

# **Informe Septiembre**

## **Clase Espejo - Aprendizaje de Máquina**



**CORHUILA - ING. SISTEMAS**

**7/10/2022**

**Julian Andres Quimbayo**

**Eilen Lorena Perez Montero**

## Tabla de contenido

1. Actividad 1: Análisis Exploratorio Descriptivo .....	4
1.1. Exploración de variables del Instrumento de diagnóstico.....	4
1.2. Validez del instrumento.....	8
1.3. Balance del dataset y Normalización .....	9
1.4. Test de Normalidad - Shapiro - wilk.....	12
1.5. Correlaciones .....	13
2. Actividad 2 Y 3: Creación de modelos clasificación y pruebas.....	15

## Lista de Figuras

Figura 1 Uso de sistema de puntos o créditos en la clase espejo .....	5
Figura 2 Preparación de la clase espejo por parte del docente.....	5
Figura 3 Puntualidad al inicio de la clase espejo.....	6
Figura 4 Trabajo colaborativo en la clase espejo .....	6
Figura 5 Adaptación del material y actividades al contexto del estudiante .....	7
Figura 6 Se entiende lo explicado por el docente en la clase espejo .....	7
Figura 7 Formula de Alpha de Cronbach.....	8
Figura 8 Distribución del logro académico .....	9
Figura 9 Curva ROC multiclase logro académico .....	10
Figura 10 Precisión y exhaustividad por clases logro académico.....	10
Figura 11 Uso de técnica SMOTE de tipo under samplig para logro académico .....	11
Figura 12 Mejora después de aplicar balanceo de clases .....	12
Figura 13 Mapa de correlaciones para todas las dimensiones usadas .....	13
Figura 14 Nuevas correlaciones para el desarrollo de los modelos .....	15

## Lista de Tablas

Tabla 1 Escala de valoración preguntas.....	4
Tabla 2 Criterios de interpretación del coeficiente de alfa de Cronbach .....	8
Tabla 3 Relación criterio de evaluación del logro con cantidad de estudiantes .....	9
Tabla 4 Resultados test de normalidad shapiro -wilk .....	12
Tabla 5 Métricas modelos primer momento algoritmo LDA .....	16
Tabla 6 Métricas modelos primer momento algoritmo SVC .....	16
Tabla 7 Métricas modelos primer momento algoritmo RLOG Multinomial .....	16
Tabla 8 Métricas modelos segundo momento algoritmo LDA .....	16
Tabla 9 Métricas modelos segundo momento algoritmo SVC.....	17
Tabla 10 Métricas modelos segundo momento algoritmo RLOG Multinomial .....	17

## 1. Actividad 1: Análisis Exploratorio Descriptivo

Se creo el instrumento de diagnóstico el cual consta de 3 dimensiones las cuales son: D1. Demográfico, D2. Logro Académico y D3. Pedagógica. Encontrando los siguientes resultados:

1. Estudiantes del programa de ingeniería de sistemas y mecatrónica del semestre A y B de 2021.
2. Instrumento con 26 preguntas distribuidas en las siguientes dimensiones:  
D1. Dimensión demográfica (9 preguntas):  
D2. Logro Académico (1 Pregunta): Se evidencian los rangos de obtención de logro creados de manera institucional.  
D3. Dimensión Pedagógica (16 Preguntas)
3. Escala de las respuestas (1 a 5) siendo uno la mas baja y 5 la más alta.
4. Total de observaciones dentro del dataset: 103 filas \* 26 columnas = 2756 datos en total

### 1.1. Exploración de variables del Instrumento de diagnóstico

Dentro de las información demográfica se evidencia 63% de la población es de estrato 2, 21% estrato 1 y el restante 16% estrato 3, 4 y 5; por otro lado, El 63% de la población se encuentran en el rango de edad 21 a 25 años y el El 21% entre 17 -20 años y el restante 16% son mayores de 25 años, esto teniendo en cuenta que el 65% de los encuestados son hombres y el 35% son mujeres.

De igual manera la distribución de estudiantes frente al semestre es el 45% de las personas encuestadas pertenecen a decimo semestre, el 21% al noveno semestre, el 18% a septimo semestre y el 16% final a quinto semestre. Un aspecto a resaltar es el apoyo y convivencia con los padres factor social que indica la importancia de seguimiento por parte del núcleo familiar en donde 37 estudiantes del género masculino no viven con sus padres actualmente, 30 estudiantes del mismo género si conviven con sus padres. En el caso de las mujeres la relación es 17 no viven con sus padres y 19 si lo hacen actualmente. Focalizando un poco mas la relación estrato y vivencia con los padres 54 estudiantes de estratos 1 y 2 es donde se encuentran el mayor número de esta categoría.

En cuanto a la situación laboral 68 personas de los estratos 1, 2, 3, 4 y 5 no se encuentran laborando actualmente. Siendo los estratos 1 y 2 los que mayor cantidad poblacional tienen en esa situacion. Para la siguiente serie de figuras se analizaron las preguntas más relevantes de la dimensión pedagógica. Teniendo en cuenta una escala de 5 rangos los cuales se evidencian en la tabla uno y entendiendo que en el gráfico de barras 1 es el genero masculino y 2 femenino.

