

**UNIVERSIDAD LAICA ELOY ALFARO DE MANABI**

**FACULTAD DE CIENCIAS INFORMATICAS**

**CARRERA DE TECNOLOGIA DE LA INFORMACION**

**TEMA:**

SUPPORT VECTOR MACHINES

**NOMBRE:**

MACIAS PICO JOSSELYN STEFANY

**CURSO:**

SEXTO “B”

**MATERIA:**

MINERIA DE DATOS

**DOCENTE:**

ING. FABRICIO JAVIER RIVADENEIRA ZAMBRANO

**FECHA:**

14-06-2021

**MANTA-MANABI-ECUADOR**

## INDICE

INTRODUCCIÓN.....	3
SUPPORT VECTOR MACHINES.....	4
¿QUE ES?:.....	4
OBJETIVO: .....	4
¿COMO FUNCIONA? .....	4
VENTAJAS: .....	5
DESVENTAJAS: .....	6
HIPERPLANOS Y VECTORES DE SOPORTE .....	6
CODIFICACIÓN Y EJECUCIÓN DEL SCRIPT.....	7
Conclusiones.....	12
Bibliografías .....	13

## INTRODUCCIÓN

Una máquina de vectores de soporte son otro tipo de algoritmo supervisado aplicable a problemas de regresión y clasificación. Aunque es más usado como modelo de clasificación. Además de ser un modelo de aprendizaje automático supervisado que utiliza algoritmos de clasificación para problemas de clasificación de dos grupos. Después de proporcionar un modelo de SVM con conjuntos de datos de entretenimiento etiquetados para cada categoría.

Estas maquinas de vector soporte suponen una generalización de un clasificador simple denominado, los SVM. El objetivo de este algoritmo es encontrar un hiperplano que separa de la mejor forma posible de dos clases diferentes de puntos de datos. Estos vectores de soporte hacen referencia a un subconjunto de las observaciones de entrenamiento que identifican la ubicación del hiperplano de separación.

Support Vector machines pertenecen a una clase de algoritmos de Machine Learning denominados métodos de karnel, donde se puede utilizar una función de karnel para transformar las características. En este proceso los datos no se tienen que transformar explícitamente, lo que supondría una alta carga computacional.

## SUPPORT VECTOR MACHINES

### ¿QUE ES?:

Support Vector Machine (SVM) es un algoritmo de aprendizaje supervisado que se utiliza en muchos problemas de clasificación y regresión, incluidas aplicaciones medicas de procesamiento de señales, procesamiento lenguaje natural y reconcomiendo de imágenes y voz.

### OBJETIVO:

El objetivo del algoritmo SVM es encontrar un hiperplano que separe de la mejor forma posible dos clases diferentes de puntos de datos: “De la mejor forma posible” implica el hiperplano con el margen mas amplio entre las dos clases, representado por los signos mas y menos.

Los vectores de soporte hacen referencia a un subconjunto de las observaciones de entrenamiento que identifican la ubicación del hiperplano de separación. EL algoritmo SVM estándar esta formulado para problemas de clasificación binaria.

### ¿COMO FUNCIONA?

Se debe identificar el hiperplano correcto, en la imagen 1 vemos tres hiperplanos (A, B y C). En este el hiperplano B ha realizado este trabajo de manera excelente.

En el siguiente escenario imagen 2 también hay un hiperplano correcto el cual en esta parte se maximiza las distancias entre el punto de datos mas cercano y el hiperplano nos ayudara a decidir el hiperplano correcto a la cual distancia se llama **Margen**.

En la imagen 2 se puede observar que el margen para el hiperplano C es alto en comparación con A y B. Por lo tanto, nombramos el hiperplano derecho

