

DISEÑO DE ACTIVIDAD

Curso que imparto: Inteligencia de Negocios	Carrera: Ingeniería en Sistemas y Redes Informáticas
Nombre de la actividad: Algoritmos de clasificación Probabilística.	
Descripción de la actividad: Un sistema clasificación es más efectivo mientras cada elemento del espacio (instancia) quede ubicado en uno y solo una clase y disminuye su precisión cuando un elemento puede quedar ubicado en dos o más clases (ambigüedad). Dentro de la minería de datos, se pueden tener un conjunto de instancias, donde cada instancia recoge una serie de valores clasificados en atributos, de manera similar como en una tabla de una base de datos se tienen registros donde cada registro recoge una serie de valores ordenados en diferentes campos. El desarrollo del algoritmo de clasificación probabilística comprende en estimar el camino más probable que se tomará en el proceso de varios casos con combinaciones similares de factores. El algoritmo es llamado clasificador bayesiano.	
Metodología utilizada: Metodología PDI: con esta profundizan en contenidos científicos y matemáticos, utilizan de forma adecuada herramientas de información. Buscan alternativas para desarrollar lo planteado con los recursos que les ayuden a agilizar los procesos.	
Principios que la sustentan: <ul style="list-style-type: none">- Conocimientos previos: Manteniendo conceptos vistos anteriormente pueden aplicarse y tenerse en cuenta en el desarrollo.- Aprendizaje significativo: Los estudiantes relacionan los resultados con los conocimientos que han adquirido, generando sus propias conclusiones.- Procesamiento de la información:	
Niveles de pensamiento superior esperados: <ul style="list-style-type: none">- Crear: diseña en Excel las fórmulas para realizar los cálculos.- Evaluar: resultados obtenidos al aplicar un algoritmo de clasificación.	
Descripción y/o ejemplos de actividades o preguntas: Se pide evaluar una tabla de hechos y a partir de ésta revisar el contexto en que se aborda para aplicar el algoritmo y encontrar una tendencia, ejemplo de contexto: Se ha observado el tiempo en el cual una ruta de distribución de productos cumple con su tarea, notándose que en algunas ocasiones se cumplen los tiempos de entrega y en otras no. A partir de observaciones se ha determinado que el suceso de cumplir a tiempo o no con las entregas depende de condiciones ambientales externas tales como la lluvia, el día de la semana, la hora del día y el tráfico. La tabla de observaciones que toma en cuenta el tiempo de entrega como la clase:	

Día	Hora	Lluvia	Tráfico	Clase
semana	mañana	ninguna	poco	a tiempo
semana	tarde	poca	poco	a tiempo
semana	tarde	poca	poco	a tiempo
semana	tarde	copiosa	alto	tarde
sábado	noche	ninguna	normal	a tiempo
semana	madrugada	ninguna	normal	muy tarde
feriado	noche	poca	alto	a tiempo
domingo	noche	ninguna	normal	a tiempo
semana	tarde	copiosa	alto	muy tarde
semana	noche	poca	poco	a tiempo
sábado	mañana	copiosa	alto	cancelada
semana	noche	poca	alto	a tiempo
sábado	tarde	ninguna	normal	tarde
semana	noche	ninguna	alto	a tiempo
semana	tarde	copiosa	normal	muy tarde
sábado	madrugada	poca	alto	a tiempo
semana	madrugada	copiosa	poco	a tiempo
feriado	mañana	poca	normal	a tiempo
semana	mañana	ninguna	normal	a tiempo
semana	mañana	poca	normal	a tiempo

Se pide:

- 1-Identificar los atributos incluidos en la tabla.
- 2-Identificar los miembros de cada atributo.
- 3-Identificar los eventos incluidos en la tabla.
- 4-Calcular las probabilidades prioritarias de cada clase

Etapas del modelo SAMR de integración de las TIC:

Sustituir

- Los estudiantes:
Utilizan papel y calculadora para realizar los cálculos propuestos.

Proceso de sustituir:

- Antes: Los estudiantes entregan los resultados en papel.
- Ahora: los estudiantes utilizan Herramienta de hoja de cálculo para facilitar el

ordenamiento de los datos:

Los estudiantes utilizan Herramienta de hoja de cálculo para realizar los cálculos de los algoritmos, graficarlos y presentar un reporte con formato.

Nombre de la herramienta TIC y descripción de su utilización:

- EXCEL: Herramienta de hojas de cálculo donde se pueden crear y utilizar fórmulas, funciones, etc, tabular datos muy fácilmente, ordenar datos, etc.

Instrucciones para realizar la actividad:

1-Los atributos se encuentran en el título de cada columna en la tabla a excepción de "Clase".

2-Los miembros de cada atributo son los registros en cada columna.