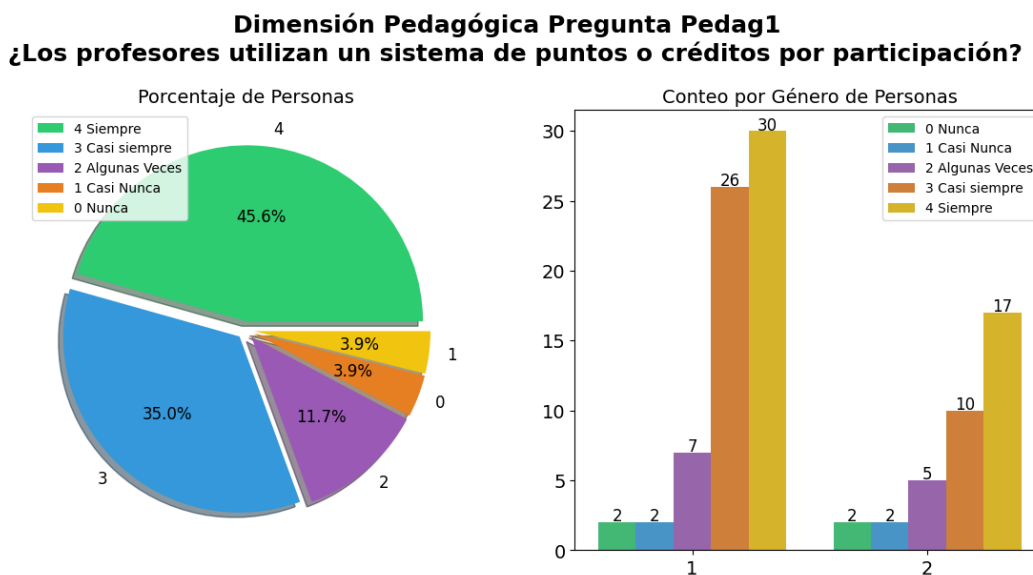
Tabla 1 Escala de valoración preguntas

Opción	Valor
Nunca	0
Casi Nunca	1
Algunas veces	2
Casi Siempre	3
Siempre	4

Fuente: Autores

Empezando por la pregunta si ¿Los profesores utilizan un sistema de puntos o créditos por participación?, como se evidencia en la figura 1, el 45,6% de los encuestados aducen que los docentes si utilizan un sistema de puntos o créditos por participación en la clase espejo. Un 8% aduce que nunca se utiliza dicho sistema.

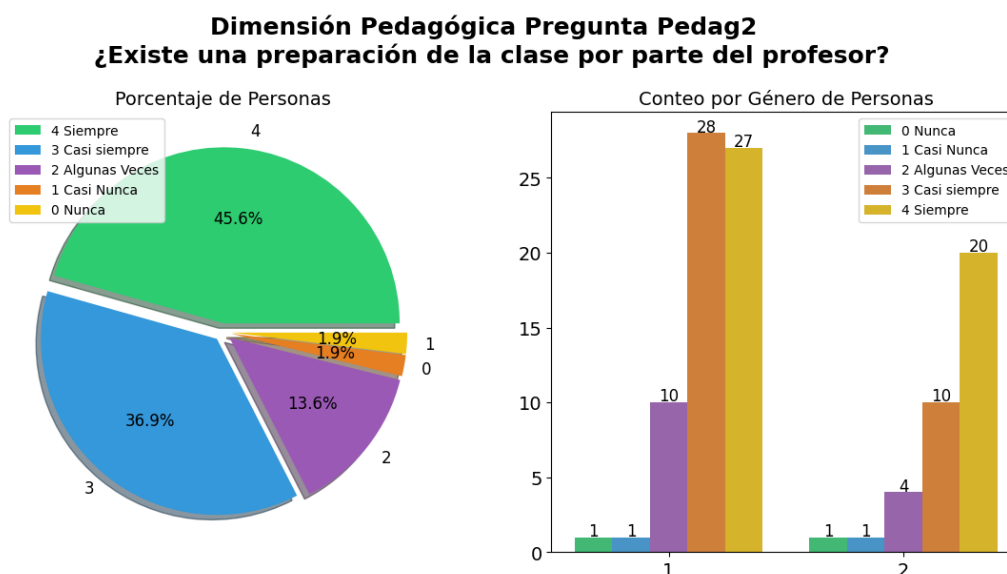
Figura 1 Uso de sistema de puntos o créditos en la clase espejo



Fuente: Autores

Frente a la pregunta si las clases espejo son preparadas por el docente como se evidencia en la figura 2, el 45,6% de los encuestados aducen que los docentes casi siempre preparan la clase espejo a impartir. Un 4% reconoce que no existe preparación para dicho encuentro.

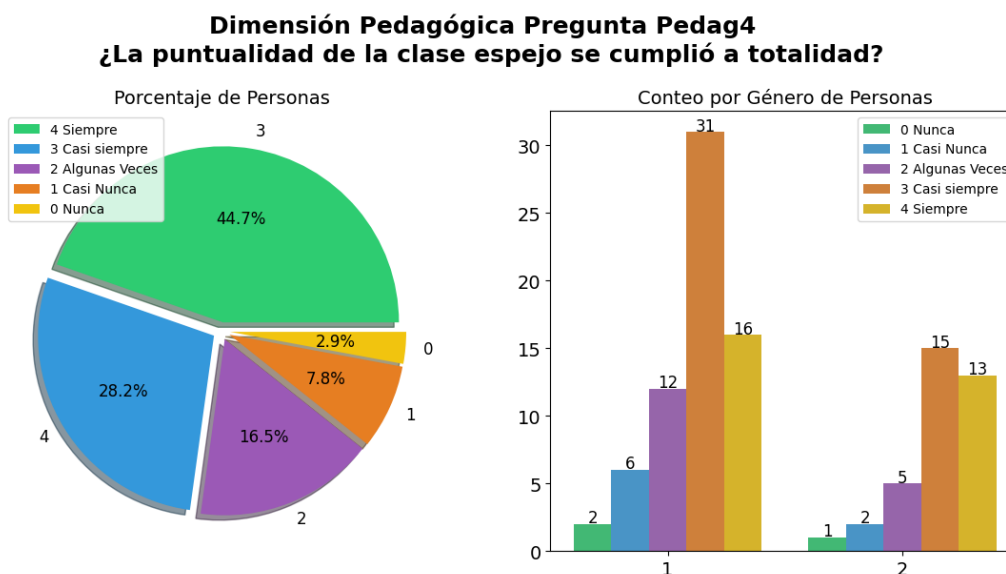
Figura 2 Preparación de la clase espejo por parte del docente



Fuente: Autores

Otro de los aspectos significativos es la puntualidad para comenzar una clase espejo más si es mediada por tecnología, a lo cual El 44,7% de los estudiantes afirman que casi siempre se cumple con el horario de la clase espejo. Un 11% aduce que no se cumple con los horarios pactados como referencia en la figura 3.

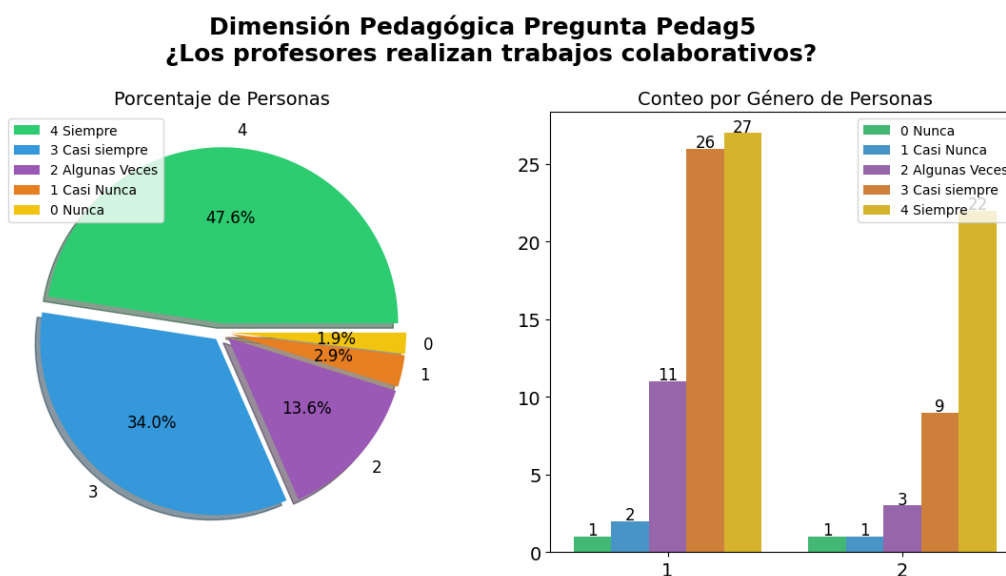
Figura 3 Puntualidad al inicio de la clase espejo