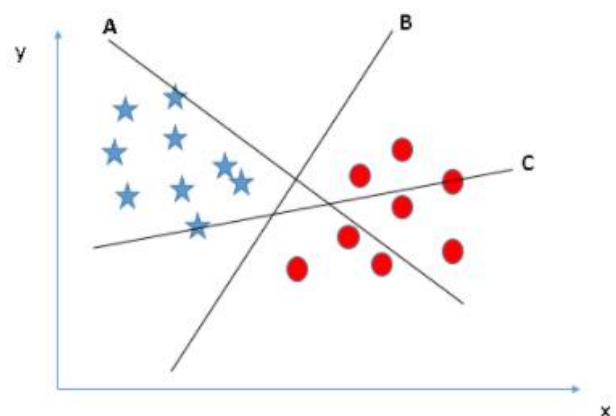
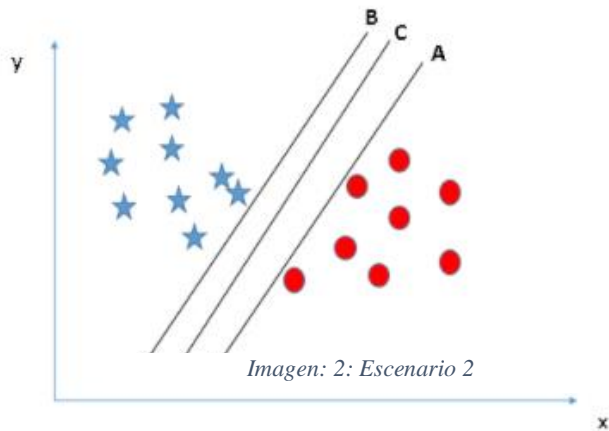


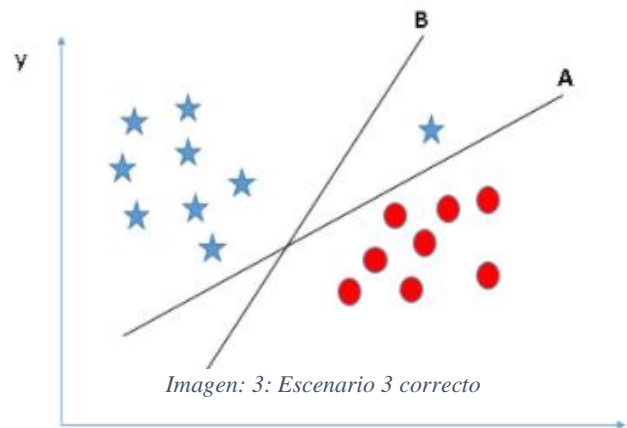
Imagen: 1: "hiperplano correcto"



como C. El plano con mayor margen es la robustez. Si seleccionamos un hiperplano que tiene un margen bajo, existe una gran posibilidad de clasificación errónea.

Como tercer escenario en la imagen 3 use las reglas discutidas anteriormente para identificar el hiperplano correcto. Algunos de ustedes pueden

haber seleccionado el hiperplano B ya que tiene un margen más alto en comparación con A. Pero, aquí está el truco, SVM selecciona el hiperplano que clasifica las clases con precisión antes de maximizar el margen. Aquí, el hiperplano B tiene un error de clasificación y A lo ha clasificado todo correctamente. Por lo tanto, el hiperplano derecho es A. (RAY, 2017)



## VENTAJAS:

- Son muy populares y logran un buen rendimiento en muchas tareas de clasificación y regresión.
- Aunque los algoritmos SVM están formulados para la clasificación binaria, los algoritmos SVM multiclase se construyen combinando varios clasificadores binarios.
- Los kernels hacen que los SVM sean más flexibles y capaces de gestionar problemas no lineales.
- Para construir la superficie de decisión, solo se requieren los vectores de soporte seleccionados a partir de los datos de entrenamiento. Una vez terminado el entrenamiento, el resto de los datos de entrenamiento es irrelevante, produciendo una representación compacta del modelo que es adecuada para generar código de forma automatizada. (MathWorks)

## DESVENTAJAS:

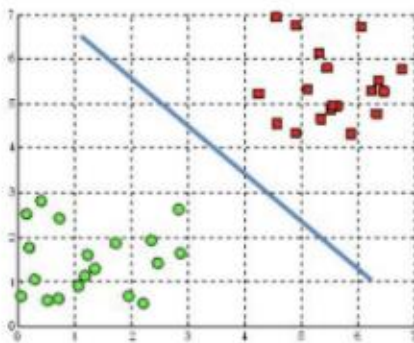
- No funciona bien cuando tenemos un gran conjunto de datos porque el tiempo de entrenamiento requerido es mayor
- Tampoco funciona muy bien cuando el conjunto de datos tiene más ruido, es decir, las clases de destino se superponen
- SVM no proporciona directamente estimaciones de probabilidad, estas se calculan mediante una costosa validación cruzada de cinco veces. Está incluido en el método SVC relacionado de la biblioteca scikit-learn de Python.

