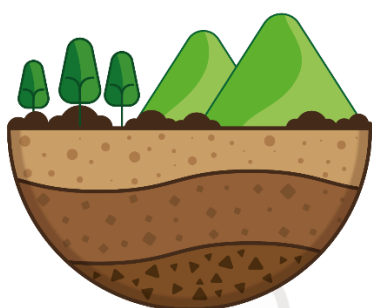


Sistema de información de suelos en el Altiplano Cundiboyacense



Sistema IRAKA

GLOSARIO DE TÉRMINOS
V1



Rural Development
Administration



KoLFACI
KOREA - LATIN AMERICA
FOOD & AGRICULTURE
COOPERATION INITIATIVE



El campo
es de todos

Minagricultura

AGROSAVIA

Corporación colombiana de investigación agropecuaria

Centro de Investigación Tibaitatá, Mosquera (Cund.)

Febrero de 2020

GLOSARIO

A

AVE (Amount of Variance Explained): Porcentaje de varianza explicado de un conjunto total de datos por medio de un número reducido de componentes o factores.

C

Coefficiente Kappa: Es un indicador de precisión en modelos de *Machine Learning* de clasificación. Es una medida de cómo se comparan los resultados de la clasificación con los valores asignados por casualidad. Toma valores entre 0 y 1. Entre más cercano a uno, más precisa es la clasificación.

Covariables: Variables que explican, caracterizan y relacionan directa o indirectamente una variable respuesta. Las covariables en *Machine Learning* pueden ser cuantitativas o cualitativas.

D

Digital Soil Mapping: Mapeo Digital de suelos o *Digital soil mapping* es la creación de una base de datos de suelos geográficamente georreferenciada en una resolución dada. El mapeo digital incluye covariables que relacionen los factores de formación del suelo y expliquen una variable de respuesta dada.

E

Error cuadrático medio de la raíz (ECMR): El error cuadrático medio de la raíz o *Root mean square error* (RMSE) es una medida para evaluar la exactitud de un modelo con una variable de respuesta continua. Este valor mide la cantidad de error que hay entre dos conjuntos de datos. De otra forma, compara entre un valor predicho y un observado o

conocido.

Error medio absoluto (MAE): El error medio absoluto o *Mean absolute error* (MAE) es una medida para evaluar la exactitud de un modelo. Este valor mide la magnitud promedio de los errores en un grupo de predicciones, sin considerar su dirección (positivo o negativo).

Estratificación: Técnica de separación de datos en dos o más grupos que se puedan distinguir.

I

Incertidumbre: Mide con que confianza un algoritmo es capaz de predecir o clasificar una nueva observación.

M

Machine Learning: El aprendizaje automático o *Machine Learning* es un área de la inteligencia artificial, en la que se les dan a los computadores las herramientas necesarias para “aprender” sin necesidad de ser programados por una persona, con el fin de que puedan predecir comportamientos para la toma de decisiones.

Métodos basados en regresión: Son algoritmos de *Machine Learning* que tienen como resultado un valor numérico, dentro de un conjunto infinito de posibles resultados.

Métodos de separación: Se encuentran dentro de las técnicas de clasificación que aprovechan la división de los atributos en el espacio. Como ejemplos se tiene *support vector machine* y *artificial neural network*.

Métodos ensamblados: Es la combinación de las predicciones de probabilidad o numéricas, de dos o más modelos de *Machine Learning*, promediándolos y ponderándolos. Esto proporciona un escenario de voto múltiple que probablemente impulse una predicción a la clase correcta en clasificación o más cerca del número correcto en los modelos de regresión.

Métodos heurísticos: Son una agrupación de técnicas de clasificación usados ampliamente como árboles de decisión, bosques aleatorios y vecinos más cercanos.

Métodos probabilísticos: Se encuentran dentro de las técnicas de clasificación que aprovechan la fórmula de Bayes y computan la posterior probabilidad para clasificar registros, *Naïve bayes*, *tree-augmented naïve bayes*.

P

Precisión general (overall accuracy): Nos indica el porcentaje de sitios mapeado correctamente en función de todos los sitios evaluados.

R

R cuadrado (R^2): se denomina como coeficiente de determinación y refleja la bondad de ajuste de un modelo a la variable que pretende explicar.

Random forest: Los bosques aleatorios (*Random forest*) son un algoritmo de *Machine Learning* utilizado en la clasificación y regresión de variables e imágenes. Este modelo se deriva de los árboles de decisión o *decision trees* para aumentar el número de datos.

Ranger: “RANDOM forest GENErator” conocido como *Ranger*, es un paquete del software estadístico R utilizado de forma eficiente como una implementación rápida de *Random forest* para grandes volúmenes de datos.

