



Universidad Nacional Experimental del Táchira - Vicerrectorado Académico - Decanato de Docencia -  
Departamento de Ingeniería en Informática - Trabajo de Aplicación Profesional - Proyecto Especial de Grado

# Diseño de un clasificador de tubérculos de papa criolla para diferentes densidades de siembra según el peso fresco por calibre empleando Bayes Naive.

Autor: Estefany Salas - [estefany.salas@unet.edu.ve](mailto:estefany.salas@unet.edu.ve)

Tutor: Rossana Timaure - [rttg@unet.edu.ve](mailto:rttg@unet.edu.ve)

Julio, 2019.

# Capítulo 1: Preliminares

- Planteamiento y formulación del problema
- Objetivos
- Aportes de la investigación

# Objetivo General

Diseñar un clasificador de tubérculos de papa criolla para diferentes densidades de siembra según el peso fresco por calibre empleando Bayes Naive.

# Objetivos Específicos

- Diagnosticar el formato de las variables de entrada y salida para el clasificador.
- Establecer el tipo de algoritmo Bayes Naive a emplear y las características para el entrenamiento.
- Implementar el algoritmo de clasificación basado en Bayes Naive.
- Realizar las pruebas de funcionamiento y la comparación estadística.

# Aportes de la investigación

- Control de las características de la papa criolla *Solanum phureja*.
- Herramienta para clasificación y evaluación de la misma.

# Capítulo 2: Fundamentos Teóricos

- Antecedentes
- Bases Teóricas

# Antecedentes

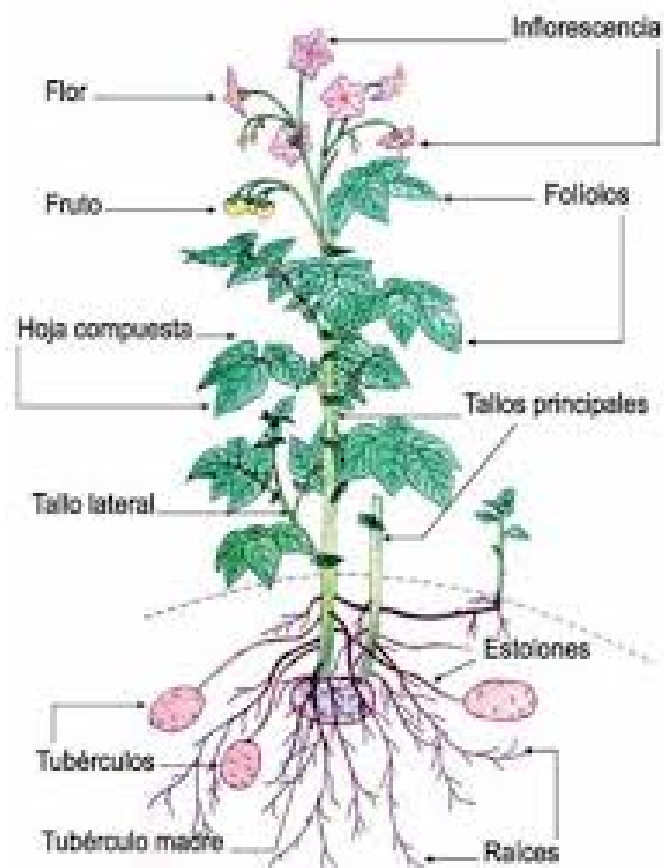
- Bacteria Identification from Microscopic Morphology using Naïve Bayes. Mohamad, N., Jusoh, N., Htike, Z., Lei Win, S. (2014).
- Classification of Selected Apple Fruit Varieties using Naïve Bayes. Misigo, R., Miriti, E. (2016).
- Evaluación del Rendimiento en Papa Criolla (*Solanum phureja*) variedad Yema de Huevo, bajo diferentes Densidades de Siembra en la Sabana de Bogotá. Arias, V., Bustos, P., Núst ez, C. (1996).
- Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. Tsangaratos, P., Ilia, I. (2016).

