



Construcción de indicadores de desempeño para el Proyecto de Discapacidad e Inclusión de la Universidad del Valle usando probabilidades

Sebastián Ospina Sabogal
Andrés Felipe Orozco Apache

Universidad del Valle
Facultad de Ingeniería
Escuela de Estadística
Cali, Colombia
2023

Construcción de indicadores de desempeño para el Proyecto de Discapacidad e Inclusión de la Universidad del Valle usando probabilidades

Sebastián Ospina Sabogal
Andrés Felipe Orozco Apache

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadística

Director: Ph.D. Jose Rafael Tovar Cuevas
Co-director: MSc Jennyfer Portilla Yela

Universidad del Valle
Facultad de Ingeniería
Escuela de Estadística
Cali, Colombia
2023

Agradecimientos

Agradecemos al profesor José Rafael Tovar Cuevas, por su acompañamiento, dedicación y paciencia con nosotros en el transcurso de la carrera y en el presente trabajo.

Agradecemos al Proyecto de Discapacidad e Inclusión por su apoyo en este proceso especialmente a la profesional Norida Bravo Meneses

Agradecemos a nuestros familiares por su apoyo incondicional en nuestra formación profesional

Agradecimientos especiales a Sara Milena Gutierrez, Rodrigo Luna, Karen Paola Villalobos, y Valentina Córdoba Latorre quienes nos alentaron a seguir adelante en este proyecto.

Dedicatoria

Este trabajo es dedicado a cada una de las personas que confiaron, apoyaron y creyeron en nuestro proceso formativo, que de alguna manera pusieron su mano en nuestra espalda y nos animaron a seguir adelante
Shinzou wo sasageyo.

Resumen

Este trabajo de grado propone evaluar el desempeño del Proyecto de Discapacidad e Inclusión de la Universidad del Valle utilizando probabilidades. Se propone un índice para medir el desempeño del proyecto, basado en los informes de seguimiento. Los resultados indican un buen desempeño del proyecto. Se recomienda estandarizar la recolección de información y fortalecer la capacitación del personal.

Palabras clave: Desempeño, Procesamiento de lenguaje natural, tokenización, lematización, stopwords, indicadores.

Abstract

This degree project aims to assess the performance of the Proyecto de Discapacidad e inclusión de la Universidad del Valle using probabilities. An index is proposed to measure the project's performance, based on the follow-up reports. The results indicate a satisfactory performance of the project. It is recommended to standardize the data collection process and strengthen staff training to ensure consistent and reliable information. Additionally, continuous monitoring and evaluation are advised to track the project's progress and identify areas for further improvement.

Keywords: Performance, Natural Language Processing, Tokenization, Lemmatization, Stopwords, Index.

Contenido

Resumen	v
1 Introducción y planteamiento del problema	4
1.1 Introducción	4
1.2 Planteamiento del problema	6
1.3 Justificación	8
1.4 Objetivos	9
1.4.1 Objetivo General	9
1.4.2 Objetivos específicos	9
2 Antecedentes	10
2.0.1 Antecedentes contextuales	10
2.0.2 Antecedentes estadísticos	11
3 Marco teórico	14
3.1 Marco contextual	14
3.1.1 Normativa internacional	14
3.1.2 Normativa nacional	15
3.1.3 Normativa a nivel institucional	15
3.2 Marco teórico Estadístico	16
3.2.1 Variable latente	16
3.2.2 Procesamiento del Lenguaje Natural	16
3.2.3 Unigrama y Bigrama	17
3.2.4 Índice	18
3.2.5 Distribución Multinomial	18
3.2.6 Distribucion Dirichlet	19
4 Metodología	21
4.1 Lista general de estudiantes del Proyecto de Discapacidad e Inclusión	21
4.1.1 Estructura de la información	21
4.2 Delimitación de la muestra	22
4.2.1 Procesamiento de documentos	23
4.2.2 Creación de unigramas y bigramas	28
4.2.3 Creación de categorías y sub-categorías	29

4.2.4	Clasificación de unigramas y bigramas	31
4.2.5	Método de Elicitación predictivo	33
4.3	Modelos Estadísticos	34
4.3.1	Experimento	34
4.3.2	Modelo estadístico 1	34
4.3.3	Modelo estadístico 2	37
4.3.4	Modelo estadístico 3	40
4.4	Estimación usando OpenBUGS	43
4.5	Creación del índice heurístico	43
4.5.1	Niveles de desempeño	44
4.5.2	Evaluación de robustez y convergencia	44
5	Resultados	47
5.1	Delimitación de la muestra	47
5.2	Clasificación de los Unigramas	47
5.3	Modelos estadísticos	49
5.4	Índice General	50
6	Conclusiones, recomendaciones y limitaciones	52
6.1	Conclusiones:	52
6.2	Recomendaciones:	53
6.3	Limitaciones:	53
7	Anexos	54

Índice de tablas

4-1	Cantidad de estudiantes esperados en grupo 1 partiendo de elicitación por profesionales del proyecto.	33
4-2	Cantidad de estudiantes esperados en grupo 2 partiendo de elicitación por profesionales del proyecto.	33
4-3	Cantidad de estudiantes esperados en grupo 3 partiendo de elicitación por profesionales del proyecto.	34
4-4	Proceso de simulación para cada modelo	45
4-5	Número de simulaciones de probabilidades con su respectivo calculo del índice	46
5-1	Frecuencia de los unigramas entre documentos, intra documentos y sus respectivas categorías asignadas por los profesionales del Proyecto	48
5-2	Parámetros de las distribuciones a priori no informativas y distribuciones posteriores	49
5-3	Parámetros de las distribuciones a priori informativas y distribuciones posteriores	49
5-4	Estimaciones medias dadas por las distribuciones posteriores no informativas e informativas usando OpenBugs.	51
7-1	Categorías, grupos y sub-categorías creadas por los profesionales del Proyecto de Discapacidad	57

Declaración

Nos permitimos afirmar que hemos realizado el presente trabajo de grado de manera autónoma y con la única ayuda de los medios permitidos y no diferentes a los mencionados en el propio trabajo de grado. Todos los pasajes que se han tomado de manera textual o figurativa de textos publicados y no publicados, los hemos reconocido en el presente trabajo. Ninguna parte del presente trabajo se ha empleado en ningún otro tipo de tesis o trabajo de grado.

Igualmente declaramos que los datos utilizados en este trabajo están protegidos por las correspondientes cláusulas de confidencialidad.

Santiago de Cali, 06.07.2023

(Sebastián Ospina Sabogal y Andrés Felipe Orozco Apache)

1 Introducción y planteamiento del problema

1.1. Introducción

Según la Organización Mundial de la Salud (2014), se define una persona con discapacidad “aquellas que tienen deficiencias físicas, mentales, intelectuales o sensoriales a largo plazo que, en interacción con diversas barreras, pueden obstaculizar su participación plena y efectiva en la sociedad en igualdad de condiciones con los demás”. En el ámbito de la educación superior, dichas barreras van desde las actitudes y estigmas negativos, ya que los estudiantes pueden sufrir algún tipo de discriminación y prejuicios por la falta de comprensión del estudiantado y profesorado; falta de capacitación del personal universitario y limitaciones de algunos programas académicos, puesto que estos pueden estar diseñados sin tener en cuenta las diversas necesidades y capacidades de todos los estudiantes; hasta la inadecuada infraestructura de las universidades. Todos estos aspectos mencionados anteriormente van de la mano con el propio financiamiento de la educación superior, además que un mal financiamiento limita la obtención de apoyos como materiales en formatos accesibles o materiales tecnológicos.

Internacionalmente, las limitaciones de la discapacidad en la educación han sido tratadas a lo largo de los años con la Declaración de Salamanca, Informe Delors o La Convención Interamericana sobre la eliminación de todas las formas de discriminación. En adición, la ONU presenta un número de declaraciones y normas en las cuales sustentan desde hace más de 50 años la educación superior inclusiva, que inicia desde la declaración de los derechos humanos (*Comisión de Derechos Humanos*, 1948) hasta la convención sobre los derechos de las personas con discapacidad (*Organización de las Naciones Unidas*, 2006). En Colombia, existen normativas nacionales que buscan promover la inclusión educativa, la igualdad de oportunidades y el acceso equitativo a la educación para todos los estudiantes con discapacidad en el país, además de ser uno de los 17 países que trabajó de manera ágil para adquirir normativas en el contexto inclusión en la educación (Carvajal, 2015).

La Universidad Nacional y la Universidad del Valle, en el 2012 y 2013 respectivamente, firmaron sus correspondientes acuerdos de política inclusiva; la inclusión educativa de las personas con discapacidad en la Universidad Nacional de Colombia (*Universidad Nacional*

de Colombia, 2012) se enfoca en los estudiantes con discapacidad con el fin de generar buenas condiciones, accesibilidad académica, estructural, curricular y formación de docentes (Carvajal, 2015). Es por esto que la Universidad del Valle ha tomado la iniciativa de abordar esta temática creando una política de discapacidad e inclusión en colaboración con miembros de la comunidad universitaria, donde la corresponsabilidad de todos los actores involucrados, el enfoque formativo y la colaboración continua han sido identificados como las tres características principales de dicha política (Carvajal, 2015). Presentando un compromiso con todos sus estudiantes, el cual aumenta gradualmente al referirse a estudiantes con discapacidad, manteniendo como prioridad el acompañamiento que reciben estos estudiantes, teniendo en cuenta su diversidad funcional para asegurarles una mejor orientación como Institución y así contribuir a una educación digna y de alta calidad.

La “Política Institucional de Discapacidad e Inclusión de la Universidad del Valle” establecida en el 2013 reconoce la existencia de las personas con discapacidad, y como parte de su compromiso por contribuir a la construcción de una sociedad democrática, justa e igualitaria, se ha unido al deseo de disminuir la situación de exclusión social de las personas con discapacidad e igualar sus oportunidades respecto a las demás personas (*Consejo Superior de la Universidad del Valle*, 2013).

Es por esto que surge el Proyecto de Discapacidad e Inclusión de la Universidad del Valle, atribuido a la Vicerrectoría de Bienestar Universitario, donde buscan promover todo lo referido a la educación inclusiva mencionado anteriormente. El proyecto comenzó en 2016 y su objetivo principal es Apoyar la implementación de la Política de Discapacidad e Inclusión de la Universidad del Valle. La universidad implementa una variedad de actividades y estrategias que pueden ir desde cambiar el conocimiento general sobre la discapacidad y el entorno para facilitar su acceso, proporcionar tecnología adecuada y dispositivos de asistencia, brindar monitorias, hasta enfocarse en nutrir el conocimiento y las habilidades de los docentes para promover una cultura inclusiva y brindar los ajustes razonables a los estudiantes con discapacidad.

La universidad destina recursos económicos para que el Proyecto de Discapacidad e Inclusión pueda apoyar en procesos de acceso y permanencia de la población con discapacidad, así como asegurar avances en la implementación del Acuerdo N. 004 - Política de Discapacidad e Inclusión, además de asegurar el debido funcionamiento del Proyecto y garantizar las mejores condiciones de trabajo. Por tanto, resulta justo y necesario construir indicadores que permitan conocer el desempeño de dicho proyecto.

El proyecto involucra a diversos profesionales donde uno de los objetivos es velar por el bienestar de los estudiantes vinculados. Estos profesionales asumen diversas responsabilidades como realizar entrevistas iniciales a los estudiantes interesados en formar parte del

proyecto y asignar a un monitor o monitores para acompañar al estudiante con las diferentes actividades curriculares en caso de ser necesario. De igual manera, se encargan de recopilar información periódica a través de reuniones para supervisar el estado de los estudiantes vinculados, identificar posibles dificultades y determinar si requieren mayor acompañamiento durante su trayectoria universitaria.

En este trabajo se considerarán y utilizarán las reuniones de seguimiento como fuente de información. Se dará un enfoque ampliado a esta metodología, enfatizando su importancia en el análisis estadístico. El problema radica en la naturaleza densa y compleja de la información recopilada en los seguimientos debido a la falta de uniformidad en la generación de informes por parte de diversos profesionales, además de no contar con indicadores que permitan medir la efectividad de los objetivos del proyecto. Esta situación genera una dificultad significativa al emplear una metodología estadística precisa para evaluar los resultados de dicha información recopilada en los informes de seguimiento periódicos.

El presente trabajo se tomo una muestra de 41 estudiantes con discapacidad que se encuentran vinculados al proyecto de Discapacidad e Inclusión de la Universidad del Valle, sobre los cuales se hablará posteriormente más a detalle.

1.2. Planteamiento del problema

En el año 2005 el DANE reportó una tasa en la cual se estimaba que el 33.3 % de la población con discapacidad no ingresó al bachillerato y que tan sólo el 2.34 % ingresaron a un nivel de educación superior, siendo el dato más preocupante que únicamente el 1 % de quienes ingresaron a educación superior terminaron sus estudios, estos datos fueron Tomados del censo general hecho en el año mencionado anteriormente por el Departamento Administrativo Nacional de Estadística. Cabe recalcar que el Gobierno de Colombia se encuentra trabajando en la promoción de la educación y el fortalecimiento de los derechos de los estudiantes en situación de discapacidad. Sin embargo, se menciona que “Se encontró que la educación superior inclusiva es una visión educativa que ha ganado cada vez mayor fuerza en el mundo y en Colombia, pero que demanda todavía esfuerzos para su realización plena, particularmente en los aspectos de accesibilidad, practicas pedagógicas incluyentes y ajustes razonables”. (Cobos, 2014, p.1)

En el marco contextual del presente trabajo se abordara la normativa con algunas de las leyes y decretos del país. Además, todos los estudiantes, en especial los que presentan alguna condición de discapacidad son propensos a necesitar apoyos, ya que dependiendo de sus limitaciones, el estudiante requerirá acceso a ciertos servicios como medicamentos, fisioterapia, apoyo psicológico, apoyo académico, guía, tutores, profesores, mediaciones en el acto pedagógico, Entorno arquitectónico e incluso también algún tipo de beneficio económi-

co tales como el reconocimiento de pensiones de invalidez, subsidios y asistencia técnica. Lamentablemente, a pesar de los esfuerzos del gobierno colombiano en la promoción de la educación y los derechos de los estudiantes con discapacidad, aún quedan muchos desafíos en el proceso de garantizar su acceso a una educación de calidad, como en muchos países alrededor del mundo (Armstrong, 1999).

El Proyecto de Discapacidad e Inclusión de la Universidad del Valle, entidad que se encuentra comprometida con la educación superior, promueve acciones de inclusión para facilitar la óptima participación de los estudiantes con discapacidad en el desarrollo de su aprendizaje y su crecimiento personal articulándose con docentes y monitorias denominadas especiales. Además de otras acciones de impacto institucional. El objetivo del Proyecto de Discapacidad e Inclusión es lograr que los estudiantes que acceden a él, experimenten la vida universitaria en igualdad de oportunidades que el resto de sus pares. Para ello se tiene en cuenta información de diferentes fuentes, entre ellas; la historia clínica, una caracterización, diálogo con docentes y análisis del contexto, estos datos se recolectan mediante la aplicación de una entrevista inicial, además de recopilar los requerimientos y necesidades del estudiante con discapacidad para posteriormente realizar un seguimiento exhaustivo de su proceso.

El Proyecto de Discapacidad e Inclusión de la Universidad del Valle hasta la segunda mitad del año 2022 contaba con 91 estudiantes activos que hacen parte de diferentes sedes. Cada estudiante recibe seguimiento de acuerdo a la duración de su participación en el proyecto. Hoy por hoy, los seguimientos contienen una amplia gama de información escrita sobre las tareas llevadas a cabo por el proyecto y su equipo, así como las actividades y procedimientos involucrados en la atención de los estudiantes con discapacidad.

