

Consejo Asesor Técnico del Sistema Estadístico Nacional (CASEN) 2023-2025
Sala Especializada para la Modernización Tecnológica de la Producción Estadística
Acta No. 4

Ciudad: Bogotá D.C.

Lugar: sala DANE.

Tema: Recibir realimentación y recomendaciones por parte de los miembros de la sala.

Hora: 2:00 p. m. a 4:00 p. m.

Fecha: 10/05/2024.

Dependencia responsable: Secretaría Técnica del CASEN.

Participantes

Miembros de la Sala Especializada para la Modernización Tecnológica de la Producción Estadística del CASEN

Mario Linares Vásquez.

Nicolás Cardozo Álvarez.

León Darío Parra.

Departamento Administrativo Nacional de Estadística (DANE)

Julieth Alejandra Solano Villa, directora técnica de la Dirección de Regulación, Planeación, Estandarización y Normalización (DIRPEN) y secretaria técnica CASEN.

Luis Martin Barrera Pino, jefe Oficina de Sistemas (OSIS).

Natalia Ximena Arteaga Gutiérrez, asesora subdirección.

Andrea Roncancio, coordinadora Grupo Interno de Trabajo (GIT) Prospectiva y Analítica de Datos (PAD), DIRPEN.

Andrea Catherine Neira Bustamante, designada OSIS.

Alejandro Sandoval Pineda, designado OSIS.

Daniel Mauricio Montenegro Reyes, designado OSIS.

Luis Fernando Barajas Duarte, designado OSIS.

Victoria Adriana González Ramírez, designada DSCN.

Juan José Rubio Mesa, designado DIMPE.

Secretaría Técnica del CASEN - DIRPEN

Julieth Alejandra Solano Villa, directora técnica DIRPEN y secretaria técnica CASEN.

Elizabeth Moreno Barbosa, asesora.

Yennifer Castillo Murcia, Grupo Interno (GIT) Planificación y Articulación Estadística (PAE).

Sandra Yaneth Cortés Gamba, responsable general del CASEN, GIT PAE.

Derly Vivian Lizarazo García, responsable de la Sala Especializada para la Modernización Tecnológica de la Producción Estadística.

Agenda

Tiempo	Actividad	Responsable
2:00 p. m. a 2:05 p. m.	Instalación, verificación del quorum y registro fotográfico.	Derly Lizarazo, responsable de la sala.
2:05 p. m. a 2:10 p. m.	Apertura del evento.	Julieth Solano, directora técnica DIRPEN y secretaria técnica del CASEN.
2:10 p. m. a 2:15 p. m.	Síntesis de la reunión anterior y compromisos.	Derly Lizarazo, responsable de la sala.
2:15 p. m. a 3:00 p. m.	Presentación de la línea 1.2. <i>Proyectos de ciencia de datos, Índice de noticias</i> y espacio de realimentación por parte de los miembros de la sala.	Equipo índice de noticias – GIT Prospectiva y Analítica de Datos (PAD)
3:00 p. m a 3:20 p. m.	Presentación del artículo <i>Colombian agricultural sector's early estimator of Gross Domestic Production using Nowcasting and Big Data methods</i> .	León Darío Parra, experto de la sala.
3:20 p. m. a 3:40 p. m.	Concertar el detalle del evento de transferencia de conocimiento denominado <i>Taller de cultura de la arquitectura</i> , así como definir los demás productos de la sala.	Derly Lizarazo, responsable de la sala.
3:40 p. m a 3:45 p. m.	Compromisos de la reunión.	Derly Lizarazo, responsable de la sala.
3:45 p. m a 4:00 p. m.	Conclusiones y cierre.	Julieth Solano, directora técnica DIRPEN y secretaria técnica del CASEN.

