

CAMPO TÉCNICO DE LA INVENCIÓN

La presente invención pertenece al campo técnico de Tecnologías de la información y la comunicación, específicamente en el campo técnico de gestión del aprendizaje aplicando machine learning, con arreglos basados en modelos de computación específicos, reglas para modelar estilos de aprendizaje individuales y la automatización de procesos administrativos mediante aplicaciones informáticas y métodos adaptables al aprendizaje para e-learning.

EL ARTE PREVIO

El estado de la técnica relacionado con el presente desarrollo se refiere en general a gestión del aprendizaje aplicando técnicas de tratamiento de variables categóricas, modelos de machine learning, y en particular un modelo predictivo del progreso en el aprendizaje de los estudiantes en las pruebas estandarizadas.

Las pruebas estandarizadas están relacionadas con la evaluación de aprendizajes que permite determinar qué tan bien asimilan el contenido del curso los participantes, midiendo el grado en que cambian los conocimientos, habilidades y/o actitudes como resultado de la educación que recibe y produce un estudiante. Teniendo claro este concepto, se puede establecer que la evaluación de aprendizaje es un proceso sistemático que proporciona como resultados indicadores para medir el grado de eficacia del programa de formación y la calidad de esta.

Así mismo, se realiza una prueba estándar de conocimiento impartida por el estado que monitorea el desarrollo de las competencias básicas en los estudiantes de educación básica antes de ingresar a la educación superior, entendiendo una competencia como los conocimientos flexibles que pueden actualizarse en diferentes contextos y la capacidad de aplicar los conocimientos a situaciones fuera

del entorno de aprendizaje, dando el seguimiento a la calidad del sistema educativo.

Estos resultados de pruebas estandarizadas brindan información a los investigadores de la calidad educativa sobre las destrezas, habilidades y valores que los estudiantes desarrollan durante la trayectoria escolar, las pruebas estandarizadas son impartidas en cada estado de acuerdo con una política educativa.

También estas evaluaciones tienen dos factores altamente influyentes, por un lado, el conocimiento y alcance del profesorado de las competencias específicas y generales de la titulación para realizar la práctica docente, y por otro, el modo en que el docente sepa valorar en el estudiante los resultados que ha obtenido del proceso de aprendizaje.

Consultado el estado del método, en el estado de la técnica se encuentra una diversidad de invenciones en el desarrollo de propuestas para mejorar la calidad de la enseñanza, métodos, y modelos de predicción de progresión de acuerdo con los datos de evaluación (CN105139313A), dispositivos y sistemas que analizan y recopilan información relacionada con la efectividad del sistema de evaluación educativa (CN108805448A), datos de rendimiento de los estudiantes, estilos de aprendizaje individuales, capacidad integral del estudiante (CN111598744A) analizar los aspectos débiles de cada estudiante y la situación esperada del estudio; y algunos utilizan etiquetas y agrupamientos de los atributos relacionados con el sistema de análisis de la capacidad del alumno (JP2012194507A). En esta misma línea se encuentra en otro de los documentos analizados (WO2019125106) una invención que proporciona el sistema de una plataforma educativa para el desarrollo de trayectorias de aprendizaje, por grado, por concepto o por ciclo escolar.

También se encuentran desarrollos, como el que se presenta en el documento KR20210012791A el cual proporciona un aparato, sistema y método capaz de operar un modelo de aprendizaje automático de manera más eficiente y precisa al determinar cuándo se requiere volver a aprender de un modelo de predicción de aprendizaje automático de acuerdo con los cambios ambientales.

Otro de los documentos de patente WO2019155511A1 explica como mediante la aplicación de técnicas de aprendizaje automático, un estudiante de control predictivo de modelo inverso resuelve un problema de optimización para actualizar la función objetivo y las restricciones, basado en la diferencia de características comparadas.

El desarrollo US20110177480 proporciona un método y un aparato para una plataforma de gestión del aprendizaje. Basado en la información del perfil del estudiante y la determinación de que existen similitudes entre los estudiantes, un motor de experiencia de aprendizaje proporciona una recomendación de aprendizaje individualizada para un estudiante.

