

Proyecto de Integración y Automatización de Datos para IA para el proyecto Brújula Uni: Facilitando Decisiones Informadas en la Educación Superior.

Alexander Victoria¹, Elthon Brayan Marquez², Maria Alejandra Fandiño³

¹Facultad de Ingeniería y Ciencias Básicas

Universidad Central

Maestría en Analítica de Datos

Automatización e Integración de Datos para IA

Bogotá, Colombia

{¹avictoria, ²emarquezb}@correo1.com, ³mfandinom@correo2.com

October 7, 2023

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (<i>Primera entrega</i>)	3
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	5
2.2.1	Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>)	5
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	5
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	6
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	6
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	7
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	7
3.2	¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	8

3.3	¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (<i>Primera entrega</i>)	8
3.4	¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	8
4	Diseño de integración y Automatización de Datos para IA (Diagrama) (<i>Primera entrega</i>)	9
5	Bibliografía	10

1 Introducción (Max 250 Palabras) - (*Primera entrega*)

En la actualidad, el acceso a la educación superior es una decisión crítica que puede determinar el futuro de los estudiantes. Sin embargo, la elección de una universidad y una carrera se ha vuelto compleja y abrumadora debido a la diversidad de opciones y la abundancia de información disponible sobre carreras, universidades, etc. Los estudiantes se enfrentan a una marea de información, desde resultados académicos hasta costos de matrícula y opiniones de otros estudiantes. Surge así la necesidad de crear una herramienta que simplifique este proceso y permita a los futuros estudiantes tomar decisiones educativas más informadas.

La elección de una carrera universitaria es una de las decisiones más significativas en la vida de un estudiante que quiera continuar con la educación superior, implicando una inversión considerable de tiempo y recursos, así como la construcción de un camino hacia metas personales y profesionales. La pregunta clave que enfrentan los estudiantes es: ¿Cómo puedo tomar una decisión tan importante de manera informada y precisa?

Para abordar esta pregunta, se propone el desarrollo de un proyecto de automatización centrado en la creación de un comparador de universidades por carrera. Esta plataforma busca proporcionar a los futuros estudiantes una herramienta integral y fácil de usar que les permita evaluar de manera efectiva las opciones disponibles en función de múltiples factores clave. Desde el desempeño académico de las universidades hasta los detalles específicos de los programas de estudio y las experiencias de otros estudiantes, esta herramienta busca ofrecer una visión completa y precisa de las alternativas educativas que actualmente dispone nuestro país.

En un mundo impulsado por la información y la tecnología, es esencial contar con herramientas que simplifiquen la toma de decisiones y ayuden a los estudiantes a plantear una perspectiva clara de la educación superior. El presente proyecto busca abordar este desafío, brindando a los futuros estudiantes una brújula confiable para explorar sus opciones académicas y tomar decisiones que estén en consonancia con sus objetivos y aspiraciones.

2 Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (*Primera entrega*)

El proyecto de investigación titulado "Brújula Uni: Facilitando Decisiones Informadas en la Educación Superior" se enfoca en el desarrollo de un comparador de universidades por carrera que hace uso de técnicas de Integración y Automatización de Datos. Su objetivo principal es proporcionar a los estudiantes una herramienta

robusta y de fácil acceso que les permita evaluar de manera efectiva las múltiples opciones disponibles en función de una amplia variedad de factores clave, que van desde el desempeño académico de las universidades hasta detalles específicos de los programas de estudio y las experiencias de otros estudiantes.

Para lograr este propósito, se han definido objetivos específicos que guiarán el desarrollo y la implementación de la plataforma "Brújula Uni". En primer lugar, se llevará a cabo la obtención y almacenamiento exhaustivo de datos actualizados sobre las carreras de educación, incluyendo resultados ponderados de las pruebas Saber Pro hasta el año 2022, costos de pregrado, duración de semestre, modalidad de estudio, entre otros. Este proceso se realizará a través de fuentes de datos abiertas y de acceso público, garantizando la legalidad y ética en el uso de la información, adicionalmente se implementará un proceso de web scraping para obtener información no estructurada sobre los planes de estudio de las carreras y las opiniones de estudiantes. Esta técnica permitirá enriquecer la información disponible y proporcionar una visión más completa de cada programa académico.

En segundo lugar, se diseñará y aplicará un sistema de ponderación que evaluará y asignará valores a las diversas características de las carreras y las universidades. Esto permitirá a los futuros estudiantes tener una visión equilibrada y objetiva de las opciones disponibles, facilitando así la comparación y toma de decisiones.

En tercer lugar, se creará una página web orientada al usuario, intuitiva y de fácil navegación. Esta plataforma será el punto de acceso principal para los usuarios, donde podrán seleccionar una carrera y comparar universidades en función de las características ponderadas. Se prestará especial atención al diseño de la interfaz para asegurar que la experiencia del usuario sea óptima y efectiva.