En el marco del proyecto y la política de discapacidad e inclusión, se realizan diferentes tipos de actividades de acuerdo con cada caso específico y el entorno en el que se mueve el estudiante beneficiario. Hasta la fecha, las profesionales del proyecto realizan informes de actividades, densos y de alto contenido técnico, lo que hace difícil hacer un seguimiento pragmático a los resultados del proceso de acompañamiento, pues, muchos de los mismos, se diluyen entre el alto volumen de actividades que deben ser realizados para cada estudiante para cada semestre en el que se encuentre matriculado (Perez, 2021).

Debido al tipo de seguimiento utilizado, los informes escritos y el traslado de información con el pasar de los años de formato físico al digital, se ha recopilado un repositorio de reportes de cada estudiante con discapacidad que ha sido admitido en la universidad.

Lo anteriormente mencionado, referente a la metodología de recolección de información en el proyecto, no permite contar con un sistema de información que proporcione indicadores de gestión o desempeño de las actividades e intervenciones realizadas desde el proyecto y

orientadas a los estudiantes con discapacidad. Lo anterior evidencia la necesidad de iniciar la construcción de indicadores numéricos para así, poseer disponibilidad de información concreta respecto al desempeño, lo cual permitirá el establecimiento de objetivos y la evaluación de su cumplimiento, con el objetivo de apoyar y reforzar los procesos de toma de decisiones basadas en datos a niveles administrativos y económicos. En consecuencia, se aplica la primera fase del trabajo, la cual consiste en el tratamiento de la información, donde se exploran técnicas estadísticas como el “análisis de texto” o la “minería de texto”, esta etapa permitirá posteriormente avanzar a una segunda fase.

En la segunda fase, basados en la experiencia y conocimiento de los profesionales se procede a construir categorías y sub-categorías que abarcan tres tipos de información, la académica, la personal y el acompañamiento brindado desde el proyecto. Posteriormente se utiliza metodología estadística donde se determina la existencia de las palabras contenidas en el corpus que constituye el documento de cada estudiante en cada categoría/ posteriormente a partir de dos sub-categorías se establecen grupos independientes en los cuales se diferencia la presencia de una o ambas sub-categorías y se agrupan los estudiantes en estos grupos. Lo anterior, con el objetivo de por medio de la metodología estadística construir el índice estadístico que pueda evaluar el desempeño del proyecto en la vida académica de los estudiantes beneficiados por el mismo.

1.3. Justificación

La inversión de la Universidad del Valle en este proyecto demuestra su compromiso de mejorar las oportunidades de los estudiantes con discapacidad adoptando la Política de discapacidad e inclusión da Universidad del Valle.

El proyecto de discapacidad e inclusión actualmente avanza en brindar adecuadas condiciones para la participación de las Personas con Discapacidad en la vida universitaria, esto conlleva una serie de gestiones, organización e inversiones. Por lo cual se evidencia la necesidad de contar con un índice que permita evitar interpretaciones de la información subjetivas o sesgos personales debido a la forma en que se presenta a día de hoy. Por lo anterior la presentación de resultados de manera objetiva es una parte valiosa y necesaria del desarrollo de la actividad propia del proyecto. La toma de decisiones requiere una base sólida en la cual pueda realizar un análisis objetivo, comparar resultados e implementación de mejoras y dado el caso poder sostener con argumentos la necesidad de una mayor inversión. En este trabajo de grado se propone una metodología para la creación de un índice basado en la experiencia y seguimientos elaborados por las profesionales de acompañamiento.

1.4. Objetivos

1.4.1. Objetivo General

Proponer una metodología estadística usando probabilidades para evaluar el desempeño del proyecto de discapacidad e inclusión de la Universidad del Valle.

1.4.2. Objetivos específicos

- Construir un conjunto de datos a partir de los ajustes razonables “informe de seguimiento” que permita establecer categorías desde el conocimiento técnico de los profesionales de rehabilitación.
- Establecer las características propias del proyecto que permitan obtener información para ser usadas como indicadores de desempeño.
- Proponer un índice que permita medir el desempeño del proyecto de Discapacidad e Inclusión.

2 Antecedentes

2.0.1. Antecedentes contextuales

1. La inclusión académica de los estudiantes en situación de discapacidad en la Universidad del Valle 2009

El trabajo realizado por el estudiante Andrés Mauricio Giraldo Naranjo (2009) para obtener el grado como sociólogo en la Universidad del Valle presentó en el capítulo tres una descripción de 20 estudiantes con discapacidad a los cuales se les tomó una encuesta y construyó una caracterización resaltando dos puntos; el ingreso y la permanencia. En el primero describe el número de intentos que los estudiantes realizaron para conseguir ingresar a la universidad, el motivo de ingreso, la presentación del examen de admisión, si este tuvo apoyo para presentar el examen, el programa de estudios al que ingresó y la influencia de su discapacidad a la hora de seleccionar un programa académico.

Para la permanencia del estudiante, construyó una caracterización de eventos dirigida a las actividades y espacios universitarios tales como; dificultad para la asistencia a clases, problemas de desplazamiento en el campus Melendez y San fernando, inasistencia a clases por tipo de discapacidad, problemas para cursar asignaturas por su discapacidad, ayudas pedagógicas tales como como la utilización de video Beam, la aplicación de talleres en clase y demás, barreras arquitectónicas, cumplimiento con las obligaciones académicas y la frecuencia con la que se presentaban dos tipos de dificultad en aula, los cuales son: explicación del docente y movilidad en el aula. Además de seis tipos de características dirigidas a las relaciones interpersonales en la universidad. Cabe mencionar que el trabajo mencionado es un punto contextual interesante pues la información presentada da una imagen del entorno de los estudiantes con discapacidad antes del acuerdo N.o. 004 de julio 12 del 2013 donde se adopta la “Política institucional de discapacidad e inclusión de la Universidad del Valle” incluso antes de la creación del Programa de Rediversia y la asignación de nuevos recursos distintos a los brindados por recursos humanos. (Giraldo, 2009)

2. Caracterización de los perfiles de apoyo de estudiantes con discapacidad vinculados al programa de Rediversia de la Universidad del Valle

Trabajo de grado realizado por el estudiante Brayan Steven Pérez Pérez con el fin de graduarse del pregrado en estadística de la Universidad del Valle. Este trabajo es un punto de partida contextual, pues Rediversia es un programa de la Escuela de Rehabilitación Humana encabezado actualmente por la profesora Mónica María Carvajal.

La practica de Rediversia realiza una labor en el proceso de formación enfocado en el acompañamiento de los estudiantes que entraron a la Universidad con algún tipo de discapacidad o eventualmente dentro del proceso formativo en la universidad fue diagnosticado con una.

Rediversia cuenta con practicantes que semestralmente hacen seguimiento a el estudiante y hacer informes técnicos, pues estos son necesarios para demostrar la finalidad de su práctica. En este trabajo se tomó las fichas de seguimiento de doce estudiantes y se llevó a cabo una caracterización considerando tres variables; tipo de apoyos, intensidad de apoyo y actividades de la vida universitaria. A partir de esto ver las relaciones entre ellas mismas llevando a resultados donde muestra que las actividades mas realizadas son las académicas en donde envuelve la realización de tareas o trabajos. Además, el apoyo que se fue brindado y con mayor intensidad fue el apoyo de personas que les ayuda en sus tareas. (Perez, 2021)

2.0.2. Antecedentes estadísticos

3. Análisis de temas utilizando Twitter: Una aplicación del modelo LDA al caso Colombiano.

Realizado por Manuel Alejandro Díaz Rubiano de la Universidad Santo Tomas (2022). El estudio resaltó la importancia actual de las tecnologías que nos permiten mantenernos conectados a las personas de manera instantánea, desde familiares hasta desconocidos. En el trabajo mencionado se seleccionó la red social Twitter de entre una lista de redes representativas como Facebook, Instagram y Snapchat, la cual a la fecha es una de las redes más utilizadas contando con más de 300 millones de usuarios, en donde se publican de manera escrita pensamientos y opiniones de los usuarios respecto a gran variedad de temas. Otra de las razones por la cual esta red social fue seleccionada fue el diseño del API, el cual proporciona acceso a la base de datos de tweets de los usuarios, lo cual la hace más susceptible de análisis de una manera más sencilla.

Los datos proporcionados mediante el acceso al API pueden tener distintos usos, un

ejemplo de uso es el análisis exploratorio sobre la opinión de las personas en temas relacionados a la política, utilizando filtros por áreas. Teniendo en cuenta que la base de datos está conformada por opiniones escritas, se utilizaron distintas técnicas de (NLP) Procesamiento de lenguaje natural, como la limpieza de palabras, palabras vacías “Stop Word”, espacios, emotes, signos, puntuación y lematización, además de otras técnicas de agrupamiento. La asignación de Dirichlet Latente (LDA) es un modelo estadístico generativo que permite explicar conjuntos de observaciones mediante grupos no observados que explican por qué algunas partes de los datos son similares (Díaz, 2022)

4. Implementación de un modelo de análisis de sentimientos con respecto a la JEP basado en minería de datos en Twitter.

Trabajo realizado por Erika Paola Paez y Andres Felipe Monroy (2020) Universidad católica de Colombia. El trabajo tuvo como interés el análisis de sentimientos con el objetivo de caracterizar la posición política de los usuarios de la red social Twitter respecto a la Jurisdicción Especial para la Paz (JEP) y basados en la caracterización obtenida, determinar el algoritmo con mejor desempeño en este proceso de obtención. Aunque el trabajo se dividió en seis fases nos centramos en las de nuestro interés, las cuales fueron: la construcción de la base de datos, el procesamiento de lenguaje natural (NLP) y la extracción de características. También es importante resaltar las limitaciones presentadas en el desarrollo del trabajo, puesto que las características utilizadas son calificaciones a los sentimientos de las personas, las cuales en este caso se limitan a tres posibles: positivos, neutros y negativos. (Paez, 2020)

5. Eliciting Dirichlet and Connor-Mosimann prior distributions of multinomial models.

Artículo realizado por Fadlalla Elfadaly y Paul Garthwaite en el año 2013. Consiste en obtener los hiperparámetros de una distribución previa informativa para modelos multinomiales por medio de la distribución beta. La distribución de Dirichlet presenta dificultades con respecto a su flexibilidad. Uno de sus inconvenientes es el número limitado de parámetros, que sus variables siempre están correlacionadas negativamente por lo cual no se puede especificar independencia de sus valores medios y dificulta representar con exactitud la información del experto.

A partir de las dificultades de la distribución Dirichlet se crea la distribución Connor-Mosimann, la cual comparte algunas de las propiedades deseables de la distribución estándar de Dirichlet siendo conjugada a modelos multinomiales y presenta un mayor

número de parámetros $2(k-1)$. La función de densidad de la distribución Connor-Mosimann es:

$$\pi(p_1, p_2, \dots, p_k) = \prod_{i=1}^{k-1} \left\{ \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} p_i^{a_i-1} \left(\sum_{j=1}^k p_j \right)^{b_{i-1}-(a_i+b_i)} \right\} p_k^{b_{k-1}-1} \quad (2-1)$$

Donde $\sum p_i=1$, $0 \leq p_i \leq 1$, $a_i > 0$ y $b_i > 0$ para $i=2, \dots, k-1$. y b_0 es arbitrario. Si se hace $b_{i-1} = a_i + b_i$, entonces la ecuación 2-1 se reduce a la función de densidad de la distribución dirichlet con $\alpha_i = a_i$ y $\alpha_k = b_{k-1}$ para $i = 1, \dots, k-1$.

3 Marco teórico

3.1. Marco contextual

3.1.1. Normativa internacional

Datos de la Organización Panamericana de la Salud

- En total, se estima que casi el 12 % de la población de América Latina y el Caribe vive con al menos una discapacidad, lo que representa alrededor de 66 millones de personas
- 17 países han firmado la Convención Interamericana para la Eliminación de Todas las Formas de Discriminación contra las Personas con discapacidad.
- Las personas con discapacidad tienen de 2 a 4 veces más probabilidades de morir en desastres y emergencias que las personas sin discapacidad.
- Muchos establecimientos y servicios de salud son inaccesibles para las personas con discapacidad. Entre las barreras a las que se enfrentan las personas con discapacidad figuran los espacios de salud inaccesibles, las barreras de comunicación, la falta de formación de los profesionales y las barreras financieras.

Actualmente existen muchas leyes y normativas internacionales que protegen los derechos de las personas en condición discapacidad en la sociedad. Algunas de las más destacadas son:

- Convención sobre los Derechos de las Personas con Discapacidad (CDPD): Es un acuerdo internacional sobre los derechos de las personas con discapacidad. “...Hacer posible que las personas con discapacidad participen de manera efectiva en una sociedad libre.” (*Organización de las naciones unidas*, 2006).
- Declaración de Salamanca: Publicada por la UNESCO en 1994 y respaldada por 92 países, esta aboga por la educación inclusiva como un medio para garantizar el acceso a una educación de calidad.
- Informe Delors: El informe de la Comisión Internacional sobre Educación para el Siglo XXI encabezada por Jacques Delors, publicado por la UNESCO en 1996, enfatizó la importancia de la educación inclusiva y el respeto por las diferencias en la educación.

Estas son algunas de las normativas internacionales que se enfatizan en brindarle importancia a la educación inclusiva y de calidad a las personas con discapacidad, facilitando su participación en la sociedad y el validando el ejercer de sus derechos.

3.1.2. Normativa nacional

En Colombia, la normativa nacional que protege los derechos de las personas con discapacidad en el ámbito educativo incluye las siguientes leyes y decretos:

Decreto 1421. *“Por el cual se reglamenta en el marco de la educación inclusiva la atención educativa a la población con discapacidad”*: Este decreto regula la atención educativa para la población con discapacidad visual, auditiva, sordoceguera, entre otras. Establece las pautas para la prestación de servicios educativos especializados, incluyendo la adaptación de materiales y la formación de docentes especializados (*Presidencia de la república de Colombia*, 2017).

Ley Estatutaria 1618. *“Por medio de la cual se establecen las disposiciones para garantizar el pleno ejercicio de los derechos de las personas con discapacidad”*: Esta ley establece el marco legal para la garantía de los derechos de las personas en situación de discapacidad, incluyendo el derecho a la educación inclusiva. Establece medidas para promover la inclusión educativa y fomentar la igualdad de oportunidades en el sistema educativo (*Presidencia de la república de Colombia*, 2013)

Decreto 366. *“Por medio del cual se reglamenta la organización del servicio de apoyo pedagógico para la atención de los estudiantes con discapacidad y con capacidades o talentos excepcionales en el marco de la educación inclusiva”*: Este decreto regula el Sistema Nacional de Educación para el Trabajo y el Desarrollo Humano. Establece que las instituciones educativas deben ofrecer programas de formación profesional inclusivos que respondan a las necesidades de las personas con discapacidad (*Presidencia de la república de Colombia*, 2009).

3.1.3. Normativa a nivel institucional

Acuerdo No. 004. *“Por la cual se adopta la Política Institucional de Discapacidad e Inclusión de la Universidad del Valle”* (*Consejo Superior de la Universidad del Valle*, 2013).

Resolución No. 040. *“Por la cual se reglamentan las Monitorías en la Universidad del Valle”* (*Consejo Superior de la Universidad del Valle*, 2002).