Si bien es cierto, las patentes citadas muestran desarrollos en torno a sistemas, métodos, y modelos algunos usando modelos de predicción de progresión y en su mayoría encaminados a promover la optimización de la orientación del alumno y hasta definir un modelo de evaluación de la capacidad integral del estudiante. Todo lo anterior para realizar en algunos casos, una evaluación subjetiva y una evaluación objetiva de los estudiantes o medir el crecimiento de los estudiantes o brindar una educación de calidad en algunas áreas específicas u obtener los parámetros de finalización del aprendizaje para mejorar la calidad integral de la gestión, algunos en los niveles de secundaria o en la educación universitaria. No se advierte, que las patentes estudiadas cuenten con un modelo que permita predecir la relación del progreso del aprendizaje de los estudiantes en centros educativos universitarios y de sus competencias en las pruebas de estado estandarizadas de

salida, con respecto a las pruebas estandarizadas de entrada a la educación superior, a partir de la caracterización de los estudiantes aplicando técnicas de tratamiento a las variables categóricas (características psicosociales, sociodemográficas, socioeconómicas y personales) de los estudiantes y técnicas de Machine Learning.

De esta manera, el presente desarrollo proporciona una solución al problema de la calidad educativa de los estudiantes que ingresan a las universidades sin las competencias mínimas para culminar con éxito una carrera profesional y cerrar a sí las brechas educativas y el impacto social que conlleva esto. La metodología desarrolla modelos predictivos asociados a los procesos de enseñanza-aprendizaje en las Instituciones de Educación Superior, que facilitan la gestión académico-administrativa de los programas, al generar una caracterización y diagnóstico de los estudiantes que permite llevar a cabo una intervención temprana y focalizada, aplicando técnicas de tratamiento a las variables categóricas (características psicosociales, sociodemográficas, socioeconómicas y personales) de los estudiantes y modelos de machine learning, los cuales brindan una metodología para determinar el crecimiento en las competencias genéricas en las pruebas estandarizadas de salida de la educación universitaria. Las técnicas de tratamiento de variables categóricas aplican los índices MCA (Análisis de Correspondencias Múltiples) y PCA (Análisis de Componentes Principales). Los algoritmos de machine learning utilizados para el modelo predictivo, son regresión logística, redes neuronales y máquinas de vector soporte, los cuales predicen el resultado de una variable dependiente en función de las variables independientes o predictoras. El desarrollo que se presenta se centra en un modelo predictivo del progreso en el aprendizaje de los estudiantes de universidad en unas competencias genéricas, logrando así, el cierre de las brechas de los aprendizajes.

BREVE DESCRIPCIÓN DE LA INVENCION

Un procesador configurado para realizar un procedimiento para predecir el progreso de las competencias de estudiantes a partir de la data relacionada con los resultados de pruebas estandarizadas de entrada y salida que aplican las instituciones educativas competentes, con técnicas de tratamiento a las variables categóricas, que corresponden a características psicosociales, sociodemográficas, socioeconómicas y personales de los estudiantes.

El objetivo principal es describir la metodología para desarrollar modelos predictivos asociados a los procesos de enseñanza-aprendizaje en las Instituciones de Educación Superior, que permitan, desde la gestión académico administrativa de los programas, llevar a cabo una intervención temprana y focalizada.

El modelo predictivo del progreso en el aprendizaje de los estudiantes en las pruebas estandarizadas y las técnicas de Machine Learning, brindan una metodología para determinar el crecimiento en las competencias genéricas en las pruebas estandarizadas de salida de la educación universitaria. Las técnicas de tratamiento de variables categóricas aplican los índices Análisis de Correspondencias Múltiples - MCA y Análisis de Componentes Principales- PCA. Los algoritmos de machine learning utilizados para el modelo predictivo, son regresión logística, redes neuronales y máquinas de vector soporte, los cuales predicen el resultado de una variable dependiente en función de las variables independientes o predictoras.

Esta metodología centrada en el desarrollo de un modelo predictivo del progreso en el aprendizaje de los estudiantes de universidad, enfatiza en las competencias de razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés, dentro de las pruebas estandarizadas de salida de la educación superior, a partir de la información suministrada por las instituciones que aplican pruebas

estandarizadas de entrada, y la información institucional de las universidades asociada a las variables independientes del modelo.

La metodología del modelo predictivo del crecimiento en el aprendizaje cuenta con tres pasos:

1) Análisis descriptivo, donde se realiza el pre-procesamiento que contiene una limpieza de datos, una normalización de variables, unos estadísticos, unos gráficos y selección de variables de la data suministrada por las instituciones que aplican pruebas estandarizadas de entrada y salida, así como, la información institucional de las universidades referida a los profesores, la infraestructura y la caracterización de estudiantes, asociada a las variables independientes del modelo.