Objetivo

Recibir realimentación por parte de los miembros de la sala sobre la línea de investigación No. 1.2. *Proyectos de ciencia de datos*, referente con el *índice de noticias* y presentación del artículo *Colombian agricultural sector's early estimator of Gross Domestic Production using Nowcasting and Big Data methods*, así como concertar la propuesta del evento de transferencia de conocimiento y de los demás productos de la sala.

Desarrollo

1. Instalación, verificación del quórum y apertura de la reunión

Se realizó la verificación del quórum, contando con la participación de los tres miembros de la sala, los designados de las Oficina de Sistemas, del GIT Prospectiva y Analítica de Datos de la DIRPEN, de la Dirección de Síntesis y Cuentas Nacionales (DSCN) y de la Dirección de Metodología y Producción Estadística (DIMPE), así como de la secretaría técnica del CASEN y su equipo de trabajo. Posterior a ello, se realizó el registro fotográfico y se le dio la palabra a Julieth Solano, directora técnica de la Dirección de Regulación, Planeación, Estandarización y Normalización (DIRPEN) y secretaría técnica del CASEN, quien realizó la apertura a la reunión.

2. Síntesis de la reunión anterior y compromisos

Derly Lizarazo, responsable de la sala, realizó una breve síntesis de la reunión anterior y señaló la realimentación y las recomendaciones por parte de los miembros de la sala, sobre el documento: resumen de los aspectos teóricos y conceptuales de la línea ciclo de vida; presentación de la *línea 1. Ciencia de datos*, 1.1. *Ciclo de vida de modelos de proyectos de ciencia de datos*, 1.1.2. *Exploración de herramientas y plataformas para cada etapa del ciclo de vida de los modelos de ciencia de datos*, y de la *línea 1.2. Proyectos de ciencia de datos*, App diversa y espacio de realimentación y recomendaciones por parte de los miembros de la sala.

También hizo referencia a los compromisos adquiridos: envío del acta de la reunión del 26 de abril (9 de mayo), remisión del material de trabajo y asunto de consulta para los siguientes temas: Reunión 10 de mayo: línea de investigación No.1. *Ciencia de datos*, 1.2. *Proyectos de ciencia de datos* / 1.2.1. *Índice de noticias* (remitido el 3 de mayo) y avances en la propuesta del taller: *Cultura de la arquitectura de datos*.

3. Presentación de la línea 1.2. Proyectos de ciencia de datos, Índice de noticias y espacio de realimentación por parte de los miembros de la sala

La presentación la realizó Sebastián Salazar, del GIT Prospectiva y Analítica de Datos de la DIRPEN, donde resaltó el uso de fuentes alternativas de información a través del índice de noticias de las actividades

económica y en el asunto de consulta a los miembros del CASEN realizaron las siguientes preguntas: ¿Cómo se brinda calidad a este tipo de modelos que utilizan fuentes alternativas de información? ¿Cómo se evalúan este tipo de modelos que utilizan fuentes alternativas de información? ¿Cómo definir la métrica de desempeño del modelo del índice de noticias? ¿Qué variables (y qué ponderación de ellas) se deben tener en cuenta? y ¿Cómo incluir etapas de validación en cada una de las fases del proceso?

Estas fueron las recomendaciones generales de los tres miembros de la sala:

- En la mejora de los modelos se pueden usar técnicas no supervisadas par la clasificación de las noticias en lenguaje natural. Otra opción es usar un modelo semántico donde periódicamente se pueden ir refinando los términos de la búsqueda, esta opción busca subir la calidad.
- No usar solo el concepto, el título, sino también el texto y buscar similitud entre las noticias y hacer un *cluster* donde el sentimiento general (positivo negativo) para no quedarse con una fuente en particular.
- Antes de correr un modelo es aconsejable hacer un análisis cualitativo con una muestra de manera reiterada y después arreglar la parte semántica. Además, es fundamental revisar la calidad del dato antes de aplicar cualquier modelo.
- Es importante incluir entre el procesamiento y el análisis un paso donde se revise la noticia o el tópico con un filtrado automático de noticias y se analice dentro del conjunto de noticias a fondo en cada uno de esos grupos.
- Es importante identificar la fuente más fidedigna (noticias, twitter, blogs) a utilizar dependiendo del tipo de análisis que se va a realizar y la temporalidad.
- Se pueden tomar los titulares, agrupar los que mencionan el mismo tema y por volumen se puede o no controlar un sesgo en particular, dependiendo los porcentajes; esto permite mirar lo positivo y lo negativo y tomar la decisión dependiendo de eso.

