición, riesg, participación, entre otras que se relacionan con los temas propios del enfoque de este plan; En el PND Salud se observan otras palabras importantes como asegur, sistem, model, medicament, lo cual se relaciona con los objetivos del plan enunciados anteriormente.

Ilustración 6. Conteo de palabras (diez más frecuentes) y nube de palabras PDSP (izquierda) y PND Salud (derecha)



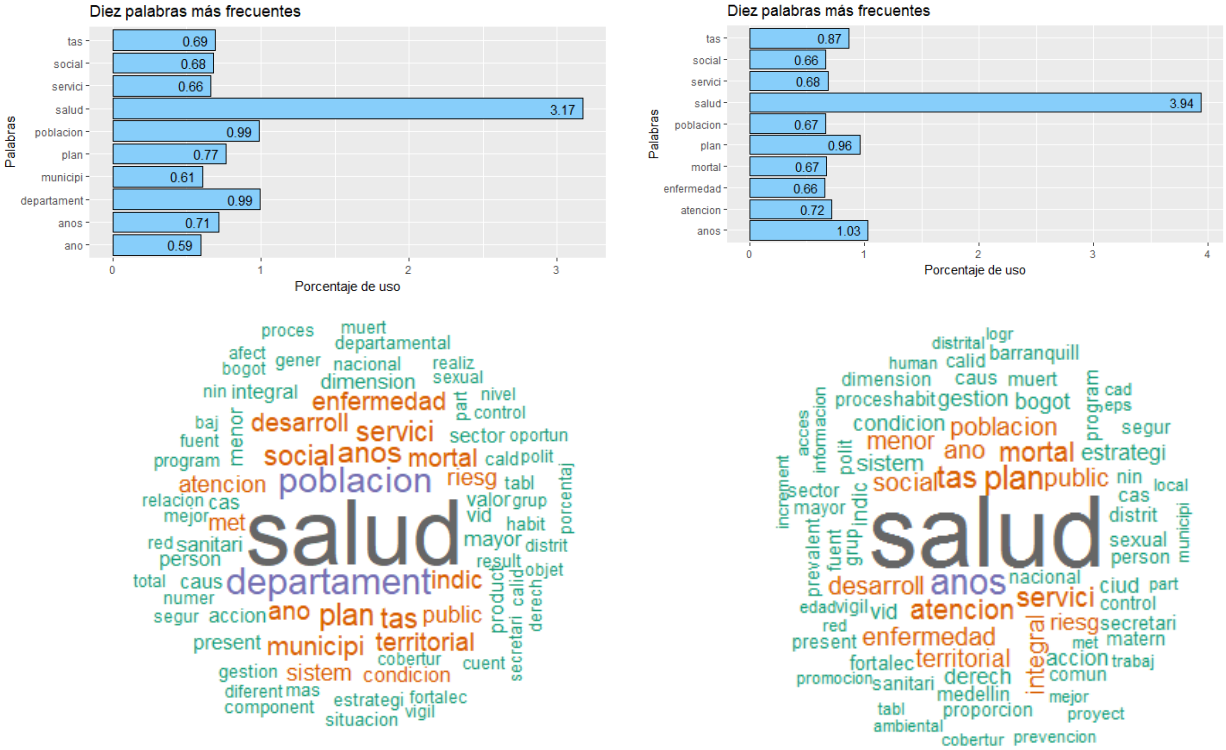
Fuente: Elaboración propia

Por su parte, cuando se analizan dos conjuntos de PTS territoriales, por un lado los PTS de distritos y departamentos y por otro lado los PTS de las 4 ciudades más importantes del país que representan aproximadamente el 30% de la población colombiana, se observan dos diferencias importantes respecto a los planes nacionales. Primero, se observa un énfasis de estos planes en temas más puntuales de la salud pública, lo cual se infiere por las raíces *tas, servici, mortal, enfermedad, atención.* Esto es coherente con el papel más específico en temas de salud pública de estos planes en comparación con los nacionales.

