



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Computación Científica

**Predicción de fechas óptimas para la evaluación de
tizón tardío de papa usando algoritmos de árboles de
decisión y bosques aleatorios**

TESIS

Para optar el Título Profesional de Licenciado en Computación
Científica

AUTOR

Omar Eduardo BENITES ALFARO

ASESOR

Edinson Raúl MONTORO ALEGRE

Lima, Perú

2015

Resumen

PREDICCIÓN DE FECHAS ÓPTIMAS PARA LA EVALUACIÓN DETIZÓN TARDÍO DE PAPA USANDO ALGORITMOS DE ÁRBOLES DE DECISIÓN Y BOSQUES ALEATORIOS

Omar Eduardo, Benites Alfaro

Asesor : Edinson Raúl Montoro Alegre

Título Obtenido : Licenciado en Computación Científica

En esta tesis presentamos el uso de algoritmos de *Arboles de Decisión* y *Bosques Aleatorios* como instrumentos matemáticos y estadísticos-heurísticos para la predicción de fechas optimas en evaluación de *Tizón Tardío*. Dichos algoritmos utilizan los *índices de ganancia* de información (entropía de la información) y los *índices de Gini* para ajustar al máximo la predicción.

Para el desarrollo y análisis de los resultados de los árboles de decisión se utilizaron las implementaciones conocidas como *C4.5* y *CART*; mientras que para los bosques aleatorios se empleo *RandomForest*.

Palabras Clave: Árboles de decisión
Bosques Aleatorios
Tizón tardío
Ganancia de información
Indice de Gini
Poda

Abstract

Prediction of optimal evaluation dates for late blight of potato using decision trees and random forest algorithms

Omar Eduardo Benites Alfaro

November - 2015

In this work, we propose the use of *Decision trees* and *Random Forest* algorithms as a mathematical and statistical-heuristic instrument to forecast optimal dates of evaluation for potato late blight diseases. The mentioned algorithms use three types of indexes like *Information Gain Index*, *Information Gain Ratio* and *Gini Index*, to fit best the prediction.

To develop and analyze the results we employ language R, specifically, well-know r-implementations packages as *C4.5* y *CART* for decision trees; while for Random Forest we use *RandomForest* r-package.

Keywords: Decision Trees
Random Forest
Late Blight
Information Gain
Gini Index
Prune, Pruning