**UNIVERSIDAD LAICA ELOY ALFARO DE MANABI**

**FACULTAD DE CIENCIAS INFORMATICAS**

**CARRERA DE TECNOLOGIA DE LA INFORMACION**

**TEMA:**

U2, DATA MINING

**NOMBRE:**

MACIAS PICO JOSSELYN STEFANY

**CURSO:**

SEXTO “B”

**MATERIA:**

MINERIA DE DATOS

**DOCENTE:**

ING. FABRICIO JAVIER RIVADENEIRA ZAMBRANO

**FECHA:**

30-06-2021

**MANTA-MANABI-ECUADOR**

Índice

[Introducción 3](#_Toc75446819)

[Diferencia entre algoritmo “supervisado” y “no supervisado” 4](#_Toc75446820)

[Algoritmo Supervisado 4](#_Toc75446821)

[Algoritmo no supervisado 4](#_Toc75446822)

[Técnica o un algoritmo de Aprendizaje NO Supervisado 5](#_Toc75446823)

[Codificación en lenguaje R 5](#_Toc75446824)

[Conclusiones 10](#_Toc75446825)

[Bibliografías 11](#_Toc75446826)

# Introducción

Dada la disponibilidad de datos sin precedentes y recursos informáticos, existe una amplia renovación en aplicar métodos de aprendizaje automático basados ​​en datos a problemas para los que el desarrollo de métodos convencionales. Las soluciones de ingeniería se ven desafiadas por el modelado o las deficiencias algorítmicas. Este documento proporciona una introducción de alto nivel a los conceptos básicos del aprendizaje supervisado y no supervisado. Ejemplificándolas aplicaciones a las redes de comunicaciones son discutidas por distinguir las tareas realizadas en el borde y en el segmento de la nube de la red en diferentes capas de protocolos, con énfasis en la capa física.

El algoritmo de maximización de expectativas (EM), y Q-learning, con una serie de algoritmos modernos avances, incluidas nuevas técnicas de regularización y Horarios de ritmo de aprendizaje adaptativo. Si el éxito se basa en la disponibilidad de datos sin precedentes y recursos informáticos en muchos dominios de la ingeniería. Mientras la nueva ola de promesas y avances en torno al aprendizaje automático podría decirse que se queda corto, al menos para ahora, de los requisitos que impulsaron la investigación temprana de IA, los algoritmos de aprendizaje han demostrado ser útiles en una serie de aplicaciones importantes, y más ciertamente en camino.

La presentación está organizada en torno a la descripción de conceptos técnicos generales. Posteriormente se proporciona redes.

# Diferencia entre algoritmo “supervisado” y “no supervisado”

## Algoritmo Supervisado

En el aprendizaje supervisado, los algoritmos trabajan con datos etiquetados, intentado encontrar una función que, dadas las variables de entrada, les asigne la etiqueta de salida adecuada. El algoritmo se entrena con un histórico de datos y así aprende a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida. (Simeone, 2018)

Por ejemplo, un detector de spam analiza el histórico de mensajes, viendo qué función puede representar, según los parámetros de entrada que se definan el remitente, si el destinatario es individual o parte de una lista, si el asunto contiene determinados términos etc, la asignación de la etiqueta “spam” o “no es spam”. Una vez definida esta función, al introducir un nuevo mensaje no etiquetado, el algoritmo es capaz de asignarle la etiqueta correcta. (Santos, 2017)

El aprendizaje supervisado se suele usar en:

* Problemas de clasificación: identificación de dígitos, diagnósticos, o detección de fraude de identidad.
* Problemas de regresión: predicciones meteorológicas, de expectativa de vida, de crecimiento etc.

Estos dos tipos principales de aprendizaje supervisado, clasificación y regresión, se distinguen por el tipo de variable objetivo. En los casos de clasificación, es de tipo categórico, mientras que, en los casos de regresión, la variable objetivo es de tipo numérico.