Cuando se observan las nubes de palabras se hace más evidente la aparición de palabras relacionadas con el seguimiento a indicadores específicos de salud

pública en los PTS territoriales; se observan raíces *como muert, tabl, cobertur, tas, met.*

Ilustración 7. Conteo de palabras (diez más frecuentes) y nube de palabras PTS departamentos y distritos (izquierda) y PTS 4 ciudades (derecha)



Fuente: Elaboración propia

Coeficientes y árboles de asociación

Los coeficientes y los árboles de asociación nos permiten ampliar el análisis de los conteos de palabras, nos indican las palabras asociadas a aquellas de más alta frecuencia representadas anteriormente.

La Tabla 2 muestra las palabras asociadas a salud, para cada uno de los instrumentos de planeación. Si bien salud es siempre la palabra más utilizada en los planes, cuando se oberva las palabras asociadas es posible decir que no se refieren al mismo concepto en cada uno de ellos. En el PDSP la palabra salud, está asociada a social*, determinan, public, result*, que es coherente con el enfoque en determinantes sociales de este instrumento. En PND Salud, la palabra salud, está asociada a basic, diferencial, gestión, respuest, que es coherente con los objetivos planteados en este plan, donde el énfasis está en el acceso, la atención básica y diferencial y la respuesta a las necesidades en

salud de la población. En departamentos, distritos y las cuatro ciudades más grandes del país se observan asociaciones con palabras *public, servici, atencion, integral.* Servicios e integrales son conceptos que no habían aparecido tan asociados en salud en los planes nacionales, pero sí lo hacen en los planes territoriales, lo cual refleja un interés de las administraciones locales en este tema.

Tabla 2. Palabras asociadas a salud por instrumento de planeación

Salud				
PDSP	social	determin	public	result
	0,55	0,53	0,53	0,46
PND Salud	basic	diferencial	gestion	respuest
	0,64	0,6	0,54	0,54
Deptos y distritos	public	servici	social	atencion
	0,48	0,48	0,48	0,45
4 Ciudades	integral	public	atencion	servici
	0,54	0,54	0,5	0,45

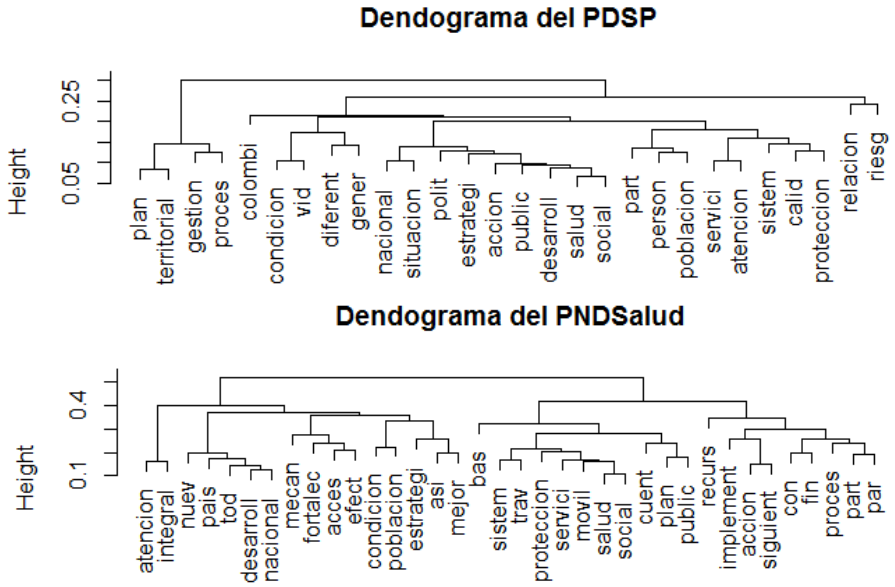
Fuente: Elaboración propia.