Fuente: Autores

En cuanto a los aspectos colaborativos, durante la clase espejo se manifiesta que Algunas veces los profesores realizan trabajos colaborativos durante la clase espejo con un 13,6% de los estudiantes lo afirman, y se denota en la figura 4.

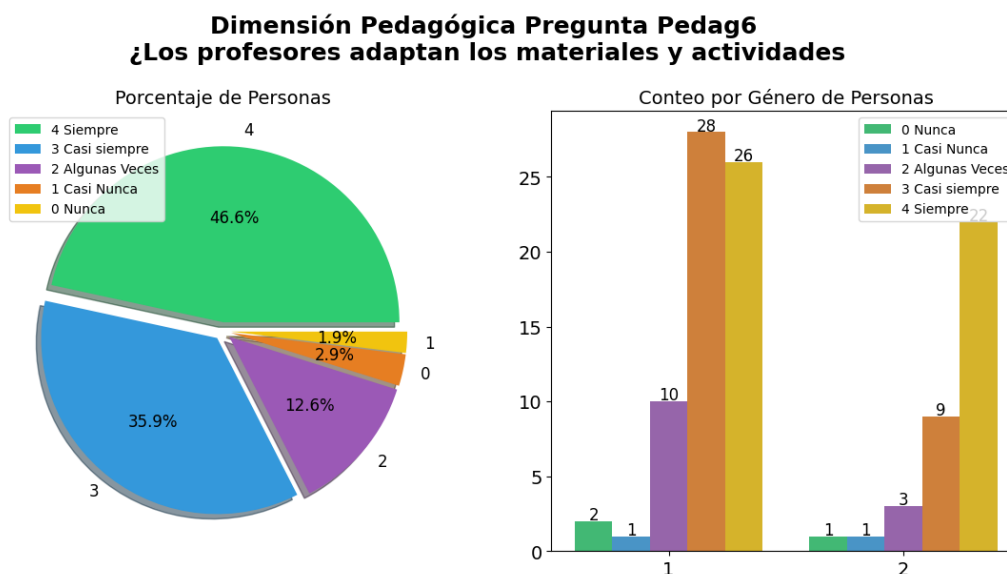
Figura 4 Trabajo colaborativo en la clase espejo



Fuente: Autores

En cuanto al material y actividades que se utilizan en la clase espejo y como se adaptan al contexto del estudiante. Se encontró que el 35,9% de las personas reconocen que casi siempre los docentes adaptan el material a los intereses de los estudiantes, un 12,6% afirma que casi nunca, evidenciado en la gráfica 5.

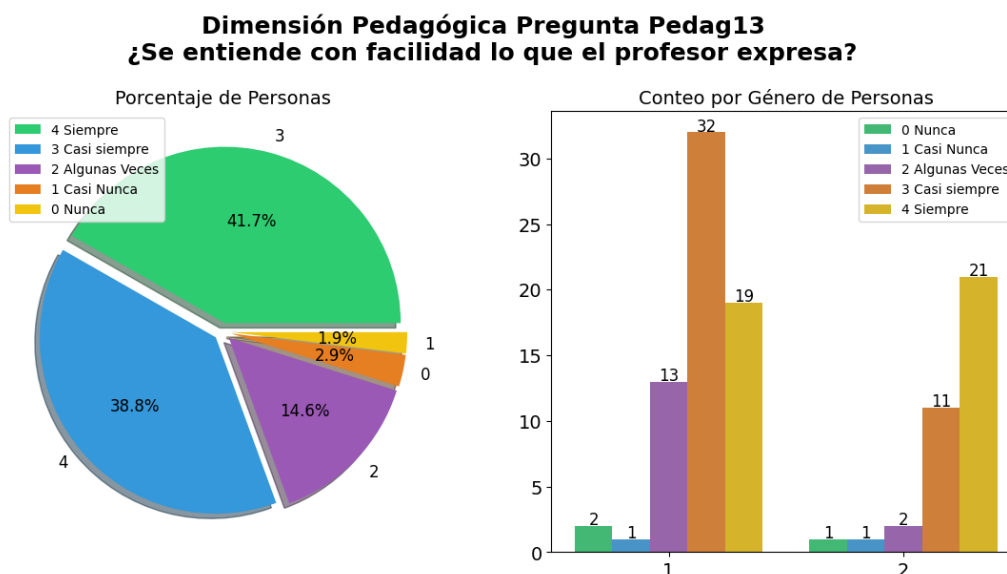
Figura 5 Adaptación del material y actividades al contexto del estudiante



Fuente: Autores

Por otro lado, en cuanto a si se entiende lo explicado y llega el mensaje con claridad de la clase espejo, se evidenció que el 41,7% reconoce que casi siempre se entiende lo que el profesor expresa. Un 18% reconoce que no se entiende lo explicado por el docente, referenciado en la figura 6.

Figura 6 Se entiende lo explicado por el docente en la clase espejo



Fuente: Autores

Como entendimiento inicial se evidencia que la clase espejo para la facultad de ingeniería los estudiantes se encuentran familiarizados y activos con la participación de las

temáticas desarrolladas. Los aspectos para analizar es que un porcentaje bajo de docentes no preparar su clase espejo siendo un momento diferente y complementario. Se evidencia la falta de inmersión y flexibilidad de los contenidos acordes a las temáticas propias de cada ingeniería y finalmente, reconocer que es un ambiente medido por TIC donde captar la atención y participación por periodos de tiempo prolongados es muy complejo. Como preliminar esto evidencia la falta de preparación de estudiantes, administrativos, docentes y directores de programa de las instituciones inmersas en una estrategia que implica empatía, aceptación, disrupción y sobre todo voluntad de cambio y aprendizaje.

