

MODELO PREDICTIVO SOBRE LA DESERCIÓN EN EDUCACIÓN MEDIA POR DEPARTAMENTO EN COLOMBIA

Juan Julio, Camilo Sinning
djulioj@uninorte.edu.co, cjsinning@uninorte.edu.co

1. Introducción

El presente informe se hizo con el objetivo de dejar evidencia sobre el procedimiento llevado a cabo en el momento de desarrollar el modelo predictivo. La base de datos seleccionada es MEN ESTADÍSTICAS EN EDUCACIÓN EN PREESCOLAR, BÁSICA Y MEDIA POR MUNICIPIO ubicada en el sitio web gubernamental https://www.datos.gov.co/Educaci-n/MEN_ESTADISTICAS_EN_EDUCACION_EN_PREESCOLAR-B-SICA/nudc-7mev. La mencionada base de datos nos presenta información estadística correspondiente a los niveles educativos de preescolar, básica y media por municipios, con información oficial recolectada entre los años 2011, 2019 y datos preliminares del año 2020[1].

2. Metodología

Primeramente, haciendo uso de la página gubernamental Datos Abiertos, se seleccionó la base de datos más adecuada para el proyecto, ya con la base de datos seleccionada se realizó un análisis gráfico de los datos y las relaciones existentes entre las variables que este posee, teniendo toda esta información entonces se procedió a plantear una hipótesis para ser comprobada haciendo uso del modelo. Finalmente, se desarrolló el modelo predictivo y se usó este para la comprobación de la hipótesis.

Luego, mediante el uso de la librería seaborn disponible para Python, se graficó la relación entre las columnas relevantes para el modelo. La gráfica mostró que existía una posible relación entre APROBACION_MEDIA Y DESERCIÓN_MEDIA y, ya con esto se orientó el análisis en un sentido en el que se relacionarían estas variables.