3.2. Marco teórico Estadístico

3.2.1. Variable latente

En el Proyecto de Discapacidad e Inclusión todo se registra de manera verbal, debido a esto el uso de variables latentes es fundamental. En general, todo lo que es medible se puede ver afectado de varios fenómenos que no son observables, así mismo, en la práctica existen fenómenos no observables que son afectados y están relacionados por otros que sí son observables y pueden medirse con precisión bajo ciertas condiciones, esto es el concepto de variable latente (Latorre, 2005). Un ejemplo común de una variable latente es la inteligencia, ya que esta no se puede medir directamente, pero se pueden usar varias pruebas o medidas (variables observables), como el coeficiente intelectual, el rendimiento académico, la capacidad de resolución de problemas, y más. En pocas palabras, una variable latente es una construcción o concepto hipotético que describe un fenómeno o proceso que no se puede medir directamente, sino que se mide o determina mediante variables observables.

3.2.2. Procesamiento del Lenguaje Natural

El Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) es una rama de la inteligencia artificial y la lingüística computacional que se ocupa de la interacción entre los seres humanos y las computadoras a través del lenguaje natural. El objetivo principal del NLP es permitir a las computadoras comprender, interpretar, analizar y generar el lenguaje humano de manera similar a como lo hacen los seres humanos (Silge, 2017).

El NLP involucra una amplia gama de tareas y técnicas, que incluyen:

- **Análisis morfológico y sintáctico:** El primero se enfoca en la estructura interna y la construcción de las palabras, mientras que el análisis sintáctico se enfoca en las relaciones gramaticales entre las palabras en oraciones que solo el lenguaje natural puede comprender y expresar.
- **Reconocimiento de entidades nombradas:** Reconoce y clasifica entidades como nombres de personas, nombres de organizaciones y nombres de lugares en palabras.
- **Análisis semántico:** Busca comprender el significado de palabras y oraciones, relacionarlas con ideas y expresar el propósito o la intención de las palabras.
- **Análisis de sentimientos:** Identifica y evalúa ideas y conceptos por escrito. Se utiliza para determinar si un propósito de un texto es positiva, negativa o neutral en áreas como la extracción de opiniones y el análisis de contenido de las redes sociales.
- **Generación de lenguaje natural:** Se refiere a la capacidad de las computadoras para generar texto o habla de manera coherente y comprensible para los seres humanos.

- **Preprocesamiento de los documentos:** Es la fase donde se eliminan los stopwords, se realiza la lematización o el stemming y los documentos quedan listos para su análisis.
- **Interpretación de los resultados:** Explora palabras y las relaciona para comprender y nombrar temas puntuales. También puede mostrar los resultados como gráficos o matrices de temas y palabras.

El NLP es utilizado en una amplia gama de aplicaciones, como asistentes virtuales, traducción automática, análisis de sentimientos en redes sociales, resumen automático de textos, recuperación de información, chatbots, entre otros. A través del procesamiento del lenguaje natural, las computadoras pueden interactuar de manera más natural y efectiva con los humanos, comprendiendo y generando lenguaje en diversas formas y contextos (Ignatow, 2016).

Preprocesamiento de los documentos

Stopwords

También conocidas como palabras vacías, son las palabras frecuentes dentro del documento que no agregan significado al texto y, por lo tanto, se excluyen del análisis estadístico del texto (Vernazza, 2021).

Tokenización

Según Sobrino (2018), el objetivo es dividir el texto en partes más pequeñas llamadas tokens. El cual es una palabra, número, signo de puntuación o cualquier otro elemento dentro del texto. La Tokenización es una parte importante de muchos proyectos de NLP porque ayuda a convertir palabras en patrones que se pueden procesar mediante aprendizaje automático (Hernández, 2019).

Lematización

Su objetivo es reducir las palabras a sus formas o términos básicos denominados raíz o lema, el cual es la forma canónica de una palabra que indica su propósito principal. Por ejemplo, una palabra como “mesas” quedaría como “mesa”, del mismo modo, palabras como “amigas”, “amigos”, “amiguitas” y “amiguitos”, tienen en común la raíz “amigo” (Jurafsky, 2009).

3.2.3. Unigrama y Bigrama

Un unigrama, también conocido como “monograma” o “uni-término”, es una unidad básica del procesamiento del lenguaje natural. Un unigrama es una secuencia de un elemento, que

generalmente denota una sola palabra. Por otra parte, bigrama hace referencia a dos palabras (Sobrino 2018).

3.2.4. Índice

El término índice se usa para denotar, comparar o evaluar algún aspecto en específico de un conjunto de datos (Silva, 1997). En nuestro caso, es una medida resumida y cuantitativa que proporciona información relevante y fácilmente interpretable sobre una característica o fenómeno específico.

Existen diferentes tipos de índice utilizados en diferentes campos. Así mismo existen diferentes maneras construir un índice dependiendo el contexto y objetivos del estudio. algunos métodos pueden ser:

- Método de ponderación simple
- Método de ponderación basada en la varianza.
- Método de ponderación basada en la correlación.
- Método de ponderación basada en análisis factorial.
- Método de ponderación basada en análisis de componentes principales.
- Método de ponderación basada en regresión.
- Método de ponderación basada en análisis de preferencias.

Para este trabajo se utilizara el método de ponderación basada en análisis de preferencias en este procedimiento se utiliza un enfoque de elección discreta, donde a cada variable se le estable un peso en función de la importancia o influencia que presenta según el caso estudiado. (Posada, 2015)

3.2.5. Distribución Multinomial

La distribución multinomial es una distribución de probabilidad discreta que describe el número de ocurrencias de diferentes categorías en un experimento con múltiples resultados posibles. Es una generalización de la distribución binomial, que se aplica cuando hay más de dos resultados posibles (Johnson, 1997).

En la distribución multinomial, se tienen n ensayos independientes y cada ensayo puede tener uno de k resultados posibles, donde k es un número finito. Cada resultado tiene una probabilidad fija de ocurrencia en cada ensayo, y la suma de las probabilidades de todos los resultados es igual a 1.

La función de masa probabilidad de la distribución multinomial se puede expresar de la siguiente manera:

$$f(x_1, x_2, \dots, x_k; n, p_1, p_2, \dots, p_k) = \begin{cases} \frac{n!}{x_1! * x_2! * \dots * x_k!} * p_1^{x_1} * p_2^{x_2} * \dots * p_k^{x_k} & \text{cuando } \sum_{i=1}^k x_i = n \\ 0 & \text{en otro caso} \end{cases} \quad (3-1)$$

Donde:

- $x_1, x_2, \dots, x_k$ son las cantidades observadas de cada resultado.
- $\sum_{i=1}^k p_i = 1$
- $p_1, p_2, \dots, p_k$ son las probabilidades de ocurrencia de cada resultado.

La distribución multinomial se utiliza en diversos campos, como estadística, ciencias sociales, biología, genética y economía, cuando se desea modelar situaciones en las que hay múltiples categorías posibles con probabilidades de ocurrencia conocidas.

Además la forma de la esperanza y varianza son:

- $E(x_i) = n * p_i$ Donde representa el numero esperado de veces que cada categoría es observada en todos los ensayos.
- $Var(n_i) = np_i(1 - p_i)$

3.2.6. Distribucion Dirichlet

La distribución de Dirichlet es una relevante distribución continua multivariada en el ámbito de la probabilidad y estadística. Se considera una generalización multivariada de la distribución beta y se utiliza de manera natural en el modelado de datos de composición y medidas de proporciones. (Lin, 2016)

Sea $Q = \{Q_1, Q_2, \dots, Q_k\}$ una función de probabilidad aleatoria, que cumple: $0 \leq Q_i \leq 1$ para $i = 1, 2, 3, \dots, k$ y $\sum_{i=1}^k Q_i = 1$. igualmente el supuesto que $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$, con $\alpha_i > 0$;

$$\sum_{i=1}^k Q_i \alpha_i = 1$$

Entonces, se dice que Q tiene una distribución Dirichlet con vector de parámetros α , denotándose como $Q \sim \text{Dirichlet}(\alpha)$ (Frigyik and Gupta, 2010)

Función de masa de probabilidad

$$f(q; \alpha) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k q_i^{\alpha_i-1} ; \sum_{i=1}^k q_i = 1, 0 \leq q_i \leq 1, \alpha_0 = \sum_{i=1}^k \alpha_i$$

A partir de lo anterior se conoce que la media y la varianza de la distribución Dirichlet están dadas por 4-4 y 4-5, respectivamente (Lin, 2016).

$$E[q_i] = \frac{\alpha_i}{\alpha_0}$$

$$Var[q_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)} ; \text{ para } i = 1, 2, \dots, k$$

En estadística bayesiana, la distribución de Dirichlet es especialmente útil, ya que actúa como una priori conjugada para la distribución multinomial.

4 Metodología

4.1. Lista general de estudiantes del Proyecto de Discapacidad e Inclusión

El Proyecto proporcionó un listado general donde se encontraban 123 estudiantes desde el año 2016 hasta 2020, este listado contiene información sobre los estudiantes activos, activos sin acompañamiento, retirados, egresados, aspirantes, no interesados y las diferentes sedes donde se encuentran matriculados. La información presentada en este listado contiene información personal como: Tipo de discapacidad, diagnóstico, origen de la discapacidad, características de la discapacidad, funciones en las que presenta dificultades, programa académico, sexo y demás información relacionada con la identidad de la persona.

Posteriormente se realizaron estadísticas descriptivas para tener información usando las principales características que permitan delimitar la posible población de estudio. El estado del estudiante, para observar la cantidad de estudiantes activos, retirados y egresados del proyecto; Sede, para ver dónde el proyecto tiene mayor cantidad de estudiantes; nivel educativo, puesto que dentro del proyecto de discapacidad e inclusión se encuentran tanto estudiantes de pregrado como de posgrado; y tipo de discapacidad, para identificar la condición de salud, posibles barreras y facilitadores de participación que los/las estudiantes tienen en la vida universitaria.

4.1.1. Estructura de la información

Dentro del Proyecto de Discapacidad e Inclusión se llevan a cabo diferentes tareas y actividades para con los estudiantes con discapacidad, las cuales se registran y almacenan de diferentes maneras, esto implica la existencia de diferentes formatos, documentos de trabajo y registros de los estudiantes vinculados en el proyecto. Uno de estos componentes se encuentran al realizar una entrevista inicial.

Componente uno

La entrevista de caracterización inicial se realiza de manera verbal y se registra de manera escrita a los estudiantes que entrarán a formar parte del proyecto, en una plantilla creada por el mismo proyecto (y que posteriormente fue revisada y ajustada por un profesor de

estadística), recolectando información personal, económica y académica, además de agregar información como la percepción sobre la discapacidad y características de la misma, donde se busca esclarecer las dificultades que presenta el estudiante frente a diferentes actividades y los tipos de apoyo que podría necesitar. Los datos relevantes pasan al listado general mencionado anteriormente, por lo que este documento con el paso del tiempo pasó a ser en formato digital Word para facilitar el traspaso de datos entre el documento de caracterización inicial y el listado.

Componente dos

Algunos estudiantes con discapacidad necesitan ciertos apoyos, aquí entra en acción el proyecto, que sirve como intermediario entre el estudiante y el programa académico. Estos apoyos pueden ir desde una solicitud para que determinado profesor revise la metodología a implementarse o el sistema de evaluación a realizar con el estudiante, hasta solicitar el cambio de aula si el estudiante presenta barreras motrices o incluso asignar uno o varios monitores que realizan acciones de apoyo acorde a la caracterización y plan de ajustes razonables del semestre en el que se encuentra el estudiante. Para el segundo caso, los monitores asignados tienen diferentes deberes y uno de estos es registrar cada actividad que se realice con el estudiante, la información es almacenada en una plantilla creada por el proyecto en formato digital Excel.

Componente tres

En el presente trabajo se hará énfasis en los ajustes razonables, los cuales son reuniones periódicas con los estudiantes vinculados para llevar a cabo un seguimiento de los mismos, documentando cualquier cambio que presente el estudiante, cualquier solicitud (tanto como el componente uno como el componente dos). Estas sesiones son diligenciadas por distintos profesionales del proyecto y después guardadas en documentos digitales Word. Unos de los principales problemas con este componente, es que no se le realiza a todos los estudiantes o se pierden por la falta de un sistema de información al migrar los documentos de formato físico a digital.

4.2. Delimitación de la muestra

La siguiente fase del trabajo contó con la ayuda de tres especialistas en el Proyecto de Discapacidad e Inclusión, las cuales son profesionales de acompañamiento y cuentan con conocimiento en terapia ocupacional, políticas públicas, educación superior y desarrollo humano. Con su apoyo se establecieron diferentes criterios de inclusión para delimitar los estudiantes que se incorporarían en la base de datos del presente trabajo de grado. Dichos criterios fueron los siguientes:

- El estudiante debe tener al menos un documento de ajuste razonable.
- El/los documentos de ajuste razonable deben estar entre los periodos académicos 2018-1 y 2022-1.
- Debe ser estudiante de pregrado y pertenecer a la sede Meléndez.

Los períodos académicos establecidos se basan en la disponibilidad de documentos de ajustes razonables digitalizados, comenzando con el período 2018-1 y finalizando con el período 2022-1, que representa el último documento encontrado al momento de realizar los criterios de inclusión.

El proceso de recolección de información se llevó a cabo manualmente, accediendo al Google Drive perteneciente al Proyecto donde se almacena toda la información de los estudiantes en carpetas individuales. Se verificó individualmente el cumplimiento de los criterios accediendo a cada carpeta estudiantil y se buscó el período académico específico para comprobar la existencia de documentos de ajustes razonables; luego se consolidó el conjunto de documentos. En total, se obtuvieron 71 documentos de ajustes razonables de 41 estudiantes pertenecientes a diferentes períodos académicos. Cabe destacar que la mitad de los estudiantes se les encontró un documento de ajustes razonables y el estudiante que más documentos tuvo contó con cinco.

4.2.1. Procesamiento de documentos

Para realizar el procesamiento de los documentos con el software estadístico R versión 4.2.2. se transformaron los formatos digitales Word a PDF, ya que las librerías a emplear funcionaban con dicho formato.

Debido a la variabilidad en el número de documentos por estudiante, se optó por llevar a cabo una unión de documentos para cada estudiante mediante iLovePDF, una aplicación online encargada a la manipulación del formato PDF; uniendo los documentos de ajustes razonables por estudiante en un único documento de seguimiento. Por lo cual, cada estudiante tendrá un documento de seguimiento donde se encontrarán los diferentes ajustes razonables. La librería usada para leer los formatos PDF en R fue *pdftools* (Ooms, 2023), además de la librería de Procesamiento de lenguaje natural *NPL* (Hornik, 2020) como una dependencia de *tm* (Feinerer, 2023) que permitió tokenizar los documentos.

Prueba piloto

Con el fin de identificar posibles inconvenientes en el procesamiento del lenguaje natural tales como los algoritmos de limpieza, lematización y organización, además de detectar in-

convenientes con la estructura del formato con el cual se realizaba el ajuste razonable, se realizó una prueba piloto agrupando los periodos académicos de la siguiente manera:

- **Pre-pandemia:** Corresponde a los periodos 2018-1, 2018-2, 2019-1 y finalmente 2019-2
- **Pandemia:** Periodos académicos 2020-1, 2020-2 y 2021-1
- **Post-Pandemia:** Los periodos 2021-2 y 2022-1

Adicionalmente, se agruparon los estudiantes según los periodos académicos donde se les realizó un seguimiento con el fin de utilizar el estudiante con la mayor cantidad de palabras en cada agrupación. Resultando una muestra de cinco estudiantes para la prueba piloto, cada uno dentro de los siguientes agrupaciones:

- Pre-Pandemia y Pandemia
- Pandemia
- Pandemia y Post-pandemia
- Post-pandemia
- Pre, Durante y Post-Pandemia

Código utilizado en la prueba piloto

```
#seleccionar la ruta de trabajo y el tipo de documentos
my_dir <- "C:/Users/EJEMPLO/DE/RUTA"
setwd(my_dir)
files <- list.files(path = my_dir, pattern = "pdf$")

#Leer y tokenizar los documentos
corp <- Corpus(URISource(files, encoding = "latin1"),
               readerControl = list(reader = readPDF,
                                   language = "es-419"))

#Limpieza
corp <- tm_map(corp, content_transformer(removePunctuation))
#Elimina puntuacion
corp <- tm_map(corp, content_transformer(tolower))
#poner todo en minusculas
corp <- tm_map(corp, stripWhitespace)
#Eliminar doble espaciado
corp <- tm_map(corp, removeWords, stopwords())
```


nombre” entre otras) mayor a la que realmente debería tener el estudiante. Por tanto se eliminaron manualmente los títulos, subtítulos y demás caracteres mencionados inherentes al diligenciar el seguimiento.