## 1.2. Validez del instrumento

Finalmente el instrumento fue validado con el indicador de alpha de cronbach el cual permite entender la confiabilidad del instrumento, generando una relación entre el número de ítems, la varianza observada de las puntuaciones y la varianza observada del componente i para la actual muestra como lo evidencia la figura 7.

Figura 7 Formula de Alpha de Cronbach

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum_{i=1}^k \sigma_{y_i}^2}{\sigma_x^2} \right)$$

K = Numero de ítems

$\sigma_x^2$  = varianza de observada de las puntuaciones

$\sigma_{y_i}^2$  = varianza observada del componente i para la actual muestra

Según George y Malley el Alpha de Cronbach se puede interpretar de la siguiente manera, como se evidencia en la tabla 2.

Tabla 2 Criterios de interpretación del coeficiente de alfa de Cronbach

Valores del Alfa	Interpretación
0.90 - 1.00	Muy Satisfactorio
0.80 - 0.89	Adecuado
0.70 - 0.79	Moderado
0.60 - 0.69	Bajo
0.50 - 0.59	Muy Bajo
< 0.50	No Confiable

Fuente: Autores

El resultado del instrumento para su Alpha de Cronbach fue de 0.87, se utilizó por medio de la librería pingouin en Python con un índice de confianza del 95%, lo cual se considera como un instrumento adecuado para el desarrollo de la recopilación de la información. Finalmente se procedió a normalizar la data ya que existían datos distantes unos entre otros con el objetivo de evidenciar las mejores variables para la clasificación, esto se realizó con el procedimiento de mínimos y máximos de cada columna buscando colocar los datos entre cero y uno, y así reduciendo el ruido entre los mismos. Teniendo en cuenta la distribución de los datos de la variable escogida para predecir que en este caso es Logro Académico, se toma la nota promedio de carrera al momento de realizar la encuesta y se



convierte a variable categórica teniendo en cuenta los criterios de evaluación de logro institucionales, los cuales son: 4.9 a 5.0 se cumple plenamente, 4.2 a 4.8 se cumple en alto grado, 3.0 a 4.1 se cumple aceptablemente, 2.0 a 2.9 se cumple en bajo grado y de 0.0 a 1.0 se cumple insatisfactoriamente. Para este caso en particular los 103 encuestados se encontraban en los rangos de cumplimiento en alto grado, aceptable y bajo grado como lo evidencia la tabla 3.

Tabla 3 Relación criterio de evaluación del logro con cantidad de estudiantes

Criterio de evaluación logro	Cantidad de estudiantes
Se cumple en alto grado	26
Se cumple aceptablemente	70
Se cumple en bajo grado	7

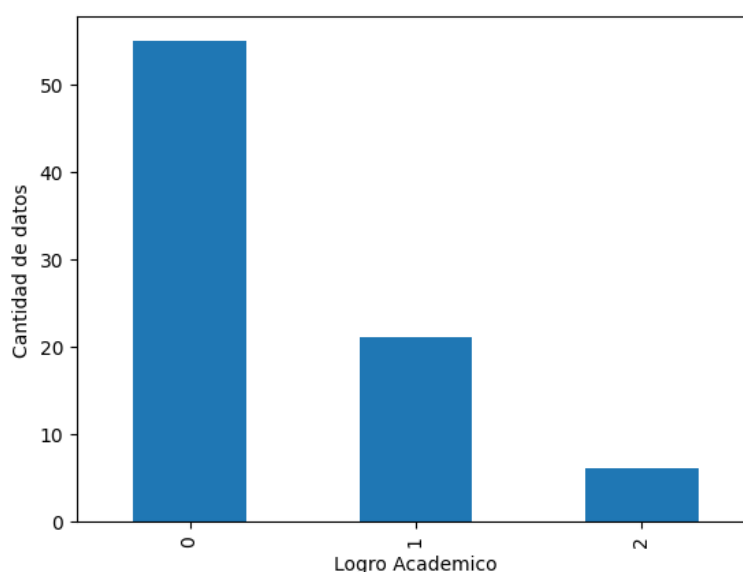
Fuente: Autores

Debido a esta distribución desbalanceada se hace necesario emplear técnicas que permitan nivelar la data hacia arriba o hacia abajo, esto se conoce como over sampling o under sampling. Con el objetivo que los algoritmos usados puedan realizar un trabajo eficiente mas no sesgado a la distribución actual.

### 1.3. Balance del dataset y Normalización

Para esta fase dentro del primer paso es convertir las etiquetas de la categoría LogroAca de texto a numero con el fin de que se pueda utilizar el balanceo de la data ya que no permite que se ejecute sobre valores de tipo texto quedando de la siguiente manera como se evidencia en figura 8. Donde se observa el valor cero (0) para se cumple aceptablemente, el valor de uno (1) para se cumple en alto grado y el valor de dos (2) se cumple en bajo grado, dicha conversión se conoce como label encoder que permite sobre una columna de tipo categórico (texto), convertir las distribuciones a formato numérico, allí se utilizó la librería scikit learn con su modulo preprocessing label encoder versión 1.1.2 con Python 3.10.

Figura 8 Distribución del logro académico



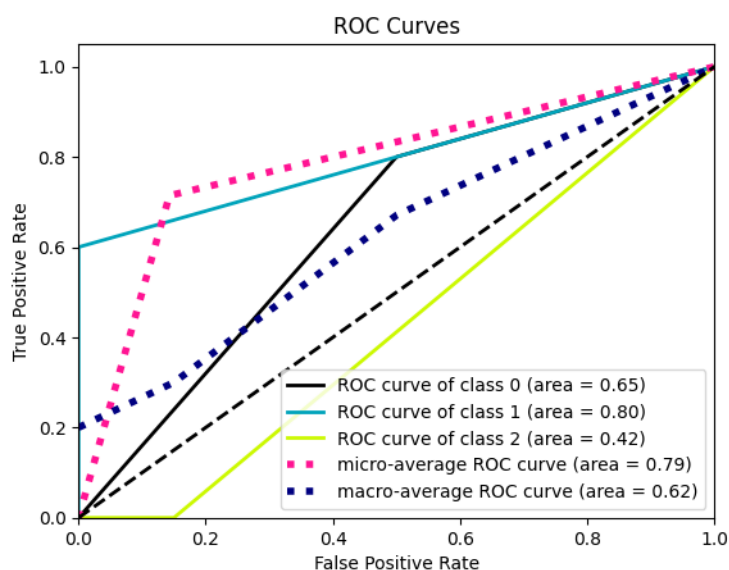
Fuente: Autores

Luego de realizar dicha conversión, se procede a particionar la data con la librería test\_train\_split de scikit learn usando 80% de entrenamiento, 20% de testeo y un random