En estos se aprenden funciones, relaciones que asocian entradas con salidas, por lo que se ajustan a un conjunto de ejemplos de los que conocemos la relación entre la entrada y la salida deseada. Este hecho incluso llega a proporcionar una de las clasificaciones más habituales en el tipo de algoritmos que se desarrollan, así, dependiendo del tipo de salida, suele darse una subcategoría que diferencia entre modelos de clasificación, si la salida es un valor categórico (por ejemplo, una enumeración, o un conjunto finito de clases), y modelos de regresión, si la salida es un valor de un espacio continuo.

## Algoritmo no supervisado

Los modelos de aprendizaje no supervisado son aquellos en los que no estamos interesados en ajustar pares de entrada y salida, sino en aumentar el conocimiento estructural de los datos disponibles, como posibles datos futuros que provengan del mismo fenómeno, por ejemplo, dando una agrupación de los datos según su similaridad (clustering), simplificando las estructura de los mismos manteniendo sus características fundamentales (como en los procesos de reducción de la dimensionalidad), o extrayendo la estructura interna con la que se distribuyen los datos en su espacio original (aprendizaje topológico).

Muchos de los algoritmos no supervisados se reservaban para tareas de preprocesamiento de datos integrados en metodologías más amplias. Este hecho se debe, principalmente, a una cadena de factores.

## Diferencia

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.

## Técnica o un algoritmo de Aprendizaje NO Supervisado

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

## Codificación en lenguaje R

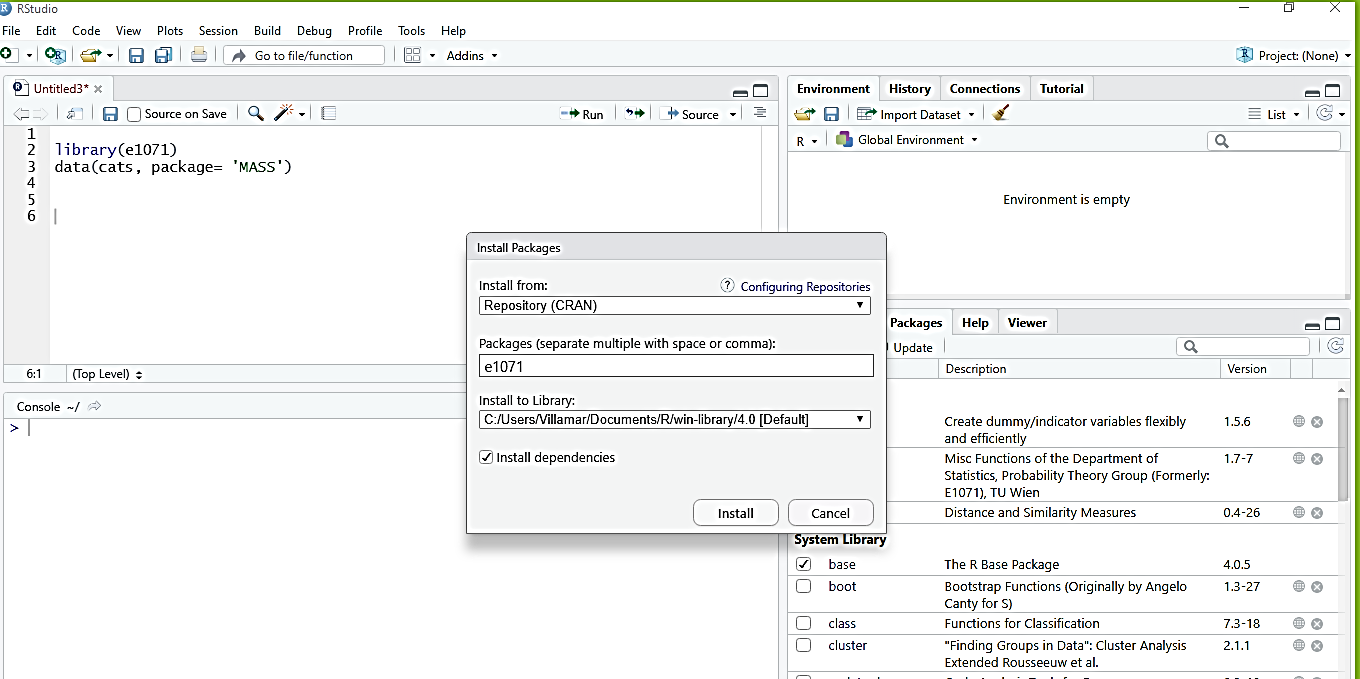
1. Instalar la librería **e1071** esto se realiza en el apartado de **Packages** y luego en **Install** se ubica el nombre del paquete y se recurre a instalar.