2) Análisis predictivo, el cual se desarrolla a partir de la construcción de índices que corresponden al tratamiento de las variables categóricas para ser utilizadas en modelos predictivos, la medición del crecimiento en las pruebas estandarizadas de salida, la definición de variables para los modelos predictivos, la definición conjuntos de datos de entrenamiento y prueba para los modelos predictivos, la evaluación de métricas y la selección del modelo más eficiente de acuerdo con los objetivos propuestos por una universidad o centros de estudios de educación superior.

3) Análisis prescriptivo, una vez definido el modelo más eficiente, se aplican nuevas datas. Con el modelo, se obtiene la predicción de crecimiento en las pruebas estandarizadas de salida de los estudiantes, facilitando así, la implementación de estrategias de acompañamiento de manera focalizada y personalizada, para la obtención de mejores resultados.

Su aplicabilidad puede determinar la probabilidad de deserción, ausentismo, demoras para alcanzar el logro del aprendizaje de un curso, entre otros procesos

académicos de los estudiantes, así como, la probabilidad de que un estudiante continúe su formación profesional a nivel posgradual.

DESCRIPCIÓN DE LAS FIGURAS

5

Fig. 1. Enseña el modelo predictivo del progreso en el aprendizaje de los estudiantes en las pruebas estandarizadas aplicando técnicas de Machine Learning.

10

Fig. 2. Ilustra un diagrama de flujo de un análisis descriptivo (A)

Fig. 3. Ilustra un diagrama de flujo de un análisis predictivo (B)

Fig. 4. Ilustra un diagrama de flujo de un análisis prescriptivo (C)

15

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

Un procesador configurado para realizar el siguiente procedimiento: Un análisis descriptivo (A), un análisis predictivo (B), un análisis prescriptivo (C).

20

En el análisis descriptivo (A) tiene dos entradas, donde la primera corresponde a la data (10) de las instituciones de educación superior, la cual está compuesta de la data (11) de profesores de instituciones de educación superior (Entrada 3A), data (12) infraestructura de instituciones de educación superior (Entrada 2A) y la data (13) la caracterización de estudiantes de instituciones de educación superior (Entrada 1A) y la segunda entrada corresponde a la data (20) de las pruebas estandarizadas de entrada y salida, la cual está compuesta por la data (21) de los resultados de las pruebas estandarizadas de entrada a la educación superior (Entrada 4A) y la data (22) de los resultados de las pruebas estandarizadas de salida de la educación superior (Entrada 5A).

30

Para determinar la correspondencia entre el año de la presentación de la pruebas estandarizada de entrada a la educación superior y la pruebas estandarizada de salida, se establece el cruce (23) de las datas pruebas estandarizadas de entrada a la educación superior (21) y pruebas estandarizada de salida (22), el cual permite obtener los datos asociados a cada estudiante en las pruebas estandarizadas de entrada a la educación superior y pruebas estandarizada de salida. Los resultados de las pruebas estandarizadas de entrada a la educación superior y pruebas estandarizada de salida se deben normalizar (24) para garantizar una métrica común a la hora de determinar el crecimiento en las pruebas estandarizada de salida.

Posteriormente se realiza un preprocesamiento (31) de las datas (11), (12), (13) y (24), proceso donde se pueden corregir datos incorrectos, incompletos, duplicados o erróneos en un conjunto de datos. Con la información del preprocesamiento (31), se establecen los estadísticos y gráficos (32) descriptivos para las variables de estudio. De igual manera, con la información del preprocesamiento (31) se hace la selección de variables (33) de interés para el modelo (dependientes e independientes), las cuales se dividen en variables categóricas para índices (34) en Salida 1A y variables numéricas (35) en Salida 2A.

El análisis predictivo (B) tiene dos entradas, donde la primera corresponden a la data de las variables categóricas para índices (34) desde la salida 1A del análisis descriptivo (A) a través de la entrada 1B, la cual corresponde a la entrada del módulo de índices variables (40) y la segunda entrada corresponde a la data de las variables numéricas (35) desde la salida 2A del análisis descriptivo (A) a través de la entrada 2B, la cual corresponde a la entrada del módulo para determinar crecimiento en PES (50).