state de 100. Para así usar un modelo que permitiera identificar las métricas futuras se usó DecisionTreeClasifier para obtener la métrica de característica del operador receptor (ROC), que permite comparar la sensibilidad y especificidad de los modelos de clasificación cuando son binarios o multi clase como en este caso. Dichos valores van desde 0 a 1 siendo el mas cercano a 1 el mejor valor a obtener. Cabe resaltar que la sensibilidad es la capacidad del algoritmo de poder clasificar los valores de cada clase comparando lo actual con lo futuro es decir los verdaderos positivos y la especificidad es realiza el mismo procedimiento, pero enfocado a buscar la parte negativa es decir los verdaderos negativos. Para este caso sería ver cuales tienen un logro académico alto y bajo ósea la clase 1 y 2 versus como clasifica el modelo la clase 0 personas con logro académico aceptable. Tal como se evidencia en la figura 9, la métrica ROC revela que para la clase cero (0) el balance entre sensibilidad y especificidad es de 0.65, si lo llevamos a la situación problema quiere decir que puede clasificar de 103 alumnos solo 67 de manera correcta cuando queremos buscar en logro académico aceptable.

Por otro lado, para la clase uno (1), la curva ROC revela 0.80, quiere decir que de 103 alumnos 82 pueden ser clasificados como estudiantes de alto grado de logro de manera correcta. Finalmente, para la clase dos (2), la métrica indica un 0.42 es decir de 103 alumnos 43 pueden ser clasificados de manera correcta como logro de bajo grado.

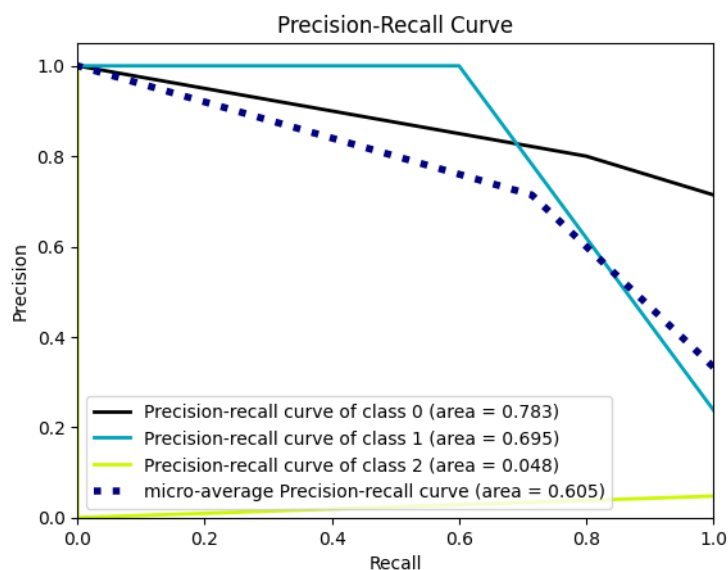
Figura 9 Curva ROC multiclase logro académico



Fuente: Autores

De igual manera la curva de precisión versus la exhaustividad, entendiendo la precisión como la calidad del modelo a la hora de ejecutar las tareas y la exhaustividad la cantidad de personas que es capaz de clasificar según su clase. Así pues, como se evidencia en la figura 10, el modelo sin balancear es capaz de clasificar mejor la clase 0 de logro académico en aceptable con un 78,3% es decir 3,1 de cada 4 estudiantes pueden ser clasificados según el logro aceptable. En segundo lugar, la clase 1 con un 69.5%, esto quiere decir que 2,8 d cada 4 estudiantes se pueden clasificar correctamente como logro de alto grado y finalmente la clase 2 con un 4,8% que indica que casi el modelo no detecta esta clasificación de bajo grado, esto por los pocos datos recopilados de esta clase y se manifiesta el problema de desbalance para el modelo.

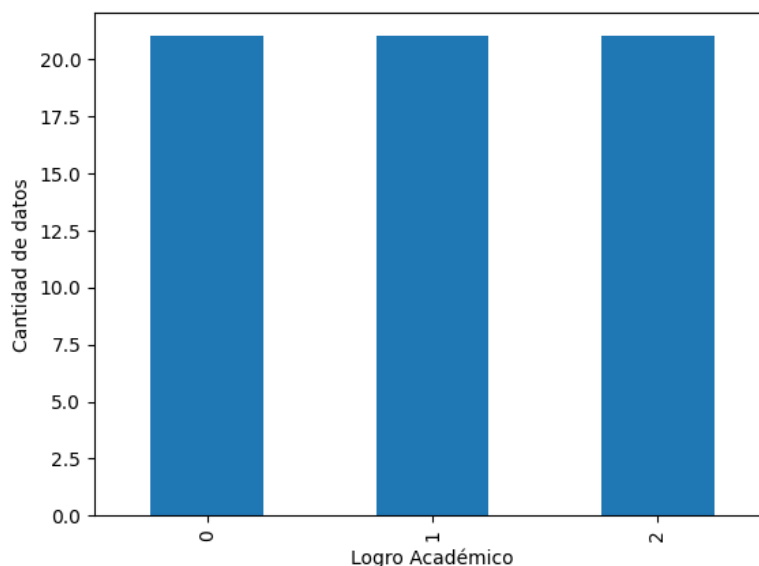
Figura 10 Precisión y exhaustividad por clases logro académico



Fuente: Autores

Para solventar dicho problema de clasificación se procede a utilizar la técnica denominada Técnica sintética de sobremuestreo de minorías (SMOTE), que genera con las clases inferiores unos puntos sintéticos o centroides para a partir de allí crear data "sintética" que complemente las demás clases. Se uso para este proceso la librería imbalanced-learn en su versión 0.9.1 para Python 3.10. Generando un balanceo de tipo under y nivelando la carga para las tres clases objeto de estudio, como se evidencia en la figura 11.

