

Taller de Minería de Datos

Modelo Naive Bayes

Clasificación de Solicitantes Crédito

Alejandro Peña P., I.M., M.Sc., Ph.D.

japena@eafit.edu.co

Área de Gestión de la Información y Riesgos

Escuela de Administración

Distinguished International Associates

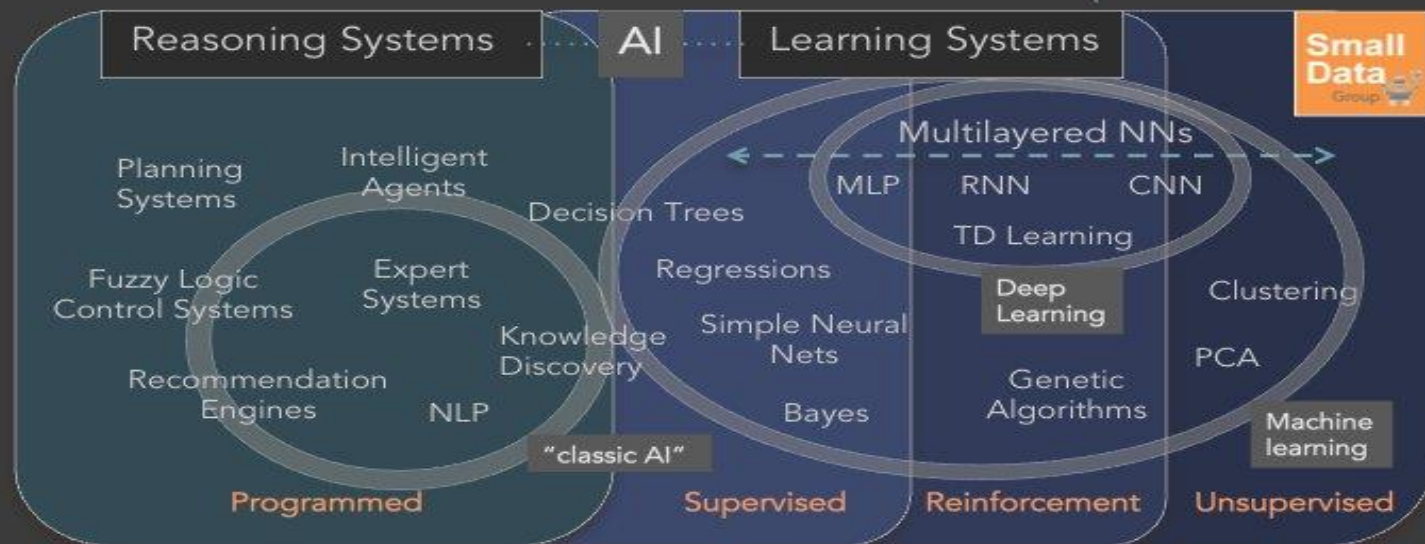
[Royal Academy – Legacy Award](#)

UNIVERSIDAD
EAFIT[®]

www.eafit.edu.co

INSPIRA CREA TRANSFORMA

Taxonomy of AI + ML



© Allen Bonde 2018 / Small Data Group

- ✓ **FINTECH**, Desarrollo de productos financieros soportados en la IA – *Banca Cognitiva*.
- ✓ **INSURTECH**, Asegurar nuevos conceptos promovidos IOT-IOB – *Riesgos Paramétricos*.

Sostenibilidad Ambiental y Financiera



NUBANK, EL BANCO MÁS GRANDE DE OCCIDENTE, EN NÚMEROS

iPro^{UP}



nu bank

2013

Año de su fundación en San Pablo, Brasil.

1.400

Empleados posee a nivel global.



#1

1° en innovación en América Latina según Fast Company.



5.000.000
De tarjetas emitidas en Brasil.

nu conta

1.500.000
De usuarios de cajas de ahorro (NuConta).

u\$s180 MILLONES

Inversión recibida del gigante chino Tencent.

u\$s4.000 MILLONES

Es su valor de mercado.

u\$s170 MILLONES

Factura anualmente.

FUENTE: Elaboración propia

iproup.com

- El *Modelo Naive Bayes* es un modelo Machine Learning (ML) por aprendizaje supervisado para la clasificación de los datos en términos de una variable de referencia.