# Bases Teóricas

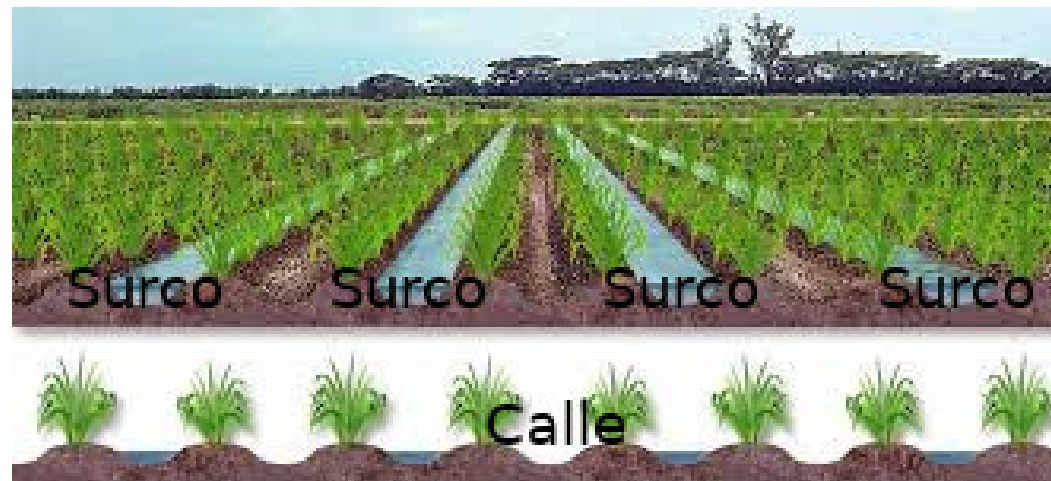
- Python
- PyCharm
- Regresión Espacial
- Probabilidad Condicional
- Teorema de Bayes
- Inferencia Bayesiana
- Clasificador Bayes Naive
- Bayes Naive Bernoulli
- Bayes Naive Multinomial
- Bayes Naive Gaussiano
- Curva Característica Operativa del Receptor
- Papa criolla (*Solanum Phureja*)
- Glosario



## Morfología de la papa criolla (*Solanum Phureja*)



## Disposición de las plantas en una siembra



# Capítulo 3: Fundamentos Metodológicos

- Enfoque de la investigación: cuantitativo
- Tipo o nivel de investigación: de campo y correlacional
- Diseño de la investigación: no experimental cuantitativa y transeccional correlacional-causal
- Metodología

# Metodología: CRISP-DM (Cross Industry Standard Process for Data Mining)

1. Comprensión del problema
2. Comprensión de datos
3. Preparación de datos
4. Modelado
5. Evaluación

# Capítulo 4: Desarrollo

# 1. Comprensión del problema

1.1. Determinar los objetivos del problema

1.2. Evaluación de la situación

1.3. Determinación de los objetivos del proyecto propuesto

1.4. Producción de un plan del proyecto

## 2. Comprensión de datos

2.1. Recolección de datos iniciales

2.2. Descripción de los datos

2.3. Exploración de datos

2.4. Verificación de la calidad de los datos

## 2.2. Descripción de los datos

	Densidad	PD1	PD2	PD3	PD4
Conteo	2839	2839	2839	2839	2839
Media	1.8742	74.1536gr	226.3262gr	213.3114gr	18.9698gr
Mínimo	1	0gr	0gr	0gr	0gr
Máximo	3	390.0000gr	1080.0000gr	1290.0000gr	775.9459gr
Mediana	2	62.0547gr	195.0000gr	164.7058	0.0000gr
Varianza	0.6582	2892.7971	26386.3844	36986.5868	3486.4316
Moda	1	0gr	0gr	0gr	0gr
Distribución	N/A	normal	normal	normal	normal



## 2. Comprensión de datos

2.1. Recolección de datos iniciales

2.2. Descripción de los datos

2.3. Exploración de datos

2.4. Verificación de la calidad de los datos

# 3. Preparación de datos

3.1. Selección de datos

3.2. Limpieza de los datos

3.3. Estructuración de los datos

3.4. Integración de los datos

3.5. Formato de los datos

# 4. Modelado

## 4.1. Cálculo de media y varianza

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

# 4. Modelado

## 4.2. Cálculo de probabilidades previas

$$Prob\ Previa(c) = \frac{ocurrencias\ de\ la\ clase\ c}{Nro\ total\ de\ datos}$$

# 4. Modelado

## 4.3. Cálculo de probabilidades posterior

$$P(x_i|c) = \frac{1}{\sqrt{2\pi\sigma_{x_i,c}^2}} e^{-\frac{(x_i - \bar{x}_{x_i,c})^2}{2\sigma_{x_i,c}^2}}$$