Es importante destacar que este proyecto se centra en simplificar la toma de decisiones en el ámbito de la educación superior y no pretende reemplazar la orientación profesional o académica proporcionada por expertos en el campo. La plataforma "Brújula Uni" será una herramienta complementaria que busca empoderar a los estudiantes con información valiosa y precisa para que puedan tomar decisiones educativas informadas y alineadas con sus objetivos y aspiraciones.

2.1 Título del proyecto de investigación (Max 100 Palabras) - (Primera entrega)

Brújula Uni: Facilitando Decisiones Informadas en la Educación Superior.

2.2 Objetivo general (Max 100 Palabras) - (*Primera entrega*)

Desarrollar un comparador de universidades por carrera que permita a los futuros estudiantes tomar decisiones informadas y fundamentadas sobre su educación superior, por medio de un sistema automatizado que facilitará la evaluación de las diferentes opciones educativas al considerar múltiples factores, desde resultados académicos hasta características específicas de los programas de estudio, localización, puntajes y opiniones de estudiantes.

2.2.1 Objetivos específicos (Max 100 Palabras) - (*Primera entrega*)

- Obtener y almacenar de manera exhaustiva datos actualizados sobre las carreras de educación, incluyendo resultados ponderados de las pruebas Saber Pro hasta 2022, costos de pregrado, duración de semestre, modalidad de estudio, entre otras.
- Diseñar y aplicar un sistema de ponderación que evalúe y asigne valores a las diversas características de las carreras y las universidades, teniendo en cuenta su relevancia en la toma de decisiones de los estudiantes.
- Crear una página web orientada al usuario intuitiva y de fácil navegación que permita a los futuros estudiantes seleccionar una carrera y comparar universidades en función de las características ponderadas, proporcionando así una visión clara y objetiva de las opciones disponibles.
- Implementar un proceso de web scraping para obtener información no estructurada sobre los pensums académicos de las carreras y opiniones de estudiantes, enriqueciendo de esta manera la información disponible y brindando una visión más completa de cada programa académico.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

El alcance de este proyecto se concentra en el desarrollo de "Brújula Uni", una plataforma web diseñada con la finalidad de simplificar la toma de decisiones en el ámbito de la educación superior, su objetivo principal es permitir a futuros estudiantes tomar decisiones informadas al evaluar una amplia variedad de factores, como logros académicos, características de los programas de estudio, ubicación geográfica, costos y valoraciones de otros estudiantes. Para lograr esta misión, se llevará a cabo la recopilación y el almacenamiento exhaustivo de datos detallados sobre programas académicos y universidades. Se aplicará un sistema de ponderación que valorará estos factores de manera estratégica. Además, se creará una plataforma web intuitiva que permitirá a los usuarios comparar opciones de manera eficaz, la implementación de esta plataforma se logrará mediante la puesta en marcha de una máquina virtual que facilite la

conexión a las bases de datos mencionadas previamente, automatizando así los procesos que respaldan la funcionalidad de la página web, adicionalmente, se empleará la técnica de web scraping para enriquecer aún más la información disponible con detalles sobre planes de estudio y experiencias estudiantiles.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿En qué medida puede la automatización de datos, la integración de información sobre carreras y universidades, y el uso de técnicas de web scraping mejorar la calidad de la información disponible para los estudiantes que buscan orientación en la elección de su educación superior?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

Con este proyecto, se espera demostrar que una manera condensada de mostrar la información por medio de una plataforma web las diferentes puntuaciones que obtienen las universidades según las variables a estudiar por parte del usuario que en este caso serían los estudiantes, de este modo se realizaría un análisis más detallado por parte del usuario que le permitiera tomar decisiones informadas acerca de las universidades y/o carreras disponibles en Colombia.

3 Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (*Primera entrega*)

Para llevar a cabo este proyecto, se recurrió a diversas fuentes de información, todas ellas de acceso público, ya que fueron proporcionadas por distintas entidades gubernamentales. Inicialmente, se empleó la base de datos con los resultados

anonimizados de la prueba Saber Pro del período comprendido entre los años 2018 a 2022. Posteriormente, se utilizó otra base de datos que contiene a

los estudiantes matriculados en educación superior para Colombia en 2022; esta base es suministrada por el Ministerio de Educación Nacional del país basada en los reportes de las propias instituciones de educación superior a través del Sistema Nacional de Información de la Educación Superior (SNIES). Finalmente, se incorporó la base de datos del SNIES que proporciona información

sobre los programas académicos ofrecidos en el área de ciencias de la educación en el país. Sobre todas las bases de datos, se realizó un filtro a la información

seleccionando únicamente los programas profesionales de ciencias de educación. Es importante destacar que durante el proceso de análisis, se identificaron

datos faltantes en algunos registros, particularmente en los resultados de las pruebas Saber Pro y en la información de los costos de matrícula de algunas universidades.

