A picture containing logo

Description automatically generated

Tesis para “Maestría en explotación de datos y descubrimiento del conocimiento”

Estudio de impactos de sequía en rendimientos de cultivos agrícolas mediante métodos de aprendizaje automático

Reporte de avance para Taller de Tesis I

Alumno: Santiago Luis Rovere (Ingeniero en informática, UBA)

Director: Andrés Farall (Doctor en ciencias de la atmósfera, UBA)

17 de julio de 2021

1. Sinopsis

El presente documento constituye el Reporte de avance n° 1 correspondiente a la Tesis de Maestría a ser llevada a cabo por el autor. Este trabajo de tesis forma parte de un proyecto de investigación y desarrollo denominado “Diseño e implementación inicial de un Sistema de Información sobre Sequías para el Sur de América del Sur (SISSA)”, el cual es financiado por el programa de Bienes Públicos Regionales del Banco Interamericano de Desarrollo (Cooperación Técnica RG-T3308 [1]).

El documento “Plan de Trabajo” presentado previamente incluyó una breve descripción del Proyecto SISSA, la descripción del tema de abordar en la tesis y los objetivos buscados. También se presentó un del plan de trabajo a llevar a cabo para concretar el trabajo de tesis en el término de un año, así como también un cronograma de actividades tentativo. Por tal motivo, esta información no será replicada en este documento, salvo aquello que sea necesario para el desarrollo del presente reporte.

Dado que la naturaleza de este documento es presentar el grado de avance del trabajo de investigación, se comenzará por realizar un resumen del flujo de tareas presentadas en el plan de trabajo. Luego, para cada una de las tareas se detallarán las actividades desarrolladas. Para aquellas tareas que aún no hayan sido abordadas, se presentarán las ideas tentativas que se tienen en mente para su ejecución en el futuro.

Por lo expuesto previamente, se proseguirá con un resumen de las tareas vinculadas al plan de trabajo (Sección 2). El resto de las secciones (Sección 3 a Sección 8) se corresponderán con las principales tareas del plan de trabajo. Finalmente, en la Sección 9 se indicarán las actividades más inmediatas a ser llevadas a cabo.

1. Plan de trabajo

Para llevar a cabo este trabajo de tesis han propuesto una serie de actividades basadas en al diagrama de la Fig. 1. Para poder abordar el estudio de impactos de sequías en rendimientos de cultivos, se deberán buscar regiones agrícolas importantes dentro del área abarcada por el CRC-SAS (Argentina, Bolivia, Brasil – debajo de 10°S, Chile, Paraguay y Uruguay). En estas regiones se determinarán ubicaciones puntuales para las cuales existan registros históricos largos (de al menos 30 años) de temperatura y precipitaciones.

Una vez definidas las ubicaciones puntuales que se utilizarán para el presente estudio, se deberá recabar información acerca de las actividades agrícolas más importantes, los manejos típicos y los tipos de suelos predominantes para cada zona. Esta información permitirá caracterizar y parametrizar las actividades agrícolas y los cultivos cuyos ciclos de vida serán simulados haciendo uso de DSSAT a partir de numerosas series temporales de precipitación y temperatura.

Como se mencionó en el párrafo previo, es necesario contar con numerosas series temporales de precipitación y temperatura para poder ejecutar las simulaciones de los ciclos de vida de los cultivos. Para ello se generarán series temporales estocásticas de precipitación y temperatura que tengan las mismas propiedades estadísticas que las series históricas originales. Esto se llevará a cabo haciendo uso del paquete de R *gamwgen* [2] que fue desarrollado, en parte, por integrantes del proyecto SISSA y de otros proyectos anteriores.

Haciendo uso de las series temporales estocásticas generadas y las parametrizaciones de actividades agrícolas (cultivos, manejos, tipos de suelo, etc.), se ejecutarán las simulaciones de los ciclos de vida de los cultivos correspondientes. Esto significa que para cada serie temporal de una campaña agrícola se asociará un rendimiento resultante, producto de la simulación.

Cada una de las series temporales generadas deberá transformarse en un conjunto de atributos que permitan definir condiciones de sequía para cada momento del ciclo de vida del cultivo. Este proceso se realizará mediante el cálculo de eventos basados en índices de sequía actualmente utilizados por el SISSA [3]. A través de este proceso de *feature engineering* se podrá construir un conjunto de datos tabular con atributos y resultados.

Una vez que se haya logrado construir un conjunto de datos tabular con atributos y resultados, será posible aplicar diversos modelos estadísticos y de aprendizaje automático que permitan vincular los atributos (los cuales definen condiciones de sequía) con los rendimientos asociados. Como todo proceso de *data mining*, deberá ser llevado a cabo de manera iterativa e interactiva hasta lograr los objetivos propuestos.

Fig. 1. Pipeline conceptual del plan de trabajo propuesto

|  |
| --- |
| Diagram  Description automatically generated |
|  |

1. Definición de zonas de estudio
2. Generación de series sintéticas de clima
3. Parametrización de tipos de suelo y manejos de cultivos
4. Simulación de rendimientos de cultivos
5. Caracterización de series sintéticas de clima
6. Vinculación de condiciones de sequía con rendimientos

Referencias

[1] RG-T3308: Diseño e Implementación Inicial de un Sistema de Información sobre Sequías para el sur de América del Sur, <https://www.iadb.org/es/project/RG-T3308>.

[2] Paquete *gamwgen*: <https://github.com/CRC-SAS/weather-generator>

[3] Índices de sequía: <https://sissa.crc-sas.org/monitoreo/indices-de-sequia/>