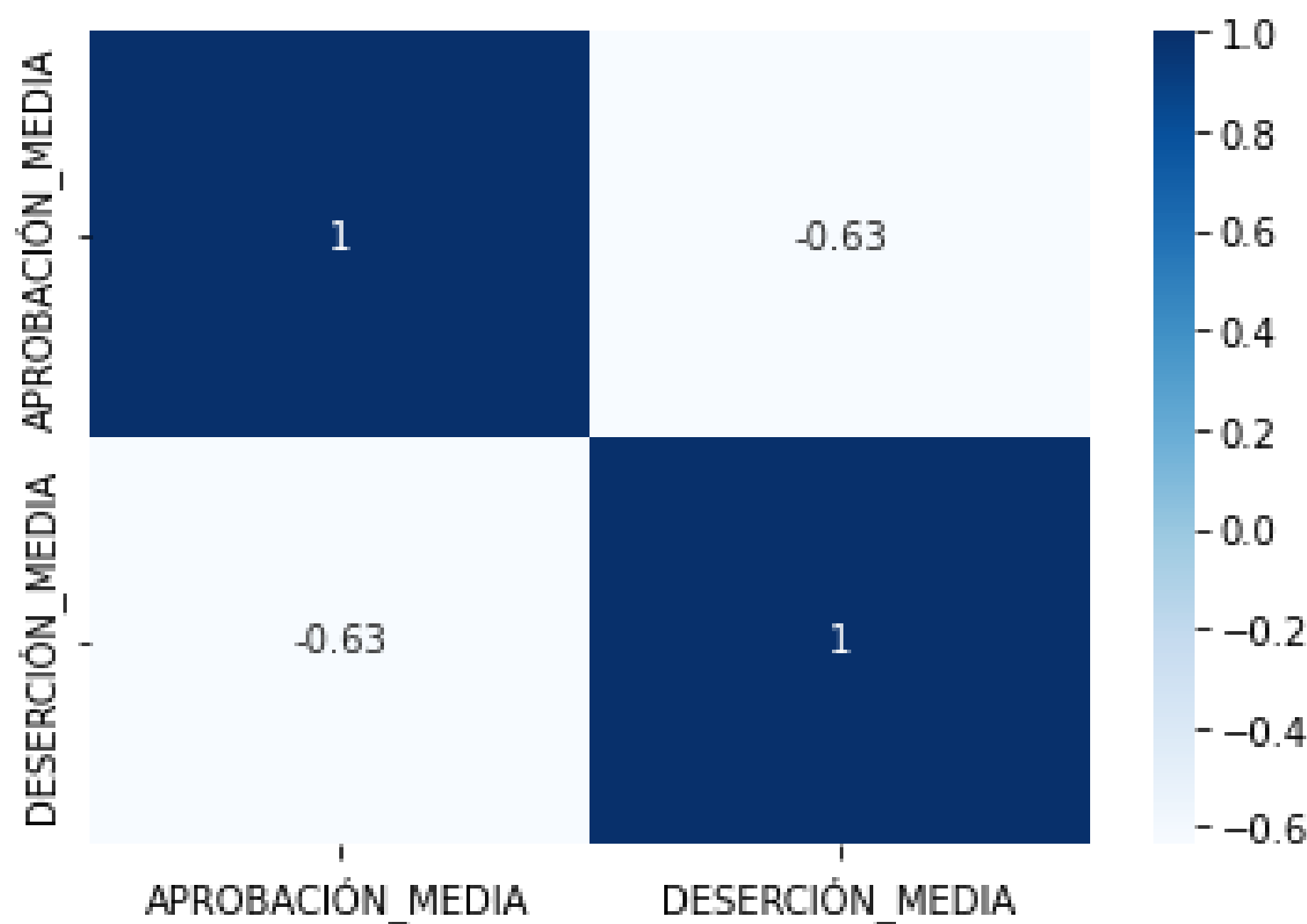


Figura 1: Gráfico de correlación entre la columna aprobación y la columna de deserción

3. Modelo

El modelo predictivo desarrollado fue limitado a analizar y predecir datos referentes a la educación media, Siendo más concreto, se usa la aprobación y el departamento en cuestión para predecir una deserción baja o alta en cierto municipio. Para el desarrollo del modelo se usó la librería pandas para cargar la base de datos y poder modificarla y usarla en Python.

3.1. Forma

El modelo hecho puede ser representado como se observa en la figura 2. En este se representan las entradas y la salida que representa 1 si se considera deserción alta o 0 si está es baja.

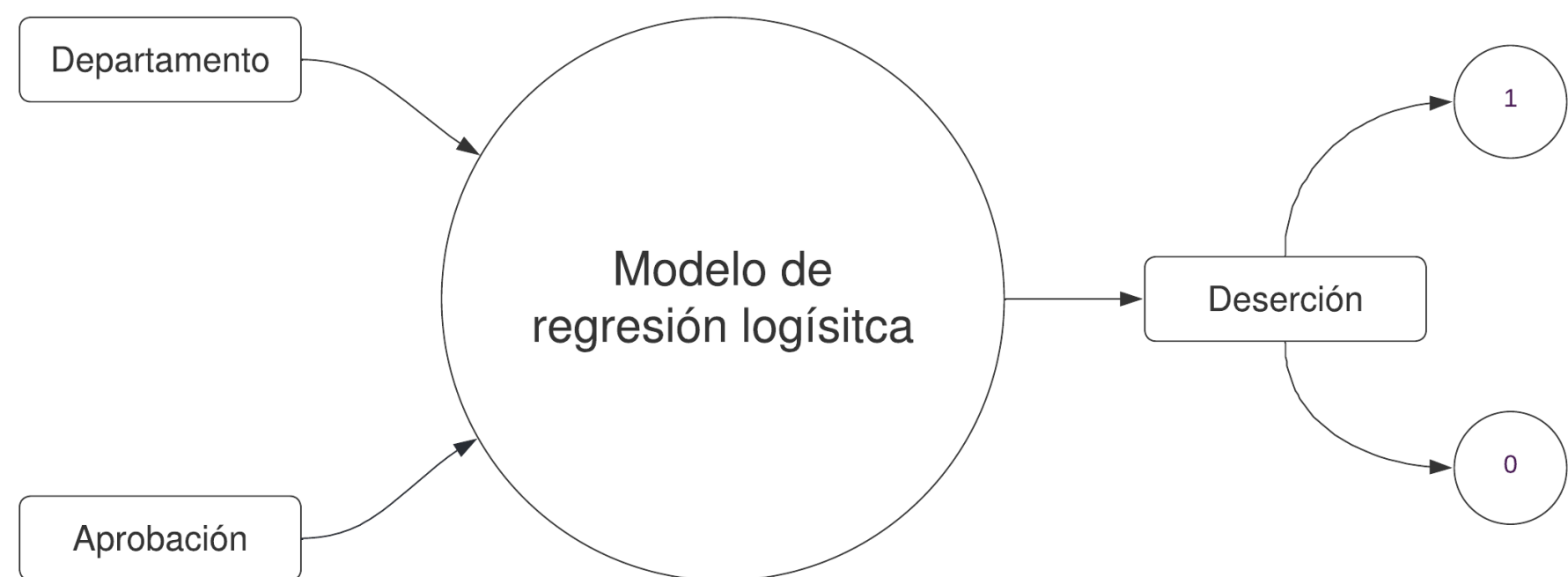


Figura 2: Modelo de regresión logística

4. Resultados

El modelo se entrenó y válido mil veces, obteniendo que el puntaje promedio que obtiene este es de un 75 % y la varianza de un 4 %, como se observa en la figura 3.