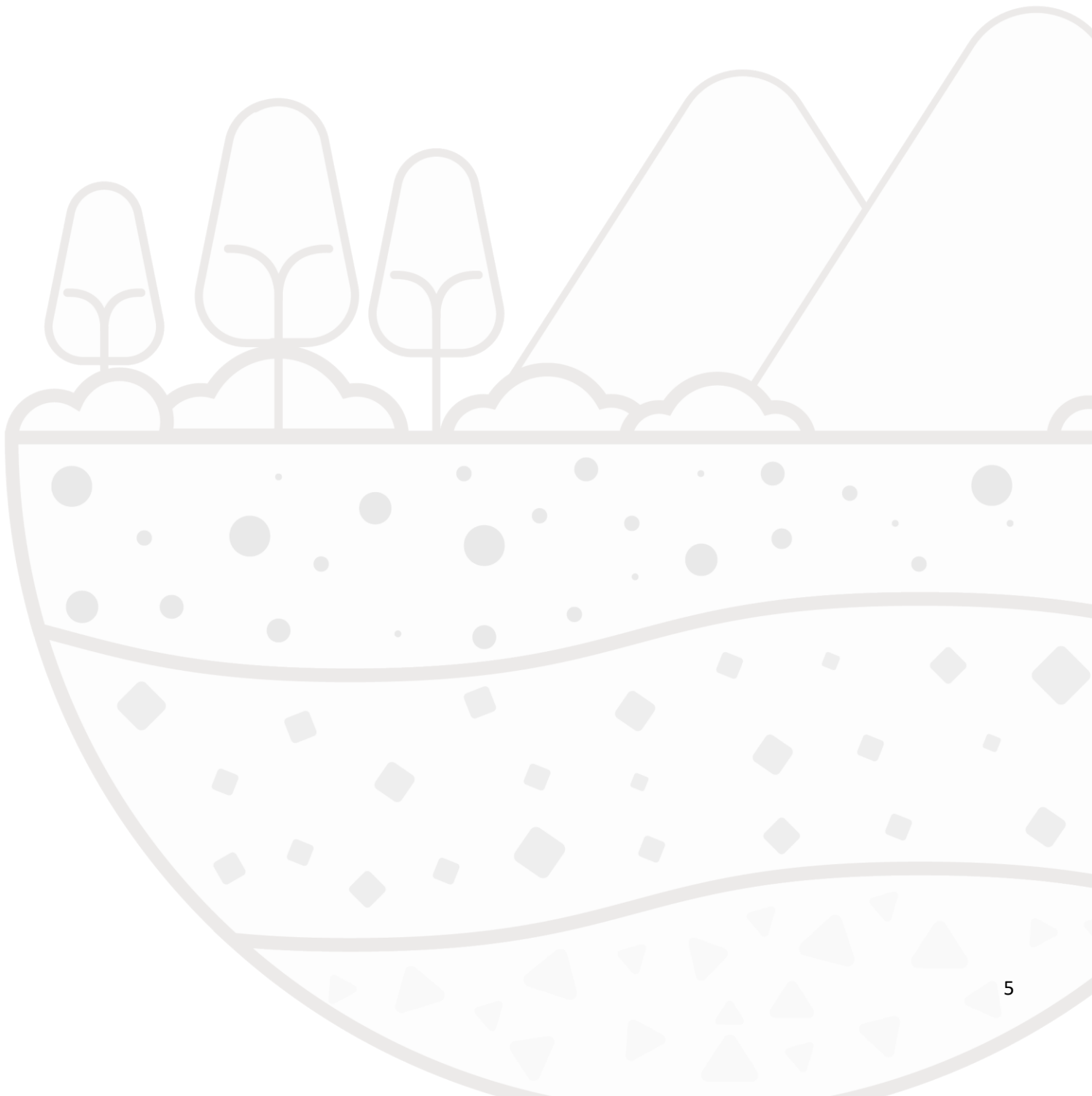
S

Spline: Técnica para homogeneizar los valores de profundidad de muestras de suelos y su posterior modelación en mapeo digital de suelos.

Support vector machine: Las máquinas de soporte vectorial (*Support vector machine*) son un algoritmo de clasificación-regresión dentro de *Machine Learning*. Estas se fundamentan en el concepto de hiperplano, el cual presenta una frontera entre las diversas clases. El proceso de clasificación solo busca saber de qué lado de la frontera está el dato y de esa manera realizar la asignación de la etiqueta.

T

Técnicas de clasificación: Son algoritmos de *Machine Learning* que tienen como resultado una clase o una categoría. Algunos de estos se apoyan de probabilidades para la asignación de un individuo a una categoría.



Control de versiones

Número	Fecha	Autor	Cambio
0	23/10/2019	Camilo I. Jaramillo-Barrios	Redacción del glosario
1	14/02/2020	Gustavo A. Araujo-Carrillo	Ajuste de redacción y disposición de figuras