Caso de Estudio:

- Una Fintech quiere clasificar sus solicitantes de crédito tomando como referencia la variable *Preaprobación*.
- La *variable Pre-Aprobación*, es una variable binaria, la cual toma los valores 0 (Pre-Negado) y 1 (Pre-Aprobado).
- Para este modelo, la entidad financiera quiere utilizar como variables de entrada: Edad, Ingresos, Egresos, Monto.

Valor de Pertenencia (VP):

- Se define como la *afinidad* que un valor de una variable tiene con respecto a su *estructura estadística*.
- Esta función de pertenencia se denota y define:

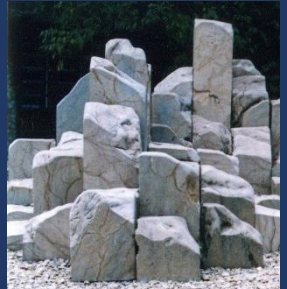
$$VP = \text{Exp} \left(-\frac{1}{2} \cdot \left(\frac{x - \mu}{\sigma} \right)^2 \right)$$

Donde:

VP : Valor de pertenencia $VP \in (0\%, 100\%)$

μ : Indica la media de la variable aleatoria

σ : Desviación estándar de la variable aleatoria.



Contenido

Analítica y Ciencia Datos

Medidas de Tendencia Central y Dispersión
Coeficiente de Correlación
Normalización
Programación - POO

Modelos Machine Learning

Introducción al Machine Learning
Modelo Naive Bayes

Referencias

Modelo Naive Bayes

- El *Modelo Naive Bayes* es un modelo Machine Learning (ML) por aprendizaje supervisado para la clasificación de los datos en términos de una variable categórica de referencia: Pre-Aprobación, Nivel de Renta, Nivel de Estudios, Número de Hijos.

Valor de Pertenencia (VP):

- Se define como la *afinidad* que un valor de una variable tiene con respecto a su *estructura estadística*.
- Esta función de pertenencia se denota y define:

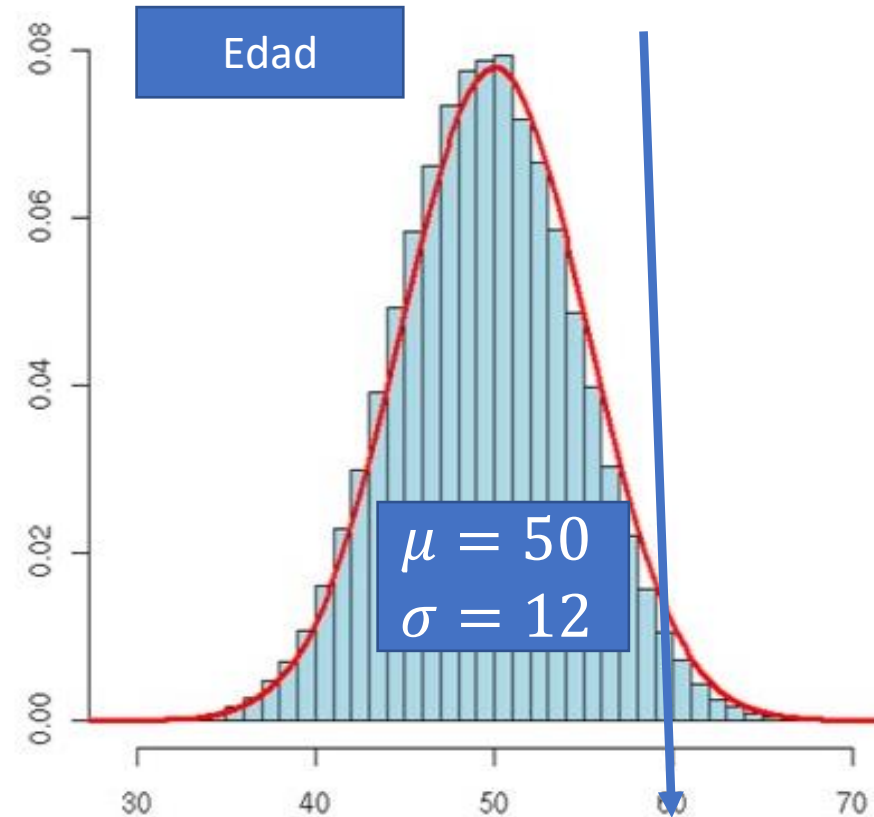
$$VP = \text{Exp} \left(-\frac{1}{2} \cdot \left(\frac{x - \mu}{\sigma} \right)^2 \right)$$