## HIPERPLANOS Y VECTORES DE SOPORTE

Los hiperplanos son límites de decisión que ayudan a clasificar los puntos de datos. Estos que caen a ambos lados del hiperplano se pueden atribuir a otras clases. Además, la dimensión del hiperplano depende del número de características. Si el número de entidades de entrada es 2, entonces el hiperplano es solo una línea. Si el número de entidades de entrada es 3, entonces el hiperplano se convierte en un plano bidimensional.

Los puntos de datos que caen a ambos lados del hiperplano se pueden atribuir a diferentes clases. Además, la dimensión del hiperplano depende del número de características. Si el número de entidades de entrada es 2, entonces el hiperplano es solo una línea. (Gandhi, 2018)

A hyperplane in  $\mathbb{R}^2$  is a line



A hyperplane in  $\mathbb{R}^3$  is a plane

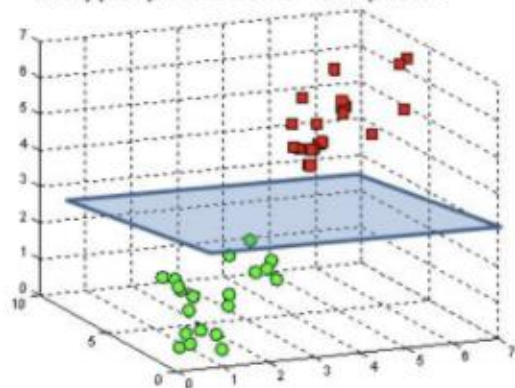
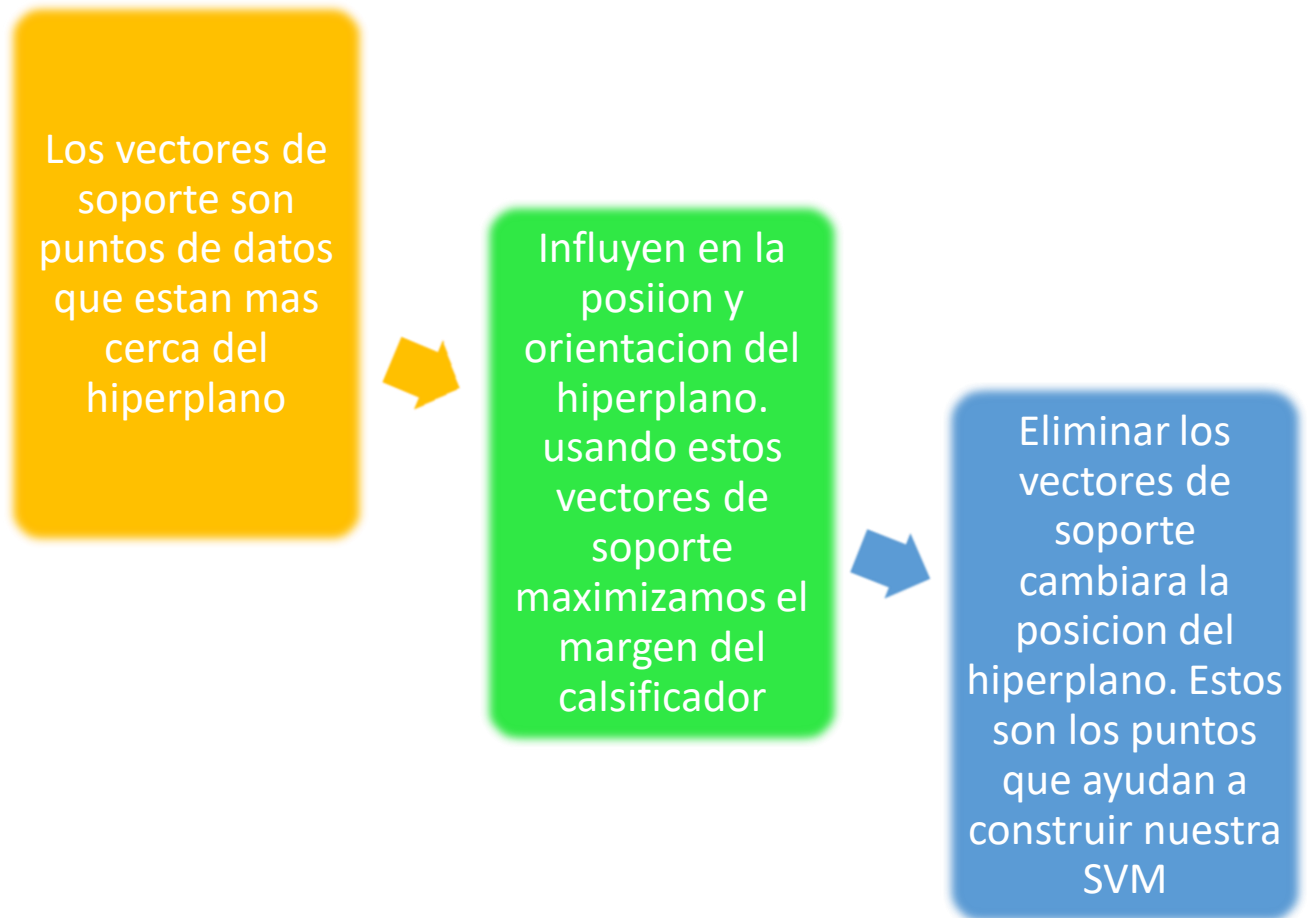


