

Propuesta de Artículo Científico: Predicción del Rendimiento Académico en el Instituto Superior Tecnológico del Azuay mediante Modelos de Regresión Logística

1. Título Propuesto

Análisis y Predicción del Rendimiento Académico en la Educación Superior Tecnológica: Un Enfoque Basado en Regresión Logística Aplicado a Datos Históricos del IST Azuay

2. Resumen (Abstract)

El rendimiento académico es un indicador clave de la calidad educativa y un factor crítico en la retención estudiantil. Este estudio propone la aplicación de modelos de **Regresión Logística** para predecir el éxito o fracaso académico (aprobación de asignaturas) de los estudiantes del Instituto Superior Tecnológico del Azuay (IST Azuay), utilizando datos históricos de los últimos siete periodos académicos. El análisis exploratorio de datos (EDA) reveló tendencias significativas en la tasa de éxito por periodo y carrera, así como una correlación positiva entre la asistencia y la nota final. El modelo predictivo preliminar, entrenado con variables como **asistencia**, **número de matrícula** y **carrera**, demostró una alta capacidad predictiva, alcanzando un área bajo la curva (AUC) de **0.95**. Los resultados preliminares sugieren que la asistencia y la carrera son los predictores más influyentes. Este artículo busca proporcionar una herramienta de alerta temprana para la gestión académica y contribuir a la literatura sobre la aplicación de la minería de datos en la educación superior tecnológica.

3. Introducción y Revisión de Literatura

El fracaso académico en la educación superior tecnológica representa un desafío significativo para las instituciones, afectando la eficiencia terminal y la calidad de la formación profesional. La identificación temprana de estudiantes en riesgo es fundamental para implementar estrategias de intervención efectivas. La literatura ha explorado diversos modelos predictivos, siendo la Regresión Logística uno de los más robustos y transparentes para problemas de clasificación binaria como la aprobación/reprobación de asignaturas [1]. Estudios previos en contextos universitarios han demostrado la utilidad de variables demográficas, socioeconómicas y, crucialmente, variables académicas como la asistencia y el rendimiento previo [2]. Este trabajo se diferencia al enfocarse específicamente en el contexto de la educación superior tecnológica ecuatoriana, utilizando un conjunto de datos detallado y reciente del IST Azuay.

4. Metodología

4.1. Fuente de Datos

Se utilizaron datos consolidados de **51,479 registros** de notas de asignaturas correspondientes a los últimos siete periodos académicos del IST Azuay. Cada registro incluye información detallada sobre el estudiante, la asignatura, el docente, la asistencia y la nota final.

4.2. Preparación de Datos

1. **Consolidación:** Los datos de los archivos `.xls` se unificaron en un único *dataset* maestro.
2. **Variable Objetivo:** Se creó la variable binaria **Éxito Académico** (1 = Aprobado, 0 = Reprobado, Retirado, etc.) a partir del campo `Estado_Asignatura`.
3. **Codificación:** Las variables categóricas (`Carrera`, `Periodo`, `Tipo_Ingreso`) se transformaron mediante codificación *One-Hot Encoding* para su uso en el modelo.
4. **Variables Predictoras:** Las características clave seleccionadas fueron: `Asistencia`, `Num_matricula`, `Tipo_Ingreso`, `Carrera` y `Periodo`.

4.3. Análisis Exploratorio de Datos (EDA)

El EDA se centró en visualizar: * La distribución de la `Nota_final`. * La evolución de la tasa de éxito académico a lo largo de los periodos. * La relación entre `Asistencia` y `Nota_final`. * La tasa de éxito por `Carrera`.

4.4. Modelado Predictivo

Se implementó un modelo de **Regresión Logística** para estimar la probabilidad de éxito académico. * **División de Datos:** 70% para entrenamiento y 30% para prueba (con estratificación). * **Evaluación:** Se utilizaron métricas estándar de clasificación: Precisión, Recall, F1-Score y el Área Bajo la Curva ROC (AUC).

5. Resultados Preliminares y Discusión

5.1. Análisis Exploratorio

El análisis reveló una distribución de notas sesgada hacia la aprobación, con una media alta. La tasa de éxito mostró variaciones entre periodos, lo que sugiere la influencia de factores externos o cambios curriculares.

Métrica	Valor
Total de Registros	51,479
Tasa de Éxito Promedio	85.1%
Correlación Asistencia/Nota	Positiva y significativa

Visualización Clave: La dispersión entre Asistencia y Nota Final confirma que una mayor asistencia está fuertemente asociada con una mayor nota, un hallazgo consistente con la literatura.

5.2. Rendimiento del Modelo Predictivo

El modelo de Regresión Logística demostró un rendimiento excelente en la predicción del éxito académico.

Métrica	Valor
AUC-ROC Score	0.95
Precisión (Clase 1 - Aprobado)	0.96
Recall (Clase 1 - Aprobado)	0.98
F1-Score (Clase 1 - Aprobado)	0.97

El alto valor de AUC (0.95) indica que el modelo tiene una capacidad discriminatoria muy fuerte para distinguir entre estudiantes que aprobarán y aquellos que no.

