

CAPITULO III

3. DISEÑO DEL APLICATIVO WEB

La página web que se ha desarrollado para el aprendizaje de análisis discriminante permite probar 5 métodos que se expusieron en el capítulo I.

Aunque el análisis discriminante puede trabajar con elementos determinados por n variables ya sean categóricas o numéricas, esta herramienta se restringe a trabajar con elementos determinados por 2 variables numéricas. Esto se debe a que ésta es una herramienta gráfica que permite visualizar cada elemento según su clase en 2

dimensiones. Puesto que cada elemento tiene 2 variables o atributos pueden ser representados como puntos (pares ordenados) en un plano bidimensional.

Aunque en un mismo plano bidimensional pueden ser visualizados elementos de más de 2 clases, esta página sólo trabaja con datos provenientes de 2 clases posibles.

Al iniciar este aplicativo web se presenta un plano bidimensional con un conjunto de puntos de 2 clases junto con una sección donde el usuario puede manipular las distintas opciones que se ofrecen. La figura 3.1 presenta esta pantalla inicial.

La primera opción que ofrece la página es el método que se desea aplicar, la figura 3.2 presenta este detalle, donde la opción “Regresión” se refiere al método de los mínimos cuadrados ordinarios. A continuación se expondrá la manera de utilizar las otras opciones presentes en la página.

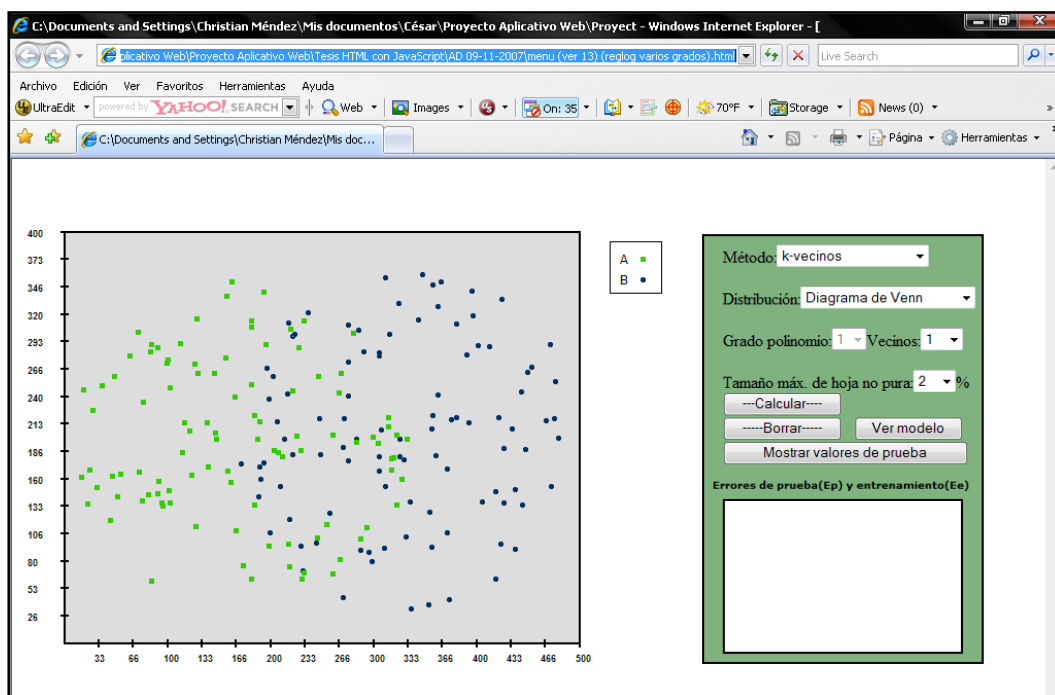


FIGURA 3.1 PANTALLA INICIAL DEL APLICATIVO WEB

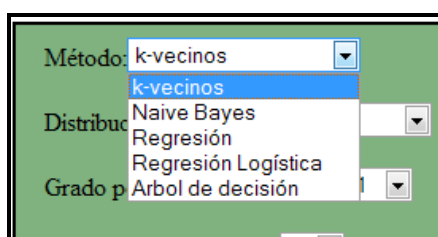


FIGURA 3.2 SELECCIÓN DEL MÉTODO

3.1 Distribución del conjunto de datos

En el plano bidimensional se presentan un conjunto de puntos distribuidos de una manera específica. Existen 3 distribuciones que el usuario puede seleccionar (figura 3.3):