Imagen: 2 Hyperplanes in 2D and 3D



La SVM es un algoritmo supervisado puede ser de clasificación y regresión ayuda a resolver problemas de clasificación o regresión este algoritmo trata de buscar el hiperplano que separe lo mejor posible las clases es decir con el mayor margen posible. Originalmente se ideo para resolver problemas binarios, pero puede resolver problemas con mas de dos clases. (Suarez, 2013)

## **CODIFICACIÓN Y EJECUCIÓN DEL SCRIPT**

Ejemplo de clasificación entre gato y gata con el peso completo y el peso del corazón.



Imagen: 3: Ejercicio a resolver

1. Instalar la librería **e1071** esto se realiza en el apartado de **Packages** y luego en **Install** se ubica el nombre del paquete y se recurre a instalar.

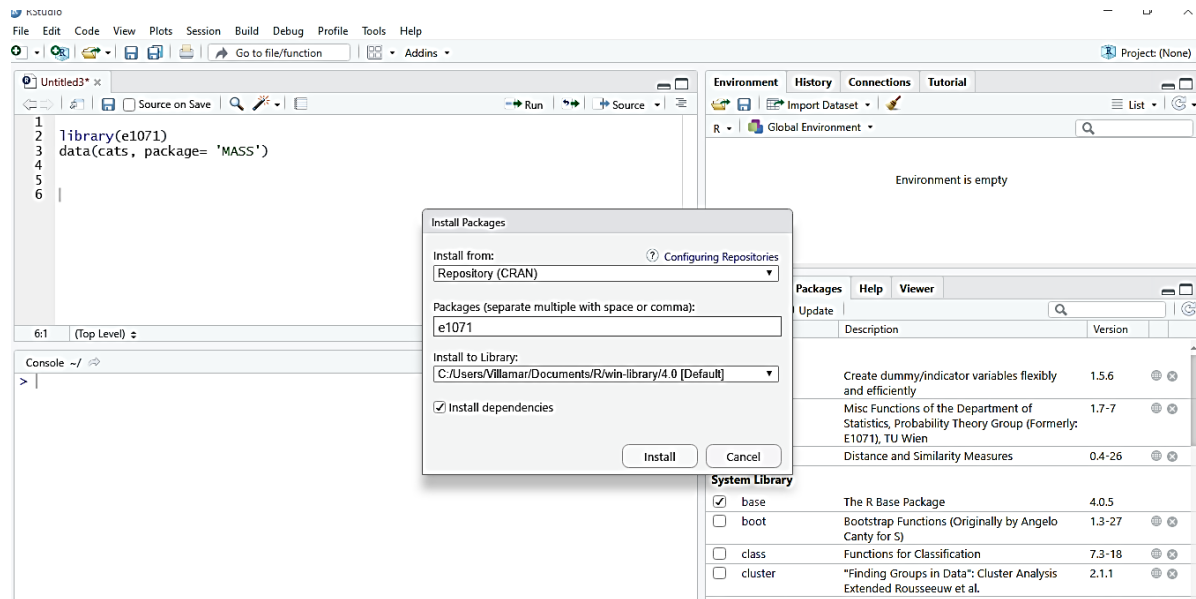


Imagen: 4: Instalación del paquete e1071

2. Se espera que se termine la instalación y se ejecuta la siguiente línea de código **data(cats, package= 'MASS')** si nos damos cuenta en el apartado de la parte derecha se mostrara en valores la variable cats. Por lo que ya estará guardados todos los datos de los gatos.