Donde:

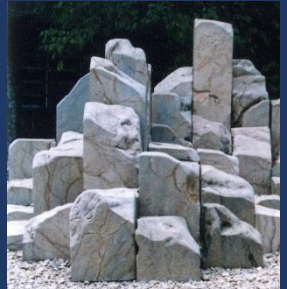
VP: Valor de pertenencia $VP \in (0\%, 100\%)$

μ : Indica la media de la variable aleatoria

σ : Desviación estándar de la variable aleatoria.



$$VP = \text{Exp} \left(-\frac{1}{2} \cdot \left(\frac{60 - 50}{12} \right)^2 \right) : VP = 0.706648$$



Contenido

Analítica y Ciencia Datos

Medidas de Tendencia Central y Dispersión
Coeficiente de Correlación
Normalización
Programación - POO

Modelos Machine Learning

Introducción al Machine Learning
Modelo Naive Bayes

Referencias

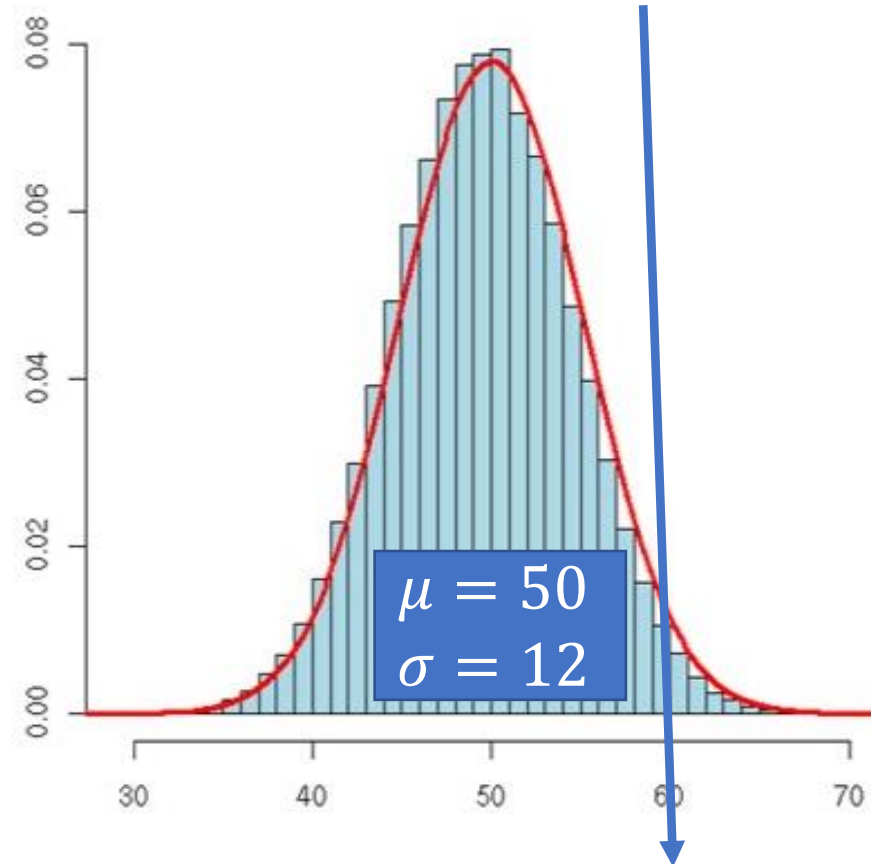
Modelo Naive Bayes

- El *Modelo Naive Bayes* es un modelo Machine Learning (ML) por aprendizaje supervisado para la clasificación de los datos en términos de una variable de referencia.

