

Detección de la roya por medio de árboles de decisión

Esteban Gonzales Tamayo
Universidad Eafit
Colombia
egonzalez@eafit.edu.co

David Felipe Garcia Contreras
Universidad Eafit
Colombia
dfgarcia1@eafit.edu.co

Mauricio Toro
Universidad Eafit
Colombia
mtorobe@eafit.edu.co

RESUMEN

El café es una de las plantas que requiere un cuidado especial para evitar la roya que es una de las enfermedades más nocivas en el café, porque como lo plantea (Consuelo, Oscar, & Roberto) “requiere condiciones climáticas específicas para su producción, su cultivo exige condiciones especiales de suelo, temperatura, precipitación y altitud sobre el nivel del mar”, por lo cual cultivar café exige unas condiciones específicas donde es necesario el uso de la tecnología para medir las condiciones de los cultivos y así poder alertar en qué momentos se puede presentar la roya en el café y poder tomar acción temprana a este problema ya que si no se realiza la roya podría generar grandes pérdidas a los caficultores.

Palabras clave

Arboles de decisión; Roya; Algoritmo; Complejidad; Pila; Patrones; Analisis

Palabras clave de la clasificación de la ACM

Mathematics of computing → Information theory → Coding theory

Theory of computation → Design and analysis of algorithms → Mathematical optimization → Continuous optimization → Stochastic control and optimization

Theory of computation → Theory and algorithms for application domains → Algorithmic game theory and mechanism design → Network formation

1. INTRODUCCIÓN

La planta de café es una de las plantas más importantes para el ser humano, tanto así que su producción es en masa viéndose a Colombia como una de las tres mayores exportadoras, las cuales se ven afectadas por la plaga de la roya en el café lo cual involucra problemas no solo económicos donde también se ven afectados los cultivadores. Las pérdidas en solo América Latina se ven afectados en un 30% de las cosechas. Retomando lo anterior dicho por lo

cual en este trabajo se busca dar una solución con los datos tomados de la roya del café. Se quiere dar una solución óptima en la cual se pueda detectar patrones en la plaga que afecta los cultivos.

2. PROBLEMA

La roya en el café es una situación en la cual muchos cultivadores tienen que pasar por diferentes condiciones. Este problema no se ve como algo nuevo ya que en el año 1868 en la isla de Ceilán se cultivaba el café, pero por culpa de la plaga de la roya cambiaron el café por el té. Si se mueve este problema a Colombia la plaga en los cultivos del café trasciende mucho más ya que si no es tratado se puede ver afectado cerca del 50% del cultivo transformándose este problema en algo más avanzando involucrando factores biológicos. Basándonos en lo anterior se busca a través de un análisis sobre las condiciones de los cultivos de café se pretende detectar patrones o condiciones determinadas que son las causantes de la roya en el café, para poder controlar o prevenir la infestación de este hongo en las plantas de café y así evitar posibles pérdidas materiales en la cultivación del café al igual que la escasez de este cultivo en muchas zonas del país.

3. TRABAJOS RELACIONADOS

3.1 Nodo C5.0

Es un algoritmo usado para generar árboles de decisión, dividiendo información para generar mejores resultados, solamente puede predecir un solo objetivo. El algoritmo como su función principal busca dividir el problema en varias subdivisiones hasta que no hay maneras de dividir luego de realizar esta operación se comienza a eliminar las subdivisiones innecesarias. El algoritmo C5.0 al poder predecir un solo objetivo esta predicción se considera exacta, no solo limitándose a Árboles de decisión por lo que se puede generar además conjuntos de reglas para retener la mayor información posible si se cumple a un registro específico

3.2 Algoritmo C4.5

C4.5 es un algoritmo utilizado para generar árboles de decisión, este se caracteriza porque sus árboles pueden ser

utilizado para clasificación estadística. El algoritmo elige cada atributo y considera todas las posibles pruebas para determinar para así poder crear nuevas subdivisiones a partir del atributo base lo cual resulta más eficiente para dividir los datos cuando se analizan los atributos, al igual que este determina con las pruebas cuál dato obtuvo mayor ganancia de información donde este será utilizado como factor o parámetro de decisión. Este algoritmo puede solucionar problemas basados en hechos o condiciones como realizar una actividad ya sea caminar, correr, etc. basándose en los datos que proceso para tomar una decisión de si es conveniente realizarlo o no