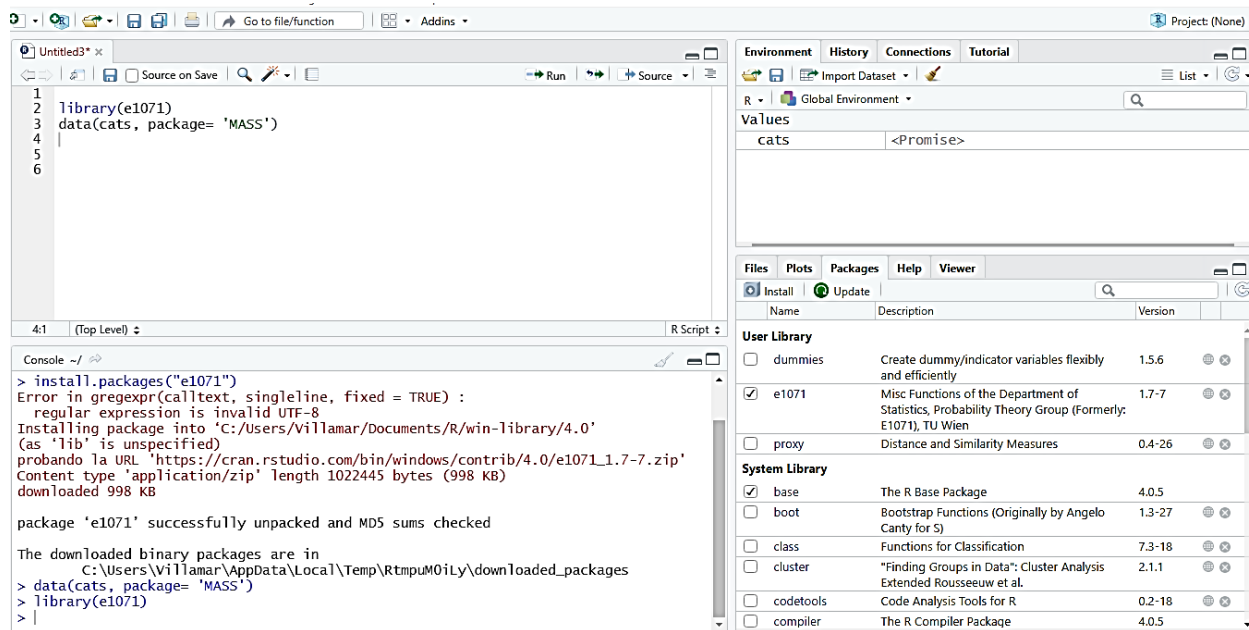


Imagen: 7: Después de instalar paquete

- A continuación, se ejecutará solamente la variable **cats** en la cual se mostraran todos los datos de los gatos, por lo que se mostraran todos los datos disponibles.