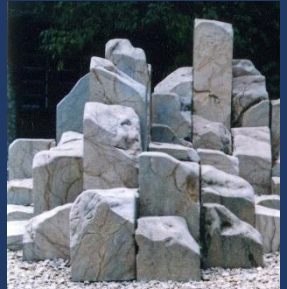
Reto- Challenge:

Modelo Naive Bayes Multivariado:

- Crear un modelo *Naive Bayes* para la preaprobación de créditos para la totalidad de los solicitantes del crédito de la base de datos de referencia:
- Las *variables socioeconómicas* que describen un solicitante de crédito son las siguientes: Edad, Hijos, Personas a Cargo, Estrato, Ingresos, Egresos, Monto, Plazo y Cuota



$$VP = \text{Exp}\left(-\frac{1}{2} \cdot \left(\frac{60 - 50}{12}\right)^2\right) : VP = 0.706648$$



Contenido

Analítica y Ciencia Datos

Medidas de Tendencia Central y Dispersión
Coeficiente de Correlación
Normalización
Programación - POO

Modelos Machine Learning

Introducción al Machine Learning
Modelo Naive Bayes

Referencias

- El *Teorema de Bayes* expresa la probabilidad de que ocurra un evento A dado que a ocurrido un evento B, el teorema de Bayes se puede expresar:

$$P(A/B) \cdot P(B) = P(A) \cdot P(B/A): P(A/B) = \frac{P(A) \cdot P(B/A)}{P(B)}$$

Caso de Estudio:

- Una empresa tiene una fábrica que dispone de tres máquinas A,B,C, que produce tres envases.
- Se sabe que la máquina A produce el 40%, la máquina B el 30% y la máquina C un 30%.
- Se sabe que cada máquina produce envases defectuosos,. La máquina A produce el 2%, la máquina B un 3% y la máquina C un 5%.
- ¿Cuál es la probabilidad de que el envase sea defectuoso cuando fue producido por la máquina A?

Solución al Problema

- Probabilidad que un envase sea del mismo tipo: $P(A) = 0.4, P(B) = 0.3, P(C) = 0.3$.
- La probabilidad de que los envases sean defectuosos:

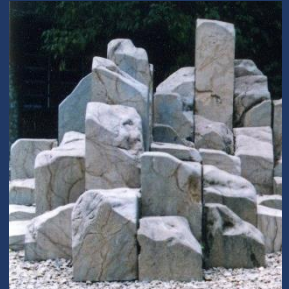
$$P(D/A) = 0.02, P(D/B) = 0.03, P(D/C) = 0.05$$

- ¿Cuál es la probabilidad de que el envase sea defectuoso?

$$P(D) = [P(A) \cdot P(D/A)] + [P(B) \cdot P(D/B)] + [P(C) \cdot P(D/C)]$$

- La probabilidad de que una máquina produzca un envase defectuoso:

$$P(A/D) = \frac{P(A) \cdot P(D/A)}{P(D)}$$



Contenido

Analítica y Ciencia Datos

Medidas de Tendencia Central y Dispersión
Coeficiente de Correlación
Normalización
Programación - POO

Modelos Machine Learning

Introducción al Machine Learning
Modelo Naive Bayes

Referencias

- El *Modelo Naive Bayes* es un modelo Machine Learning (ML) por aprendizaje supervisado para la clasificación de los datos en términos de una variable de referencia.

Caso de Estudio:

- Una Fintech quiere clasificar sus solicitantes de crédito tomando como referencia la variable *Preaprobación*.
- La variable preaprobación, es una variable binaria, la cual toma los valores 0 (PreNegado) y 1 (Preaprobado).
- Para este modelo, la entidad financiera quiere utilizar como variables de entrada: Edad, Ingresos, Egresos, Monto.

Valor de Pertenencia (VP):

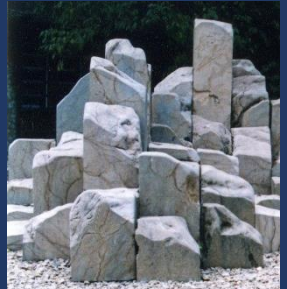
- Para el caso de estudio, el modelo Naive Bayes se puede expresar:

$$P(Ap) = \frac{P(Ap) \cdot P(Ed/Ap) \dots P(Monto/Ap)}{Evidencia}$$

$$P(Ap) = \frac{P(Ap) \cdot P(Ed/Neg) \dots P(Monto/Neg)}{Evidencia}$$

$$Evidencia = Numerador 1 + Numerador 2$$

- El modelo *Naive Bayes* asume que las variables explicativas son estadísticamente independientes.



Contenido

Analítica y Ciencia Datos

Medidas de Tendencia Central y Dispersión
Coeficiente de Correlación
Normalización
Programación - POO

Modelos Machine Learning

Introducción al Machine Learning
Modelo Naive Bayes

Referencias

Inspira Crea Transforma

Muchas Gracias