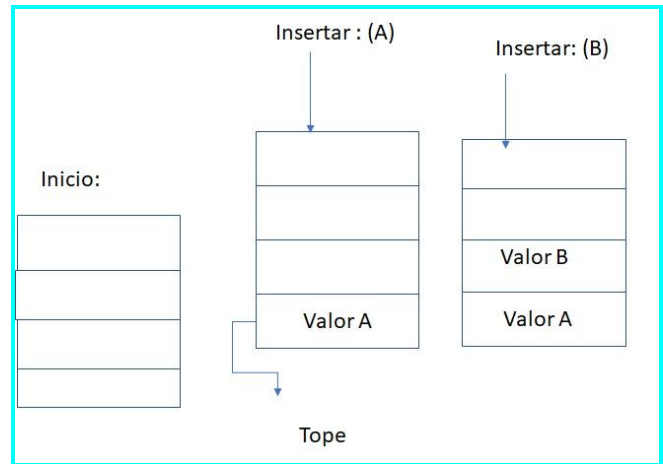
3.3 Algoritmo CART

Es un algoritmo que permite variables de todo tipo, las reglas son de tipo binario permitiendo una estructura de mayor profundidad en los árboles de decisión Para la clasificación se usa la entropía, el índice de Gini, el criterio de Twoing. La función principal de este algoritmo es hallar una estructura de árbol de decisión los más compleja posible, midiendo la complejidad de este por medio de sus nodos terminales, donde se eliminan distintas ramas de manera que combina el riesgo o coste de predicción por lo cual de manera inteligente se usa para eliminar las ramas más débiles, dando así al subárbol más óptimo, permitiendo el alargamiento de los problemas para encontrar por la manera que tome más variables al mismo tiempo , para que su resultado sea el mejor para la utilización dentro del problema.

3.4 Algoritmo ID3

Es un algoritmo que se basa en la búsqueda de una hipótesis o reglas basándose en un conjunto de ejemplos que emplea el algoritmo para determinar sus hipótesis y así poder clasificar todas las instancias basándose en el ejemplo según el tipo de valor que tenga el ejemplo este creará el árbol de decisión. El árbol contendrá ciertos elementos llamados Nodos: Atributos Arcos: valores posibles del nodo principal Hojas: Nodos que clasifican los ejemplos como negativos y positivo.

4. Estructura de datos



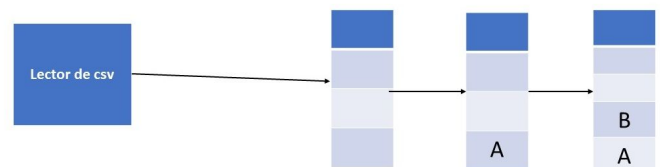
Gráfica 1: Pila.

4.1 Operaciones de la estructura de datos

Diseñen las operaciones de la estructura de datos para solucionar el problema eficientemente. Incluyan una imagen explicando cada operación

Como estructura de datos se implementará la estructura de las pilas, para almacenar la información obtenida previamente, la cual será implementada para ser analizada por el árbol, la pila almacena por cada inserción los datos obtenidos en la lectura del archivo csv los cuales son: ph, soil_temperature,soil_moisture,illumiance,env_temperatur e,env_humidity.

Al igual que al momento de analizar la información solamente se necesitará obtener la información previamente recolectada para luego ser procesada y analizada, sin embargo la información almacenada será almacenada en una copia para trabajar con los datos futuramente.



Grafica 2: Almacenamiento de datos obtenidos del csv.

4.2 Criterios de diseño de la estructura de datos

Nosotros implementamos esta estructura ya que su complejidad es $O(1)$ para realizar push y pop lo cual es útil para obtener la información procesada del archivo csv para

ser procesada en el árbol de decisión para identificar la roya en el café, adicionalmente esta información tendrá una copia de respaldo para conservar la información leída si llega a ser necesitada a futuro, la idea de la pila es enviarle toda la información al árbol para analizarlas y también implementamos esta estructura dinámica precisamente para no tener un espacio fijo de memoria para almacenar la información, si no que se pueda ingresar la información que sea necesaria y también al momento de realizar pop será útil ya que el tamaño o el consumo de la pila disminuye.

4.3 Análisis de Complejidad

Metodo	Complejidad
Insertar (push)	$O(1)$
Eliminar(pop)	$O(1)$
Lectura csv	$O(n)$

Tabla 1: Tabla de complejidad

REFERENCIAS

CONSUELO MONTES R, OSCAR ARMANDO P, & ROBERTO AMILCAR CADENA, Infestación E Incidencia De Broca, Roya Y Mancha De Hierro en Cultivo De Café Del Departamento Del Cauca, Biotecnología En El Sector Agropecuario y Agroindustrial,2012

Wikipedia. ID3 algorithm, mayo 22, 2019. https://en.wikipedia.org/wiki/ID3_algorithm

Amir Ali, 13 Julio 2018

<https://medium.com/machine-learning-researcher/decision-tree-algorithm-in-machine-learning-248fb7de819e>

Jaime Cárdenas L, Oscar Rodrigo S, & Francisco Orozco M,Royadelcafétero,2018

<https://www.croplifela.org/es/plagas/listado-de-plagas/roya-del-cafeto>

IBM,NodoC5.0,2019

https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/clementine/c50node_general.html

Wikipedia. C4.5, abril 9, 2018. <https://es.wikipedia.org/wiki/C4.5>

Jorge Martin A, Data Mining con Árboles de Decision <https://web.fdi.ucm.es/posgrado/conferencias/JorgeMartinslides.pdf>