Frente a lo expuesto por los diferentes participantes de la sala, Julieth Solano, desde la Secretaría Técnica, realizó los siguientes comentarios:

- Se llamo la atención para tener en cuenta todas las líneas exploratorias y el análisis cualitativo y respetando la línea del proyecto y el tiempo, en el marco de la conversación se mencionaron elementos para fortalecer la metodología y generar otras líneas de trabajo.
- Las ideas claves: análisis de *cluster*, poder usar el texto y conseguir más información de contexto que permita aportar a la positividad.

- La pestaña de Google News, puede permitir un directorio, el cual ya se tiene y esto garantiza que son de una fuente idónea.
- En este ejercicio se está simulando lo que hace un analista del DANE, cuando a partir de todos los datos que recibe para el sector agrícola le está dando x resultado al valor agregado del PIB y hay que explicarle al país porque está subiendo o porque está bajando y este proceso se hace revisando las noticias para explicar las dinámicas que se están presentando.

4. **Presentación del artículo *Colombian agricultural sector's early estimator of Gross Domestic Production using Nowcasting and Big Data methods***

El profesor León Darío Parra, contó cómo se realizó la construcción del artículo, cuyo porte es hacia la toma de decisiones y a convertirlo como un instrumento que puede tener como diferentes matices. Igualmente, esto permitió que el artículo fuera aceptado en el Journal of Technology Management and Innovation tipo tres indexado, donde será publicado para el segundo semestre de 2024 o en 2025.

5. **Concertar el detalle del evento de transferencia de conocimiento denominado *Taller de cultura de la arquitectura*, así como definir los demás productos de la sala**

Se presentó la propuesta para el evento que incluía el título, el objetivo y los temas a tratar. Se estableció como fecha del evento el 20 de junio en el auditorio del DANE y se indicó que se conseguirán los expositores de la Universidad de Los Andes, la Universidad Nacional de Colombia y la Universidad EAN. Además, se concertó que estará dirigido a funcionarios del DANE, del DNP y de entidades del Plan Nacional de Infraestructura de Datos (PNID).

6. **Compromisos de la reunión**

Se indagó sobre el tema para la elaboración de la nota estadística y se postularon los siguientes: interoperabilidad importancia para el SEN, Inteligencia Artificial para el SEN y procesamiento en lenguaje natural. La OSIS propuso como tema el X-road. El tema se definirá en una próxima reunión.

Compromisos

Tarea	Envío de acta de la presente reunión para revisión y aprobación de los miembros y participantes de la sala.
Responsable	Responsable de la sala: Derly Lizarazo.
Fecha entrega	27/05/2024.

Tarea	Remitir material de trabajo y asunto de consulta para los siguientes temas: 1. Reunión 31 de mayo: línea de investigación <i>1. Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencias de datos, No. 1.1. Revisión de prácticas de Ingeniería de Software para Machine Learning</i> . Se remitirá lo correspondiente el 22 de mayo.
Responsable	Responsable de la sala: Derly Lizarazo.
Fecha entrega	22/05/2024
Tarea	Reunión para ajustar la propuesta del evento de cultura de la arquitectura de datos.
Responsable	Derly Lizarazo, Sandra Cortés y Mauricio Valencia.
Fecha entrega	17/05/2024

Próxima reunión

Responsable de convocar: Secretaría Técnica del CASEN.

Fecha: 31 de mayo de 2024 de 2:00 p. m. a 4:00 p. m. (virtual).