

## PRACTICA PROFESIONALIZANTE III

### MODELO SVM PARA LA PREDICCIÓN DE DESERCIÓN REAL

Escuela de Artes Plásticas – Termas de Río Hondo

---

#### 1. OBJETIVO GENERAL Y MOTIVACIÓN

La deserción escolar representa uno de los principales desafíos contemporáneos dentro del ámbito educativo,

especialmente en instituciones donde las trayectorias escolares se ven afectadas por factores académicos, sociales,

familiares y territoriales. Este proyecto propone el diseño e implementación de un modelo predictivo basado en

Máquinas de Vectores de Soporte (SVM), capaz de anticipar casos reales de abandono escolar mediante el análisis

integrado de variables cuantitativas y cualitativas.

El objetivo es proporcionar una herramienta práctica para la detección temprana del riesgo académico y social,

permitiendo así orientar intervenciones institucionales más eficientes.

---

#### 2. PROBLEMA IDENTIFICADO

Las escuelas de la ciudad de Las Termas de Río Hondo presentan desafíos vinculados a la retención escolar.

La falta de un sistema de alerta temprana dificulta:

- La identificación de estudiantes en riesgo antes de que ocurra la deserción.
- La intervención oportuna por parte de docentes, tutores y equipos de orientación escolar.
- La toma de decisiones basada en evidencia cuantitativa.

En particular, la Escuela de Artes Plásticas registra dificultades asociadas a la irregularidad en la asistencia,

situaciones familiares complejas y un contexto socioeconómico diverso que influye directamente en las trayectorias de los estudiantes.

—

### **3. RELEVANCIA PARA LA ENTIDAD**

La construcción de un modelo predictivo posibilita:

- Reconocer patrones que anteceden a la deserción real.
- Optimizar el trabajo docente y del equipo directivo mediante datos concretos.
- Implementar estrategias de acompañamiento en alumnos que aún no han desertado, pero presentan señales de alerta.
- Alinear la política institucional con prácticas de gestión basadas en evidencias.

---

### **4. IMPACTO ESPERADO**

El modelo desarrollado contribuye significativamente a:

- Fortalecer la equidad educativa, al priorizar a estudiantes con trayectorias frágiles.
- Reducir el abandono escolar mediante alertas predictivas.
- Integrar herramientas tecnológicas y analíticas dentro de la gestión educativa.
- Profesionalizar el seguimiento académico mediante indicadores y métricas objetivas.

---

### **5. ALCANCE Y CONTEXTO**

El análisis se realizó sobre datos reales recolectados mediante formularios institucionales, correspondientes

a alumnos de distintos cursos de la Escuela de Artes Plásticas.

Las variables incluidas permiten observar múltiples dimensiones:

- Académica: asistencia, inasistencia, promedio.
- Familiar: tipo de conformación familiar.
- Territorial: distancia entre el hogar y la escuela.
- Sociocultural: nivel socioeconómico, conectividad, actividades extracurriculares, trabajo, etc.

---

## 6. ESTRUCTURA DE DATOS

El dataset utilizado (\*Artes\_Plasticas\_bbdd.csv\*) contiene una matriz heterogénea de información. Las variables

más significativas incluidas fueron:

- Asistencia %, Inasistencia %, Promedio académico.
- Situación Familiar: madre sola, padre solo, ambos padres, otro familiar a cargo.
- Distancia a la escuela: cerca, medio, lejos.
- Nivel socioeconómico.
- Conectividad.
- Edad y sexo.
- Actividad extracurricular y condición laboral.

---

## 7. VARIABLE OBJETIVO: DESERCIÓN REAL

La deserción real fue definida combinando criterios académicos y sociales: Criterios académicos (obligatorios):

- Asistencia < 60%
- Inasistencia > 40%
- Promedio “Regular”

Criterios sociales/territoriales de riesgo:

- Situación familiar vulnerable (madre sola, padre solo, otro familiar a cargo).
- Distancia “Lejos”.

Este enfoque multicausal refleja las trayectorias escolares reales del contexto institucional.