- Diagrama de Venn
- Cuadrados traslapados
- Exponencial

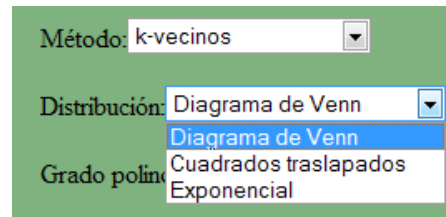


FIGURA 3.3 OPCIONES DE DISTRIBUCIÓN DE PUNTOS

Cuando se inicia la página la distribución por defecto es la diagrama de Venn como su nombre lo indica esta opción distribuye los puntos como un diagrama de Venn. La figura 3.4 muestra el esquema de esta distribución.

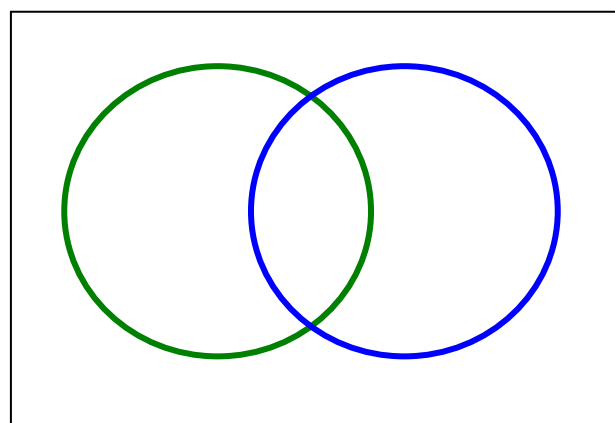


FIGURA 3.4 DISTRIBUCIÓN DE DIAGRAMA DE VENN

En cada círculo del esquema anterior se generarán puntos de una clase determinada, es decir, en el círculo verde caerán puntos o elementos (x,y) de la clase A y en el otro círculo los puntos de la clase B. Sin embargo dentro de cada círculo los puntos estarán distribuidos al azar. Como se observa en la figura 3.4 esta distribución de puntos sí tiene solapamiento, la intersección de los círculos es el área en donde hay puntos de ambas clases. La figura 3.1 muestra como luce este tipo de distribución. También en la figura 3.5 se ha sobrepuesto los círculos del diagrama a un conjunto de puntos generados con esta distribución.

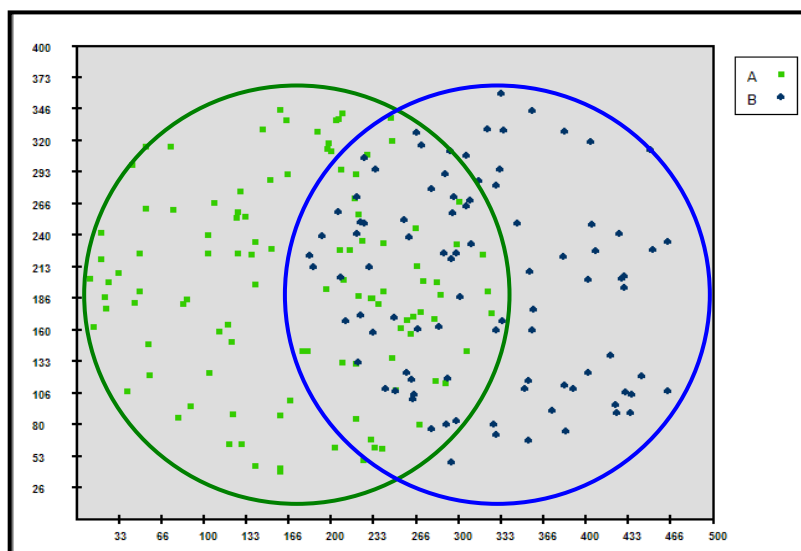


FIGURA 3.5 CONJUNTO DE PUNTOS DISTRIBUIDOS COMO DIAGRAMA DE VENN

Otra de las distribuciones llamada “cuadrados traslapados” genera cada tipo de puntos en un cuadrado (o rectángulo) correspondiente, estos rectángulos se superponen en una de sus esquinas. Un conjunto de puntos generados según la distribución “cuadrados traslapados” se muestra en la figura 3.6. Los cuadrados que están sobrepuestos en la figura permiten visualizar el esquema de la distribución.