- La lematización: Las librerías utilizadas para este proceso tienen integrado unos diccionarios que permiten identificar las palabras y sus raíces. La prueba piloto fue hecha con librerías que contienen diccionarios en ingles, por esta razón se trabajó con la librería *udpipe* (Wijffels, 2023) que cuenta con un diccionario en español.
- Palabras personalizadas: Se construyó un diccionario que contuviera nombres, apellidos, meses entre otras palabras comunes y coloquiales que son frecuentes dentro del seguimiento pero su aporte no es significativo al texto (Se agregaron 2988 caracteres).

Código con correcciones utilizado

```
#seleccionar la ruta de trabajo y el tipo de documentos
my_dir <- "C:/Users/EJEMPLO/DE/RUTA"
setwd(my_dir)
files <- list.files(path = my_dir, pattern = "pdf$")

#Leer los documentos
corp <- Corpus(URISource(files, encoding = "latin1"),
               readerControl = list(reader = readPDF,
                                   language = "es-419"))

#Limpieza
corp <- tm_map(corp, content_transformer(str_replace_all),
               pattern="http\\S*", "") #elimina paginas http
corp<-tm_map(corp, content_transformer(str_replace_all),
             "www\\S*", "") #elimina paginas www
corp<-tm_map(corp, content_transformer(str_replace_all),
             "meet\\S*", "") #elimina meets
corp <- tm_map(corp, content_transformer(gsub),
               "[[:punct:]]", "") #elimina simbolos
corp <- tm_map(corp, content_transformer(gsub),
               "[[:digit:]]", "")
corp <- tm_map(corp, content_transformer(removePunctuation))
#Elimina puntuacion
corp <- tm_map(corp, content_transformer(tolower))
```

```

#poner todo en minusculas
corp <- tm_map(corp, stripWhitespace)
#Eliminar doble espaciado
corp <- tm_map(corp, removeWords, stopwords())
#Eliminar palabras en espanol
corp <- tm_map(corp, removeNumbers)
#Eliminar numeros

#Se inserta el diccionario creado
nnn <- read_excel("DICCIONARIO_PERSONAL.xlsx")
#borrar nombres, apellidos, meses y
                                otros caracteres personalizados
mywords <- c(nnn$Nombres, nnn$Apellidos, nnn$Meses)
corp <- tm_map(corp, removeWords,
c(stopwords("spanish"), mywords))
corp <- tm_map(corp, content_transformer(gsub), "a'", "a")
corp <- tm_map(corp, content_transformer(gsub), "e'", "e")
corp <- tm_map(corp, content_transformer(gsub), "i'", "i")
corp <- tm_map(corp, content_transformer(gsub), "o'", "o")
corp <- tm_map(corp, content_transformer(gsub), "u'", "u")

#lematizacion (convertir palabras a su raiz)
tdm <- TermDocumentMatrix(corp)
# term document - terminos van en filas y
                                documentos en columnas
n<-sort(rowSums(as.matrix(tdm)), decreasing=TRUE)
df_tdm <- data.frame(word=names(n), freq=n,
freqrel=round(100*n/sum(n), 2))

#udpipe_download_model("spanish")
model <- udpipe_load_model(file =
"spanish-gsd-ud-2.5-191206.udpipe")
df_lemma <- udpipe_annotate(model,
x=df_tdm$word, doc_id =df_tdm$word)
df_lemma <- as_tibble(df_lemma)

df_lemma_select <- df_lemma |>
  filter(token_id==1) |>
  select(doc_id, lemma)

```

```

df_tdm2 <- left_join(df_tdm, df_lemma_select ,
                     by=c("word"="doc_id"))
sum(is.na(df_tdm2$lemma))
df_tdm2$word=NULL

df_tdm2 <- df_tdm2 |>
  group_by(lemma) |>
  summarise(freq=sum(freq))

#Creacion de la nube de palabras
nb1<-wordcloud2(df_tdm2, size=0.75, backgroundColor="White"
, color="random-light", shuffle=FALSE)

```

La correcta limpieza de los documentos permite reducir el ruido y mejorar la eficiencia del análisis, es por esto que se deben eliminar links personalizados, símbolos, puntuación, doble espaciado, palabras en español, números, además de insertar el diccionario de elaboración propia y transformar las vocales acentuadas en vocales no acentuadas. De igual manera, los stopwords también forman parte de este componente, ya que consisten en palabras no relevantes. En español son palabras como artículos, conjunciones y preposiciones, por ejemplo: él, ella, para, con, es, entre otros. Estas últimas fueron desechadas con la librería *stopwords* (Benoit, 2021).

4.2.2. Creación de unigramas y bigramas

En la figura 4-2 se muestra la frecuencia de una unidad básica de procesamiento del lenguaje natural conocida como unigrama, aumentando el tamaño del unigrama entre mayor sea su frecuencia. Dependiendo el contexto cada palabra puede dar información, de modo que el contexto para la interpretación del unigrama es una pieza importante para la información que este contiene. Por esta razón se construyeron los bigramas, que son conformados por dos palabras y así complementar el contexto, teniendo un equilibrio sobre la cantidad de n-gramas necesarios para el significado contextual que se presenta.

La tabla 5-1 refleja la frecuencia de las palabras entre documentos e intra documentos. Las frecuencias de unigramas intra documentos se calcula usando el algoritmo para todos los documentos, mientras que las frecuencias entre documentos se calcula usando el algoritmo para cada uno de los documentos; extrayendo las frecuencias de unigramas para cada estudiante para después usar el algoritmo en todos estos resultados, finalizando con la frecuencia de



Figura 4-2: Nube de palabras con los respectivos ajustes

unigramas entre documentos, que en términos más simples significa las veces que aparece al menos una vez un determinado unigrama en todos los documentos.

4.2.3. Creación de categorías y sub-categorías

Los profesionales del Proyecto de Discapacidad e Inclusión de la Universidad del Valle diseñaron en conjunto una serie de categorías y sub-categorías para clasificar las palabras en términos de la información según el contexto. Las categorías están presentadas a continuación con las definiciones de las respectivas sub-categorías.

- Proyecto de Discapacidad
- Programa académico
- Sujeto

Proyecto de discapacidad:

Procesos administrativos: El componente administrativo en el marco de un proyecto se define como el conjunto de actividades realizadas por los expertos con el objetivo de garantizar el funcionamiento del proceso de apoyo dirigido a los estudiantes libre de contratiempos, incluyendo dentro de sus funciones la gestión del recurso financiero y los procesos de apoyo como la gestión del talento humano, realización de agendas y verificación del cumplimiento de las mismas, entre otros.

Procesos de caracterización: Este componente está constituido por procesos y actividades realizadas con la intención de caracterizar al individuo que accede al proyecto o programa, es decir, busca brindar un conjunto de características representativas del individuo en el instante previo al inicio del acompañamiento, abarca desde su llegada, hasta la realización de una entrevista complementaria. Para llevar a cabo una intervención adecuada, estos procesos deben ir orientados a recopilar información precisa y profunda sobre las necesidades, recursos y contexto del estudiante.

Proceso de ajuste razonable: Son actividades realizadas para reducir las barreras a los que se enfrenta el alumno durante su formación. La información proporcionada por el sujeto (estudiante) como parte del proceso de caracterización —la historia clínica, la narración del estudiante y los relatos de las personas de su entorno— se utiliza para realizar estos ajustes razonables.

Con lo anterior, se realizan los ajustes correspondientes. Por ejemplo, se puede hablar con un profesor para informarle que un estudiante necesita más tiempo para hacer un parcial, solicitar la asignación de un monitor para el estudiante o indicar que el estudiante no puede ver clases en un segundo piso, entre otros casos. Dichos ajustes son implementados con el fin de garantizar la accesibilidad y adaptación al entorno de todos los estudiantes.

Procesos de seguimiento: Comprende todos los procesos que realiza el Proyecto de Discapacidad e Inclusión relacionados a la verificación y comprobación de la Implementación de las recomendaciones realizadas por los profesionales de acompañamiento.

Por ejemplo, si un estudiante con discapacidad cognitiva tiene tres asignaturas en un mismo día y siente que tiene carga académica pesada que está afectando su rendimiento, al acudir con el profesional, este analiza la situación y recomienda disminuir la carga académica o la cancelación de semestre, entonces en las siguientes reuniones de seguimiento se verifica si implemento la recomendación.

Programa académico:

Actividad de aprendizaje: Comprende todas las actividades o hechos que el estudiante realiza como parte de su proceso académico en la institución, tales como clases, seminarios, supervisión académica de diversas materias, participación en diversos eventos, trabajos académicos, exposiciones y proyectos grupales, entre otros.

Contexto de aprendizaje: Se entiende que incluye todos los procedimientos del programa académico que tienen como objetivo mejorar la experiencia académica de un estudiante, como por ejemplo proporcionar un salón de clases acorde a sus necesidades, gestiones de accesibilidad física (rampa) que faciliten la movilidad en caso de ser necesario, o un proceso con el profesor asignado al curso para contextualizar el caso particular del estudiante, de tal

manera que se pueda establecer los ajustes en el aula correspondientes.

Sujeto:

Condición de salud: Se define como el conjunto de características psicológicas, fisiológicas y socioculturales que determinan el completo bienestar de una persona. La salud no se refiere solamente al estar libre de enfermedades sino al grado máximo que se puede lograr cuando se cumplen con todas las características mencionadas anteriormente. Cabe resaltar que el goce de este es uno de los derechos fundamentales de todos los seres humanos sin distinción de religión, raza, ideologías, entre otras variables (*Organización Mundial de la Salud*, 2014).

Curso de vida: El enfoque del curso de vida se entiende como un enfoque que aborda diferentes momentos a lo largo de la vida del ser humano, teniendo en cuenta su desarrollo y los efectos en la salud que se deben a diferentes factores, entre ellos la exposición a riesgos físicos y sociales, así como a las influencias positivas y negativas durante su trayectoria, esta última abarca las diferentes etapas, es decir, la gestación, la infancia, la adolescencia, la juventud y la vida adulta, todo ello generando a lo largo del tiempo un impacto en los resultados de salud de la persona y la población (minsalud).

4.2.4. Clasificación de unigramas y bigramas

Para posteriormente clasificar los bigramas y los unigramas en las diferentes categorías y sub-categorías, se diseñó una regla de selección donde se cumplieran tres criterios:

- La frecuencia del unigrama o bigrama entre documento fuera mayor a 20 (Que equivale al 50 % de los documentos aproximadamente)
- La frecuencia del unigrama o bigramas intra documentos fuera mayor a 25.
- La palabras sea representativa dentro del concepto de los profesionales del proyecto

La construcción de este criterio se debe a las altas cantidades de unigramas y bigramas en los resultados. Pues, aunque hubiera una alta frecuencia de palabras diferentes, su frecuencia era muy baja y hacía pesado el trabajo de los profesionales al momento de clasificar. Por ende, se armó de tal manera que, al ser implementada, mostrara los unigramas y bigramas que, al ser más frecuentes entre y dentro de los documentos, serían los más representativos. Como parte final, fueron seleccionados según criterio técnico y profesional.

El procedimiento anterior se hizo de manera general, esto quiere decir que se tomó todos los seguimientos. Partiendo de lo anterior, con las sub-categorías establecidas y clasificados los unigramas con ayuda de los bigramas, se construye un algoritmo clasificador, el cual consiste

en aplicar el Análisis textual para cada estudiante y posteriormente clasifica la frecuencia de cada unigramas dentro de su respectiva categoría y sub-categoría. Todo esto mientras hace el recorrido por cada estudiante. **4-3**

Cod	Tipo de discapacidad	Grupo	Cantidad de documentos	Sexo	Proyecto			Programa académico			Sujeto	
					Procesos administrativos	Procesos de caracterización	Procesos de ajustes razonables	Procesos de seguimiento	Actividad de aprendizaje	Contexto académico	Condición de salud	Curso de vida
D10	Visual	Durante, Pre y Pospandemia	3	Femenino	25	12	10	13	27	19	15	31
D11	Visual	Pandemia	1	Femenino	22	12	10	12	26	19	15	29
D12	Múltiple	Pandemia y Pospandemia	2	Femenino	7	7	6	9	14	9	3	20
D13	Múltiple	Prepandemia y Pandemia	2	Masculino	9	4	6	5	9	5	6	9
D14	Espectro autista	Durante, Pre y Pospandemia	4	Masculino	17	3	9	11	21	11	8	22
D17	Cognitiva/Intelectual	Pandemia y Pospandemia	2	Masculino	8	5	6	9	15	8	3	18
D18	Psicosocial	Pandemia	1	Masculino	4	3	2	5	6	5	3	7
D19	Psicosocial	Pospandemia	1	Femenino	2	2	3	2	3	2	1	0
D2	Psicosocial	Pandemia	1	Masculino	3	0	4	4	11	6	1	8
D24	Visual	Prepandemia y Pandemia	2	Masculino	24	12	10	13	27	19	15	31
D25	Psicosocial	Pandemia	1	Masculino	21	11	9	10	18	15	15	22
D26	Auditiva	Pandemia y Pospandemia	2	Masculino	11	5	7	9	19	10	4	14
D28	Cognitiva/Intelectual	Pandemia	1	Masculino	19	10	8	9	16	11	12	21
D29	Visual	Prepandemia y Pandemia	3	Masculino	5	3	1	5	3	7	1	3
D30	Visual	Pandemia	1	Masculino	24	12	10	13	27	19	15	31
D33	Psicosocial	Pandemia	1	Femenino	24	11	9	12	21	16	14	24
D36	Espectro autista	Prepandemia y Pandemia	3	Femenino	16	7	8	9	23	11	5	24
D37	Física	Pandemia	1	Femenino	2	1	0	2	5	2	1	4
D38	Psicosocial	Pandemia	1	Femenino	13	6	1	8	8	8	11	10
D39	Visual	Pospandemia	2	Femenino	24	11	10	12	27	19	15	31