Una forma de ampliar el análisis de las palabras y los conceptos involucrados es utilizando árboles de asociación. Los árboles de asociación permiten entender la estructura del documento al ubicar de manera gráfica las palabras asociadas, utilizando *clusters* jerárquicos. De esta manera se puede observar (Ilustración 8) que el PDSP presenta dos grandes conceptos que lo caracterizan; por un lado hay una idea importante que gira alrededor de la gestión y los procesos del plan territorial; por otra parte, involucra más ideas y es más complejo, habla de calidad en los servicios y la atención del sistema, lo cual incluye también protección y participación. En esta misma rama se muestra un concepto asociado a condiciones de vida y enfoques (*diferent, género*), que se puede asociar a los enfoques del PDSP. Por último, se observa un concepto en la parte central del árbol asociado a la política, las estrategias y la acción pública para el desarrollo social.

Teniendo en cuenta la estructura del PDSP, se pueden asimilar estos dos grandes conceptos mencionados anteriormente a las dimensiones prioritarias y transversales del plan, donde las transversales se refieren a gestión y procesos principalmente y se encuentran diferenciadas de las otras ideas, más específicas de cada una de las dimensiones prioritarias, donde la atención en salud es importante así como lo son los enfoque del PDSP y condiciones de vida de la población. En el PND, por su parte, se pueden identificar dos grandes ideas, una relacionada con el fortalecimiento del acceso efectivo de la población a los servicios de salud y otro relacionado con recursos y con el sistema de protección social. En los PTS de departamentos y distritos y ciudades grandes,

se observan dos grupos de ideas; el primero relacionado con el cumplimiento de metas puntuales, tasas, tablas, indicadores, mortalidad, y el segundo relacionado con las condiciones de vida de la población y la gestión de la salud pública.

Ilustración 8. Árboles de asociación planes de salud del nivel nacional



Fuente: Elaboración propia

Modelo de clasificación binaria y modelo de temas

Una vez analizada la información de los PTS de manera descriptiva, se utilizan dos modelos estadísticos para profundizar en su estructura y responder a la pregunta de qué se parece más a qué en este conjunto de información. El primer ejercicio que se realizó fue utilizar un modelo supervisado de clasificación binaria para determinar la semejanza entre los PTS territoriales y aquellos de carácter nacional, cómo ha podido observarse en esta sección, el PDSP y el PND Salud tienen importantes diferencias en cuanto a los temas en donde hacen énfasis y el enfoque con el cual se abordan los problemas de salud. Ante este panorama, lo que se hizo fue utilizar un modelo de Bayes Ingenuo para determinar a qué se parecen más los PTS departamentales y distritales (se hizo también para algunos PTS municipales y los resultados no cambian sustancialmente). La forma en la que se implementó esta metodología fue la siguiente:

En primer lugar se estableció de manera aleatoria un conjunto de entrenamiento (80% de los dos textos unidos) y un conjunto de prueba (20% de los dos textos unidos) para el contenido del PDSP y el PND Salud. Utilizando el conjunto de

prueba se realizó una validación interna de los criterios de clasificación de textos encontrando los resultados mostrados en la Ilustración 9; los niveles 1 (PND Salud) y 0 (PDSP) diferencian los dos instrumentos de planeación del nivel nacional.

Lo que se presenta es una tabla de contingencia que indica que de las siete hojas del PND Salud, que cayeron en el conjunto de prueba fueron clasificadas correctamente como pertenecientes al PND Salud. Por su parte, 44 de las 47 hojas del PDSP que hacen parte del conjunto de prueba fueron calificadas efectivamente como PDSP mientras que 3 hojas del PDSP, fueron clasificadas incorrectamente como pertenecientes al PND Salud cuando no correspondían. El error en este caso fue cercano al 5% y se consideró tolerable. Con esta calibración del modelo se clasificaron los PTS de departamentos y distritos obteniendo los resultados presentados en la Tabla 3.

De manera categórica, el modelo indica que los PTS de departamentos y distritos se asemejan más al PDSP que al PND Salud. Este resultado era de esperarse toda vez que el lenguaje y la estructura de temas abordados por los PTS territoriales muestran una afinidad mayor con el PDSP. Adicionalmente la normatividad vigente (Resolución 1536 de 2015) insta a las Entidades Territoriales a seguir una estructura derivada del PDSP. Podemos decir que efectivamente esta metodología brinda elementos para pensar que los lineamientos entregados por el Ministerio de Salud y Protección Social han sido adoptados por las ET para la construcción de sus PTS