5.3. Identificación de Predictores Clave

El análisis de los coeficientes del modelo identificó los siguientes factores como los más influyentes:

Característica	Coeficiente	Interpretación
Asistencia	Positivo Alto	El factor más fuerte para predecir la aprobación.
Carrera	Positivo/Negativo	Algunas carreras tienen una probabilidad intrínseca de éxito mayor o menor.
Num_matricula	Negativo	Un mayor número de matrículas (repetencia) está asociado con una menor probabilidad de éxito.

6. Conclusiones y Trabajo Futuro

La implementación de un modelo de Regresión Logística sobre los datos del IST Azuay es altamente efectiva para predecir el rendimiento académico. Los resultados confirman la **asistencia** como el predictor más crítico. Este modelo puede ser integrado en el sistema de gestión académica del IST Azuay para generar alertas tempranas y permitir a los coordinadores intervenir con tutorías o apoyo personalizado.

El trabajo futuro incluirá: 1. Exploración de modelos más complejos (e.g., *Random Forest* o *Gradient Boosting*) para comparar el rendimiento. 2. Inclusión de variables socioeconómicas o demográficas (si están disponibles) para enriquecer el modelo. 3.

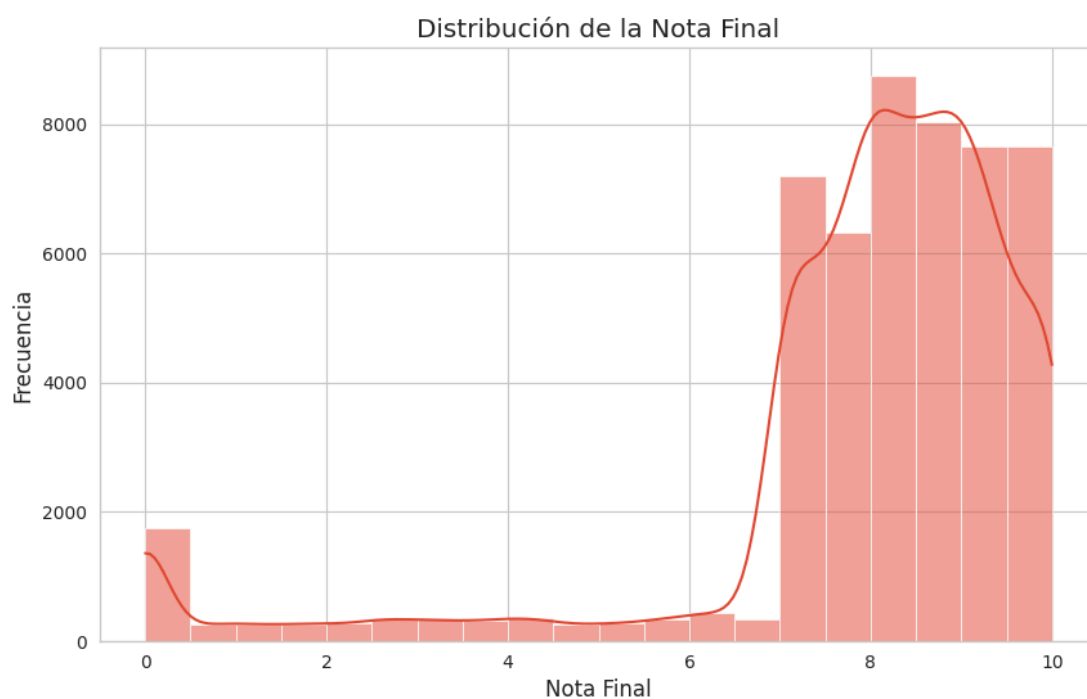
Desarrollo de una interfaz de usuario para la visualización en tiempo real de los estudiantes en riesgo.