---

## **8. METODOLOGÍA BASADA EN MACHINE LEARNING**

El proceso técnico incluyó:

- Limpieza y normalización del dataset.
- Análisis exploratorio (EDA) para identificar distribuciones y relaciones entre variables.
- Preparación del dataset para el modelo mediante:
  - One-hot encoding para variables categóricas.
  - Escalado mediante StandardScaler.
  - Balanceo de clases con SMOTE para la clase minoritaria (deserción real).
- Entrenamiento del modelo SVM con kernel RBF.
- Evaluación del desempeño mediante métricas estándar:
  - Accuracy
  - Precision
  - Recall (sensibilidad)
  - F1 Score
  - Curva ROC y AUC.

---

## **9. RESULTADOS OBTENIDOS**

El modelo SVM logró una alta capacidad predictiva, obteniendo:

- \*\*Accuracy\*\*: 0.93
- \*\*Precision\*\*: 0.66
- \*\*Recall\*\*: 1.00 (el modelo detectó correctamente todos los desertores reales)
- \*\*F1 Score\*\*: 0.80
- \*\*AUC\*\*: 1.00

Esto demuestra un desempeño excelente, especialmente considerando la escasez de datos y la complejidad social del fenómeno.

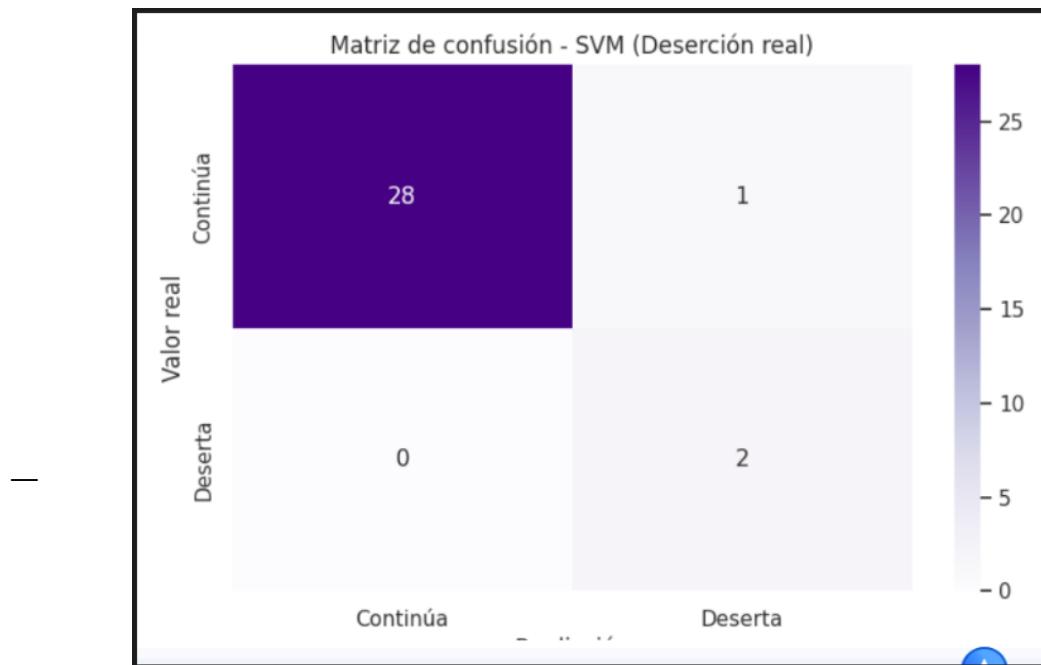
---

## 10. ANÁLISIS DE LA MATRIZ DE CONFUSIÓN

La matriz arrojó los siguientes resultados:

- 28 verdaderos negativos (continúan y el modelo lo predijo bien)
- 1 falso positivo (el modelo predijo deserción, pero no lo era)
- 2 verdaderos positivos (desertores reales correctamente identificados)
- 0 falsos negativos (ningún desertor real fue pasado por alto)

Este comportamiento es ideal para un modelo preventivo dentro del ámbito educativo.