Figura 4-3: Base de datos conformada con categorías y sub-categorías

Proyecto			Programa académico			Sujeto		Grupo 1			Grupo 2			Grupo 3		
Procesos administrativos	Procesos de caracterización	Procesos de ajustes razonables	Procesos de seguimiento	Actividad de aprendizaje	Contexto académico	Condición de salud	Curso de vida	Solo procesos administrativos	Solo procesos de caracterización	Procesos de caracterización y procesos administrativos	Procesos de ajustes razonables	Procesos de seguimiento	Procesos de seguimiento y procesos de ajustes razonables	Condición de salud	Curso de vida	Curso de vida y Condición de salud
25	12	10	13	27	19	15	31	0	0	1	0	0	1	0	0	1
22	12	10	12	26	19	15	29	0	0	1	0	0	1	0	0	1
7	7	6	9	14	9	3	20	0	0	1	0	0	1	0	0	1
9	4	6	5	9	5	6	9	0	0	1	0	0	1	0	0	1
17	3	9	11	21	11	8	22	0	0	1	0	0	1	0	0	1
8	5	6	9	15	8	3	18	0	0	1	0	0	1	0	0	1
4	3	2	5	6	5	3	7	0	0	1	0	0	1	0	0	1
2	2	3	2	3	2	1	0	0	0	1	0	0	1	1	0	0
3	0	4	4	11	6	1	8	1	0	0	0	0	1	0	0	1
24	12	10	13	27	19	15	31	0	0	1	0	0	1	0	0	1
21	11	9	10	18	15	15	22	0	0	1	0	0	1	0	0	1
11	5	7	9	19	10	4	14	0	0	1	0	0	1	0	0	1
19	10	8	9	16	11	12	21	0	0	1	0	0	1	0	0	1
5	3	1	5	3	7	1	3	0	0	1	0	0	1	0	0	1
24	12	10	13	27	19	15	31	0	0	1	0	0	1	0	0	1
24	11	9	12	21	16	14	24	0	0	1	0	0	1	0	0	1
16	7	8	9	23	11	5	24	0	0	1	0	0	1	0	0	1
2	1	0	2	5	2	1	4	0	0	1	0	1	0	0	0	1
13	6	1	8	8	8	11	10	0	0	1	0	0	1	0	0	1
24	11	10	12	27	19	15	31	0	0	1	0	0	1	0	0	1
10	4	5	6	19	8	6	14	0	0	1	0	0	1	0	0	1
22	11	9	11	19	16	14	23	0	0	1	0	0	1	0	0	1
23	11	9	9	18	15	14	23	0	0	1	0	0	1	0	0	1
5	3	2	10	16	6	3	14	0	0	1	0	0	1	0	0	1
16	7	6	8	17	13	12	22	0	0	1	0	0	1	0	0	1
24	10	10	12	26	19	15	31	0	0	1	0	0	1	0	0	1
1	1	1	0	1	1	1	2	0	0	1	1	0	0	0	0	1
10	2	2	9	9	10	2	11	0	0	1	0	0	1	0	0	1

Figura 4-4: Base de datos conformada con grupos

Finalmente, para la construcción de los indicadores, se crean tres grupos en los cuales están las sub-categorías más similares según los profesionales. Donde el primer grupo se refiere a las acciones del *Proyecto de Discapacidad e Inclusión* y está conformado por los unigramas que reflejen simultáneamente las sub-categorías “Procesos administrativos y procesos de caracterización”, el segundo evento refleja únicamente “Procesos administrativos” y el tercer evento refleja únicamente “Procesos de caracterización”; el primer evento del segundo grupo lo conforman los unigramas que estén relacionados con los “Ajustes razonables y procesos

de seguimiento” simultáneamente, el segundo evento únicamente “Ajustes Razonables” y el evento que presente unigramas relacionados con “Procesos de seguimiento”; finalmente, el último grupo se encuentra relacionado con la categoría del *Sujeto*, donde los eventos son los unigramas que contengan en conjunto a las sub-categorías “Condición de Salud y Curso de vida”, sólo “Condición de salud” y sólo “Curso de vida” (Tabla 4-4).

4.2.5. Método de Elicitación predictivo

En el proceso de elicitación se hicieron preguntas a los Profesionales del proyecto enfocadas a la obtención de información que es de interés en el parámetro haciendo referencia a la distribución previa, donde se hizo la siguiente pregunta: ¿Del 1 al 100 Entre qué cantidad pensaría que están los estudiantes en cada grupo de eventos ? Además la asignación del peso “definiendo peso como cual evento dentro del grupo es mas importante” haciendo una escala de [2 ,4], donde cuatro es el Máximo puntaje mientras que dos el mínimo, este se denota como IM_j donde $j=1,2,3$.

	Grupo 1, eventos:	Estudiantes	IM
1	El estudiante en la categoría 1 refleja las sub-categorías (1.a) y (1.b) simultáneamente.	70-80	4
2	El estudiante en la categoría 1 refleja únicamente la sub-categoría (1.a)	10-15	2
3	El estudiante en la categoría 1 refleja únicamente la sub-categoría (1.b)	10-15	3

Tabla 4-1: Cantidad de estudiantes esperados en grupo 1 partiendo de elicitación por profesionales del proyecto.

	Grupo 2, eventos:	Estudiantes	IM
1	El estudiante en la categoría 1 refleja las sub-categorías (1.c) y (1.d) simultáneamente.	60-80	4
2	El estudiante en la categoría 1 refleja únicamente la sub-categoría (1.c)	10-20	3
3	El estudiante en la categoría 1 refleja únicamente la sub-categoría (1.d)	10-20	2

Tabla 4-2: Cantidad de estudiantes esperados en grupo 2 partiendo de elicitation por profesionales del proyecto.

	Grupo 3, eventos:	Estudiantes	IM
1	El estudiante en la categoría 3 refleja las sub-categorías (3.a) y (3.b) simultáneamente.	85-100	4
2	El estudiante en la categoría 3 refleja únicamente la sub-categoría (3.a)	0-10	3
3	El estudiante en la categoría 3 refleja únicamente la sub-categoría (3.b)	0-5	2

Tabla 4-3: Cantidad de estudiantes esperados en grupo 3 partiendo de elicitación por profesionales del proyecto.

4.3. Modelos Estadísticos

4.3.1. Experimento

Se toma la información de un estudiante, obtenida a partir la caracterización realizada por los profesionales, y se observa que grupos presenta. Dado que, cada dos sub-categorías se construyen tres grupos, y se modela cada par de sub-categorías de manera independiente dado estos eventos:

A : {El estudiante se encuentra en el grupo j }

B : {El estudiante se encuentra en el grupo j }

D : {El estudiante se encuentra en el grupo j }

Donde j se mueve de 1 a 3 y representa los eventos de cada modelo mencionado anteriormente.

4.3.2. Modelo estadístico 1

Indicadores

En este caso se tiene tres indicadores de interés.

θ_1 Probabilidad que el estudiante pertenezca a la sub-categoría procesos administrativos.

θ_2 Probabilidad que el estudiante pertenezca a la sub-categoría procesos de caracterización.

θ_3 Probabilidad que el estudiante pertenezca a la sub-categoría procesos administrativos y procesos de caracterización.

Eventos

Pertenencia del estudiante i en la sub-categoría j

$A_1 = \{\text{El estudiante estuvo expuesto a procesos administrativos}\}$, $A_2 = \{\text{El estudiante estuvo expuesto a procesos de caracterización}\}$,
 $A_3 = \{\text{El estudiante estuvo expuesto a procesos administrativos y procesos de caracterización}\}$

Variables aleatorias

Sea $x_1, \dots, x_n$ observaciones de una muestra aleatoria, y sean los eventos antes planteados:

A_1, A_2 y A_3

Entonces $X_j \sim \text{Bernoulli}(\theta_j)$; $j = \{1, 2, 3\}$, donde:

$$x_1 = \begin{cases} 1, \text{Si ocurre } A_1 \\ 0, \text{En otro caso} \end{cases}$$

$$x_2 = \begin{cases} 1, \text{Si ocurre } A_2 \\ 0, \text{En otro caso} \end{cases}$$

$$x_3 = \begin{cases} 1, \text{Si ocurre } A_3 \\ 0, \text{En otro caso} \end{cases}$$

$$\text{Luego: } X_1 = \sum_{i=1}^n x_{i1} \quad X_2 = \sum_{i=1}^n x_{i2} \quad X_3 = \sum_{i=1}^n x_{i3}$$

$$(X_1, X_2, X_3 \mid n, \theta_1, \theta_2, \theta_3) \sim \text{Multinomial}(n, \theta_1, \theta_2, \theta_3)$$

Donde $\theta_1, \theta_2, \theta_3$ corresponden a los parámetros de la distribución a usar.

Función de verosimilitud

Dado que $(X_1, X_2, X_3 \mid n, \theta_1, \theta_2, \theta_3) \sim \text{Multinomial}(n, \theta_1, \theta_2, \theta_3)$ para todo $j = 1, 2, 3$ donde $n = \sum_{i=1}^3 x_i$ y θ_j es la probabilidad que un estudiante pertenezca a un grupo j , el cual es el parámetro a estimar.

$$L(\theta_1, \theta_2, \theta_3 \mid x_1, x_2, x_3) = \frac{n!}{x_1! x_2! x_3!} \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3} \propto \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}$$

Distribución a priori no informativa

Es la especificación en forma de distribución de probabilidad que busca saber que tan probable es que la variable sea afectada, a esta la llamamos distribución a priori. Hay diferentes criterios para identificar la distribución a priori, para este caso en específico distribución Dirichlet con parámetros: $\alpha_1, \alpha_2, \alpha_3$

$$(\theta_1, \theta_2, \theta_3 \mid \alpha_1, \alpha_2, \alpha_3) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3)$$

Con densidad de probabilidad:

$$\pi(\theta_1, \theta_2, \theta_3 \mid \alpha_1, \alpha_2, \alpha_3) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1) + \Gamma(\alpha_2) + \Gamma(\alpha_3)} \theta_1^{\alpha_1-1} \theta_2^{\alpha_2-1} \theta_3^{\alpha_3-1}$$

$$\alpha_1, \alpha_2, \alpha_3 \geq 0; \quad \sum_{i=1}^3 \theta_i = 1$$

En este caso, se emplea la distribución previa de Bayes-Laplace, la cual se caracteriza por simplificar la distribución de Dirichlet a una constante y puede ser interpretada como una distribución uniforme. Al tomar $\alpha_i = (1, 1, 1)$, se obtiene una distribución equiprobable en todas las categorías (Gil, 2019).

Distribución posterior no informativa

Utilizando el teorema de Bayes se calcula la distribución a posterior:

$$f(\theta_1, \theta_2, \theta_3 \mid (x_1, x_2, x_3, \alpha_1, \alpha_2, \alpha_3)) = \frac{L(\theta/x)f(\theta/\alpha)}{\int L(\theta/x)f(\theta/\alpha)d\theta}$$

$$= \frac{\frac{n!}{x_1!x_2!x_3!} \prod_{i=1}^3 \theta_j^{x_j} \frac{1}{B(\theta)} \prod_{i=1}^3 \theta_j^{\alpha_j-1}}{\int \frac{n!}{x_1!x_2!x_3!} \prod_{i=1}^3 \frac{1}{B(\theta)} \prod_{i=1}^3 \theta_j^{\alpha_j-1} d\theta} \propto \prod_{i=1}^3 \theta_i^{\alpha_i+x_i-1}$$

A partir de lo anterior se puede determinar que se distribuye Dirichlet con parámetros $\alpha + X = (\alpha_1 + x_1, \alpha_2 + x_2, \alpha_3 + x_3)$. Para este trabajo en específico los parámetros de la posterior son: $1 + X = (1 + x_1, 1 + x_2, 1 + x_3)$

Distribución a priori informativa

En este caso específico se toma la distribución a priori y por medio del proceso de elicitación que incluye el conocimiento de los expertos, en este caso los profesionales del proyecto de

discapacidad. Se calculan los hiperparámetros usando el método de Elfadaly (2013) y desarrollado en R por Wilson (2017). A esto lo se le llama distribución informativa. La aplicación de este enfoque podría obtener resultados con una mejor precisión en comparación con los informados en este estudio de investigación.

$$\pi(\theta_1, \theta_2, \theta_3 | \alpha_1, \alpha_2, \alpha_3) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1) + \Gamma(\alpha_2) + \Gamma(\alpha_3)} \theta_1^{\alpha_1-1} \theta_2^{\alpha_2-1} \theta_3^{\alpha_3-1}$$

La distribución a priori informativa para el vector de parámetros $\theta = (\theta_1, \theta_2, \theta_3)$ es una distribución Dirichlet con parámetros $\alpha = (9.20, 7.46, 9.23)$

Distribución posterior informativa

Anteriormente, se utilizó el teorema de Bayes para calcular la distribución posterior en el contexto de este análisis. Se distribuye Dirichlet con parámetros $\alpha + X = (\alpha_1 + x_1, \alpha_2 + x_2, \alpha_3 + x_3)$, Para este caso en específico los parámetros son:

$$\alpha_i + X = (9.52 + x_1, 7.46 + x_2, 9.23 + x_3)$$

4.3.3. Modelo estadístico 2

Indicadores

En este caso se tiene tres indicadores de interés.

δ_1 Probabilidad que el estudiante pertenezca a la sub-categoría procesos ajustes razonable.

δ_2 Probabilidad que el estudiante pertenezca a la sub-categoría procesos de seguimiento.

δ_3 Probabilidad que el estudiante pertenezca a la sub-categoría procesos ajustes razonable y procesos de seguimiento.

Eventos

Pertenencia del estudiante i en la sub-categoría j

$B_1 = \{\text{El estudiante estuvo expuesto a procesos de ajustes razonable}\}$, $B_2 = \{\text{El estudiante estuvo expuesto a procesos de seguimiento}\}$,

$B_3 = \{\text{El estudiante estuvo expuesto a procesos de procesos ajustes razonable y procesos de seguimiento}\}$

Variables aleatorias

Sea $y_1, \dots, y_n$ observaciones de una muestra aleatoria, y sean los eventos antes planteados: B_1, B_2 y B_3

Entonces $Y_j \sim \text{Bernoulli}(\delta_j)$; $j=\{1,2,3\}$, donde:

$$y_1 = \begin{cases} 1, \text{Si ocurre } B_1 \\ 0, \text{En otro caso} \end{cases}$$

$$y_2 = \begin{cases} 1, \text{Si ocurre } B_2 \\ 0, \text{En otro caso} \end{cases}$$

$$y_3 = \begin{cases} 1, \text{Si ocurre } B_3 \\ 0, \text{En otro caso} \end{cases}$$

$$\text{Luego: } Y_1 = \sum_{i=1}^n y_{i1} \quad Y_2 = \sum_{i=1}^n y_{i2} \quad Y_3 = \sum_{i=1}^n y_{i3}$$

$$(Y_1, Y_2, Y_3 \mid n, \delta_1, \delta_2, \delta_3) \sim \text{Multinomial}(n, \delta_1, \delta_2, \delta_3)$$

Donde $\delta_1, \delta_2, \delta_3$ corresponden a los parámetros de la distribución a usar.

Función de verosimilitud

Dado que $(Y_1, Y_2, Y_3 \mid n, \delta_1, \delta_2, \delta_3) \sim \text{Multinomial}(n, \delta_1, \delta_2, \delta_3)$ para todo $j = 1, 2, 3$ donde $n = \sum_{i=1}^3 y_i$ y δ_j es la probabilidad que un estudiante pertenezca a un grupo j , el cual es el parámetro a estimar.

$$L(\delta_1, \delta_2, \delta_3 \mid y_1, y_2, y_3) = \frac{n!}{y_1! y_2! y_3!} \delta_1^{y_1} \delta_2^{y_2} \delta_3^{y_3} \propto \delta_1^{y_1} \delta_2^{y_2} \delta_3^{y_3}$$

Distribución a priori no informativa

Es la especificación en forma de distribución de probabilidad que busca saber que tan probable es que la variable sea afectada, a esta la llamamos distribución a priori. Hay diferentes criterios para identificar la distribución a priori, para este caso en específico distribución Dirichlet con parámetros: $\Upsilon_1, \Upsilon_2, \Upsilon_3$

$$(\delta_1, \delta_2, \delta_3 \mid \Upsilon_1, \Upsilon_2, \Upsilon_3) \sim \text{Dirichlet}(\Upsilon_1, \Upsilon_2, \Upsilon_3)$$

Con densidad de probabilidad :

$$\pi(\delta_1, \delta_2, \delta_3 | \Upsilon_1, \Upsilon_2, \Upsilon_3) = \frac{\Gamma(\Upsilon_1 + \Upsilon_2 + \Upsilon_3)}{\Gamma(\Upsilon_1) + \Gamma(\Upsilon_2) + \Gamma(\Upsilon_3)} \delta_1^{\Upsilon_1-1} \delta_2^{\Upsilon_2-1} \delta_3^{\Upsilon_3-1}$$

$$\Upsilon_1, \Upsilon_2, \Upsilon_3 \geq 0; \quad \sum_{i=1}^3 \delta_i = 1$$

En este caso, se emplea la distribución previa de Bayes-Laplace, la cual se caracteriza por simplificar la distribución de Dirichlet a una constante y puede ser interpretada como una distribución uniforme. Al tomar $\Upsilon_i = (1, 1, 1)$, se obtiene una distribución equiprobable en todas las categorías (Gil, 2019).

