



DANE

SEN^{2.0}
Sistema Estadístico
Nacional-Colombia

**Consejo Asesor Técnico del Sistema Estadístico Nacional (CASEN) 2023-2025
Sala Especializada para la Modernización Tecnológica de la Producción Estadística**

Acta No. 6

Ciudad: Bogotá D.C.

Enlace: <https://acortar.link/iV9CMc>

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

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

Fecha: 19/07/2024

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)

Marly Esther de Moya Amaris, coordinadora
GIT Gestión de Datos, OSIS.

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.

Wilson Leonardo Guantivar Avendaño,
designado DSCN.

Juan José Rubio Mesa, designado DIMPE.

Secretaría Técnica del CASEN - DIRPEN

Elizabeth Moreno Barbosa, asesora Dirección
de Regulación, Planeación, Estandarización y
Normalización Estadística.

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

Mónica Patricia Pinzón Torres, Profesional especializado, GIT PAE.

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

Sandra Yaneth Cortés Gamba, contratista GIT PAE.

Mauricio Valencia Amaya, contratista GIT PAE.

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 3:15 p. m.	Presentación de la línea de investigación 1. Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencias de datos, No. 1.1.6 Especificación de artefactos.	Moderador: Alejandro Sandoval, designado OSIS.
3:15 p. m. a 3:35 p. m.	Revisar la programación de los productos de la sala.	Derly Lizarazo, responsable de la sala.
3:35 p. m. a 3:40 p. m.	Compromisos de la reunión.	Derly Lizarazo, responsable de la sala.
3:40 p. m. a 3:55 p. m.	Conclusiones y cierre.	Julieth Solano, directora técnica DIRPEN y secretaria técnica del CASEN.

Objetivo

Recibir realimentación de los miembros de la sala sobre la línea de investigación 1. Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencias de datos, No. 1.1.6 Especificación de artefactos y revisar la programación de los 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, de la Dirección de Síntesis y Cuentas Nacionales (DSCN) y de la Dirección de Metodología y Producción Estadística (DIMPE), así como de la secretaria técnica del CASEN y su equipo de trabajo. Posterior a ello, se realizó el registro fotográfico y se dio la palabra a Elizabeth Moreno, asesora de la dirección técnica de la Dirección de Regulación, Planeación, Estandarización y Normalización (DIRPEN), 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 que el objetivo fue recibir realimentación por parte de los miembros de la sala sobre la línea de investigación 1. Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencias de datos, No. 1.1.1 Revisión de prácticas de Ingeniera de Software para Machine Learning, y conocer los avances en la organización del seminario "Arquitectura de datos: fundamentos y buenas prácticas".

También hizo referencia a los compromisos adquiridos: envío de acta de la reunión para revisión y aprobación, remitir material de trabajo y asunto de consulta para los siguientes temas. 1. Ciencia de Datos 1.1.6 Especificación de los artefactos y Realización del Seminario Arquitectura de Datos 20 de junio 8:00 a. m a 12:30 m.

3. Presentación de la línea de investigación 1. Ciencia de datos, 1.1. Ciclo de vida de modelos de proyectos de ciencias de datos, No. 1.1.6 Especificación de artefactos.

La presentación la realizó Alejandro Sandoval, de la Oficina de Sistemas, y en ella se mostraron los artefactos definidos en el ámbito académico y empresarial que respaldan la documentación de un proyecto de ciencia de datos, se detalló el contenido de la documentación asociada a los procesos que integran las etapas del ciclo de vida de un modelo de ciencia de datos.

A su vez, se discutió la propuesta de documentación que mejor se acople a las necesidades de la Entidad respecto a la gestión e implementación de proyectos de ciencia de datos.

Para lo anterior, se propuso a los miembros de la sala las siguientes preguntas:

1. En relación con la estructura del repositorio, ¿consideran que es necesario ajustar algún aspecto para mejorar la claridad y comprensión de algún componente de este?