Imagen: 1: Instalación del paquete e1071

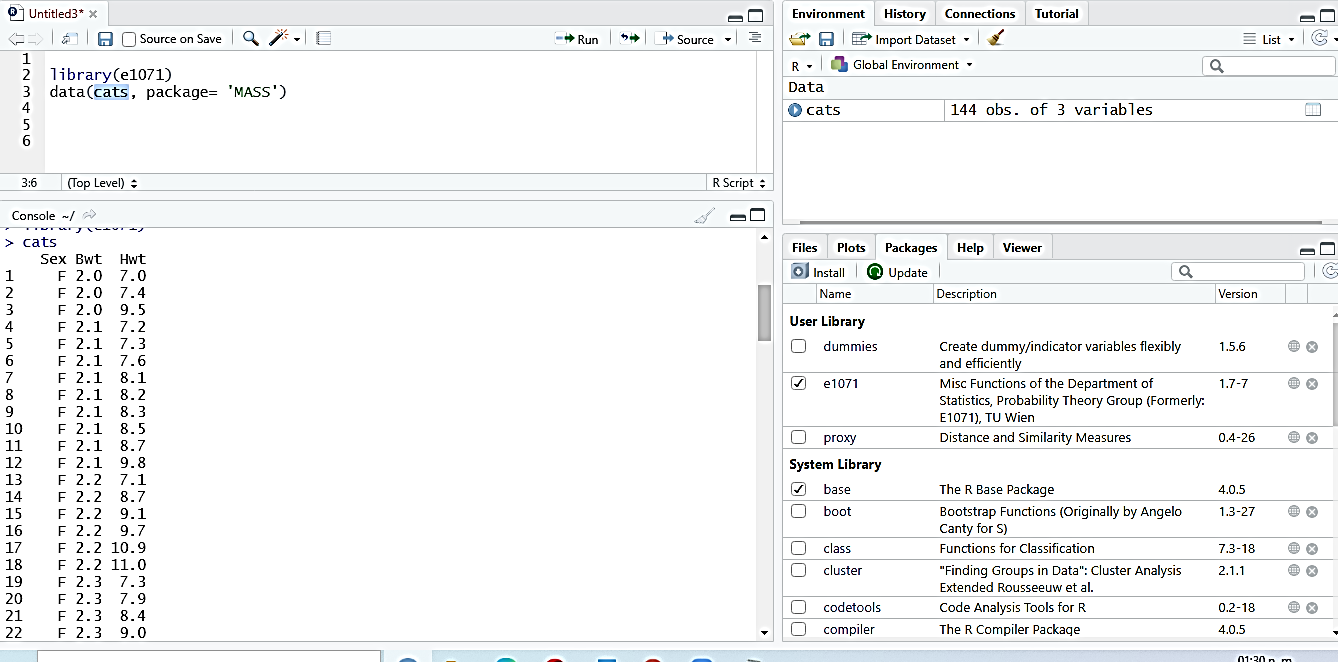
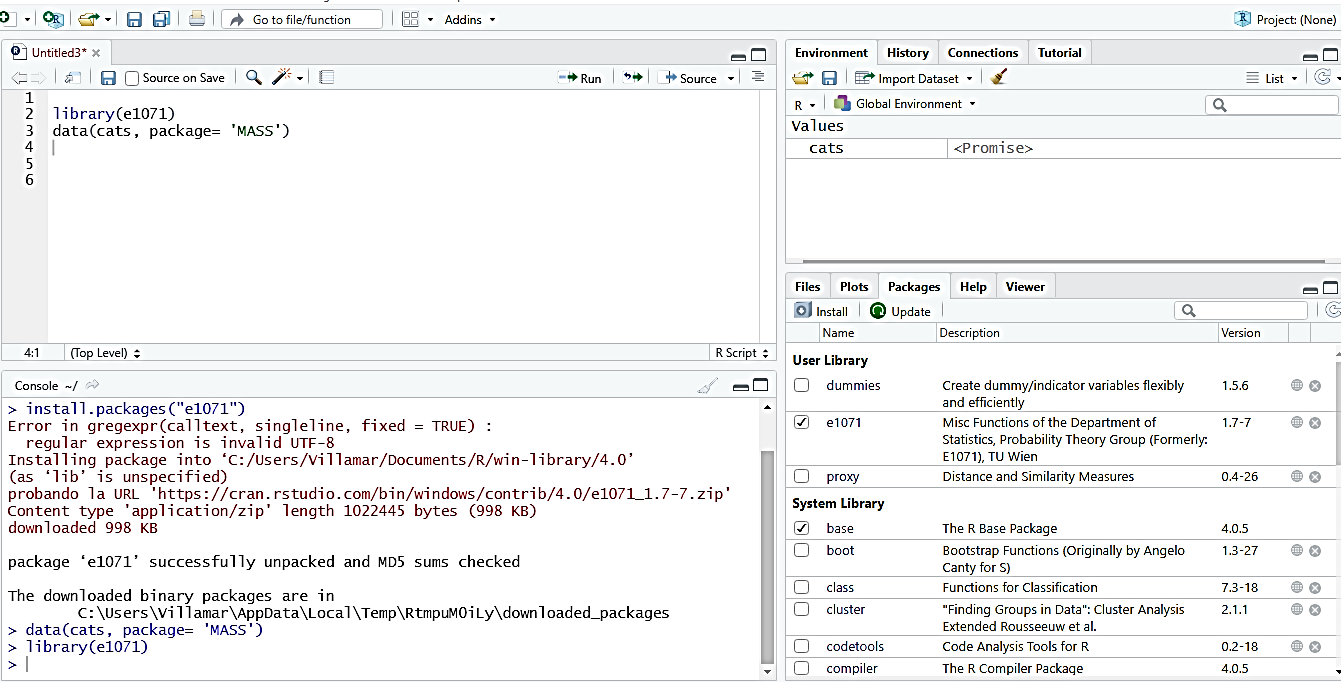
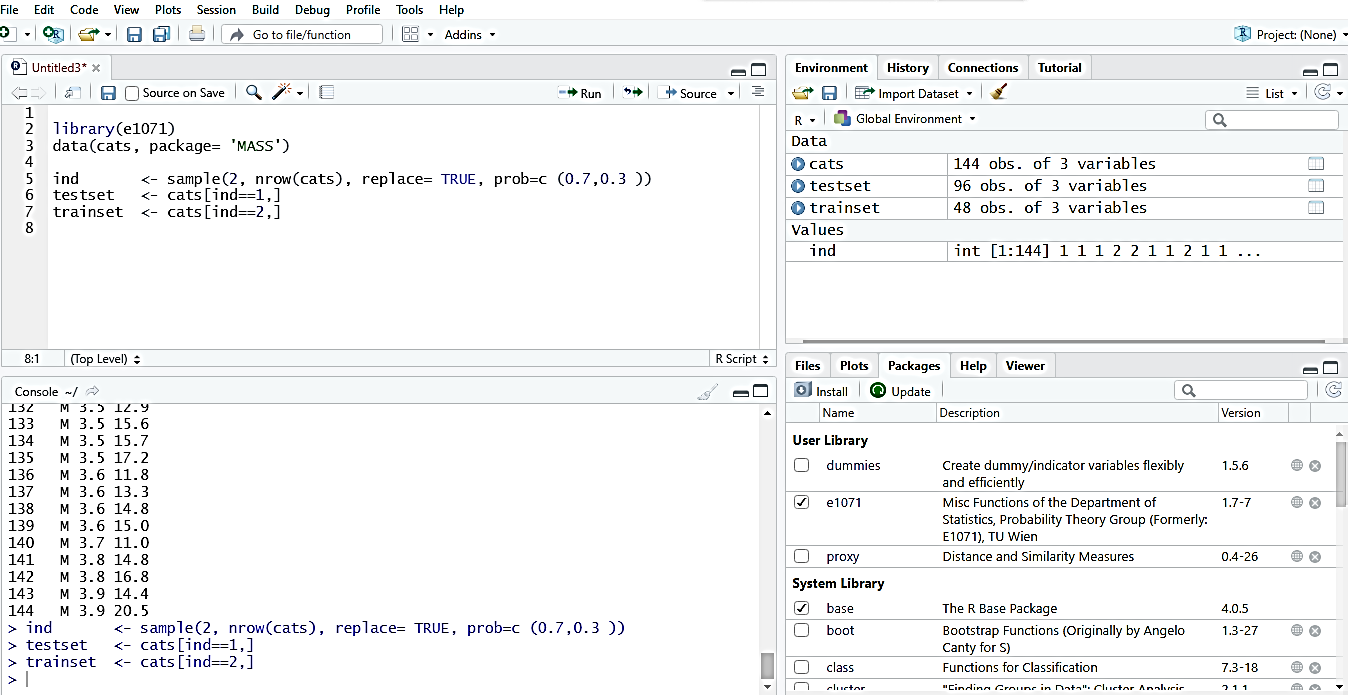
1. Se espera que se termine la instalación y se ejecuta la siguiente línea de codigo **data(cats, package= ‘MASS’)** si nos damos cuenta en el apartado de la parte derecha se mostrara en values la variable cats. Por lo que ya estará guardados todos los datos de los gatos.
2. A continuación, se ejecutará 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