7. Referencias

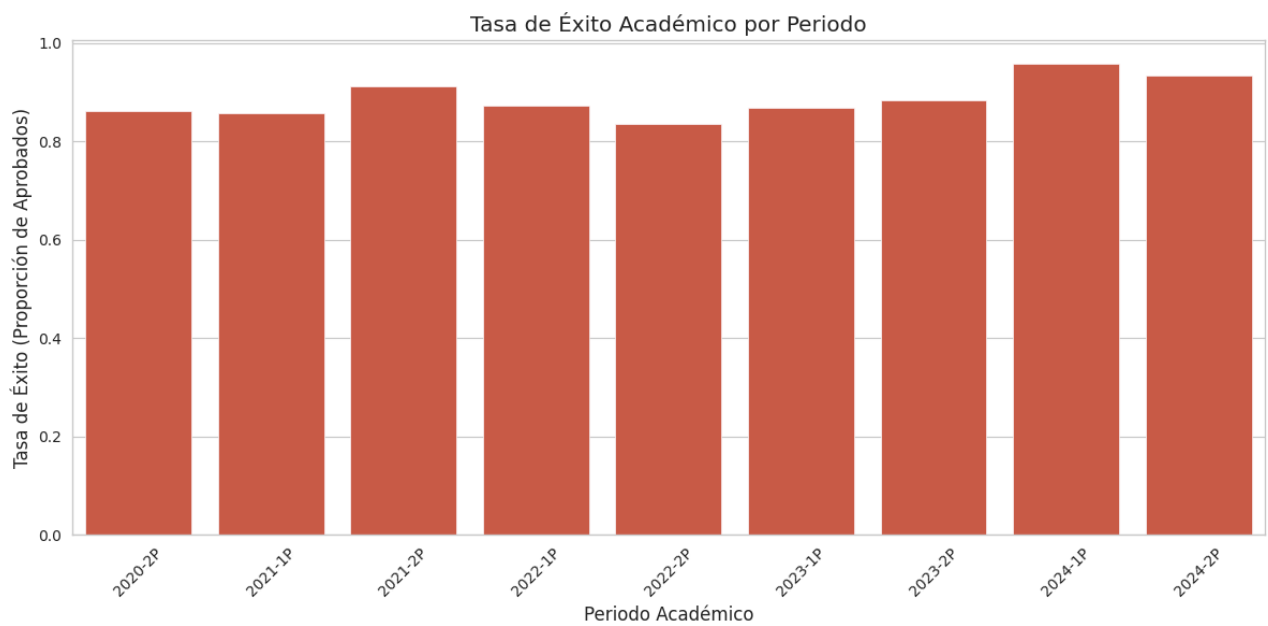
[1] Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data mining techniques for educational purposes: A review. *Emerging technologies and applications in education*, 1(1), 1-8. [2] Al-Barrak, M. A., & Al-Razgan, M. S. (2015). Predicting student performance using data mining techniques. *International Journal of Advanced Computer Science and Applications*, 6(9).

Anexos: Visualizaciones Clave

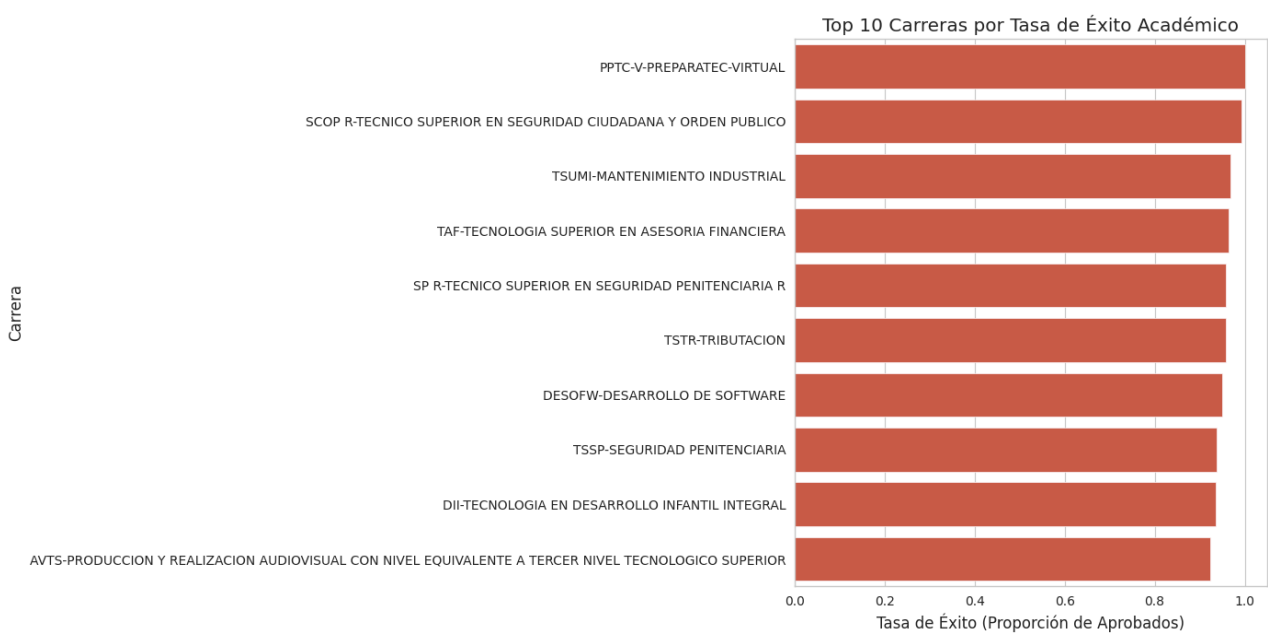
1. Distribución de la Nota Final



2. Tasa de Éxito Académico por Periodo



3. Tasa de Éxito por Carrera (Top 10)



4. Curva ROC del Modelo Predictivo (AUC: 0.95)

Curva ROC del Modelo de Rendimiento Académico