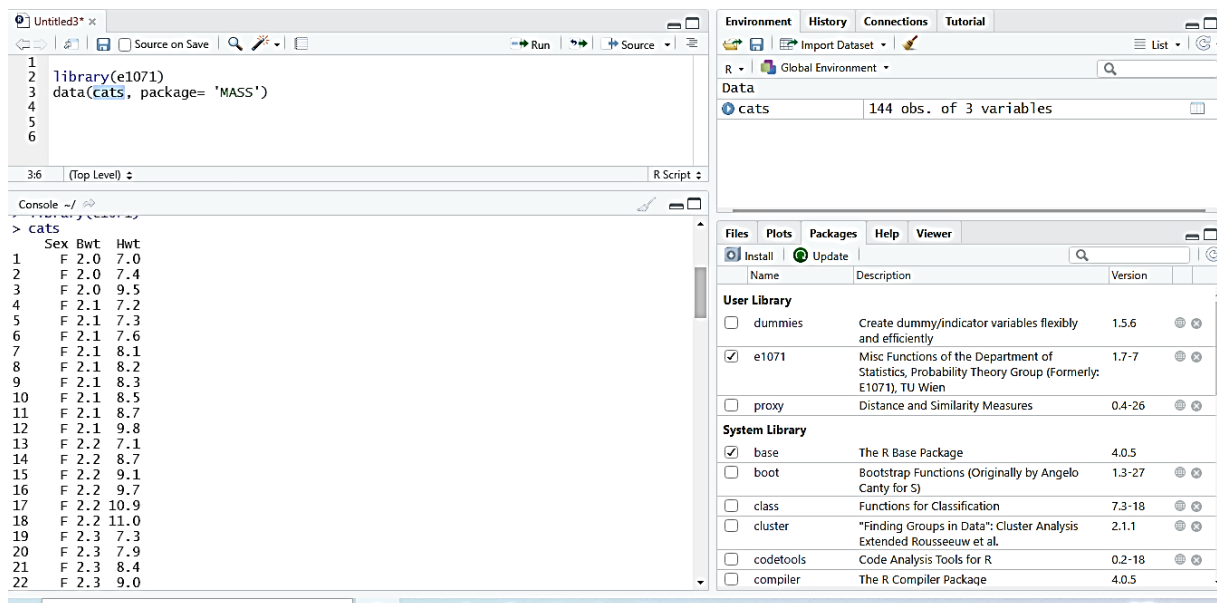


Imagen: 8: Datos de los gatos

- Se crea el conjunto de datos de enrutamiento y de test con un 70 % de las muestras para enrutamiento y las otras 30 % para test por tal razón se le incluye **prob 0.7** y **0.3** como se muestra en la siguiente imagen.

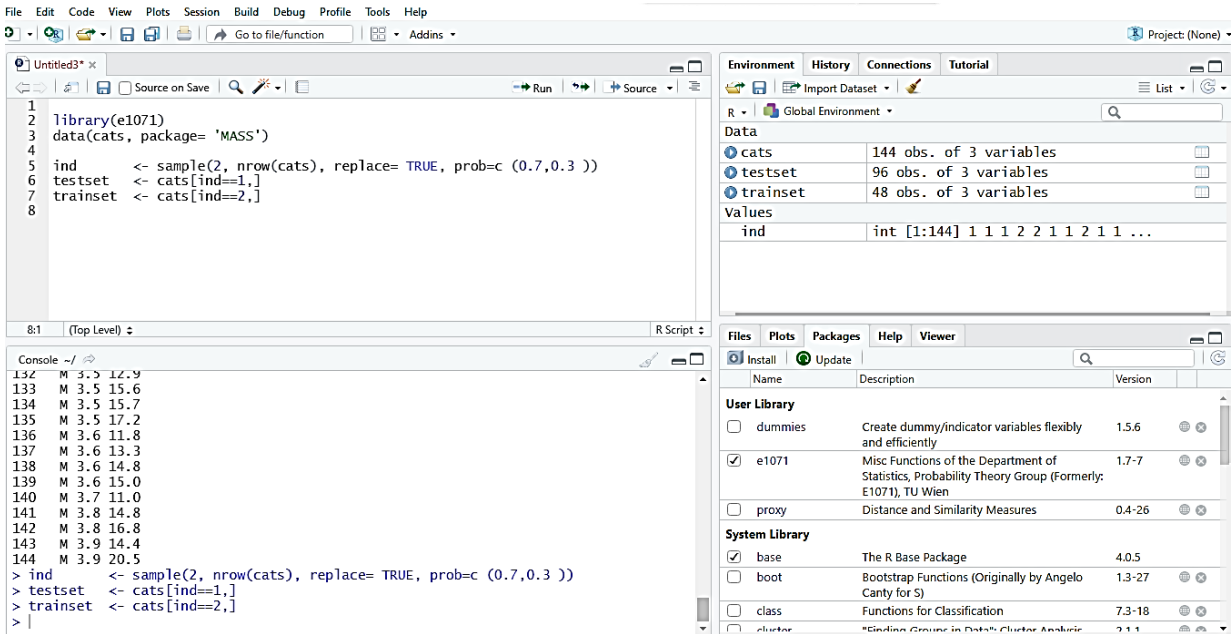


Imagen: 5: Conjunto de datos y entrentamiento

- Luego se crea el modelo de máquina de soporte vectorial donde la variable dependiente será sexo y las otras dos serán las variables independientes por defecto tendremos que el **kernel** podría ser **radial**, **lineal**, **polinomial**. Luego se crea una predicción con el modelo que se ha entrenado y el juego de datos de texto sin la variable dependiente.