Distribución posterior no informativa

Utilizando el teorema de Bayes se calcula la distribución a posterior:

$$f(\delta_1, \delta_2, \delta_3 | (y_1, y_2, y_3, \Upsilon_1, \Upsilon_2, \Upsilon_3)) = \frac{L(\delta/y)f(\delta/\Upsilon)}{\int L(\delta/y)f(\delta/\Upsilon)d\delta}$$

$$= \frac{\frac{n!}{y_1!y_2!y_3!} \prod_{i=1}^3 \delta_j^{x_j} \frac{1}{B(\delta)} \prod_{i=1}^3 \delta_j^{\Upsilon_j-1}}{\int \frac{n!}{y_1!y_2!y_3!} \prod_{i=1}^3 \frac{1}{B(\delta)} \prod_{i=1}^3 \delta_j^{\Upsilon_j-1} d\delta}$$

$$\propto \prod_{i=1}^3 \delta_i^{\Upsilon_i+y_i-1}$$

A partir de lo anterior se puede determinar que se distribuye Dirichlet con parámetros $\Upsilon + Y = (\Upsilon_1 + y_1, \Upsilon_2 + y_2, \Upsilon_3 + y_3)$. Para este trabajo en específico los parámetros de la aposterior son: $\Upsilon + Y = (1 + y_1, 1 + y_2, 1 + y_3)$

Distribución a priori informativa

En este caso específico se toma la distribución a priori y por medio del proceso de elicitación que incluye el conocimiento de los expertos, en este caso los profesionales del proyecto de discapacidad. Se calculan los hiperparámetros usando el método de Elfadaly (2013) y desarrollado en R por Wilson (2017). A esto lo se le llama distribución informativa. La aplicación de este enfoque podría obtener resultados con una mejor precisión en comparación con los informados en este estudio de investigación.

$$\pi(\delta_1, \delta_2, \delta_3 | \Upsilon_1, \Upsilon_2, \Upsilon_3) = \frac{\Gamma(\Upsilon_1 + \Upsilon_2 + \Upsilon_3)}{\Gamma(\Upsilon_1) + \Gamma(\Upsilon_2) + \Gamma(\Upsilon_3)} \delta_1^{\Upsilon_1-1} \delta_2^{\Upsilon_2-1} \delta_3^{\Upsilon_3-1}$$

La distribución a priori informativa para el vector de parámetros $\delta = (\delta_1, \delta_2, \delta_3)$ es una distribución Dirichlet con parámetros $\Upsilon = (8.11, 4.97, 9.99)$

Distribución posterior informativa

Anteriormente, se utilizó el teorema de Bayes para calcular la distribución a posterior en el contexto de este análisis. Se distribuye Dirichlet con parámetros. $\Upsilon + Y = (\Upsilon_1 + y_1, \Upsilon_2 + y_2, \Upsilon_3 + y_3)$, Para este caso en específico los parámetros son: $\Upsilon_i + y = (8.11 + y_1, 4.97 + y_2, 9.99 + y_3)$

4.3.4. Modelo estadístico 3

Indicadores

En este caso se tiene tres indicadores de interés.

τ_1 Probabilidad que el estudiante pertenezca a la sub-categoría Condición de salud.

τ_2 Probabilidad que el estudiante pertenezca a la sub-categoría Curso de vida.

τ_3 Probabilidad que el estudiante pertenezca a la sub-categoría Condición de salud y Curso de vida.

Eventos

Pertenencia del estudiante i en la sub-categoría j

$D_1 = \{\text{El estudiante estuvo expuesto a Condición de salud}\}$, $D_2 = \{\text{El estudiante estuvo expuesto a Curso de vida}\}$,

$D_3 = \{\text{El estudiante estuvo expuesto a Condición de salud y Curso de vida}\}$

Variables aleatorias

Sea $k_1, \dots, k_n$ observaciones de una muestra aleatoria, y sean los eventos antes planteados:

D_1, D_2 y D_3

Entonces $K_j \sim \text{Bernoulli}(\tau_j)$; $j = \{1, 2, 3\}$, donde:

$$k_1 = \begin{cases} 1, & \text{Si ocurre } D_1 \\ 0, & \text{En otro caso} \end{cases}$$

$$k_2 = \begin{cases} 1, \text{Si ocurre } D_2 \\ 0, \text{En otro caso} \end{cases}$$

$$k_3 = \begin{cases} 1, \text{Si ocurre } D_3 \\ 0, \text{En otro caso} \end{cases}$$

$$\text{Luego: } K_1 = \sum_{i=1}^n k_{i1} \quad K_2 = \sum_{i=1}^n k_{i2} \quad K_3 = \sum_{i=1}^n k_{i3}$$

$$(K_1, K_2, K_3 \mid n, \tau_1, \tau_2, \tau_3) \sim \text{Multinomial}(n, \tau_1, \tau_2, \tau_3)$$

Donde τ_1, τ_2, τ_3 corresponden a los parámetros de la distribución a usar.

Función de verosimilitud

Dado que $(K_1, K_2, K_3 \mid n, \tau_1, \tau_2, \tau_3) \sim \text{Multinomial}(n, \tau_1, \tau_2, \tau_3)$ para todo $j = 1, 2, 3$ donde $n = \sum_{i=1}^3 k_i$ y τ_j es la probabilidad que un estudiante pertenezca a un grupo j , el cual es el parámetro a estimar.

$$L(\tau_1, \tau_2, \tau_3 \mid k_1, k_2, k_3) = \frac{n!}{k_1! k_2! k_3!} \tau_1^{k_1} \tau_2^{k_2} \tau_3^{k_3} \propto \tau_1^{k_1} \tau_2^{k_2} \tau_3^{k_3}$$

Distribución a priori no informativa

Es la especificación en forma de distribución de probabilidad que busca saber que tan probable es que la variable sea afectada, a esta la llamamos distribución a priori. Hay diferentes criterios para identificar la distribución a priori, para este caso en específico distribución Dirichlet con parámetros: ϕ_1, ϕ_2, ϕ_3

$$(\tau_1, \tau_2, \tau_3 \mid \phi_1, \phi_2, \phi_3) \sim \text{Dirichlet}(\phi_1, \phi_2, \phi_3)$$

Con densidad de probabilidad:

$$\pi(\tau_1, \tau_2, \tau_3 \mid \phi_1, \phi_2, \phi_3) = \frac{\Gamma(\phi_1 + \phi_2 + \phi_3)}{\Gamma(\phi_1) \Gamma(\phi_2) \Gamma(\phi_3)} \tau_1^{\phi_1-1} \tau_2^{\phi_2-1} \tau_3^{\phi_3-1}$$

$$\phi_1, \phi_2, \phi_3 \geq 0; \quad \sum_{i=1}^3 \tau_i = 1$$

En este caso, se emplea la distribución previa de Bayes-Laplace, la cual se caracteriza por

simplificar la distribución de Dirichlet a una constante y puede ser interpretada como una distribución uniforme. Al tomar $\phi_i = (1, 1, 1)$, se obtiene una distribución equiprobable en todas las categorías. Gil (2019)

Distribución posterior no informativa

Utilizando el teorema de Bayes se calcula la distribución posterior:

$$\begin{aligned}
 f(\tau_1, \tau_2, \tau_3 \mid (k_1, k_2, k_3, \phi_1, \phi_2, \phi_3)) &= \frac{L(\tau/k)f(\tau/\phi)}{\int L(\tau/k)f(\tau/\phi)d\tau} \\
 &= \frac{\frac{n!}{k_1!k_2!k_3!} \prod_{i=1}^3 \tau_j^{k_j} \frac{1}{B(\tau)} \prod_{i=1}^3 \tau_j^{\phi_j-1}}{\int \frac{n!}{k_1!k_2!k_3!} \prod_{i=1}^3 \frac{1}{B(\tau)} \prod_{i=1}^3 \tau_j^{\phi_j-1} d\tau} \\
 &\propto \prod_{i=1}^3 \tau_i^{\phi_i+k_i-1}
 \end{aligned}$$

A partir de lo anterior se puede determinar que se distribuye Dirichlet con parámetros $\tau + K = (\tau_1 + k_1, \tau_2 + k_2, \tau_3 + k_3)$. Para este trabajo en específico los parámetros de la aposterior son: $1 + K = (1 + k_1, 1 + k_2, 1 + k_3)$.

Distribución a priori informativa

En este caso específico se toma la distribución a priori y por medio del proceso de elicitación que incluye el conocimiento de los expertos, en este caso los profesionales del proyecto de discapacidad. Se calculan los hiperparámetros usando el método de Elfadaly (2013) y desarrollado en R por Wilson (2017). A esto lo se le llama distribución informativa. La aplicación de este enfoque podría obtener resultados con una mejor precisión en comparación con los informados en este estudio de investigación.

$$\pi(\tau_1, \tau_2, \tau_3 \mid \phi_1, \phi_2, \phi_3) = \frac{\Gamma(\phi_1 + \phi_2 + \phi_3)}{\Gamma(\phi_1) + \Gamma(\phi_2) + \Gamma(\phi_3)} \tau_1^{\phi_1-1} \tau_2^{\phi_2-1} \tau_3^{\phi_3-1}$$

La distribución a priori informativa para el vector de parámetros $\tau = (\tau_1, \tau_2, \tau_3)$ es una distribución Dirichlet con parámetros $\phi = (6.49, 1.45, 5.01)$

Distribución posterior informativa

Anteriormente, se utilizó el teorema de Bayes para calcular la distribución a posterior en el contexto de este análisis. Se distribuye Dirichlet con parámetros. $\phi_i + K = (\phi_1 + k_1, \phi_2 + k_2, \phi_3 + k_3)$, Para este caso en específico los parámetros son: $\phi_i + K = (6,49 + k_1, 1,45 + k_2, 5,01 + k_3)$

4.4. Estimación usando OpenBUGS

Se utilizó el paquete R2OpenBUGS de R (*R Core Team*, 2020) para obtener las estimaciones de las distribuciones posteriores resultantes, ya que dicho paquete contiene las herramientas necesarias para usar el software estadístico OpenBUGS desde R (Sturtz, 2020).

4.5. Creación del índice heurístico

El índice propuesto se enfoca en la medición del desempeño del Proyecto de Discapacidad e Inclusión de la Universidad del Valle mediante los documentos de seguimiento de los estudiantes como fuente de información. El objetivo principal es evaluar la realización de acciones por parte del Proyecto, acotadas en las sub-categorías creadas por los profesionales; tales como los procesos administrativos, de caracterización, ajustes razonables y seguimiento, en adición a los procesos que reflejen la condición de salud y curso de vida del estudiante (Tabla 7-1). En este sentido, el buen desempeño del proyecto se define como la simultaneidad de todas las sub-categorías mencionadas, puesto que refleja que todos los estudiantes son expuestos a los diferentes procesos que se deben realizar de manera conjunta dentro del proyecto.

El índice se construye mediante la consideración y combinación de las diversas categorías y sub-categorías mencionadas anteriormente, las cuales representan los distintos aspectos fundamentales del proyecto. Diseñando una escala a partir de la exposición de los estudiantes a los diferentes procesos. En esta escala, se cuantifica la importancia de tener múltiples exposiciones según los profesionales, expresado como IM; asignando un valor de cuatro a quienes en cada grupo reflejen simultáneamente dos sub-categorías. Por otro lado, para los procesos que tienen menor relevancia de manera individual, se asigna un valor de dos o tres, según su orden jerárquico (Tablas 4-1, 4-2 y 4-3).

Estos niveles de desempeño se establecen mediante la ponderación de las probabilidades estimadas a cada exposición en función de su importancia relativa en el proyecto. La fórmula utilizada para calcular índice general de desempeño es la siguiente:

Índice general = $(4 * (\text{probabilidad Grupo 1. 1.a y 1.b}) + 3 * (\text{probabilidad Grupo 1. 1.b}) + 2 * (\text{probabilidad Grupo 1. 1.a})) + (4 * (\text{probabilidad Grupo 2. 1.c y 1.d}) + 3 * (\text{probabilidad Grupo 2. 1.c}) + 2 * (\text{probabilidad Grupo 2. 1.d})) + (4 * (\text{probabilidad Grupo 3. 3.a y 3.b}) + 3 * (\text{probabilidad Grupo 3. 3.a}) + 2 * (\text{probabilidad Grupo 3. 3.b}))$ (Tablas 4-1, 4-2 y 4-3)

Denotada técnicamente de la siguiente manera:

$$Indice = (4\hat{\theta}_3 + 3\hat{\theta}_2 + 2\hat{\theta}_1) + (4\hat{\delta}_3 + 3\hat{\delta}_1 + 2\hat{\delta}_2) + (4\hat{\tau}_3 + 3\hat{\tau}_1 + 2\hat{\tau}_2) \quad (4-1)$$

Donde $\hat{\theta}_i$, $\hat{\delta}_i$ y $\hat{\tau}_i$ son las probabilidades estimadas usando OpenBugs (Sturtz, 2020).

Al sumar los resultados ponderados, se obtiene un valor numérico que representa el nivel de desempeño general del proyecto. Esto permite una evaluación cuantitativa y comparativa de los distintos aspectos del proyecto en relación a los estándares establecidos, reflejando la importancia en la evaluación del desempeño del proyecto.

4.5.1. Niveles de desempeño

Los niveles de desempeño para el índice propuesto se encuentran comprendidos en una escala de [6, 12]. El mínimo muestra el peor de los casos posibles presentados, en el cual los estudiantes solo fueron expuestos a las acciones menos importantes de cada grupo. Esto implica que las probabilidades asignadas a las exposiciones de menor relevancia en cada grupo serán más altas, lo que resultará en un nivel de desempeño más bajo. En este caso, el nivel mínimo indicará que el proyecto ha tenido un desempeño deficiente, ya que no se han cumplido adecuadamente las acciones más relevantes y significativas en cada grupo. Por consiguiente el nivel mínimo se calcula a partir de la sumatoria de los IM correspondiente $(2+2+2=6)$. Por otra parte el nivel más alto informa un mejor desempeño, pues esto significa que los estudiantes fueron expuestos a todos los procesos que brinda el proyecto, lo que se calcula como la sumatoria de los IM correspondiente $(4+4+4=12)$. En este caso, las probabilidades estimadas a las exposiciones más relevantes en cada grupo serán más altas, resultando en un nivel de desempeño más elevado. Es decir, el proyecto ha logrado cumplir de manera satisfactoria con todas las acciones importantes en cada grupo, lo que demuestra un alto grado de cumplimiento de los objetivos del proyecto y un efecto positivo en los estudiantes.

4.5.2. Evaluación de robustez y convergencia

El diseño de un índice robusto es fundamental porque proporciona una medida confiable en múltiples escenarios posibles, garantizando estabilidad a los resultados frente al cambio y aumento del tamaño muestra. Las simulaciones permiten examinar diferentes combinaciones de resultados, en este caso, diferentes índices que convergen y facilitan el análisis del índice

general, obteniendo conclusiones válidas y confiables en el análisis sobre el desempeño del Proyecto de Discapacidad e Inclusión de la Universidad del Valle.