Ilustración 9. Resultados validación interna modelo de clasificación del PDSP y el PND Salud

Total Observations in Table: 54

	BaseDatos_Test$Indicador		
TextoAnalisis_Test_Pred	1	0	Row Total
1	7	0	7
	25.096	5.704	
	1.000	0.000	0.130
	0.700	0.000	
	0.130	0.000	
0	3	44	47
	3.738	0.849	
	0.064	0.936	0.870
	0.300	1.000	
	0.056	0.815	
Column Total	10	44	54
	0.185	0.815	

Fuente: Elaboración propia

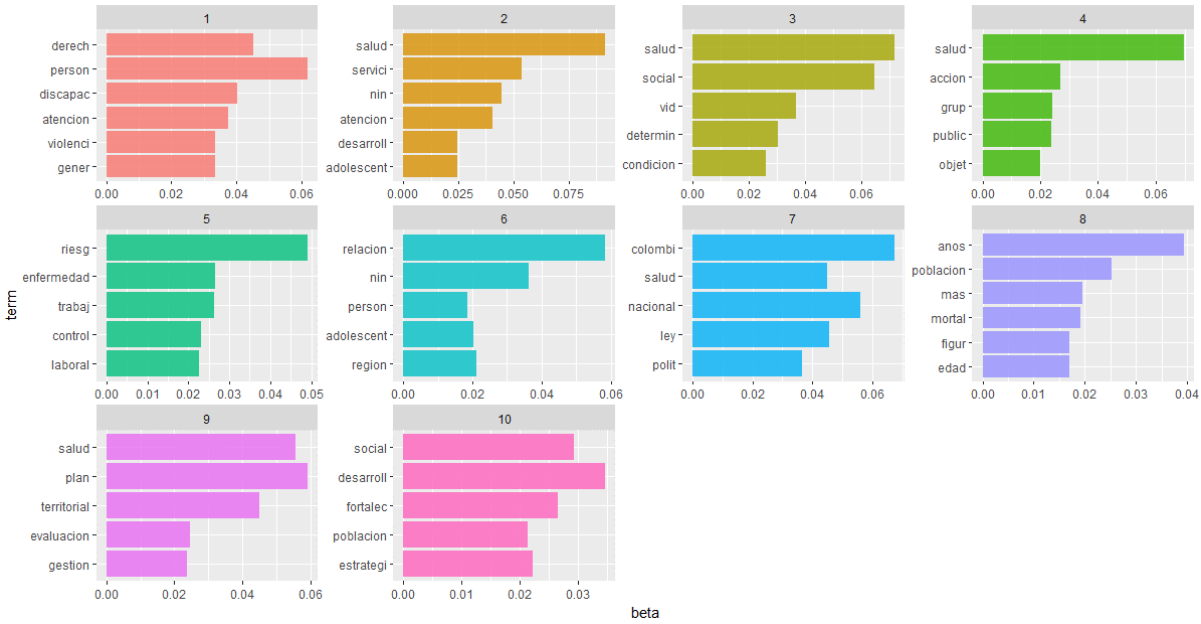
Tabla 3. Resultados Naive Bayes para Departamentos y Distritos.

Porcentaje del PTS que se clasifica como similar al PDSP	
95%	Amazonas.pdf
98%	Antioquia.pdf
100%	Barranquilla.pdf
95%	Bogota.pdf
100%	BoyacaDepto.pdf
97%	Caldas.pdf
98%	Caquetá.pdf
96%	Cartagena.pdf
98%	Casanare.pdf
100%	Cundinamarca.pdf
100%	MagdalenaDepto.pdf
100%	Nariño.pdf
95%	NorteDeSantander.pdf
100%	Putumayo.pdf
100%	Quindio.pdf
100%	Valle.pdf
100%	Vaupes.pdf
100%	Vichada.pdf

Fuente: Elaboración propia

Por último, se presenta el resultado de la implementación del modelo de temas para los instrumentos de planeación en salud en Colombia. Los modelos de temas permiten encontrar la estructura de los textos a partir de los temas más asociados a ellos, definidos por un conjunto de palabras. En el cuerpo del documento presentamos el resultado para el PDSP, los demás resultados están disponibles en los anexos. En primer lugar se observan dos temas asociados a la atención de población vulnerable, por su parte el tercer tema se relaciona con determinantes y condiciones de vida lo cual es propio del PDSP, finalmente los temas nueve y diez parecen estar más asociados a las dimensiones transversales del PDSP.