Imagen: 7: Después de instalar paquete

1. 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: 2: Conjunto de datos y enrutamiento

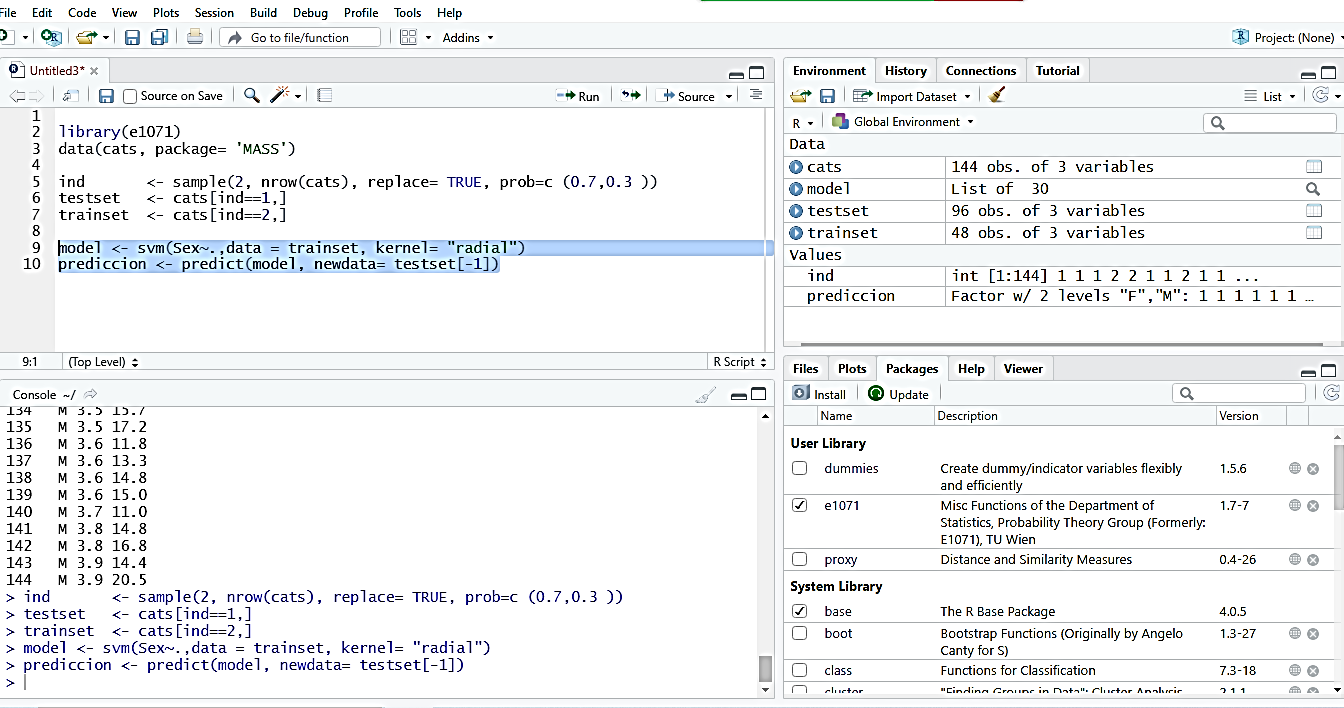
1. Luego se crea el modelo de máquina de soporte vectorial donde la variable dependiente será sexo y las otras dos serán las variables independientes por defecto tendremos que el **kernel** podría ser **radial, lineal, polinomial.** Luego se crea una predicción con el modelo que se ha entrenado y el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