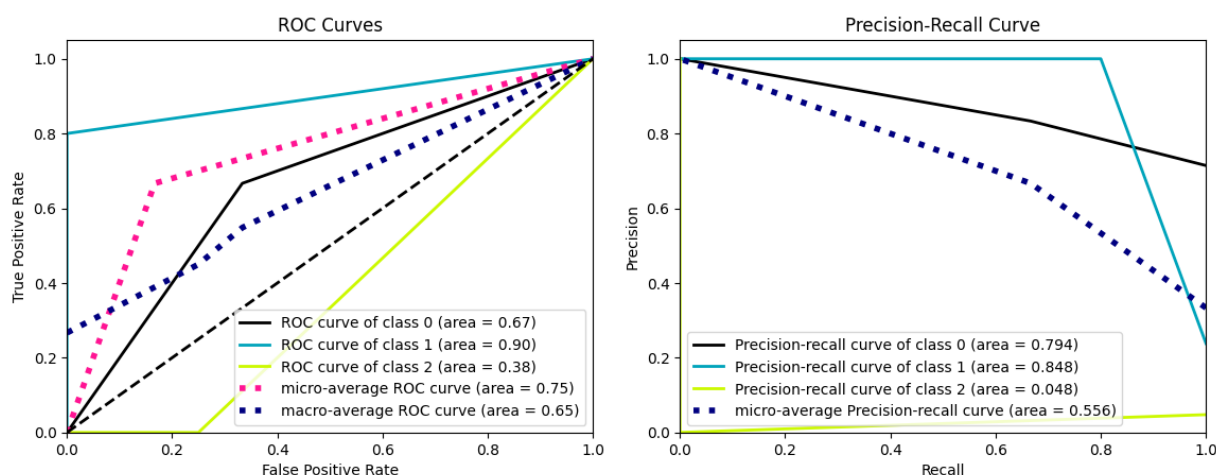
Figura 11 Uso de técnica SMOTE de tipo under samplig para logro académico



Fuente: Autores

Luego de revisar la gráfica 12 como se evidencia se generó una mejoría en la sensibilidad y especificidad de todas las clases y en cuanto a al precisión y exhaustividad la clase 1 desmejoró frente a las demás únicamente.

Figura 12 Mejora después de aplicar balanceo de clases



Fuente: Autores

#### 1.4. Test de Normalidad – Shapiro – wilk

Posteriormente se realizó la prueba de normalidad de shapiro -wilk enfocado a rechazar o aceptar la hipótesis nula  $H_0$ : Origen de los datos con distribución normal o la hipótesis alternativa  $H_1$ : Origen de los datos sin distribución normal.

Se utilizó la prueba de Shapiro - wilk para determinar la normalidad de las variables en las diferentes dimensiones del instrumento, siendo una de las pruebas más fuertes para contrastar hipótesis por su trabajo robusto frente al ruido existente en los datos.

de la distribución para determinar el valor de p, el cual si es mayor a 0.05 se asume la variable como normal de lo contrario se asume como no normal. Evidenciando que luego de la normalización efectuada ninguna variable cumple con la hipótesis nula  $H_0$  y se reconoce que la data no proviene de un universo de normalidad gaussiana como se evidencia en la tabla.

Tabla 4 Resultados test de normalidad shapiro -wilk

N°	Variable	Pregunta Asociada	Valores P
1	DemoEstrato	Nivel de estrato	1,03E-08
2	DemoEdad	¿Cuál es su edad?	2,54E-06
3	DemoGenero	¿Cuál es su género?	4,75E-03
4	DemoSemestre	¿Cuál es su semestre?	1,73E-09
5	DemoPadres	¿Vive con sus padres?	1,13E-05
6	DemoLaborando	¿En el momento en que recibió la clase espejo se encontraba laborando?	1,32E-03
7	DemoEsMental	¿Mi trabajo me exige hacer mucho esfuerzo mental?	1,63E-04
8	DemoEsFisico	¿Mi trabajo me exige hacer mucho esfuerzo físico?	9,58E-02
9	DemoTrabajo	¿Cuántas horas trabaja a la semana?	1,09E-04
10	Pedag1	¿Los profesores utilizan un sistema de puntos o créditos por participación?	1,48E-08
11	Pedag2	¿Los objetivos de la clase se plantean explícitamente?	2,83E-08
12	Pedag3	¿Existe una preparación de la clase por parte del profesor?	7,30E-07

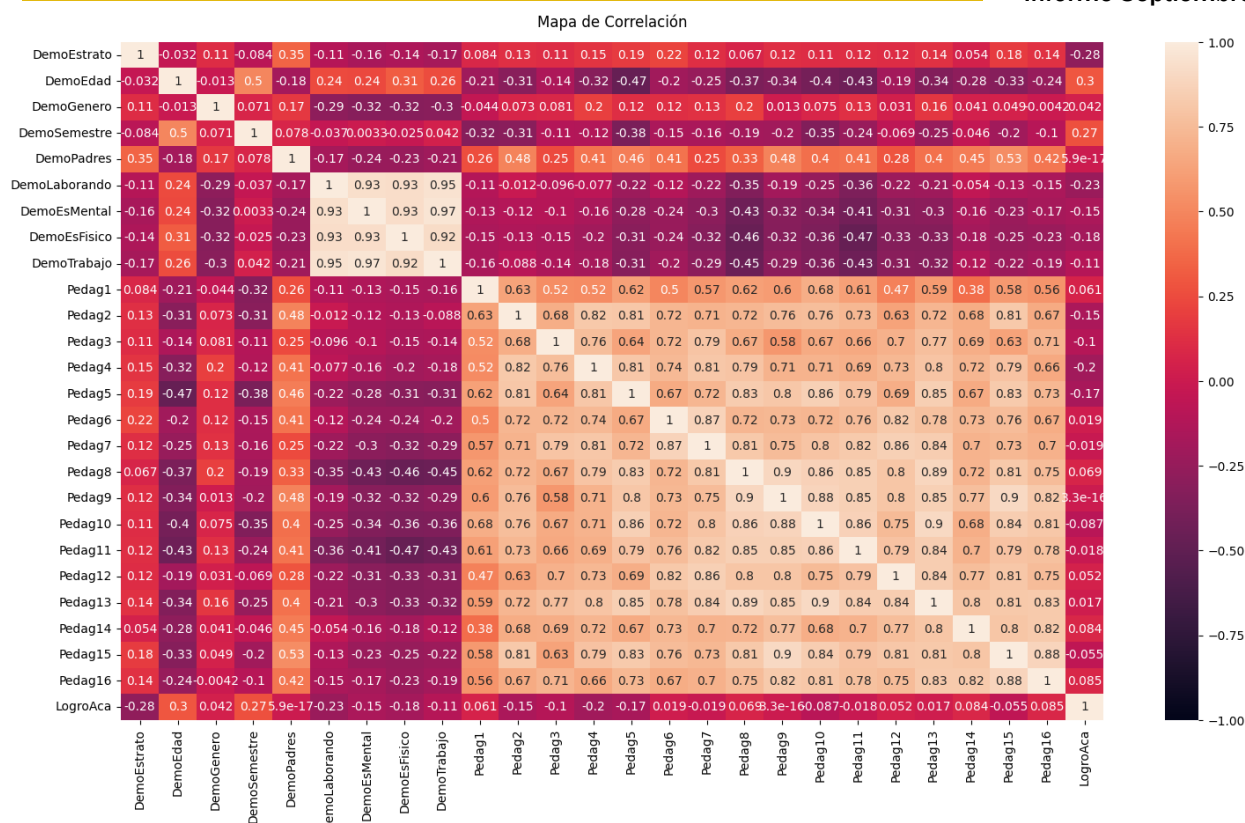
13	Pedag4	¿La clase espejo cubre las necesidades de los estudiantes?	5,26E-09
14	Pedag5	¿La puntualidad de la clase espejo se cumplió a totalidad?	3,03E-09
15	Pedag6	¿Los profesores realizan trabajos colaborativos?	1,63E-08
16	Pedag7	¿Los profesores adaptan los materiales y actividades de la asignatura a los intereses de los estudiantes?	8,57E-06
17	Pedag8	¿Los profesores explican con claridad las actividades que deben realizar los estudiantes?	1,65E-07
18	Pedag9	¿Los profesores genera instancias para que entre todos/as resolvamos las dudas que surgen durante la clase espejo?	2,12E-07
19	Pedag10	¿Los profesores atienden a todos los estudiantes de manera equitativa durante la clase espejo?	3,06E-07
20	Pedag11	¿Las actividades fuera del aula (tareas, trabajos, lecturas, etc.) nos permiten profundizar en nuestros temas de interés?	1,56E-07
21	Pedag12	En la clase espejo podemos compartir nuestras experiencias y conocer a nuestros compañeros	4,99E-08
22	Pedag13	¿Se entiende con facilidad lo que el profesor expresa?	2,71E-08
23	Pedag14	¿Los profesores realiza actividades donde los/as estudiantes que poseen más dominio acerca de algún contenido pueden ayudar a quienes aún no lo han desarrollado?	2,05E-10
24	Pedag15	¿Los profesores utilizan diversos recursos para apoyar nuestro aprendizaje?	1,08E-09
25	Pedag16	¿Los profesores disponen de un medio para comunicarnos con él/ella (web,mail, facebook, foro,etc.)?	4,10E-07