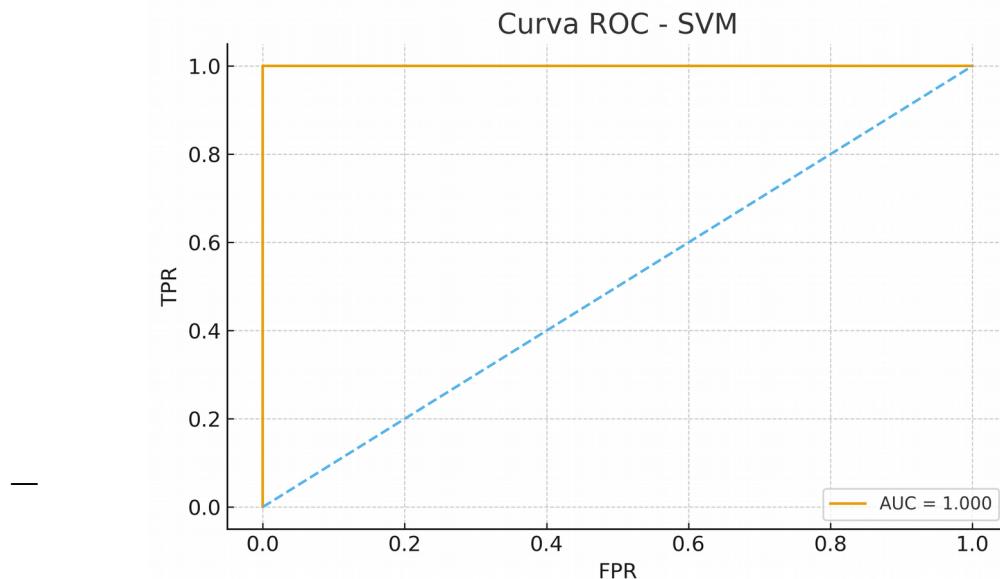


## 11. CURVA ROC Y CAPACIDAD DISCRIMINATIVA

La curva ROC mostró un área bajo la curva (AUC) de \*\*1.00\*\*, lo que implica que el modelo separa perfectamente

las clases “continúa” y “deserta”.

En términos pedagógicos, esto significa que el modelo distingue con claridad patrones de riesgo escolar real.



## 12. IMPORTANCIA DE VARIABLES (Permutation Importance)

El análisis de importancia de variables reveló que las dimensiones más influyentes fueron:

- Conectividad.
- Nivel socioeconómico.
- Edad.
- Actividad extracurricular.
- Sexo.

Estas variables funcionan como factores que explican las trayectorias educativas y deben ser consideradas

en estrategias institucionales de acompañamiento.

### **13. SOLUCIÓN DESARROLLADA**

El sistema predictivo se desarrolló con:

- Preprocesamiento de datos mediante StandardScaler.
- Balanceo con SMOTE.
- Clasificación mediante SVM (kernel RBF).
- Visualizaciones para la interpretación institucional.
- Métricas de validación que demuestran alta robustez.

Además, el proyecto ofrece una base sólida para crear futuros dashboards, sistemas de alerta y reportes automatizados.

---

### **14. CONCLUSIONES GENERALES**

- El modelo SVM es altamente eficaz para detectar deserción real con pocos datos.
- La metodología multicausal refleja fielmente las trayectorias educativas de los estudiantes.
- Los resultados permiten fundamentar intervenciones institucionales basadas en datos objetivos.
- Contribuye a la construcción de políticas educativas más inclusivas y preventivas.

---

### **15. PROYECCIÓN FUTURA Y RECOMENDACIONES**

Se recomienda:

- Integrar los resultados dentro de un panel de gestión escolar.
- Realizar nuevos relevamientos para ampliar el dataset.
- Validar el modelo con cohortes futuras.
- Incorporar indicadores emocionales y de bienestar estudiantil.
- Implementar un sistema institucional de alerta temprana.

## 16. Diagrama de flujo de trabajo