Imagen: 3: Modelo y Predicción

1. Luego mostramos el resultado de la matriz de confusión.

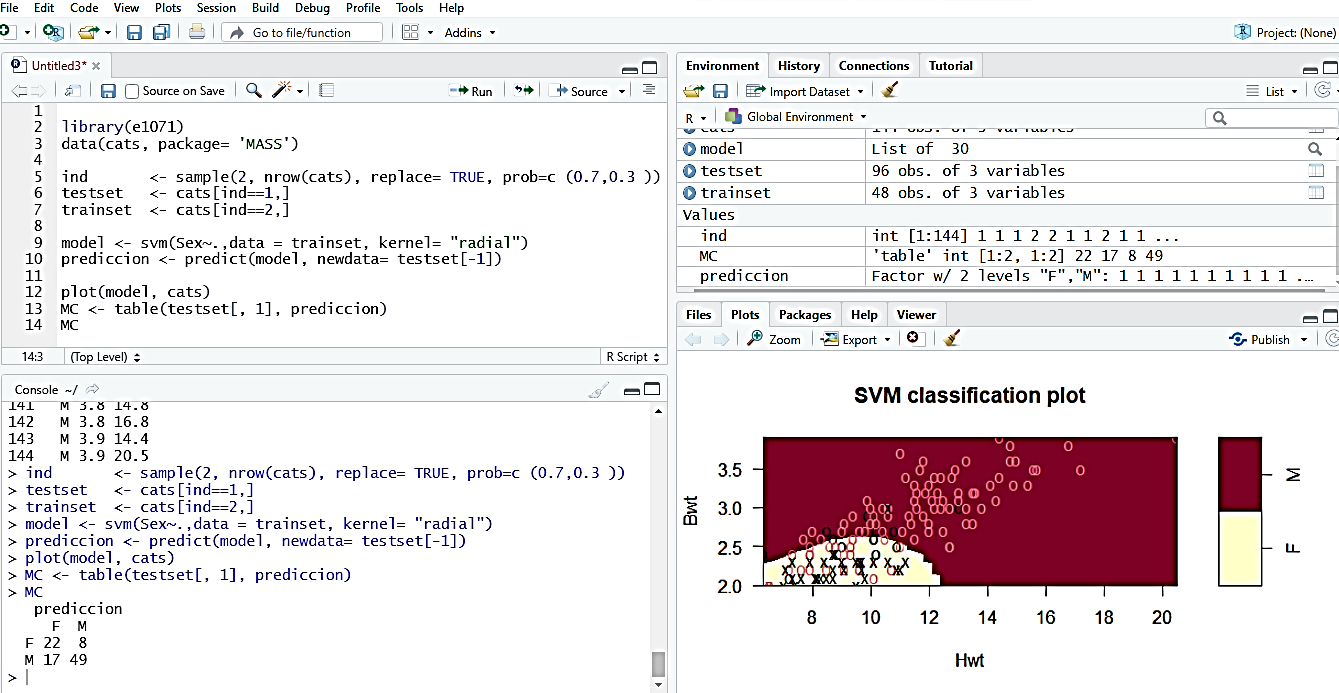


Imagen: 4: Matriz de confusión