Ilustración 10 Resultados modelo de temas para el Plan Decenal de Salud Pública



Fuente: Elaboración propia

Conclusiones

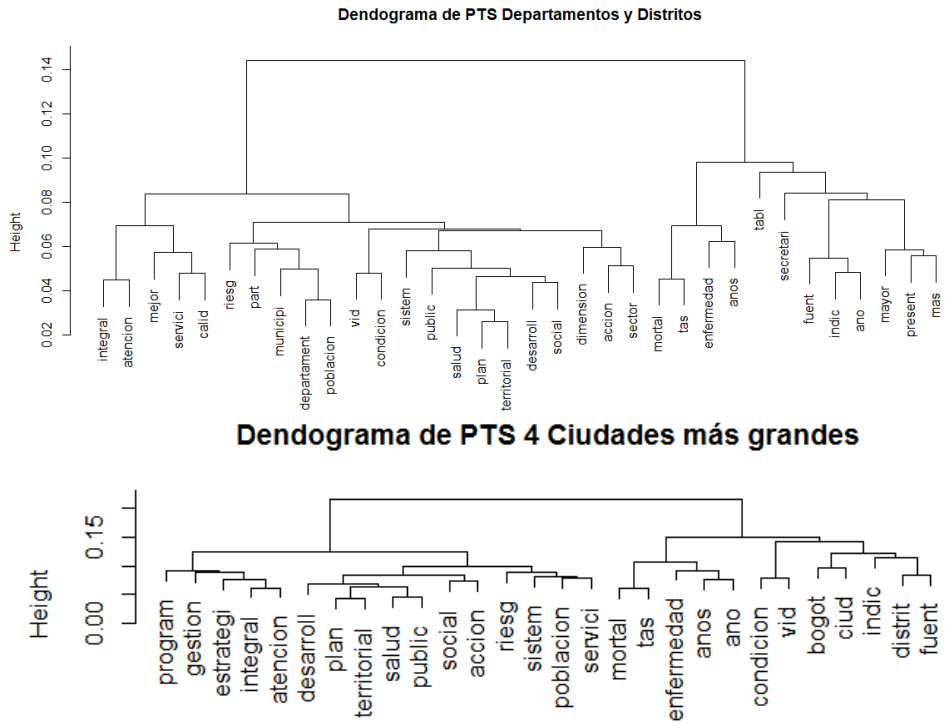
La implementación de la herramienta estadística conocida como Minería de texto arroja varias ideas alrededor del contenido de los PTS territoriales y su relación con los instrumentos de planeación en salud del nivel nacional.

La primera conclusión que se puede extraer, es que efectivamente se encuentran diferencias de enfoque en los principales instrumentos de planeación del nivel nacional, lo cual los hace documentos complementarios en cuanto a la definición de metas y estrategias sectoriales. Se evidencia en el PDSP una preocupación por los conceptos relacionados con las condiciones de vida, los determinantes sociales y la acción intersectorial. El PND Salud tiene una visión del sector que involucra otros factores tales como la sostenibilidad financiera del sector, los medicamentos (política farmacéutica), la calidad de la atención, el aseguramiento, entre otros. La herramienta es capaz de capturar estas diferencias y nos ofrece una visión de lo que está ocurriendo en los territorios. En primer lugar nos presenta los PTS como herramientas aterrizadas de planeación territorial donde el seguimiento a tasas, metas, tablas, mortalidades nos permite observar una preocupación en estas entidades por el seguimiento y los resultados en salud pública.

De la misma manera, es posible también afirmar que la herramienta minería de texto brinda elementos que soportan la idea de que los PTS territoriales efectivamente adoptaron la estructura y el lenguaje del PDSP; el modelo de clasificación binaria abrumadoramente lleva a concluir que los PTS territoriales guardan una estrecha relación con el PDSP y, adicionalmente, el modelo de temas nos permite ver que efectivamente los PTS presentan temas similares a los del PDSP pero que también se incluyen algunos referentes al PND tales como el acceso a los servicios de salud.