## 4. Modelado

4.4. Obtiene las probabilidades condicionales y las predicciones

$$P(c_i|x) = \frac{P(c_i) P(x|c_i)}{\sum_{j=1}^n P(c_j) P(x|c_j)}$$

# 5. Evaluación

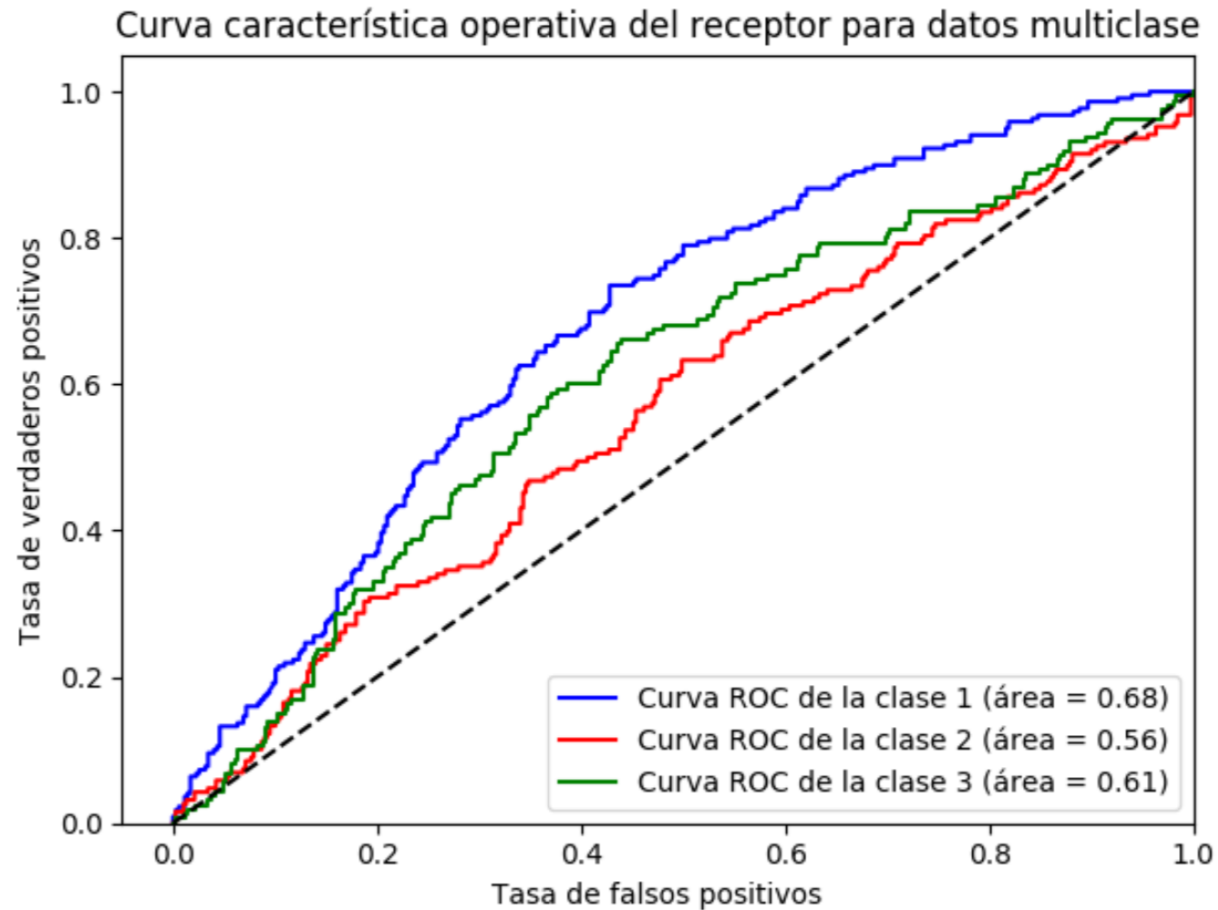
5.1. Evaluación de los resultados

5.2. Proceso de revisión

5.3. Determinación de futuras fases

# 5. Evaluación

## 5.1. Evaluación de los resultados





# 5. Evaluación

5.1. Evaluación de los resultados

5.2. Proceso de revisión

5.3. Determinación de futuras fases

# Capítulo 5: Conclusiones y Recomendaciones

Gracias.