# MACHINE LEARNING CANVAS Designed for: Designed by: SME Date: 12/10/2025 Iteration: v1 .

| TAREA DE PREDICCIÓN **Entrada:** Variables demográficas, académicas y socioeconómicas.  **Salida (Target):** **Performance** (Excellent, Good, Average, Poor en Nivel de desempeño esperado).  **Tipo de tarea:** Clasificación supervisada.  **Algoritmos candidatos:**   * Regresión Logística * SVM * Gradient Boosting (XGBoost, LightGBM) * Regresión Logística Ordinal | DECISIONES **Identificar factores sociales, educativos y personales que afectan en el desempeño de los estudiantes.**  **Diseñar tutorías específicas de acuerdo a las necesidades de cada estudiante.**  **La predicción del rendimiento también pueden ser:**   * Apoyar políticas de becas o refuerzos educativos. * Ayudar a identificar candidatos en riesgo académico. * Guiar estrategias de admisión o intervención temprana. | PROPUESTA DE VALOR **La propuesta de valor de este enfoque ofrece la posibilidad de generar información para diferentes actores del sistema educativo. Para las instituciones, el modelo permitirá identificar factores éxito académico y diseñar programas de apoyo. Para los estudiantes y sus familias, ofrecerá una visión más clara sobre el impacto real de invertir en programas de coaching, ayudando a una mejor toma de decisiones educativas. Finalmente, a nivel social, este análisis contribuirá a visibilizar posibles desigualdades en el acceso a recursos de formación, fomentando el diseño de políticas más justas y equitativas.** | RECOPILACIÓN DE DATOS **Los datos recolectados a partir de registros académicos y formularios de inscripción.**  **Los datos incluyen variables demográficas (género, casta), educativas (porcentajes de clase X y XII, tipo de institución), y familiares (ocupación de los padres).** | FUENTES DE DATOS **Los datos son obtenidos del dataset público “Student Performance on an Entrance Examination”.**  **Dataset recopilado por el Prof. Jiten Hazarika, que contiene información académica y socioeconómica de los aspirantes a universidades médicas de Assam.** |
| --- | --- | --- | --- | --- |
| SIMULACIÓN DE IMPACTO **Mayor precisión de predicción y de transparencia**  **Métricas de evaluación: Accuracy, Precision, Recall, F1-Score, Confusion Matrix** | HACIENDO PREDICCIONES **Las predicciones se hacen en bloque al finalizar las aplicaciones del examen de admisión.**  **Se pueden hacer predicciones en tiempo real para calcular la probabilidad de aprobación de los estudiantes.** |  | CONSTRUYENDO MODELOS **Construir un solo modelo de clasificación supervisada que prediga el nivel de rendimiento (Performance) de los estudiantes en categorías como *Excellent*, *Good*, Poor, Average.**   * Entrenamiento inicial con el dataset histórico. * Actualización anual con nuevos datos de ingreso. * Frecuencia: una vez por ciclo académico (anual). * Tiempo estimado para el modelado: 1-2 semanas para limpieza, entrenamiento y validación. | CARACTERÍSTICAS **Entradas que alimentan el modelo:**   * Gender * Caste * Coaching (ninguno, dentro o fuera de Assam) * Class\_ten\_education (tipo de institución) * Twelve\_education (tipo de institución) * Medium (idioma de instrucción) * Class\_X\_Percentage * Class\_XII\_Percentage * Father\_occupation * Mother\_occupation |
|  | MONITOREO **Seguimiento de métricas técnicas (Accuracy, Precision, Recall, F1-Score, Confusion Matrix)** | **Monitoreo de las diferencias de desempeño entre grupos(género, casta, desempeño previo, etc.)** | **Detectar si hay cambios en los patrones de los datos.** |  |

# 