

Fundamentos en análisis de información y minería de datos en agricultura

Fechas y lugar:

Del 23 al 27 de marzo del 2020, CIMMIT – Texcoco México.

Instructores: El taller será presentado por el Magister en sistemas agrícolas y estudiante de Ph.D en Illinois, Rodrigo Gonçalves Trevisan y Magister en Computación Hugo Andrés Dorado, ambos investigadores del Centro Internacional de Agricultura Tropical, con más de cinco años de experiencia trabajando en minería de datos aplicada a la agricultura.

Requisitos del participante: Conocimiento avanzado de estadística, Excel, R, manejo de bases de datos y preferiblemente programación. Se sugiere que en el curso participen las personas que trabajen en temas relacionados con el manejo, procesamiento o análisis de datos.

Antecedentes: Este taller se ha realizado en 5 países de Latinoamérica: Perú, Uruguay, Honduras, Nicaragua y Colombia, y ha contado con la participación de representantes del sector público y privado, tales como: INTA (Instituto Nacional de Técnica) en Argentina, INIA (Instituto Nacional de Investigación Agropecuaria) en Uruguay, FLAR (Fondo Latinoamericano para Arroz de Riego), COPECO (Comisión Permanente de Contingencias) en Honduras, FEDEARROZ (Federación Nacional de Arroceros de Colombia) y FENALCE (Federación Nacional de Cerealistas y Leguminosas de Colombia), entre otros

Objetivo: Capacitar a los asistentes en el uso de herramientas para procesamiento y análisis de datos aplicado a información comercial de cultivos. En el taller se motivará el uso de software libre tal como R y Quantum Gis. Los datos que serán utilizados provienen del proyecto con MASAGRO, <https://masagro.mx/es/>.

Requerimientos: Computador portátil y preferiblemente con los siguiente software instalados:

- Software estadístico R. Se puede descargar en <https://cran.r-project.org/> ó <https://cran.dcc.uchile.cl/>.
- R Studio. Es una interfaz para R. Se puede descargar en <https://www.rstudio.com/products/rstudio/download/>.
- Slack. Es el canal de comunicación. Hay dos opciones descargar el software o usar la versión, recomendamos esta última. En este link <https://slack.com/> pueden crear su usuario y contraseña.

Contactos: Cualquier inquietud, por favor escribir a: h.a.dorado@cgiar.org ó rodrigo7@illinois.edu

PROGRAMA.

1^{er} día, 23 de marzo

Hora	Tema
8:00 – 9:00	Presentación de los instructores y revisión de software y paquetes.
9:00 – 10:00	Ponencia: Análisis de datos de información comercial en agricultura Digital
10:00 – 10:40	Ponencia: Obtención de datos de clima y suelo por distintas fuentes de información (CRU, Daymet, Estaciones meteorológicas, Chirps, NasaPower y Soilgrid)
10:40 – 11:00	Receso
11:00 – 12:00	Ejercicio práctico de obtención de datos por fuentes de información externa.
12:00 – 1:00	Comida
1:00 – 2:00	Continuación de ejercicio práctico de obtención de datos por fuentes de información externa.
2:00 – 3:00	Ponencia: Criterios de selección, limpieza y transformación de variables agronómicas.
3:00 – 4:00	Revisión y procesamiento de datos utilizando el paquete de R dplyr.
4:00 – 5:00	Ejercicio práctico de procesamiento de datos utilizando el paquete de R dplyr.

2^{do} día, 24 de Marzo

Hora	Tema
8:00 – 9:00	Revisión del ejercicio práctico de procesamiento de datos utilizando el paquete de R dplyr.
9:00 – 10:40	Ponencia: Tratamiento de datos faltantes (K vecinos más cercanos, R Mawgen para datos climáticos, Random forest para datos mixtos y missMDA para datos multivariados)
10:40 – 11:00	Receso
11:00 – 12:00	Organización de grupos de trabajo y distribución de las bases de datos para ejercicios prácticos. Análisis exploratorio de datos
12:00 – 1:00	Comida
1:00 – 3:00	Introducción a Quantum Gis.
3:00 – 4:00	Métodos para evaluar correlación espacial y temporal.
4:00 – 5:00	Continuación de trabajo en grupo con las bases de datos (exploración, limpieza y procesamiento).

3^{do} día, 25 de Marzo

Hora	Tema
8:00 – 8:30	Discusión y repaso de conceptos sobre análisis descriptivo de datos.
8:30 – 9:30	Minería de datos (Conceptos).
9:30 – 10:40	Reducción de dimensiones, selección de atributos y clustering
10:40 – 11:00	Receso
11:00 – 12:00	Métodos supervisados, árboles de clasificación y regresión, random forest y redes neuronales
12:00 – 1:00	Comida
1:00 – 4:00	Continuación de trabajo en grupo con las bases de datos asignadas y con el soporte de los instructores.
4:00 – 5:00	Optimización basada en meta heurísticas

4^{er} día, 26 de Marzo

Hora	Tema
8:00 – 9:30	Visualización de resultados de los modelos.
9:30 – 10:40	Socialización de resultados (Aspectos para tener en cuenta en la entrega y discusión de resultados con otros investigadores o expertos locales).
10:40 – 11:00	Receso
11:00 – 12:00	Continuación de trabajo en grupo con las bases de datos asignadas y con el soporte de los instructores.
12:00 – 1:00	Comida
1:00 – 3:30	Continuación de trabajo en grupo con las bases de datos asignadas y con el soporte de los instructores.
3:30 – 4:30	Presentación de cada grupo y discusión de resultados.
4:30 – 5:00	CIERRE

Links de interés:

- <https://conceptosclaros.com/instalar-r-primeros-pasos/>
- http://cran.r-project.org/doc/contrib/rdebuts_es.pdf
- <https://github.com/bigdataciat>
- <https://www.qgis.org/es/site/forusers/download.html>