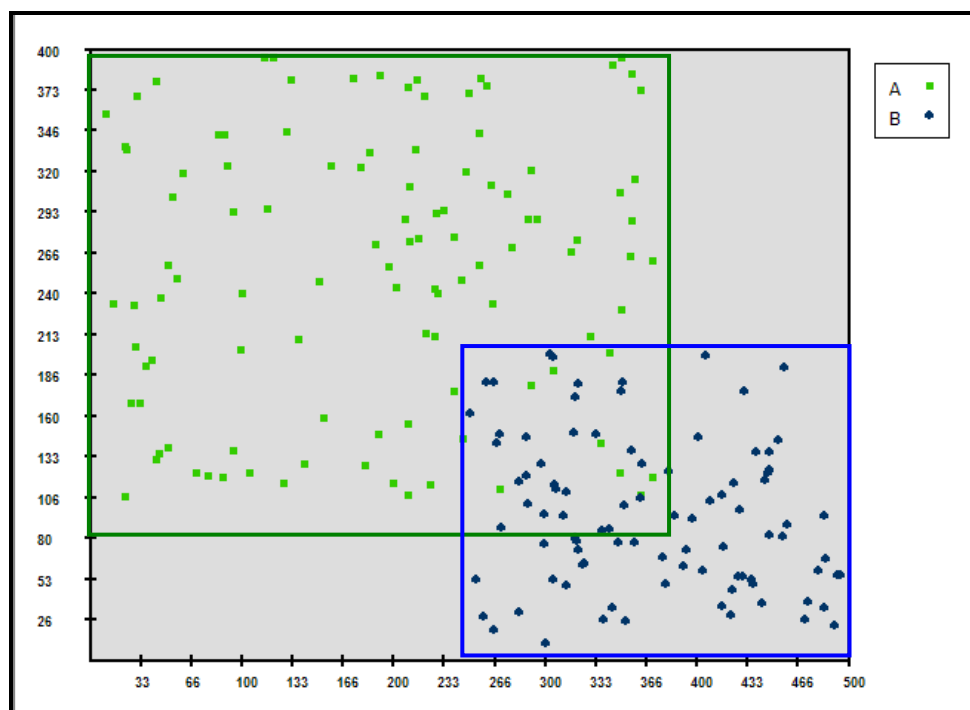


FIGURA 3.6 DISTRIBUCION DE CUADRADOS TRASLAPADOS

La última distribución que se ofrece se denomina “exponencial” porque tanto las variables X como Y de cada clase tienen función de densidad exponencial univariada $f(x)=(1/k)e^{-(x-L)/k}$ para $x \geq L$. La figura 3.7 muestra esta distribución aleatoria.

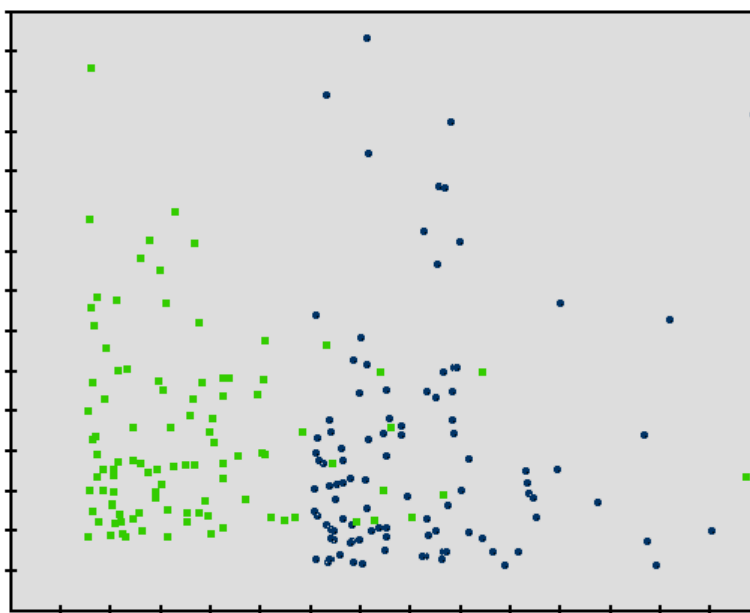


FIGURA 3.7 DISTRIBUCIÓN DE PUNTOS TIPO “EXPONENCIAL”

En esta página web el usuario tiene la posibilidad de desplazar cada punto generado para formar su propio conjunto de puntos y experimentar cualquier método. La figura 3.8 muestra uno de estos conjuntos personalizados.