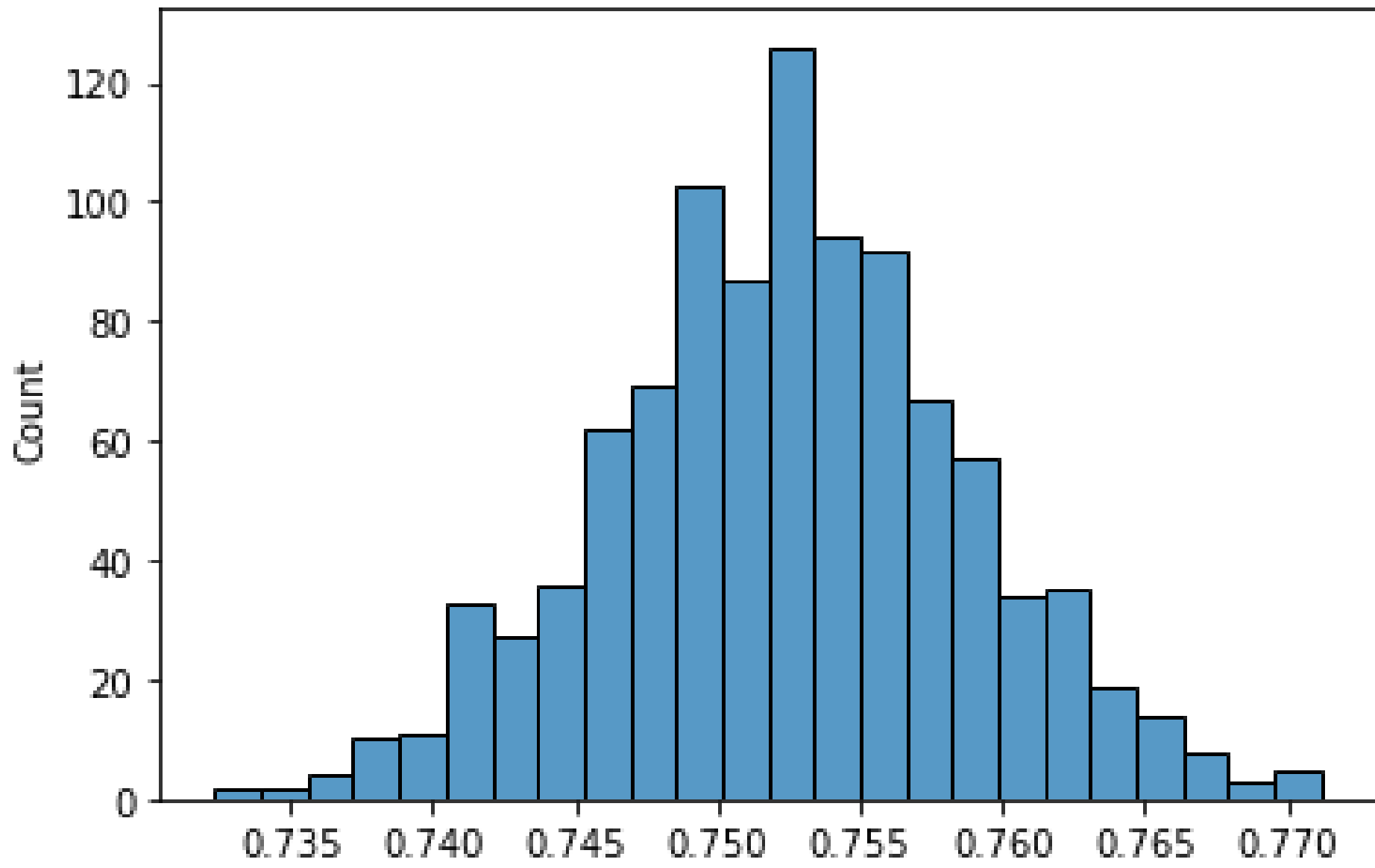


Figura 3: Gráfico de puntajes del modelo

5. Conclusión

El modelo fue desarrollado y se llegó a un resultado que se considera exitoso, ahora con este modelo entrenado se establece la veracidad de la hipótesis hecha.

Se toma como base el estudio hecho por las naciones unidas en el que se clasifican las diferentes provincias que tiene el país según su ruralidad. En la figura 4 se puede observar gráficamente esta clasificación.