3-Los eventos son cada fila de la tabla.

4-Calcular las probabilidades prioritarias de cada clase

- En la tabla se distinguen cuatro clases, se iniciará con la primera: **a tiempo**
- Se necesita algo llamado **probabilidad prioritaria** y se calcula dividiendo la cantidad de instancias de cada clase entre el número de eventos de la tabla:

Clase = a tiempo: $14/20 = 0.7$ (14 veces aparece a tiempo entre 20 filas de la tabla)

De la misma forma se calcula para las otras tres clases.

- Calcule la probabilidad de que:
 - a. La entrega se realice a tiempo en un día domingo.
 - b. La entrega se realice a tiempo en horas de la mañana.
 - c. La entrega se realice a tiempo con tráfico alto.
 - d. La entrega se realice a tiempo con copiosa lluvia.

En este caso, aplicaremos probabilidad condicional. Ejemplificando el literal a:

Dividimos todas las veces que encontramos la combinación "a tiempo – domingo" entre todas las veces que tocó domingo:

$1/1 = 1$ La probabilidad es entonces, igual a uno.

El mismo procedimiento se sigue para los otros literales.

- f) Especifique las probabilidades de ocurrencia Bayesiana para cada uno de las clases en una tabla. Especifique también en la tabla las probabilidades prioritarias de cada clase.

Las probabilidades bayesianas son las inversas de las condicionales tradicionales.

Por ejemplo, para el literal anterior, la probabilidad bayesiana sería:

Dividir todas las veces que se encontró la combinación "a tiempo – domingo" entre todas las veces que se llegó a tiempo:

$1 / 14 = 0.07$

El mismo procedimiento se sigue para las otras celdas de la tabla

g) Utilice el algoritmo de Bayes para determinar en qué clase caería una entrega que se lleve a cabo en:

- Día de semana, con tráfico alto, lluvia copiosa y en horas de la tarde.
- Día feriado, tráfico normal, poca lluvia y en horas de la mañana.

Para resolver este punto, aplicamos el siguiente procedimiento (tomaremos como ejemplo, el primer caso):

Calcular la probabilidad bayesiana en todas las clases para cada condición, así como la prioritaria de cada clase y multiplicar todas las probabilidades encontradas:

Probabilidad (día=semana | clase= a tiempo) = $9 / 14$
Probabilidad (tráfico=alto | clase= a tiempo) = $4 / 14$
Probabilidad (lluvia=copiosa | clase= a tiempo) = $1 / 14$
Probabilidad (hora = tarde | clase= a tiempo) = $2 / 14$
Probabilidad prioritaria (clase = a tiempo) = $14 / 20$
Multiplicación: $(9/14) (4/14) (1/14) (2/14) (14/20) = 0.0013$

Probabilidad (día=semana | clase= tarde) = $1 / 2$
Probabilidad (tráfico=alto | clase= tarde) = $1 / 2$
Probabilidad (lluvia=copiosa | clase= tarde) = $1 / 2$
Probabilidad (hora = tarde | clase= tarde) = $2 / 2$
Probabilidad prioritaria (clase = tarde) = $2 / 20$
Multiplicación: $(1/2) (1/2) (1/2) (2/2) (2/20) = 0.0125$

Probabilidad (día=semana | clase= muy tarde) = $3 / 3$
Probabilidad (tráfico=alto | clase= muy tarde) = $1 / 3$
Probabilidad (lluvia=copiosa | clase= muy tarde) = $1 / 3$
Probabilidad (hora = tarde | clase= muy tarde) = $2 / 3$
Probabilidad prioritaria (clase = muy tarde) = $3 / 20$
Multiplicación: $(3/3) (1/3) (1/3) (2/3) (3/20) = 0.15$

Probabilidad (día=semana | clase= cancelada) = $0 / 1$
Probabilidad (tráfico=alto | clase= cancelada) = $1 / 1$
Probabilidad (lluvia=copiosa | clase= cancelada) = $1 / 1$
Probabilidad (hora = tarde | clase= cancelada) = $0 / 1$
Probabilidad prioritaria (clase = cancelada) = $1 / 20$
Multiplicación: $(0/1) (1/1) (1/1) (0/1) (1/20) = 0$

De los cuatro productos, el mayor es 0.15, el cual corresponde a clase = muy tarde; por tanto, dicha clase es la seleccionada por el algoritmo bayesiano como respuesta del problema.

Criterios de éxito:

- Al finalizar el ejercicio se obtendrá el resultado de todas las clases (aplicando el algoritmo).
- El alumno deberá utilizar su razonamiento para determinar según los resultados de las fórmulas y encontrar la respuesta según lo descrito en el proceso de realización del ejercicio.