

CAPÍTULO REIVINDICATORIO MODIFICADO

1. Un computador CARACTERIZADO por comprender un procesador configurado, a partir de:

- Data (10) de instituciones de educación superior, compuesta por data (11) de profesores de instituciones de educación superior, data (12) de infraestructura de instituciones de educación superior y data (13) de caracterización de estudiantes de instituciones de educación superior; y
- data (20) de pruebas estandarizadas de entrada y salida de la educación superior, compuesta por data (21) de los resultados de las pruebas estandarizadas de entrada a la educación superior y data (22) de los resultados de las pruebas estandarizadas de salida de la educación superior, para ejecutar, mediante dicho procesador, las siguientes operaciones:

(A) Realizar un análisis descriptivo (A) que comprende:

- Establecer un cruce (23) entre los resultados de las pruebas estandarizadas de entrada (21) y los resultados de las pruebas estandarizadas de salida (22), para determinar la correspondencia entre el año de presentación de dichas pruebas y obtener los resultados asociados a cada estudiante en ambas pruebas;
- Normalizar (24) los resultados de las pruebas estandarizadas de entrada y de salida, para garantizar una métrica común en la determinación del crecimiento en las pruebas estandarizadas de salida;
- Realizar un preprocesamiento (31) de la información de profesores (11), infraestructura (12), caracterización de estudiantes (13) y resultados normalizados (24), corrigiendo datos incorrectos, incompletos, duplicados o erróneos en el conjunto de datos;
- Con la información del preprocesamiento (31), efectuar una selección de variables (33) de interés para el modelo, dependientes e independientes, dividiéndolas en variables categóricas para índices (34) en una salida 1A y variables numéricas (35) en una salida 2A.

(B) Ejecutar un análisis predictivo (B) que comprende:

- A partir de la salida 1A, suministrar, a través de una entrada 1B, las variables categóricas para índices (34) a un módulo de índices variables (40) que aplica los métodos estadísticos PCA (Análisis de componentes principales) y MCA (Análisis de correspondencias múltiples) para determinar índices correspondientes a las variables categóricas del modelo;
- A partir de la salida 2A, suministrar, a través de una entrada 2B, las variables numéricas (35) a un módulo para determinar crecimiento en pruebas

estandarizadas de salida (50), teniendo este módulo (50) como entradas los índices generados por el módulo (40) y las variables numéricas (35), y aplicar en el módulo (50) la ecuación (Z):

$$\rho(\text{Zona de Crecimiento}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}} ,$$

en donde la ecuación (Z) es una función en la que intervienen:

- la constante e (Euler) = 2,71728...,
- un intercepto aleatorio específico de cada institución,
- coeficientes asociados a cada variable del modelo, y
- los valores de cada variable del modelo,

para determinar el crecimiento de un estudiante en las pruebas estandarizadas de salida, obteniendo un factor del modelo aplicado a los estudiantes en las pruebas estandarizadas o la medición del crecimiento de cada estudiante, de tal forma que, cuando el crecimiento del estudiante i es mayor que uno, como se indica en la ecuación (Y):

$$Ci = (Psproi, Ps11i) = \frac{Psproi}{Ps11i} ,$$

Donde,

Ci : es el crecimiento del estudiante i ;

$Psproi$: es la transformación del resultado del estudiante i en la prueba saber pro; y

$Ps11i$: es la transformación del resultado del estudiante i en la prueba saber 11,

el estudiante se encuentra en una zona de crecimiento;

- alimentar un módulo con el resultado de la ecuación (Z) obtenido en el módulo (50), para definir variables de interés del modelo (60), que establece las variables dependientes e independientes para los modelos predictivos;
- a partir de las variables de interés definidas en el módulo (60), distribuir la información en un conjunto de entrenamiento (70) y un conjunto de prueba (71) para múltiples modelos predictivos; entrenar, mediante un módulo de entrenamiento de modelo predictivo (80), los diferentes modelos predictivos con el conjunto de entrenamiento (70); probar cada modelo predictivo mediante un módulo prueba múltiples modelos predictivos (81) empleando el conjunto de prueba (71), evaluar, mediante un módulo evaluación métricas (90) las diferentes métricas de eficiencia de los modelos predictivos y seleccionar, mediante un módulo selección modelo más eficiente (100), el modelo predictivo que se va a aplicar en el proyecto.

(C) Efectuar un análisis prescriptivo (C) que comprende:

- Recibir nueva data (210) de instituciones de educación superior, compuesta por nueva data (211) de profesores, nueva data (212) de infraestructura y nueva data (213) de caracterización de estudiantes, y recibir además nueva data (220) de resultados de pruebas estandarizadas de entrada a la educación superior;
 - Normalizar (224) los resultados de las pruebas estandarizadas de entrada a la educación superior para ser aplicados en el modelo predictivo;
 - Realizar un preprocesamiento (231) de las nuevas datas (211), (212), (213) y resultados normalizados (224), corrigiendo datos incorrectos, incompletos, duplicados o erróneos en el conjunto de datos;
 - Con la información del preprocesamiento (231), efectuar una selección de variables (233) de interés para el modelo, dividiéndolas en variables categóricas para índices (234) y variables numéricas (235);
 - Aplicar un módulo de índices de variables que utiliza los métodos estadísticos PCA y MCA para generar índices (240) a partir de las variables categóricas (234);
 - Aplicar el modelo predictivo (100), seleccionado en el análisis predictivo (B), utilizando las variables numéricas (235) y los índices (240), para obtener la predicción de crecimiento (300) de los estudiantes que aplicaron pruebas estandarizadas de entrada a la educación superior; y
 - Obtener, en una salida 1C, información (350) que identifica a los estudiantes que se predice no alcanzarán crecimiento en las pruebas estandarizadas de salida.
2. El computador de la reivindicación 1, en el que:
- El conjunto de entrenamiento (70) corresponde alrededor del 80 % del total de los datos disponibles y el conjunto de prueba (71) corresponde alrededor del 20 % del total de los datos disponibles, siendo estos porcentajes ajustables de acuerdo con los objetivos de un proyecto que se esté desarrollando; y
 - el análisis prescriptivo (C) produce la predicción de crecimiento (300) y la información (350) con los estudiantes que se predice no alcanzarán crecimiento en las pruebas estandarizadas de salida.