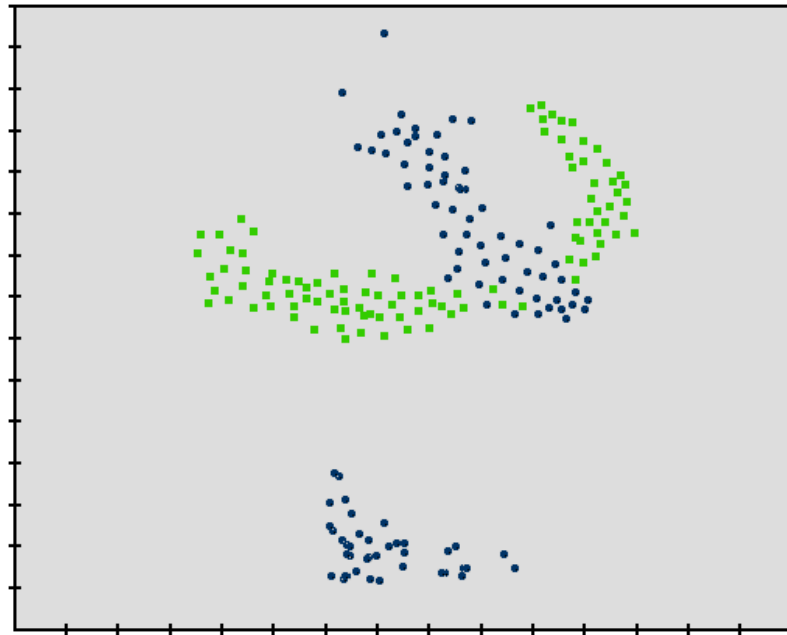


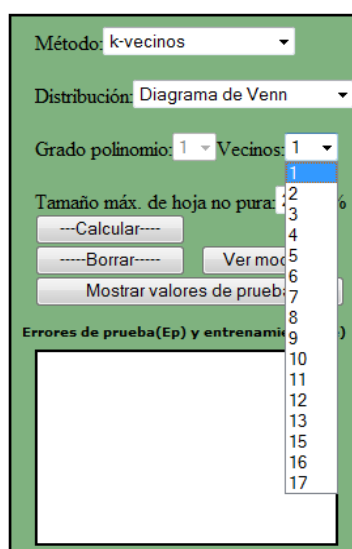
FIGURA 3.8 CONJUNTO DE DATOS PERSONALIZADO

3.2 Selección de parámetros

Cada método a excepción de Bayes Naive ofrece un parámetro el cual el usuario puede modificar antes de graficar.

El método de los kvecinos permite escoger la cantidad de vecinos, la figura 3.9 muestra esto. A su vez los métodos de regresión lineal y logística permiten escoger el grado del polinomio que utilizarán en su modelo. Por último, para el método del árbol de decisión se

permite escoger el porcentaje de elementos al cual no puede llegar un nodo para que éste pueda ser particionado, a esta opción se le ha llamado “Tamaño máximo de hoja no pura”, puesto que si una hoja no pura tuviese mayor número de elementos que el porcentaje fijado tuviese que ser particionada con lo cual dejaría de ser hoja. La figura 3.10 muestra la selección del parámetro de un árbol de decisión.



**FIGURA 3.9 SELECCIÓN DEL
NÚMERO DE VECINOS**

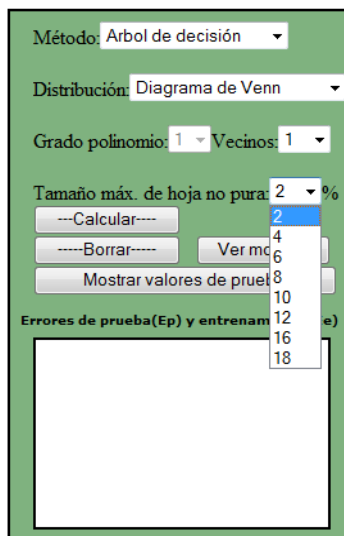


FIGURA 3.10 SELECCIÓN DEL PARÁMETRO DE PARADA DE UN ÁRBOL DE DECISIÓN

3.3 Botones de opción

El programa ofrece las opciones que se muestran en la figura 3.11. Después que el usuario ha seleccionado la distribución de puntos, el método a aplicar y el valor de su parámetro correspondiente puede presionar “Calcular” para que el método se ejecute y comience a graficarse la curva discriminante en el plano.

