Para la aprobación de este curso de posgrado se deberá elaborar un documento de carácter académico, que adquiera la forma de paper o proyecto, de acuerdo a las características y propuesta de cada individuo o grupo.

Se deberá entregar un índice y un resumen del paper hasta una semana después de su presentación formal, que tendrá lugar el último día del curso.

Existirá 1 fecha para la presentación de los proyectos que será durante la última clase del curso.

El proyecto tendrá una extensión máxima de cinco páginas a doble espacio, en letra Arial 11. Asimismo, se deberá preparar para la exposición una presentación en cualquiera herramienta audiovisual que se elija, siempre y cuando se respeten las pautas de los trabajos.

En el proyecto se deberá trabajar articulando conceptos vertidos a lo largo del curso, profundizando o polemizando con los criterios y puntos de vista expuestos durante el mismo por los profesores y/o invitados.

Los conceptos a vertirse en el proyecto deberán basarse en cualquier tema de interés de cada participante o grupo de participante y deberá contener como mínimo los siguientes:

1. Audiencia a la que se dirigirá el proyecto. Se deberá elegir una audiencia a la cual dirigir el proyecto. Se deberá aplicar la suficiente creatividad a la elección para que el proyecto sea desafiante en términos de hallazgos, patrones y explotación y visualización de datos. No se aceptarán audiencias tipo: “*profesor del curso de big data”* o “*alumnos estudiantes de big data”.*
2. Hipótesis o premisas de partida. Se deberá elegir una hipótesis que a priori resulte atractiva tanto para la búsqueda de datos como para el desarrollo del modelo.
3. Posibles Datasets a utilizar.
4. Descripción de las herramientas que se elegirían para realizar este proyecto de investigación, brindando su justificación, así como las técnicas, herramientas y métodos utilizados
5. Desarrollo del modelo de datos y explotación del mismo con una o más herramientas.

**Predicción de rinde de cultivos basado en condiciones y prácticas agropecuarias**

El siguiente trabajo está dirigido a productores agropecuarios de cualquier escala. Particularmente a aquellos que se dediquen a la siembra de los siguientes cultivos:

* Batata
* Mandioca
* Trigo
* Cebada
* Banana
* Oleaginosas

Se ofrecerá un modelo que permita estimar el rinde de cierto cultivo en función de la caída de lluvia anual y el uso de pesticidas y fertilizantes. El rinde se tomará como razón de unidad de producción por área sembrada. La hipótesis que guiará el análisis es que esas tres variables aumentan el rinde de forma directamente proporcional.

**Dataset utilizado**

Se usó el dataset “Agricultural Crop Yield in Indian States Dataset”, publicado por Akshat Gupta et al.[[1]](#footnote-2). El mismo contiene datos de producción de más de cincuenta cultivos en la India entre los años 1997 y 2020. La información está agregada por año y por estado y describe el área sembrada, la producción en toneladas métricas, la caída de lluvia anual, el uso de pesticidas en kilos, el uso de fertilizantes, también en kilos; y el rinde como razón entre la producción y el área sembrada.

Si bien esta información está tomada de la India, la idea es aplicarlo a la Argentina. Por ello se seleccionaron algunos cultivos que crecen en el país

**Objetivo**

Los datos se procesarán en python, principalmente utilizando la librería pandas. A su vez, se implementará un modelo de regresión lineal con scikit-learn. El objetivo será generar un modelo para cada tipo de cultivo seleccionado que pueda recibir los parámetros de entrada y devolver el rinde teórico.

**Análisis de datos exploratorio**

El dataset cuenta con 19689 registros y no cuenta con valores nulos ni repetidos. Cada uno tiene información sobre la producción de cierto cultivo en cierto estado de la India para un determinado año. También detalla la produccón total, el área de siembra, el rinde, que es la razón entre estas últimas; la cantidad de fertilizantes y pesticidas utilizados y, por último, la caída de lluvia en ese año.

De los

1. Disponible en <https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset> [↑](#footnote-ref-2)