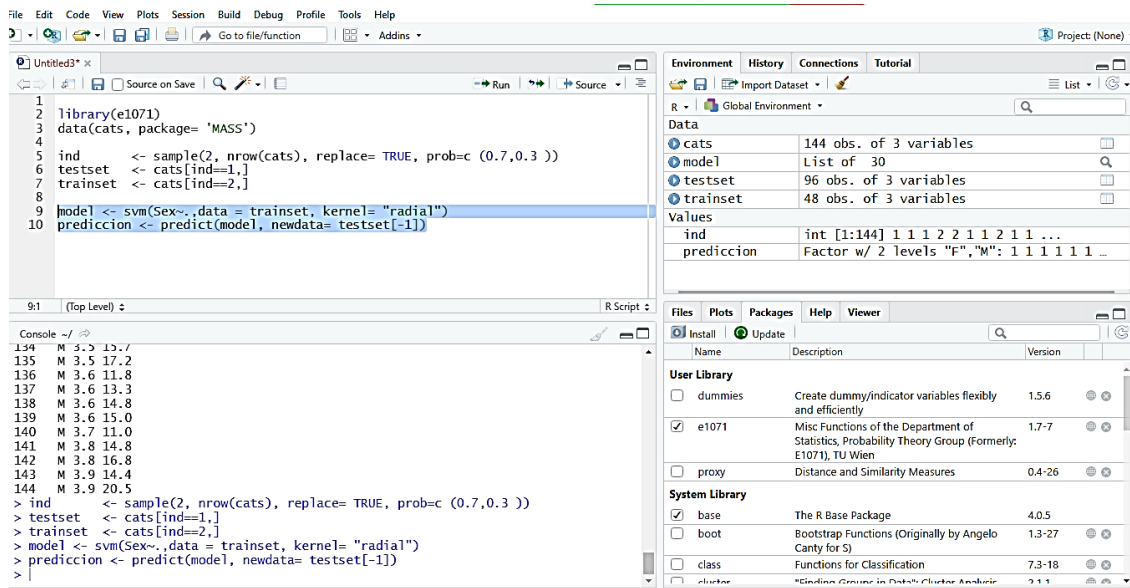


Imagen: 6: Modelo y Predicción

- Luego mostramos el resultado de la matriz de confusión.

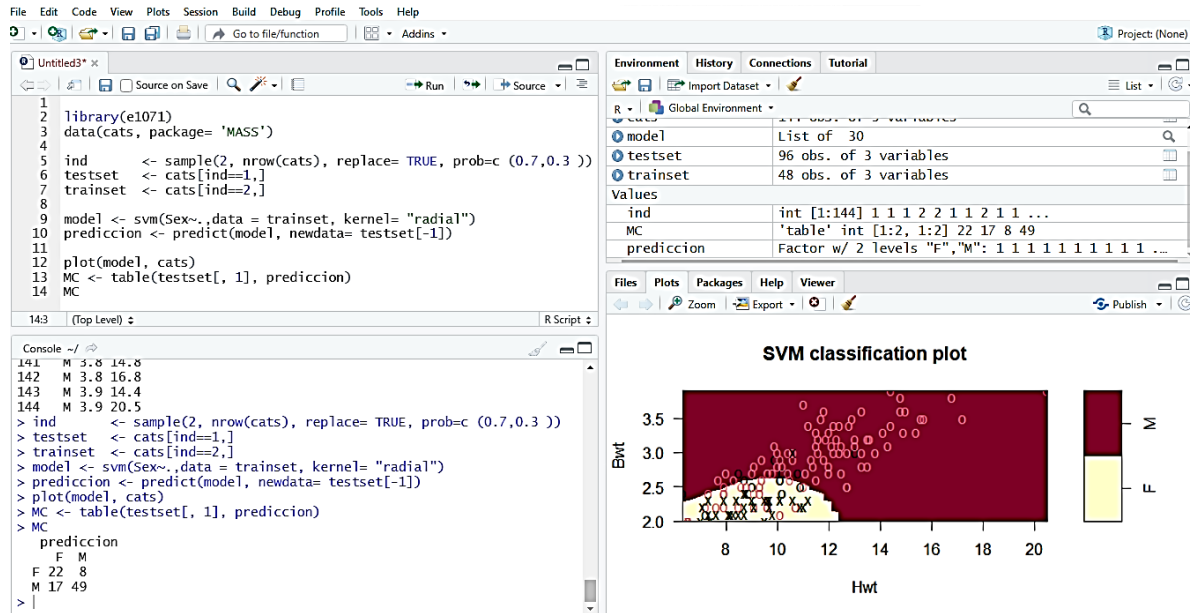


Imagen: 7: Matriz de confusión

7. Finalmente, mostramos el valor del acierto donde se va a hacer una suma entre **diag(MC)** dividido para a suma de **MC** y por ultimo mostramos el acierto.