2. En cuanto al contenido de los reportes por cada proceso clave, ¿consideran que el nivel de detalle es adecuado, o tienen alguna sugerencia para la inclusión o exclusión de información específica?
3. ¿Consideran que el contenido del README del proyecto cumple con los estándares requeridos y proporciona la información necesaria de manera adecuada?"

Recomendaciones:

A lo largo de la explicación de la OSIS, los miembros del CASEN realizaron sus aportes:

- Al momento de hacer pruebas, valdría la pena documentar con qué conjunto de datos se realizaron las pruebas, donde se encuentra ubicado, entre otras, que quede un registro de las pruebas realizadas, para dar a conocer con que conjuntos de datos se realizaron y la replicabilidad del ejercicio.
- En el documento marco se puede incluir la infraestructura teniendo en cuenta las expectativas para un buen término, los objetivos, en donde estamos y que necesitamos, este puede ser un ejercicio a cargo del Project Manager, para que sirva de referencia a la infraestructura ideal, serían las características del despliegue incluyendo presupuestos.
- En el informe resumen de datos, se podrían incluir consideraciones de seguridad, éticas y de tamaño, dependiendo el tipo de información, si se hace necesario anonimizar, si son datos sensibles (disclaimer). Incluir características de los metadatos
- Para el informe de resumen de modelos, sería importante colocar la justificación del porqué del modelo utilizado. Se sugirió también tener por separado las métricas de evaluación, la validación cruzada, las pruebas de robustez, el análisis de sesgo, varianza y las pruebas de interpretabilidad del modelo.
- En el informe resumen de despliegue, en el contenido en las consideraciones para el mantenimiento del modelo, es mejor quitar la palabra modelo y colocar del despliegue, para que sea entendible para las personas de operaciones o técnicos y detallar como sería la liberación del modelo. Por otra parte, este item se podría desglosar por ejemplo escalabilidad del modelo, la mejora continua, gestión de incidentes, toda la parte reentrenamiento y tener en consideración el monitoreo post del modelo.

A continuación, se realizó una presentación de la estructura y contenido del repositorio para conocimiento de los miembros de la sala, el cuál será compartido para su realimentación.

4. Revisión de la programación de los productos de la sala.

Derly Lizarazo, presentó los productos definidos para entrega el 2024, esto son: guía para la gestión y mantenimiento del ciclo de vida de modelos de proyectos de ciencia de datos, para entrega el 30 de agosto, recomendaciones para robustecer y precisar el modelo del índice de noticias y la app diversa, dado que estas ya se dieron, se espera que le grupo PAD presente la inclusión de las mismas, recomendaciones a la guía de anonimización e inclusión de técnicas innovadoras y emergentes y definir una acción de divulgación e implementación de la misma con participación de los miembros. Se informó que la nota CASEN y la línea de investigación 2. *Seguridad y cumplimiento*, 2.3 *Ciberseguridad y estándares de seguridad informática con alcance internacional*, serán actividades para el año 2025.

5. Compromisos de la reunión

Compromisos

Tarea	Enviar a los miembros de la sala el documento de la guía para la gestión y mantenimiento del ciclo de vida de modelos de proyectos de ciencia de datos, para recibir realimentación en control de cambios el 2 de agosto
Responsable	Derly Lizarazo, responsable de la sala.
Fecha entrega	24/07/2024.
Tarea	Remitir material de trabajo y asunto de consulta para el siguiente tema: recomendaciones sobre la guía de anonimización e inclusión de técnicas innovadoras y emergentes.
Responsable	Derly Lizarazo, responsable de la sala.
Fecha entrega	19/07/2024.
Tarea	Remitir acta de la presente reunión para la revisión y la aprobación de los miembros y los participantes de la sala.
Responsable	Derly Lizarazo, responsable de la sala.
Fecha entrega	26/07/2024.

Próxima reunión

Responsable de convocar: Secretaría Técnica del CASEN.

Fecha: 26 de julio de 2024 de 2:00 p. m. a 4:00 p. m. (virtual).