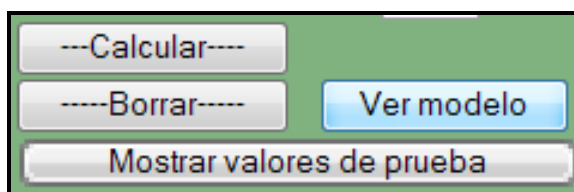


FIGURA 3.11 BOTONES DE OPCIÓN

Cada vez que el usuario presiona el botón “Calcular” se ejecuta el método que esté seleccionado en ese momento. Tal como lo muestra la figura 3.12 este aplicativo permite graficar varias curvas para discriminar el mismo conjunto de datos en el mismo plano diferenciándose éstas por su color.

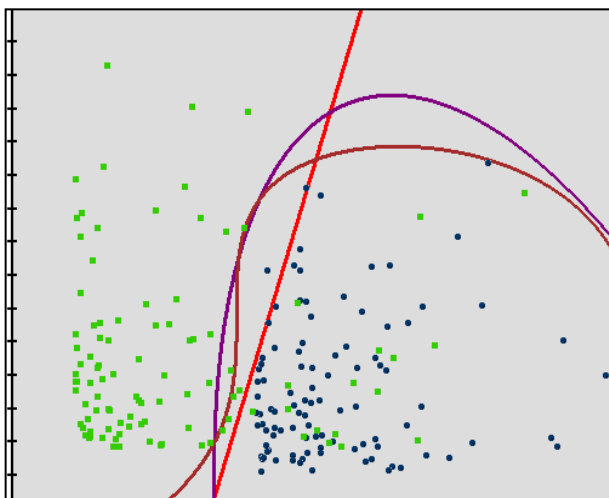


FIGURA 3.12 GRAFICACIÓN DE VARIAS CURVAS DISCRIMINANTES EN EL MISMO PLANO

Si el usuario ya no deseara ver los gráficos de los métodos anteriores bastará con presionar el botón “Borrar” que no elimina los

puntos sino los trazos de cada método. Luego de esto podrá si desea seguir graficando curvas discriminantes con el mismo conjunto de datos.

Si el usuario deseara probar otro conjunto de datos podrá hacerlo escogiendo en la lista de opciones de distribución de la cual se mencionó anteriormente. Si existiesen curvas discriminantes graficadas en el plano al escoger una nueva distribución éstas automáticamente se borrarán y se generará un nuevo conjunto de datos según la distribución seleccionada.

Los puntos que se visualizan en el plano son el conjunto de entrenamiento, en base a éstos se realiza la tarea de discriminación. El usuario podrá ver el conjunto de prueba en el cual se basa el error de prueba presionando el botón “Mostrar valores de prueba”. Las figuras 3.13 y 3.14 presentan la misma situación sin y con datos de prueba visualizados respectivamente. Se nota en la figura 3.14 que los datos de prueba podrán ocultarse presionando el mismo botón.