1. 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.

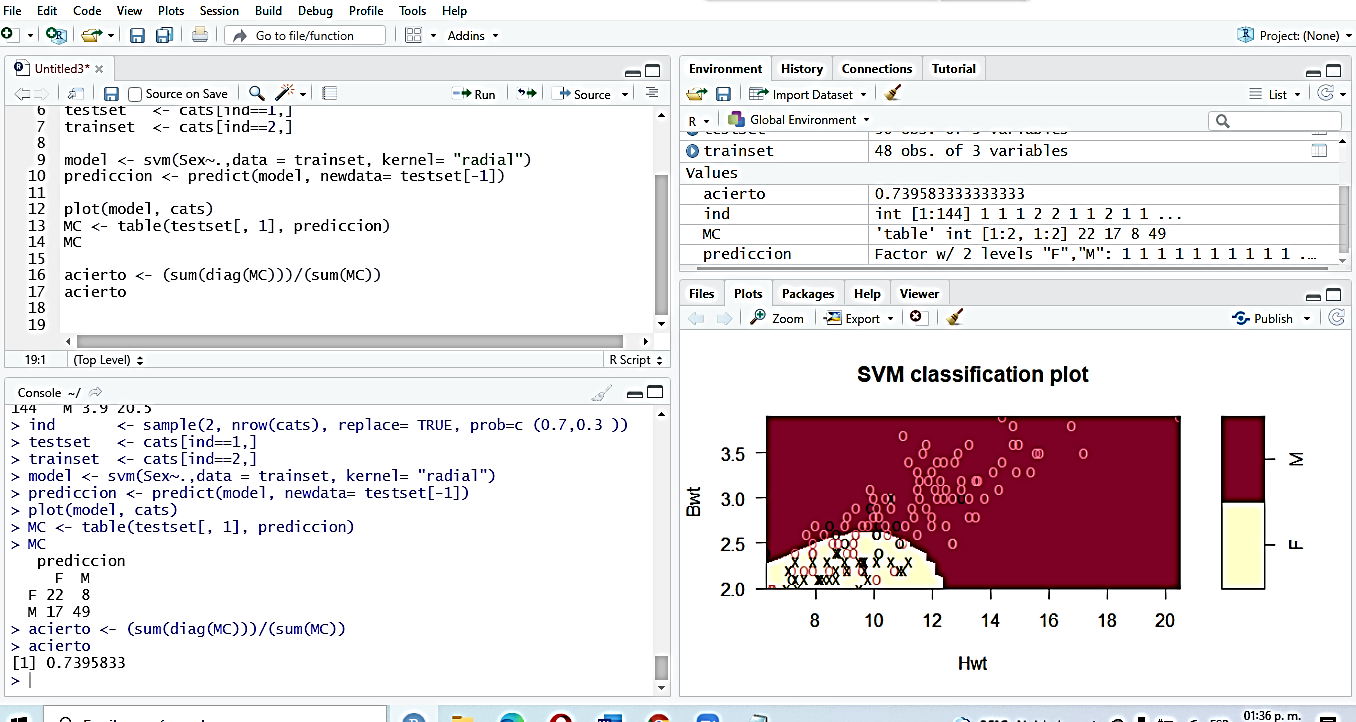


Imagen: 5: Acierto

# 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