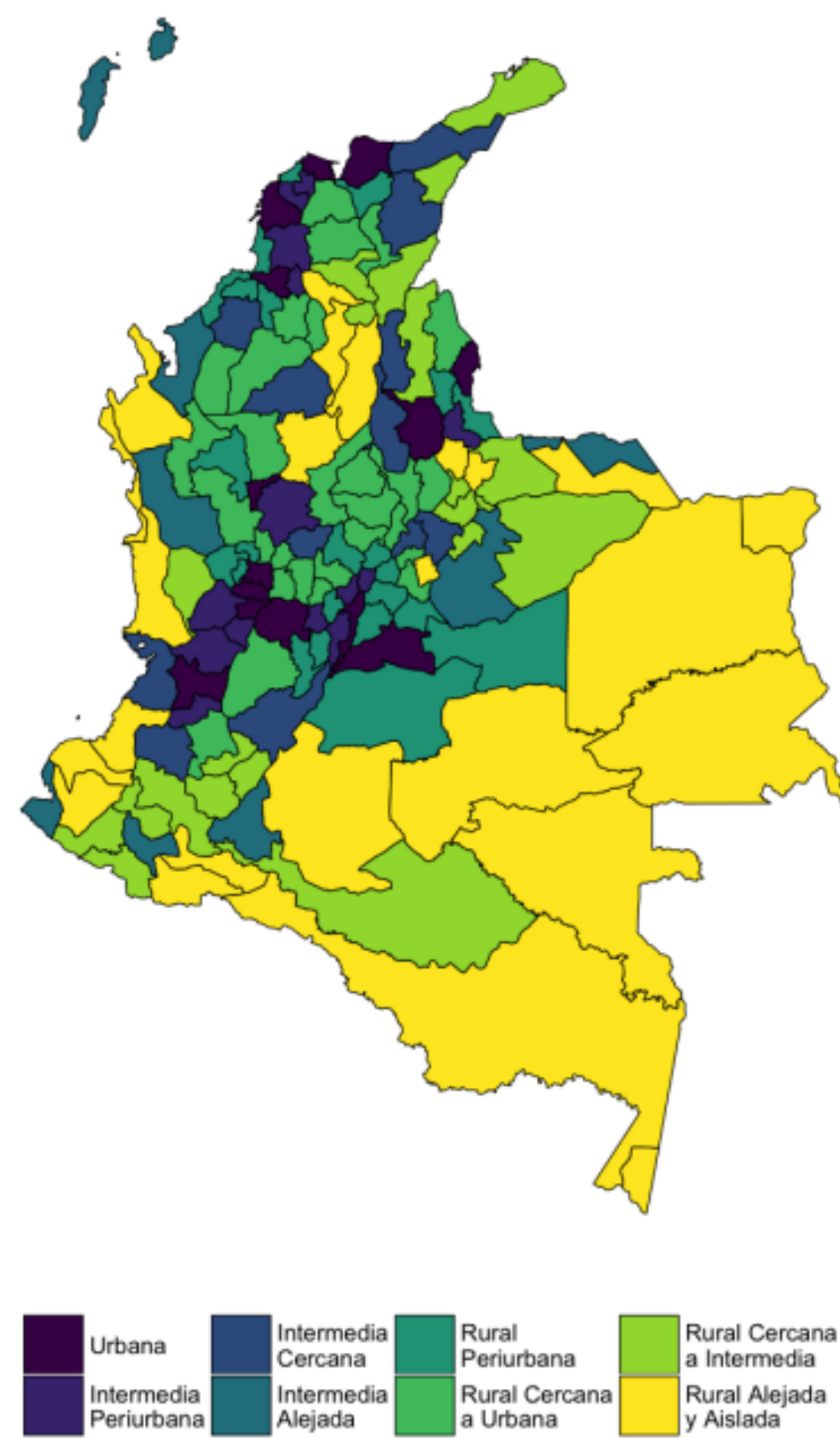


Figura 4: Clasificación de provincias por ruralidad [2]

Se tomarán los departamentos que el informe considera como rural, alejada y aisladas y se usaran municipios pertenecientes a estas para comprobar la hipótesis usando el modelo.

Se usarán los departamentos: Vichada, Guainía, Vaupés y Putumayo. Los datos indican que en un 69 % el modelo predijo que diversos municipios pertenecientes a estos departamentos tendrán deserción alta en la educación media.

Referencias

- [1] M. de Educación Nacional. Men estadísticas en educacion en preescolar, básica y media por municipio, Dec 2015.
- [2] C. E. para América Latina y el Caribe Oficina de la CEPAL en Bogotá, J. C. Rámirez, and J. M. de Aguas. Naciones Unidad, 2017.