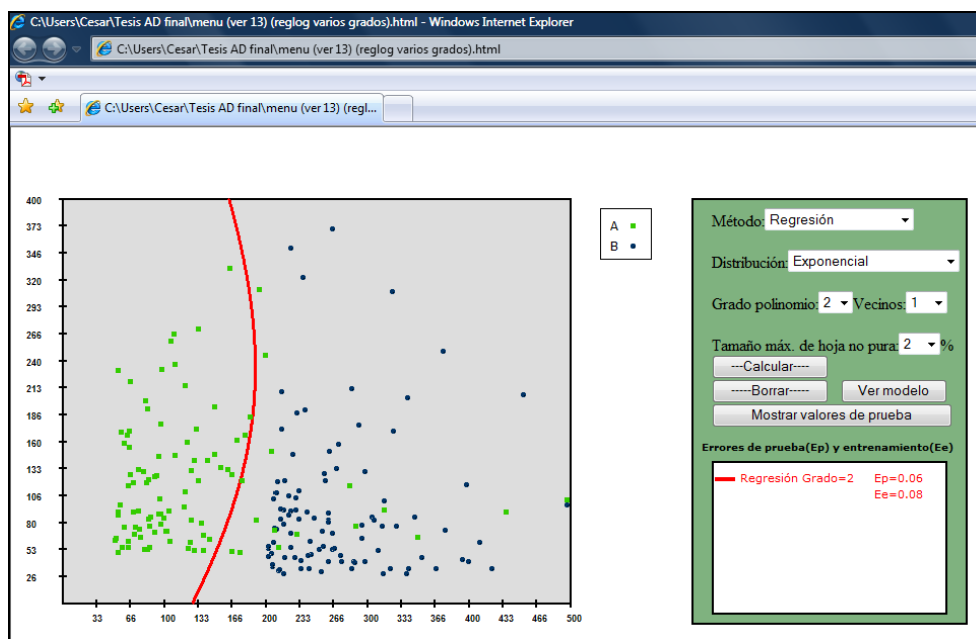


FIGURA 3.13 DISCRIMINACION SIN VISUALIZAR DATOS DE PRUEBA

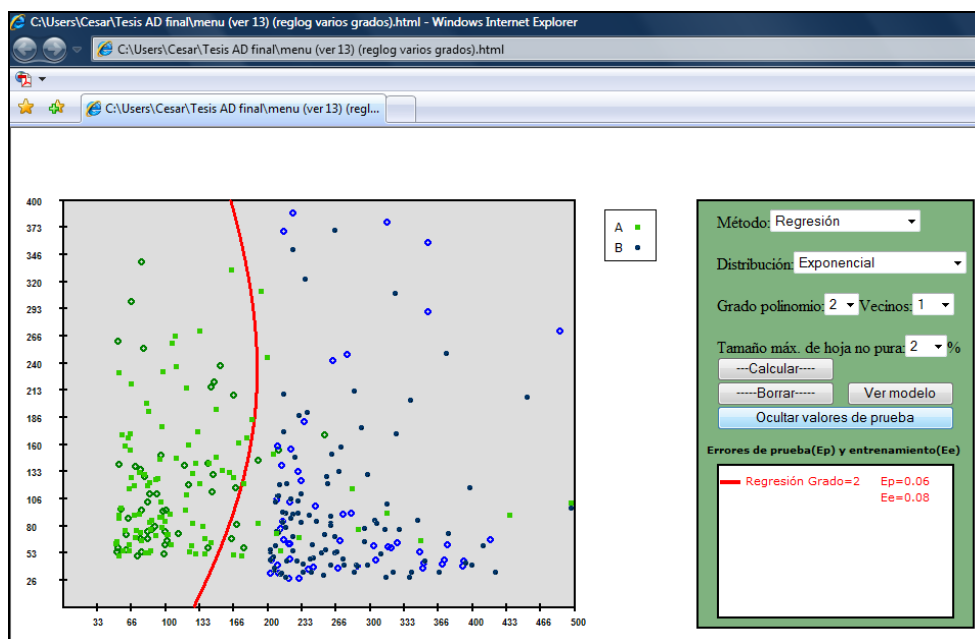


FIGURA 3.14 DISCRIMINACION VISUALIZANDO DATOS DE PRUEBA

Finalmente, el botón “Ver modelo” permitirá ver el modelo de aquellos métodos que son comprensibles, estos son: regresión lineal, regresión logística y árbol de decisión. La figura 3.15 muestra el modelo de la regresión de segundo grado aplicada en la figura 3.13.