Anexos

Anexo 1. Dendogramas PTS Departamentos y Distritos y 4 Ciudades más grandes.



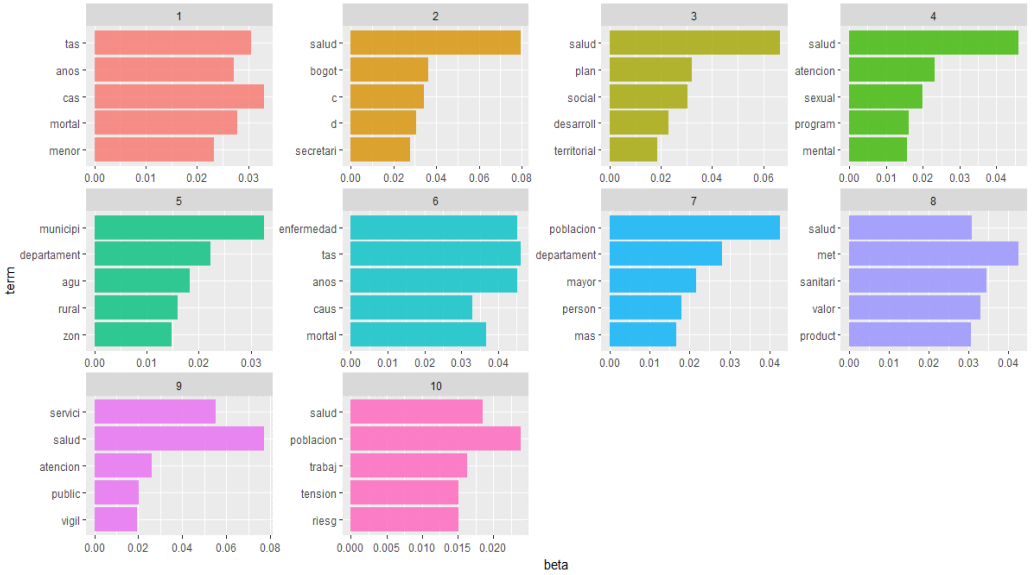
Fuente: Elaboración propia

Anexo 2. Modelo de temas para PND Salud, PTS Deptos y distritos y PTS 4 ciudades

Ilustración 11. Modelo de temas para el PND Salud

Fuente: Elaboración propia

Ilustración 12. Modelo de temas para los PTS departamentales y distritales.



Fuente: Elaboración propia

Ilustración 13. Modelo de temas para los PTS de las 4 ciudades más grandes

Fuente: Elaboración propia

Bibliografía

Congreso de la República de Colombia. (1994). *Ley 152 de 1994. Ley Orgánica de Plan de Desarrollo.*

Ford, C. (14 de Abril de 2016). *University of Virginia Library*. Obtenido de https://data.library.virginia.edu/reading-pdf-files-into-r-for-text-mining/

Ministerio de Salud y Protección Social. (2013). *Plan Decenal de Salud Pública 2012-2021.* Bogotá.

Ministerio de Salud y Protección Social. (2015). *Resolución 1536 de 2015. Por la cual se establecen disposiciones sobre el proceso de planeación integral para la salud.* Bogotá.

Pasquali, A. R. (2016). *Automatic coherence evaluation applied to Topic Models.* Facultad de Ciencias de la Unviersidad de Porto.

Plan Nacional de Desarrollo. (2015). *Plan Nacional de Desarrollo 2014-2018. Todos por un nuevo país.*

Statistics University of Michigan. (s.f.). *Courses Stat Umich*. Obtenido de http://dept.stat.lsa.umich.edu/~jerrick/courses/stat701/notes/stringmanip.html

Welbers, K., Van Atteveldt, W., & Kenneth, B. (2017). Text analysis in R. *Communication methods and measures, 11*(4), 245-265. Obtenido de https://doi.org/10.1080/19312458.2017.1387238