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (*Primera entrega*)

Las bases de datos fueron recopiladas de las siguientes fuentes de información:

- **Datos abiertos:** Según el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), la información pública se dispone en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento. (MinTIC, [2021])
- **SNIES:** Según el Ministerio de Educación Nacional, el Sistema Nacional de Información de la Educación Superior (SNIES) es un sistema de información que ha sido creado para responder a las necesidades de información de la educación superior en Colombia. En este sistema se recopila y organiza la información relevante sobre la educación superior que permite hacer planeación, monitoreo, evaluación, asesoría, inspección y vigilancia del sector. (Ministerio de Educación Nacional, [2021])

3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (Primera entrega)

Los datos abiertos son información pública que se puede reutilizar bajo licencia abierta y sin restricciones legales en Colombia. Esta regulación está respaldada por la Ley 1712 de 2014, la cual establece la obligación de las entidades públicas de poner a disposición datos y define los datos abiertos en su numeral sexto como: "Todos aquellos datos primarios o sin procesar, que se encuentran en formatos estándar e interoperables que facilitan su acceso y reutilización, los cuales están bajo la custodia de las entidades públicas o privadas que cumplen con funciones públicas y que son puestos a disposición de cualquier ciudadano, de forma libre y sin restricciones, con el fin de que terceros puedan reutilizarlos y crear servicios derivados de los mismos". (Ley 1712 de 2014, Colombia)

3.3 ¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (Primera entrega)

Los retos en la Integración y Automatización de Datos para el desarrollo del proyecto incluyen la complejidad y calidad de la información, la gestión de procesos de web scraping, la protección de datos, la integración de múltiples fuentes y el diseño preciso de ponderaciones. Estos desafíos deben abordarse para garantizar que "Brújula Uni" ofrezca a los futuros estudiantes información precisa y relevante, simplificando así la toma de decisiones sobre educación superior y proporcionando una guía confiable para sus elecciones académicas.

3.4 ¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) - (Primera entrega)

Esperamos que la automatización y la integración de datos mejoren la calidad de la información para los estudiantes que buscan orientación en la elección de su educación superior, esto simplificará la recopilación y actualización de datos, ahorrando tiempo y recursos en comparación con consultas sobre carreras efectuadas manualmente. Además, el uso de técnicas de web scraping enriquecerá la información disponible, proporcionando detalles sobre planes de estudio y experiencias estudiantiles para ofrecer una visión más completa de cada programa académico. En última instancia, esta implementación sentará las bases para la plataforma "Brújula Uni", permitiendo a los futuros estudiantes tomar decisiones educativas más informadas.

4 Diseño de integración y Automatización de Datos para IA (Diagrama) *(Primera entrega)*

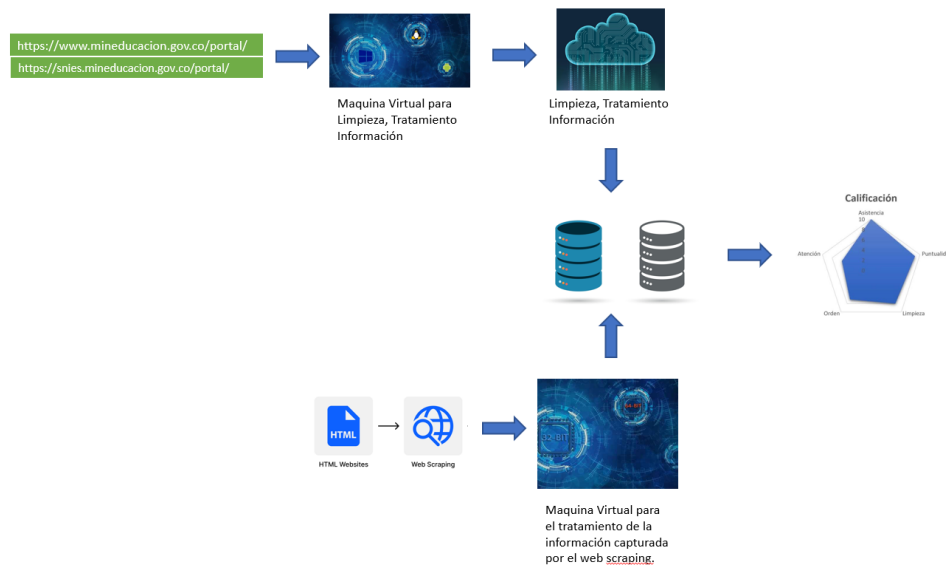


Figure 1: Diseño de integración.

5 Bibliografía