El módulo de índices variables (40) aplica los métodos estadísticos PCA (Análisis de componentes principales) y MCA (Análisis de correspondencias múltiples), para determinar los índices correspondientes a las variables categóricas del modelo. El módulo determinar crecimiento en pruebas estandarizada de salida (50) tiene como entradas la data de salida del módulo de índices variables (40) y la data de las variables numéricas (35). Para determinar el crecimiento de un estudiante en las pruebas estandarizada de salida, el cual se puede describir como la probabilidad de que un estudiante se encuentre en la zona de crecimiento, se aplica la ecuación (Z).

$$p(\text{Zona de Crecimiento}) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\dots+\beta_kX_k)}} \quad (Z)$$

Donde:

- e (Euler) = 2,71728....
- β_0 = Corresponde al intercepto aleatorio específico de cada institución
- β_i para $i = 1, \dots, k$. $k \in N$, es el coeficiente asociado a cada variable en el modelo
- X_i para $i = 1, \dots, k$. $k \in N$, corresponde a cada variable en el modelo

Con la aplicación de la ecuación se obtiene un resultado que representa un factor significativo del modelo aplicado a los estudiantes en las pruebas estandarizadas, es decir, la medición del crecimiento de cada estudiante, donde se interpreta que cuando el crecimiento del estudiante i sea mayor que uno ($C_i > 1$), como se indica en la ecuación (Y), se encuentra en zona de crecimiento.

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

- Ci : crecimiento del estudiante i.
- $Psproi$: transformación del resultado del estudiante i en la prueba saber pro.
- $Ps11i$: transformación del resultado del estudiante i en la prueba saber 11.

Este modelo se aplica para determinar la distribución de crecimiento en los resultados de las pruebas estandarizadas.

El módulo para definir las variables de interés del modelo (60) tiene como entrada la data de salida del módulo para determinar crecimiento en pruebas estandarizada de salida (50), donde se establecen las variables dependientes e independientes para los modelos predictivos de acuerdo con los objetivos en un proyecto que se esté desarrollando. Con la data de salida del módulo variables de interés del modelo (60), se realiza la distribución de la data de entrenamiento (70) y la data de prueba (71) para los múltiples modelos predictivos. El módulo entrenamiento del modelo predictivo (80) realiza el entrenamiento de los diferentes modelos predictivos con la data de entrenamiento (70), la cual corresponde al 80% del total de los datos disponibles. Este porcentaje se puede ajustar de acuerdo con los objetivos de un proyecto que se esté desarrollando. El módulo prueba múltiples modelos predictivos (81) realiza la prueba de cada modelo predictivo para tener una medida de la eficiencia del modelo entrenado, esta prueba se realiza con el 20% del total de los datos disponibles o con el porcentaje ajustado de acuerdo con los objetivos de un proyecto que se esté desarrollando. El módulo evaluación métricas (90) realiza la comparación de los resultados de las diferentes métricas de eficiencia de los modelos predictivos. El módulo selección modelo más eficiente (100), determina el modelo predictivo a utilizarse en el proyecto.

El análisis prescriptivo (C) tiene dos nuevas datas, donde la primera corresponden a la nueva data (210) de las instituciones de educación superior, la cual está compuesta de la nueva data (211) de profesores de las instituciones de educación superior (entrada 1C), nueva data (212) infraestructura de las instituciones de educación superior (entrada 2C) y la nueva data (213) la caracterización de estudiantes de las instituciones de educación superior (entrada 3C) y la segunda entrada corresponde a la nueva data (220) de las pruebas estandarizadas de entrada a la educación superior (entrada 4C). Los resultados de las pruebas estandarizadas de entrada a la educación superior se deben normalizar (224) para ser aplicadas en el modelo predictivo. Posteriormente, se realiza el preprocesamiento (231) de las nuevas datas (211), (212), (213) y (224), proceso

donde se pueden corregir datos incorrectos, incompletos, duplicados o erróneos en un conjunto de datos.

Con la información del preprocesamiento (231), se hace la selección de variables (233) de interés para el modelo (dependientes e independientes), las cuales se dividen en variables categóricas para índices (234) y variables numéricas (235).

El módulo de índices de variables aplica los métodos estadísticos PCA (Análisis de componentes principales) y MCA (Análisis de correspondencias múltiples), aplicando (240), genera los índices a partir de las variables categóricas (234). Con la información de variables numéricas (235) e índices (240) se aplican al modelo predictivo más eficiente (100), obteniendo la predicción de crecimiento (300) de los estudiantes que aplicaron pruebas estandarizadas de entrada a la educación superior, nueva data. En la salida 1C se obtiene la data con los estudiantes que se predice, no alcanzarán crecimiento en las PES (350).