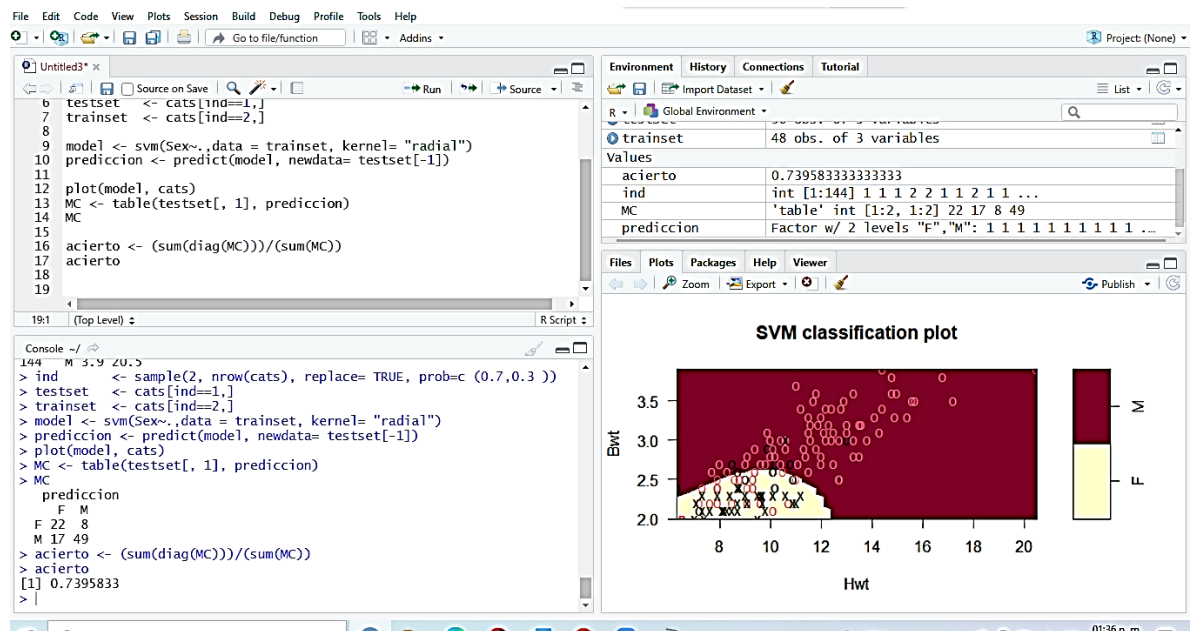


Imagen: 8: Acierto

## Conclusiones

Con el avance de las nuevas tecnologías y las posibilidades de inversión, los métodos estadísticos o de aprendizaje automático, antes reservados.

Se estudió el método de máquinas de vectores de soporte como alternativa a los modelos de regresión logística conservadores y se comparó su desempeño con los conjuntos de datos de crédito real. Especialmente en combinación con el kernel no lineal, SVM demostró ser un enfoque competitivo y proporcionó una ligera ventaja sobre el modelo de regresión logística.

A pesar de todos los hechos en contra, las máquinas de vectores de soporte siguen siendo un concepto importante desde el punto de vista educativo y teórico. También formaron una historia del aprendizaje automático, ya que fue el primer método que pudo competir con los humanos en el reconocimiento de los números escritos a mano e inspiraron muchas investigaciones posteriores. Sin embargo, su uso en la calificación crediticia no está exento de problemas y no puede recomendarse en aplicaciones reales, a menos que otro avance importante aumente aún más su rendimiento o confiabilidad.

## Bibliografías

Gandhi, R. (07 de Jun de 2018). *Support Vector Machine — Introduction to Machine Learning Algorithms*. Obtenido de <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>

*MathWorks*. (s.f.). Obtenido de <https://es.mathworks.com/discovery/support-vector-machine.html>

RAY, S. (13 de 09 de 2017). *Analytics Vidhya*. Obtenido de <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/?#>

Suarez, E. J. (17 de 11 de 2013). *Tutorial sobre maquinas de vectores soporte (SVM)*. Obtenido de Cartagena99: [https://www.cartagena99.com/recursos/alumnos/apuntes/Tema8.\\_Maquinas\\_de\\_Vectores\\_Soporte.pdf](https://www.cartagena99.com/recursos/alumnos/apuntes/Tema8._Maquinas_de_Vectores_Soporte.pdf)