Fuente: Autores

### 1.5. Correlaciones

Finalmente, el diagrama de correlación evidencia las mejores variables normales para poder realizar procesos de clasificación y determinar cuáles son las mejores variables para poder clasificar los estudiantes según su nivel de logro para la clase espejo y así reconfigurar la estrategia pedagógica y preparación previa del docente, como se evidencia en la figura 13.

Figura 13 Mapa de correlaciones para todas las dimensiones usadas



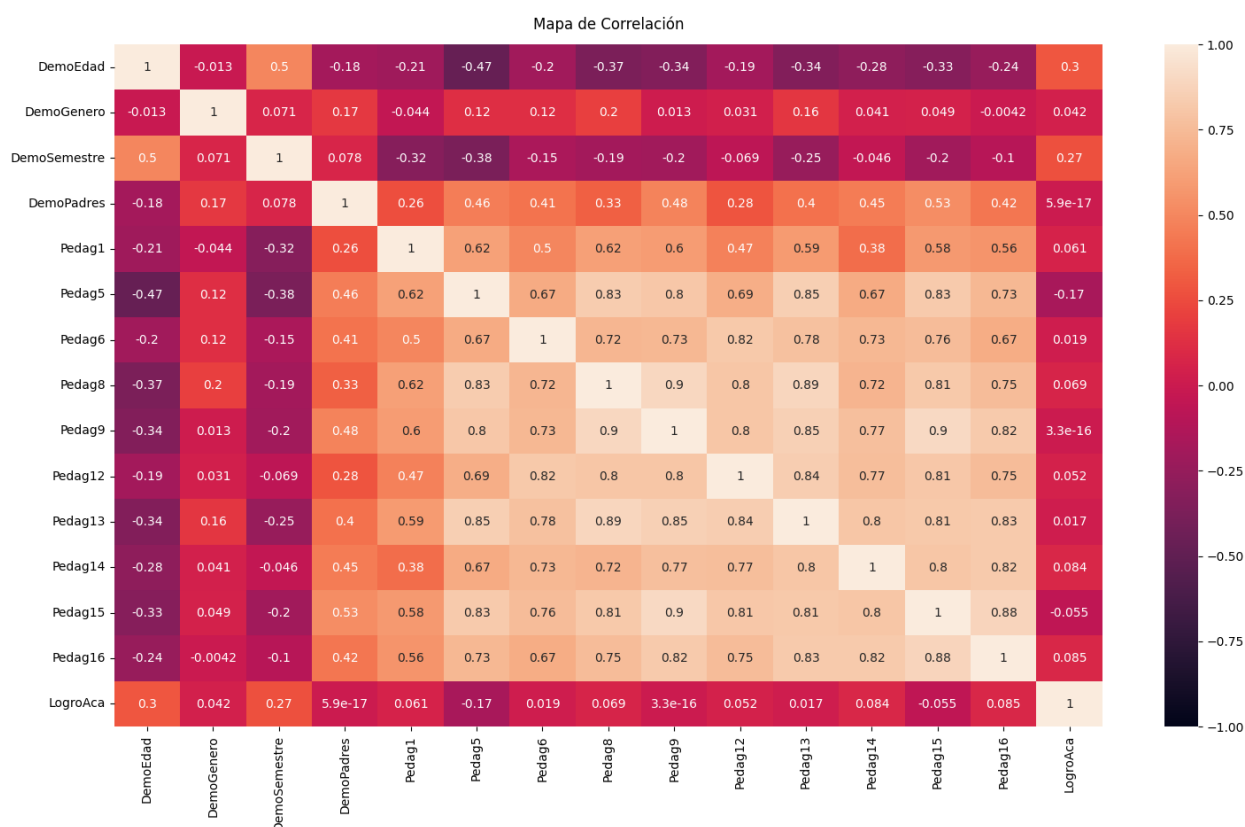
Fuente: Autores

Existe una excelente correlación entre la edad y el semestre del estudiante con un 0.58, de igual forma se evidencia la necesidad del acompañamiento de los padres en la labor pedagógicas ya que todas las preguntas pedagógicas manifestaron un rango entre 0.27 a 0.54 lo cual indica que es posible definir alguna variable pedagógica con el hecho de que los padres vivan con sus hijos. Frente al logro académico la mejor correlación se evidencia con la edad indicando la importancia del nivel de madurez y responsabilidad ya que la mayoría de los estudiantes se encuentran entre los 21 a 25 años.

Finalmente, el hecho de trabajar, el esfuerzo mental y físico son las mejores correlaciones que oscilan entre 0.89 a 0.94, se sugiere como trabajo futuro analizar las dimensiones relacionadas con el esfuerzo mental y físico en el trabajo como afectan el desarrollo académico teniendo en cuenta el horario de estudio normalmente es al final de la jornada laboral.