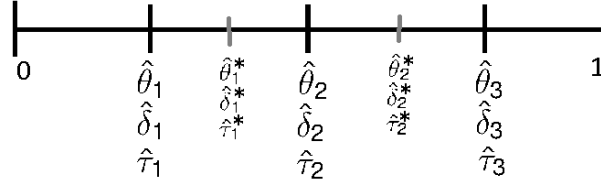


Figura 4-5: Distribución en el dominio 0-1 de estimaciones de la distribución posterior
Fuente: Elaboración propia

Donde $\hat{\theta}_1, \hat{\delta}_1, \hat{\tau}_1$ es la menor estimación de probabilidad para los eventos y $\hat{\theta}_3, \hat{\delta}_3, \hat{\tau}_3$ es la mayor. Se calcula el punto medio entre las estimaciones $\hat{\theta}_1$ y $\hat{\theta}_2$, $\hat{\theta}_2$ y $\hat{\theta}_3$ como se muestra en las ecuaciones 4-2 y 4-3 respectivamente para simular probabilidades de los distintos eventos.

$$\frac{\hat{\theta}_1 + \hat{\theta}_2}{2} = \hat{\theta}_1^* \quad (4-2)$$

$$\frac{\hat{\theta}_2 + \hat{\theta}_3}{2} = \hat{\theta}_2^* \quad (4-3)$$

Partiendo de lo anterior se simulan 1000 probabilidades por medio de la distribución uniforme para cada modelo.

Modelo 1	Modelo 2	Modelo 3
$\hat{\theta} \sim Uniforme(0, \hat{\theta}_1^*)$	$\hat{\delta} \sim Uniforme(0, \hat{\delta}_1^*)$	$\hat{\tau} \sim Uniforme(0, \hat{\tau}_1^*)$
$\widehat{p_{11k}} = (0 \geq \hat{\theta} \leq \hat{\theta}_1^*)$	$\widehat{p_{21k}} = (0 \geq \hat{\delta} \leq \hat{\delta}_1^*)$	$\widehat{p_{31k}} = (0 \geq \hat{\tau} \leq \hat{\tau}_1^*)$
$\hat{\theta} \sim Uniforme(\hat{\theta}_1^*, \hat{\theta}_2^*)$	$\hat{\delta} \sim Uniforme(\hat{\delta}_1^*, \hat{\delta}_2^*)$	$\hat{\tau} \sim Uniforme(\hat{\tau}_1^*, \hat{\tau}_2^*)$
$\widehat{p_{12k}} = (\hat{\theta}_1^* > \hat{\theta} \leq \hat{\theta}_2^*)$	$\widehat{p_{22k}} = (\hat{\delta}_1^* > \hat{\delta} \leq \hat{\delta}_2^*)$	$\widehat{p_{32k}} = (\hat{\tau}_1^* > \hat{\tau} \leq \hat{\tau}_2^*)$
$\widehat{p_{13k}} = 1 - \widehat{p_{12k}} - \widehat{p_{11k}}$	$\widehat{p_{23k}} = 1 - \widehat{p_{22k}} - \widehat{p_{21k}}$	$\widehat{p_{33k}} = 1 - \widehat{p_{32k}} - \widehat{p_{31k}}$

Tabla 4-4: Proceso de simulación para cada modelo

Posteriormente, para cada modelo i , se multiplicaron las probabilidades $\widehat{p_{ijk}}$ por su peso IM_j para generar 1000 índices. El proceso se repitió para cada modelo, generando 1000 índices por cada uno.

$$\widehat{indice}_{ik} = \sum_{j=1}^3 \widehat{p_{ijk}} IM_j \quad (4-4)$$

	Probabilidad 1	Probabilidad 2	Probabilidad 3	Índice
1	$\widehat{p_{i,1,1}}$	$\widehat{p_{i,2,1}}$	$\widehat{p_{i,3,1}}$	$\widehat{indice_{i,1}}$
2	$\widehat{p_{i,1,2}}$	$\widehat{p_{i,2,2}}$	$\widehat{p_{i,3,2}}$	$\widehat{indice_{i,2}}$
3	$\widehat{p_{i,1,3}}$	$\widehat{p_{i,2,3}}$	$\widehat{p_{i,3,3}}$	$\widehat{indice_{i,3}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1000	$\widehat{p_{i,1,1000}}$	$\widehat{p_{i,2,1000}}$	$\widehat{p_{i,3,1000}}$	$\widehat{indice_{i,1000}}$

Tabla 4-5: Número de simulaciones de probabilidades con su respectivo calculo del índice

Las probabilidades simuladas cumplen la condición $\widehat{p_{i1k}} + \widehat{p_{i2k}} + \widehat{p_{i3k}} = 1$. Con el modelo i para $i = 1, 2, 3$, y $k = 1 \dots 1000$.

La estimación media del índice es expresada por la siguiente expresión:

$$\widehat{indice}_{general} = \sum_{i=1}^3 \left(\frac{\sum_{k=1}^{1000} \widehat{indice_{ik}}}{1000} \right) \quad (4-5)$$

5 Resultados

5.1. Delimitación de la muestra

Como se mencionó anteriormente, en la fase inicial del trabajo se dispuso una base de datos general con información de los estudiantes vinculados al Proyecto de Discapacidad e Inclusión, donde los estudiantes con un tipo de discapacidad psicosocial destacaron ($n=38$, 30.89 %), seguido del tipo visual ($n=27$, 21.95 %) y física ($n=20$, 16.26 %). Por otra parte la discapacidad auditiva ($n=2$, 1.62 %) y talento excepcional ($n=1$, 0.81 %) no fueron tan comunes. Además, 91 (73.98 %) estudiantes estaban activos en la política del proyecto entre el periodo 2018-1 y 2022-1.

La mayoría de los estudiantes activos en el proyecto tenían un nivel educativo de pregrado (91.20 %). Esta fue una de las principales razones para establecer como criterio de inclusión los estudiantes de pregrado, además, obtener los documentos de ajuste razonable de los estudiantes no activos sería imposible.

En las diferentes sedes donde el Proyecto de Discapacidad e Inclusión opera, se destaca que la mayor cantidad de estudiantes se encuentran en la sede Meléndez ($n=53$) y San Fernando ($n=15$), mientras que en cada sede regional presenta una frecuencia menor a 10 estudiantes por sede (Zarzal, Buga, Pacífico, Caicedonia, Norte del Cauca, Cartago, Palmira y Tuluá). Dichos estudiantes en las sedes regionales podrían presentar una dificultad a la hora de obtener los documentos de ajustes razonables, por tanto surge la necesidad de acotar los posibles estudiantes que formarían parte de la base de datos a construir.

5.2. Clasificación de los Unigramas

La frecuencia de unigramas entre documentos sirve para determinar que unigramas se han encontrado en todos los documentos, siendo 41 el máximo que puede alcanzar este valor dado que se tienen 41 documentos; mientras que la frecuencia intra documentos se utilizará para determinar si dicho unigrama es relevante según los profesionales. La tabla **5-1** presentada a continuación refleja algunas frecuencias entre documentos e intra documentos como ejemplo.

Unigrama	Frecuencia entre documentos	Frecuencia intra documentos	Categoría	Sub-Cat
semestre	37	153	P. Académico	2.a
reunion	37	167	P. Discapacidad	1.d
proceso	37	646	P. Discapacidad	1.a
acompañamien- to	37	308	P. Discapacidad	1.c
realizar	36	403	Sujeto	3.b
apoyo	36	335	P. Académico	2.b
academico	36	238	P. Académico	2.a
proyecto	35	240	P. Discapacidad	1.a
profesional	35	304	P. Discapacidad	1.a
poder	35	329	Sujeto	3.b
⋮	⋮	⋮	⋮	⋮

Tabla 5-1: Frecuencia de los unigramas entre documentos, intra documentos y sus respectivas categorías asignadas por los profesionales del Proyecto

Finalmente, el numero total de unigramas categorizados fue 152, resultando dicha frecuencia repartida entre cada categoría y sub-categoría de la siguiente manera:

1. **Proyecto de discapacidad e inclusión:** 60 unigramas

- 1.a) Procesos administrativos: 25 unigramas
- 1.b) Procesos de caracterización: 12 unigramas
- 1.c) Ajustes razonables: 10 unigramas
- 1.d) Procesos de seguimiento: 13 unigramas

2. **Programa académico:** 46 unigramas

- 2.a) Actividad de aprendizaje: 27 unigramas
- 2.b) Contexto de aprendizaje: 19 unigramas

3. **Sujeto:** 46 unigramas

- 3.a) Condición de salud: 15 unigramas
- 3.b) Curso de vida: 31 unigramas

La construcción de la base de datos se encuentra ejemplificada en las figuras **4-3** y **4-4**. Donde establece que una categoría se ve reflejada en determinado estudiante después que este presente al menos un unigrama de dicha categoría. Como resultado, todos los estudiantes presentaron todas las categorías, puesto que en el documento de seguimiento estos son los

tres temas fundamentales que se abordan. Sin embargo, desglosando las categorías en sub-categorías y después estas en diferentes grupos de eventos mencionados anteriormente y presentados en la tabla 7-1.

La categoría *Programa Académico* no se utilizó en ningún tipo de calculo puesto que todos los estudiantes presentaron solo exposiciones de las sub-categorías “Actividad de aprendizaje” y “Contexto de aprendizaje” simultáneamente. Es decir, el proyecto de discapacidad e inclusión hace énfasis en las actividades o hechos que el estudiante realiza como parte de su proceso académico. En otras palabras, los estudiantes que se encuentran en el proyecto realizan sus deberes académicos aún con las barreras que se puedan presentar, además, la gestión con el programa académico de los estudiantes es fundamental ya que tiene como objetivo mejorar la experiencia académica del estudiante.

5.3. Modelos estadísticos

Se quiere estudiar y/o caracterizar los grupos de eventos donde los estudiantes son representados según su documento de seguimiento, puesto que estos están relacionados con el funcionamiento del Proyecto de Discapacidad e Inclusión de la Universidad.

Modelo	Distribución a priori no informativa	Distribución Posterior no informativa	DIC
1	$(\theta_1, \theta_2, \theta_3 \mid n, x_1, x_2, x_3) \sim \text{Dirichlet}(1, 1, 1)$	$1+X \sim \text{Dirichlet}(40, 2, 2)$	7.924
2	$(\delta_1, \delta_2, \delta_3 \mid n, y_1, y_2, y_3) \sim \text{Dirichlet}(1, 1, 1)$	$1+Y \sim \text{Dirichlet}(39, 2, 3)$	7.924
3	$(\tau_1, \tau_2, \tau_3 \mid n, k_1, k_2, k_3) \sim \text{Dirichlet}(1, 1, 1)$	$1+K \sim \text{Dirichlet}(40, 2, 2)$	7.924

Tabla 5-2: Parámetros de las distribuciones a priori no informativas y distribuciones posteriores

Modelo	Distribución a priori informativa	Distribución Posterior informativa	DIC
1	$(\theta_1, \theta_2, \theta_3 \mid n, x_1, x_2, x_3) \sim \text{Dirichlet}(9.52, 7.46, 9.23)$	$\alpha + X \sim \text{Dirichlet}(48.52, 8.46, 10.23)$	10.1
2	$(\delta_1, \delta_2, \delta_3 \mid n, y_1, y_2, y_3) \sim \text{Dirichlet}(8.11, 4.97, 9.99)$	$\Upsilon + Y \sim \text{Dirichlet}(46.11, 5.97, 11.99)$	13.8
3	$(\tau_1, \tau_2, \tau_3 \mid n, k_1, k_2, k_3) \sim \text{Dirichlet}(6.49, 1.45, 5.01)$	$\phi + K \sim \text{Dirichlet}(45.49, 2.45, 6.01)$	12.12

Tabla 5-3: Parámetros de las distribuciones a priori informativas y distribuciones posteriores

En las tablas 5-2 y 5-3 se observan los parámetros de las respectivas distribuciones a priori no informativas, informativas y sus formas posteriores, además de presentar el DIC para cada caso en específico. Donde se destaca que el DIC de los modelos para las distribuciones posteriores no informativas son menores en comparación a las distribuciones posteriores informativas. Esto quiere decir que los modelos a base de la información actual tienen una mejor estimación que el modelo con información elicitada.

En relación a las estimaciones de las probabilidades calculadas usando OpenBugs para los modelos con distribución a priori no informativa (Tabla 5-4) se puede observar en la figura 5-1 la distribución muestral del índice general debido a las 1000 simulaciones. Además, los estimadores representativos:

- θ_1 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados con procesos administrativos y de caracterización es 88.63 %.
- θ_2 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados únicamente con procesos administrativos es 4.55 %.
- θ_3 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados únicamente con procesos de caracterización es 6.81 %.
- δ_1 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados con procesos ajustes razonables y procesos de seguimiento es 88.63 %.
- δ_2 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados únicamente con ajustes razonables es 4.55 %.
- δ_3 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados únicamente con procesos de seguimiento es 6.81 %.
- τ_1 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados con Condición de salud y curso de vida es 88.63 %.
- τ_2 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados únicamente con condición de salud es 6.81 %.
- τ_3 : En promedio, la probabilidad de que los estudiantes presenten eventos relacionados únicamente curso de vida es 4.54 %.

5.4. Índice General

Las estimaciones de las probabilidades $\theta_1, \theta_2, \theta_3, \delta_1, \delta_2, \delta_3, \tau_1, \tau_2, \tau_3$ presentadas anteriormente (Tabla 5-4) se usan para calcular el índice propuesto como una medida del nivel de desempeño del Proyecto de Discapacidad e Inclusión. La estimación media del índice general usando las estimaciones mediante las distribuciones posteriores no informativas es de 11.04601. Esto indica un nivel de desempeño cercano a 12 siendo su máximo valor dentro de los niveles de desempeño, sugiriendo que el proyecto de discapacidad e inclusión ha mostrado un rendimiento muy favorable. En otras palabras la mayoría de estudiantes han experimentado la presencia simultánea de sub-categorías en cada uno de los procesos en relación con

Parámetros	D. Posterior no informativa	D. Posterior informativa
θ_1	0.8863160	0.7067268
θ_2	0.04558106	0.1258034
θ_3	0.06810294	0.1674699
δ_1	0.8863160	0.7195527
δ_2	0.04558106	0.09295622
δ_3	0.06810294	0.1874911
τ_1	0.8863160	0.8244520
τ_2	0.04558106	0.04556460
τ_3	0.06810294	0.1299835

Tabla 5-4: Estimaciones medias dadas por las distribuciones posteriores no informativas e informativas usando OpenBugs.

los procesos administrativos, caracterización, ajustes razonables, seguimiento, condición de salud y curso de vida; lo cual es una señal positiva de un enfoque integral y exhaustivo en la atención a los estudiantes con discapacidad.

Las numerosas actividades y gestiones sugieren que el proyecto de discapacidad e inclusión ha logrado implementar y llevar a cabo una amplia gama de acciones dirigidas a la gestión, mejoramiento y seguimiento de sus estudiantes con discapacidad, con el objetivo de mejorar la calidad educativa y brindar apoyo integral a la comunidad académica.