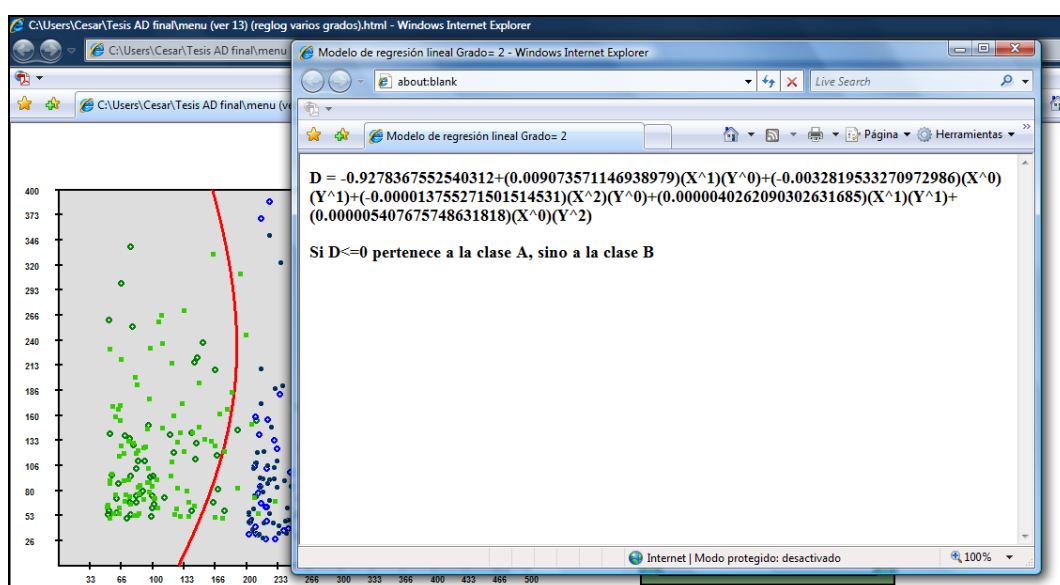


FIGURA 3.15 VISUALIZACIÓN DEL MODELO DE UN MÉTODO

3.4 Tablero de resultados

En el tablero de resultados es donde se muestran las respuestas analíticas de cada método, específicamente el error de prueba (E_p) y el error de entrenamiento (E_e). Este tablero puede contener más

de 1 método con sus respuestas aplicado al mismo conjunto de datos, estos resultados permanecerán registrados en el tablero hasta que el usuario borre las curvas discriminantes del plano. La figura 3.16 muestra las gráficas de diferentes curvas con el mismo conjunto de datos y el registro de sus errores de prueba y entrenamiento en el tablero de resultados (a la derecha de la figura).

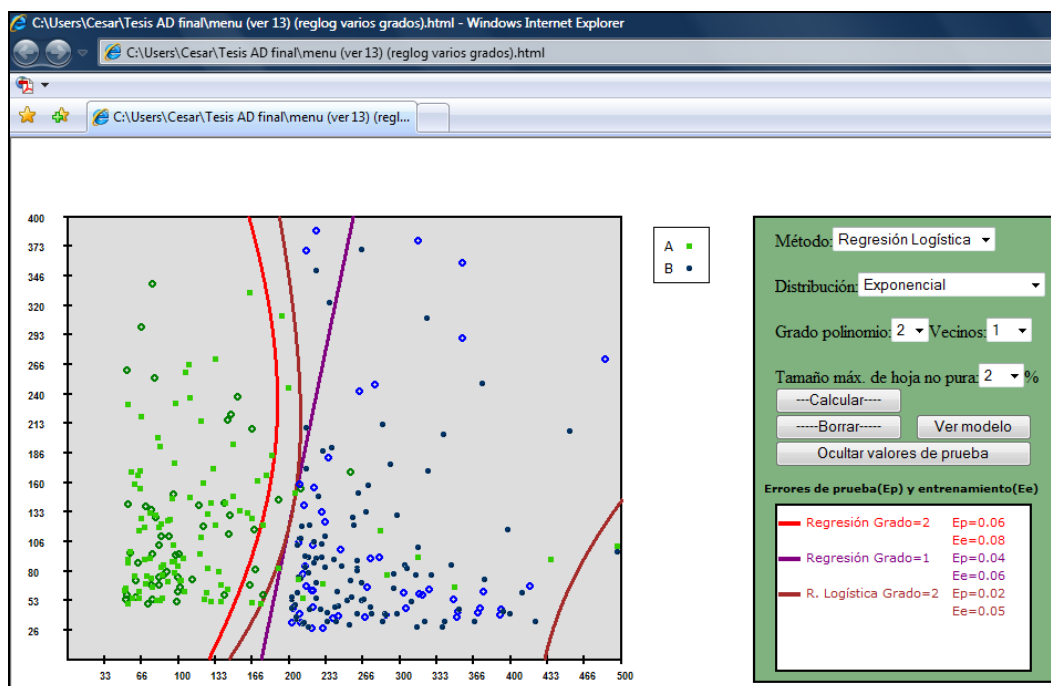


FIGURA 3.16 REGISTRO DE RESPUESTAS EN EL TABLERO DE RESULTADOS