De acuerdo con la relación de la variable dependiente que es logro se analiza la necesidad de eliminar del set de datos de correlaciones los valores demográficos como la parte laboral, esfuerzo y estrato; por el lado de las variables pedagógicas se eliminó las correlaciones más negativas en su valor de p para estas fueron las preguntas de Pedag2, Peda3, Pedag4, Pedag7, Pedag10 y Pedag11. Quedando el sistema de correlaciones como se evidencia en la figura 14.

Figura 14 Nuevas correlaciones para el desarrollo de los modelos



Fuente: Autores

## 2. Actividad 2 Y 3: Creación de modelos clasificación y pruebas

Para la siguiente fase luego del proceso de análisis de la información, normalidad y correlación se tomó los variables altamente correlacionadas y se procede a realizar la partición en datos de entrenamiento y testeo con una relación 80 – 20 dada la naturaleza de la cantidad de información. Teniendo como variable predictora la variable LogroAca (Logro Académico) para poder clasificar a que clase pertenece el estudiante si está el alto grado, aceptable o bajo grado y así mismo conformar mejor la estrategia y preparación del docente.

Comparando los modelos usados los cuales fueron Linear Discriminant Analysis(LDA), Support Vector Classifier (SVC) y Regresión Logística Multinomial (RLOG Multinomial) en dos momentos, uno con todas las variables usadas en el cuestionario y dos con las seleccionadas de acuerdo a su alta correlación y valor de p se evidenció lo siguiente para el momento uno de acuerdo a las tablas 5,6 y 7 el algoritmo SVC muestra una eficiencia general de 0.62, seguido por el algoritmo RLOG Multinomial con 0.57 y finalizando el algoritmo LDA con 0.43. El mejor algoritmo para predecir la clase cero se cumple aceptablemente es el SVC con una exhaustividad de 0.53 lo cual indica que 2 de cada 4 estudiantes es posible clasificar en este logro de acuerdo con las variables pedagógicas y demográficas que lo denotan; por otro lado, cuenta con un valor F de 0.67 esto quiere decir que existe un balance entre precisión y recall donde 67 de 100 estudiantes pueden ser clasificados para la categoría cero de manera correcta.



Tabla 5 Métricas modelos primer momento algoritmo LDA

Algoritmo Linear Discriminant Analysis (LDA)					
Clase Original	Clase Convertida	Precisión	Exhaustividad	Valor F1	Eficiencia
Cumple Aceptablemente	0	0.80	0.27	0.40	0.43
Cumple Alto Grado	1	0.57	0.80	0.67	
Cumple en bajo grado	2	0.11	1.00	0.20	

Fuente: Autores

Tabla 6 Métricas modelos primer momento algoritmo SVC

Algoritmo Support vector classifier (SVC)					
Clase Original	Clase Convertida	Precisión	Exhaustividad	Valor F1	Eficiencia
Cumple Aceptablemente	0	0.89	0.53	0.67	0.62
Cumple Alto Grado	1	0.50	0.80	0.62	
Cumple en bajo grado	2	0.25	1.00	0.40	

Fuente: Autores

Tabla 7 Métricas modelos primer momento algoritmo RLOG Multinomial

Algoritmo Regresión logística Multinomial (RLOG multinomial)					
Clase Original	Clase Convertida	Precisión	Exhaustividad	Valor F1	Eficiencia
Cumple Aceptablemente	0	0.80	0.53	0.64	0.57
Cumple Alto Grado	1	0.50	0.80	0.62	
Cumple en bajo grado	2	0.00	0.00	0.00	

Fuente: Autores

Por otro lado, el mejor algoritmo para la clase uno de cumple en alto grado es el LDA, con una exhaustividad de 0.80 lo cual indica que 3 de cada 4 estudiantes es posible clasificar en este logro de acuerdo con las variables pedagógicas y demográficas que lo denotan; por otro lado, cuenta con un valor F de 0.67 esto quiere decir que existe un balance entre precisión y recall donde 67 de 100 estudiantes pueden ser clasificados para la categoría cero de manera correcta. Finalmente, para la clase dos cumple en bajo grado el mejor algoritmo de nuevo es SVC con un balance perfecto de 1.00 y es el único que logró evaluar precisión y recall con un 0.40.

Para el segundo momento quitando las variables nombradas anteriormente los resultados se evidencian en las tablas 8, 9 y 10. En cuanto al algoritmo LDA se aumenta en un 0.05 su eficiencia, pero desmejora la clasificación de cumple en alto grado y pierde equilibrio.

Tabla 8 Métricas modelos segundo momento algoritmo LDA



Algoritmo Linear Discriminant Analysis (LDA)					
Clase Original	Clase Convertida	Precisión	Exhaustividad	Valor F1	Eficiencia
Cumple Aceptablemente	0	0.83	0.33	0.48	0.48
Cumple Alto Grado	1	0.50	0.80	0.62	
Cumple en bajo grado	2	0.14	1.00	0.25	

Fuente: Autores

Tabla 9 Métricas modelos segundo momento algoritmo SVC

Algoritmo Support vector classifier (SVC)					
Clase Original	Clase Convertida	Precisión	Exhaustividad	Valor F1	Eficiencia
Cumple Aceptablemente	0	1.00	0.47	0.64	0.62
Cumple Alto Grado	1	0.56	1.00	0.71	
Cumple en bajo grado	2	0.20	1.00	0.33	

Fuente: Autores

Tabla 10 Métricas modelos segundo momento algoritmo RLOG Multinomial

Algoritmo Regresión logística Multinomial (RLOG multinomial)					
Clase Original	Clase Convertida	Precisión	Exhaustividad	Valor F1	Eficiencia
Cumple Aceptablemente	0	0.80	0.53	0.67	0.62
Cumple Alto Grado	1	0.50	0.80	0.62	
Cumple en bajo grado	2	0.25	1.00	0.40	

Fuente: Autores

La definición del mejor modelo supone lo que se desea para la respuesta de investigación, ya que los modelos del algoritmo SVC y RLOG Multinomial han mejorado en todas sus clases y cifras pero si se desea optimizar la clasificación de acuerdo a la obtención del logro académico en función de la pedagogía y aspectos demográficos, pero si se desea clasificar en alto grado el modelo del algoritmo SVC es el mejor pero si queremos clasificar en bajo grado son mejores los números del algoritmo RLOG Multinomial. Finalmente se extraen los 3 modelos en formato .sav para ser cargados en un ambiente de producción (Página web) y revisar su rendimiento frente a un escenario real.