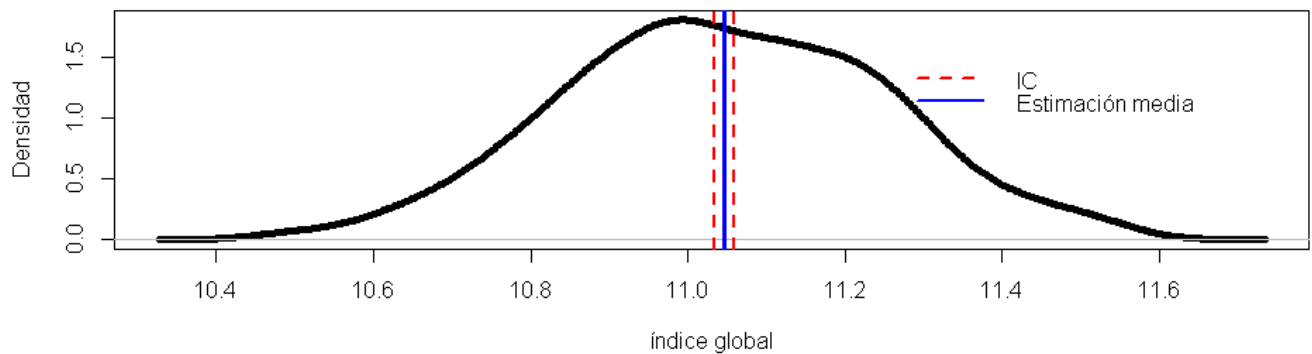


Figura 5-1: Distribución muestral del índice general y la estimación media

$$IC = (11,03317, 11,05885), \hat{Índice}_{general} = 11,04601$$

6 Conclusiones, recomendaciones y limitaciones

Con base en el desarrollo del trabajo de grado sobre el estudio del desempeño del proyecto de discapacidad e inclusión de la Universidad del Valle utilizando probabilidades, se pueden establecer las siguientes conclusiones:

6.1. Conclusiones:

En el presente trabajo de grado se presentaron diferentes obstáculos frente a la minería de textos debido a la dificultad en la complejidad de los documentos y la selección adecuada de librerías, diccionarios y otras herramientas del procesamiento del lenguaje natural. Estos inconvenientes fueron resueltos con una mayor capacitación, tomando referencias en las diferentes librerías y procedimientos utilizados en el “Taller de Minería de Textos” (Cutipa, 2023). Además se usó algunos procedimientos del curso “Introducción a la minería de textos” de la maestría en estadística aplicada USTA (2020).

Entre otras dificultades, la revisión, organización y determinación de los documentos a utilizar en la muestra formó parte de la fase del trabajo de grado prolongada, pues se necesitaba cumplir con requisitos y permisos para acceder a la información. Además, al no ser contenida en un sistema general de información y tampoco presentar un formato determinado o archivo plano, resultó tediosa y compleja la revisión de la información que tiene el Proyecto de Discapacidad e Inclusión. Por otra parte, la sincronización de tiempos disponibles entre los profesionales del proyecto y los autores del presente trabajo de grado dificultaba la creación de espacios para la realización del mismo.

El índice propuesto para el trabajo de grado se desarrolló utilizando una metodología estadística en conjunto con análisis multicriterio, planteamiento que mejora la eficacia para la construcción de un índice heurístico basado en probabilidades estimadas para las exposiciones de los diferentes eventos denotados en las tablas **4-1**, **4-2** y **4-3** para las diferentes sub-categorías y ofrecer una visión más amplia de los resultados. .

El Proyecto de Discapacidad e Inclusión de la Universidad del Valle enfrenta obstáculos al tratar de aumentar las oportunidades para los estudiantes con discapacidad. Finalmente, es

importante resaltar la afirmación de Cobos (2014) mencionada al inicio del presente trabajo, puesto que en Colombia a pesar de los esfuerzos del gobierno y de la Universidad en la búsqueda de una educación inclusiva, aún existen desafíos para garantizar un acceso a la educación superior para los estudiantes en condición discapacidad.

El presente trabajo de grado contribuye abordando esta problemática al proponer a través de una metodología estadística mediante el uso de probabilidades y el procesamiento de lenguaje natural, un índice que se enfoca en la cantidad de gestiones realizadas por el proyecto a favor de los estudiantes; proporcionando una medida numérica que refleja el grado de cumplimiento de los objetivos y la efectividad en la implementación de acciones del Proyecto de Discapacidad e Inclusión, concluyendo que el proyecto de discapacidad presenta un buen nivel de desempeño.

6.2. Recomendaciones:

Se sugiere establecer un sistema de recolección de información más estandarizado y detallado, con el fin de contar con datos más completos y precisos para futuros análisis. Esto facilitaría una evaluación más exhaustiva del desempeño del proyecto y la identificación de aspectos específicos que requieren atención. Además se recomienda realizar un seguimiento continuo y periódico del desempeño del Proyecto, utilizando el índice propuesto como una herramienta de monitoreo. Esto permitirá evaluar el desempeño de las acciones implementadas, identificar áreas de mejora y realizar ajustes en las políticas y prácticas en curso. Finalmente, se recomienda aumentar la cantidad de unigramas dentro de las categorías y sub-categorías, construir un criterio de representatividad para que a partir de las frecuencias de los unigramas se pueda establecer el grado de representatividad dentro de cada sub-categoría.

6.3. Limitaciones:

Una limitación del estudio es el enfoque exclusivo en los estudiantes de pregrado de la Universidad sede Meléndez. Para obtener una imagen más completa del desempeño del proyecto, sería necesario incluir a otros niveles educativos y sedes universitarias. Por otra parte la validez y confiabilidad del índice propuesto dependerá en gran medida de la calidad y consistencia de los datos recopilados en los informes de seguimiento. Limitaciones en la calidad de la información pueden afectar la precisión de las conclusiones y recomendaciones derivadas del análisis.

7 Anexos

Código para la creación de bigramas que se agrega antes de la lematización

```
BigramTokenizer <- function(x) {  
  NGramTokenizer(x, Weka_control(min = 2, max = 2))  
}  
  
df_tdm = TermDocumentMatrix(corp,  
control = list(tokenize = BigramTokenizer))
```

Código para el conteo de categorías y sub-categorías para cada estudiante

```
carac <- read_excel("C:/Users/RUTA/EJEMPLO")  
  
#seleccionar la direccion y el tipo de documentos  
my_dir <- "C:/Users/RUTA/EJEMPLO"  
files <- list.files(path = my_dir, pattern = "xlsx$")  
setwd(my_dir)  
  
x_file <- file.path(my_dir, files)  
  
mtx2<-matrix(0,nrow=41,ncol=8)  
  
for (i in 1:41){  
  
bse <- read_excel(x_file[i], sheet = "Sheet_1")  
  
for (j in 1:ncol(carac)){  
  var<-0  
  for (k in 1:nrow(carac)){  
  
    ifelse(grepl(carac[k,j], bse[,1])==TRUE,  
           var<-var+1, var<-var)
```

```

    }
    mtx2[i, j] <- var
  }
}

mydata <- write.xlsx(data.frame(mtx2), ".xlsx")

saveWorkbook(mydata,
              file = "C:/Users/RUTA/EJEMPLO",
              overwrite = TRUE)

```

Código para las estimaciones bayesianas

```

sink("modelo.txt")
cat("
model{
y[1]<-DATO_GRUPO_AMBAS
y[2]<-DATO_GRUPO_1
y[3]<-DATO_GRUPO_2
y[1:3]~dmulti(theta[1:3],n)
theta[1:3]~ddirch(alpha[1:3])
alpha[1]<-a
alpha[2]<-b
alpha[3]<-c
}", fill=TRUE)
sink()

datos=list(a=PARAMETRO,b=PARAMETRO,c=PARAMETRO,n=41)

parametros=list("theta")

mod1=bugs(data=datos, inits=NULL,
           parameters.to.save = parametros,
           model.file = "modelo.txt",
           n.iter=50000,n.burnin = 5000,n.thin = 1,
           n.chains = 1,debug = FALSE)

```

```
par(mfrow=c(3,6))  
#Densidad de theta1  
plot(density(mod1$sims.list$theta[,1]),  
main=expression(theta[1]),ylab="Densidad")  
abline(v=(39/41),col="red",lty=2)  
abline(v=mod1$mean$theta[1],col="green",lty=2)
```

```
#Densidad de theta2  
plot(density(mod1$sims.list$theta[,2]),  
main=expression(theta[2]),ylab="Densidad")  
abline(v=(1/41),col="red",lty=2)  
abline(v=mod1$mean$theta[2],col="green",lty=2)
```

```
#Densidad de theta3  
plot(density(mod1$sims.list$theta[,3]),  
main=expression(theta[3]),ylab="Densidad")  
abline(v=(1/41),col="red",lty=2)  
abline(v=mod1$mean$theta[3],col="green",lty=2)
```

Categoría	Grupo	Subcategorías
Proyecto de discapacidad	1	Proc. administrativo y de Caracterización
	1	Proc. administrativo
	1	Proc. de caracterización
	2	Ajuste razonable y procesos de seguimiento
	2	Ajustes razonables
	2	Procesos de seguimiento
Programa académico		Actividad y contexto de aprendizaje
		Actividad de aprendizaje
		Contexto de aprendizaje
Sujeto	3	Condición de Salud y curso de vida
	3	Condición de salud
	3	Curso de vida

Tabla 7-1: Categorías, grupos y sub-categorías creadas por los profesionales del Proyecto de Discapacidad

Bibliografía

Armstrong, F. (1999), ‘Inclusion, curriculum and the struggle for space in school’, *International Journal of Inclusive Education* .

Benoit, K. (2021), ‘Multilingual stopword lists’, *R package version 2.3* .

URL: <https://cran.r-project.org/web/packages/stopwords/index.html>

Carvajal, M. (2015), ‘Política de discapacidad e inclusión de la universidad del valle: un proceso participativo’, *Revista Sociedad y Economía* .

Cobos, M. M. . R. (2014), ‘Educación superior y discapacidad: análisis desde la experiencia de algunas universidades colombianas’, *Revista Española de Discapacidad* .

Comisión de Derechos Humanos (1948), *Declaración Universal de los Derechos Humanos* .

Consejo Superior de la Universidad del Valle (2002), *Resolución No. 040: Por la cual se reglamentan las Monitorías en la Universidad del Valle* .

Consejo Superior de la Universidad del Valle (2013), *Acuerdo No. 004: Por la cual se adopta la Política Institucional de Discapacidad e Inclusión de la Universidad del Valle* .

Cutipa, J. (2023), ‘Taller: Minería de textos’.

URL: <https://www.youtube.com/watch?v=kfsdXdlp2V8>

Díaz, A. (2022), ‘Análisis de temas utilizando twiter: Una aplicación del modelo lda al caso colombiano’, *Universidad Santo Tomás* .

Elfadaly, P. G. . F. (2013), ‘Eliciting dirichlet and connor–mosimann prior distributions for multinomial models’.

Feinerer, I. (2023), ‘Text mining package’, *R package version 0.7-11* .

URL: <https://cran.r-project.org/web/packages/tm/index.html>

Frigyik, B., K. A. and Gupta, M. R. (2010), ‘Introduction to the dirichlet distribution and related processes. department of electrical engineering’, *University of Washignton* .

Gil, L. (2019), ‘Genes relacionados con el rendimiento deportivo en atletas elite y deportistas recreativos: Una aproximación usando métodos de bayes’, *Universidad del Valle* .

- Giraldo, M. (2009), ‘La inclusión académica de los estudiantes en situación de discapacidad en la universidad del valle 2009’, *Universidad del Valle* .
- Hernández, J. (2019), ‘Traducción de texto en español a texto lsm usando aprendizaje profundo’.
URL: https://rinacional.tecnm.mx/bitstream/TecNM/1538/2/2019-149_JuanCarlos_Hernandez_Cruz.pdf4
- Hornik, K. (2020), ‘Natural language processing infrastructure’, *R package version 0.2-1* .
URL: <https://cran.r-project.org/web/packages/NLP/index.htm>
- Ignatow, R. M. . G. (2016), ‘Text mining: A guidebook for the social sciences’, *Sage Publications* .
- Johnson, N. (1997), ‘Discrete multivariate distributions’, *Wiley New York* .
- Jurafsky, J. M. . D. (2009), ‘Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition’.
URL: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
- Latorre, A. (2005), ‘Bases metodológicas de la investigación educativa’, *Barcelona: experiencia* .
- Lin, J. (2016), ‘On the dirichlet distribution’, *Queen’s University* .
- Ooms, J. (2023), ‘Text extraction, rendering and converting of pdf documents’, *R package version 3.3.3* .
URL: <https://cran.r-project.org/web/packages/pdftools/index.html>
- Organización de las naciones unidas (2006), *Convención sobre los Derechos de las Personas con Discapacidad* .
URL: <https://www.un.org/esa/socdev/enable/documents/tccconv.pdf>
- Organización Mundial de la Salud (2014), *Documentos Básicos* .
URL: <https://apps.who.int/gb/bd/PDF/bd48/basic-documents-48th-edition-sp.pdf?ua=1#page=7>
- Paez, E. (2020), ‘Implementación de un modelo de análisis de sentimientos con respecto a la jep basado en minería de datos en twitter’, *Universidad Católica de Colombia* .
- Perez, B. (2021), ‘Caracterización de los perfiles de apoyo de estudiantes con discapacidad vinculados al programa de rediversia de la universidad del valle’.
- Posada, H. (2015), ‘Evaluación multicriterio y sig. como herramientas para la gestión territorial caso de estudio ubicación del terminal de transporte en zipaquirá cundinamarca’.

Presidencia de la república de Colombia (2009), Decreto 366: Por medio del cual se reglamenta la organización del servicio de apoyo pedagógico para la atención de los estudiantes con discapacidad y con capacidades o talentos excepcionales en el marco de la educación inclusiva .

Presidencia de la república de Colombia (2013), Ley Estatutaria 1618: Por medio de la cual se establecen las disposiciones para garantizar el pleno ejercicio de los derechos de las personas con discapacidad .

Presidencia de la república de Colombia (2017), Decreto 1421: Por el cual se reglamenta en el marco de la educación inclusiva la atención educativa a la población con discapacidad .

R Core Team (2020), A Language and Environment for Statistical Computing. R Foundation for Statistical Computing .

Silge, D. R. . J. (2017), ‘Text mining with r: A tidy approach’, *O’Reilly Media* .

Silva, L. (1997), ‘Cultura estadística e investigación científica en el campo de la salud: Una mirada crítica’, *Editorial Diaz de Santos, Madrid* .

Sobrino, J. (2018), ‘Análisis de sentimientos en twitter’, *Universitat Oberta de Catalunya* .

Sosa, J. (2020), ‘Curso: Introducción a la minería de textos’.

URL: <https://github.com/jstats1702/Text-Mining-in-R-2020-1>

Sturtz, U. L. S. (2020), ‘R2openbugs: a package for running openbugs from r.’, *R package version 3.2-3.2.1* .

URL: <https://cran.r-project.org/web/packages/R2OpenBUGS/index.html>

Universidad Nacional de Colombia (2012), Acuerdo 036: Por el cual se establece la política institucional para la inclusión educativa de las personas con discapacidad en la Universidad Nacional de Colombia .

Vernazza, J. V. . E. (2021), ‘Discursos presidenciales en uruguay enfoque desde el análisis estadístico de texto’, *Cuadernos Del CIMBAGE* .

URL: <https://ojs.econ.uba.ar/index.php/CIMBAGE/article/view/2054>

Wijffels, J. (2023), ‘Tokenization, parts of speech tagging, lemmatization and dependency parsing with the ’udpipe’ ’nlp’ toolkit’, *R package version 0.8.11* .

URL: <https://cran.r-project.org/web/packages/udpipe/index.html>

Wilson, E. (2017), ‘Fitting a modified connor-mosimann distribution to elicited quantiles of multinomial probabilities’.