



Departamento de Ciencias Básicas e Ingenierías

PROTOCOLO

Sistema Unificado de Recomendación Escolar

Presentado por

170300075 - Kenneth Díaz González

180300366 - Juan Antonio Yam Lino

180300326 - Saúl Alberto Durán Gómez

Programa Educativo

Ingeniería en datos e inteligencia organizacional

Titular de la Asignatura

Dra. Mirbella Teresa Gallareta Negrón

FECHA

Cancún, Quintana Roo a 14 de septiembre del 2022

Contenido

1. Marco Contextual	1
1.1. Antecedentes	2
1.2. Situación actual	3
1.3. Problema	6
2. Marco Teórico	7
2.1. Web scraping	7
2.2. Selenium	7
2.2.1. Selenium Web Driver	8
2.3. Selenium Grid	9
2.4. Shiny	9
2.4.1. Biblioteca bs4Dash	10
2.4.2. Biblioteca ggplot2	10
2.5. FastApi	11
2.6. Azure - infraestructura como servicio (Azure Ubuntu VM)	12
2.6.1. Microsoft Azure para SURE	12
2.7. Azure - Plataforma como servicio (CosmoDB)	13
2.8. APP Web (HTTP, HTML)	13
2.8.1. HTTP	13
2.8.2. HTML	14
2.9. diagramas de sirena (mermaid diagrams)	14
2.10. Base de Datos NoSQL (mongoDB)	15
2.10.1. Mongo DB	15
3. Estado del Arte	16
3.1. Using Association Rules for Course Recommendation	16

3.2.	Recommendation System for Continuing Education Courses	17
3.3.	Courses Recommendation Algorithm Based On Performance Prediction In E-Learning	17
3.4.	Predicción del rendimiento en una asignatura empleando la regresión logística ordinal	18
4.	Descripción del proyecto	19
4.1.	Propuesta	20
4.2.	Justificación	20
4.3.	Objetivos	21
4.3.1.	Objetivo General	21
4.3.2.	Objetivos Específicos	21
4.4.	Hipótesis	21
4.5.	Metodología	22
5.	Plan de trabajo	23
6.	Riesgos	24
7.	Referencias	25
8.	Anexos	29

1. MARCO CONTEXTUAL

La Universidad del Caribe es una institución de educación superior que tiene como misión formar a profesionales que tengan los conocimientos y habilidades necesarias adquiridas en las aulas con la finalidad de desarrollarse como buenos profesionales (Universidad del Caribe, n.d.). Para lograrlo, se han ido implementando mecanismos que le han valido una serie de reconocimientos de calidad por organismos tanto nacionales como internacionales. Uno de los mecanismos desarrollados ha sido la implementación de su Sistema de Gestión de Calidad con base en la norma ISO 9001:2015. Su objetivo se centra en desarrollar y proveer servicios educativos de calidad mediante una serie de procesos internos en el que se incluye el Programa de Acompañamiento Estudiantil (PAE), dicho programa está diseñado para evitar, en medida de lo posible, la deserción estudiantil a través de la asignación de tutorados a los estudiantes que se encuentran en alguna situación de riesgo académico. Hasta el momento no existe un sistema que actúe de manera preventiva, por lo que hay una necesidad latente de una herramienta auxiliar que permita a los alumnos recibir recomendaciones curriculares para disminuir el índice de reprobación, y con ello, los casos críticos que requieren tutoría individual.

1.1 Antecedentes

Desde su fundación, la universidad estableció un modelo educativo flexible y centrado en aprendizajes que permita al estudiantado ajustarse a sus dimensiones personales, profesionales, sociales y ambientales (Comité Académico de la Universidad del Caribe, 2003, pág. 2). Esta flexibilidad permite a los estudiantes diseñar una carga curricular que corresponda a diferentes horarios, trayectorias, preespecialidad y pudiendo matricular de tres a nueve asignaturas simultáneamente en el semestre. Cabe resaltar que los programas educativos están diseñados para que idealmente se concluyan en 4 años, pero este es un caso que puede o no seguirse a la perfección dado que existen factores que pueden afectar negativamente el tiempo estimado de egreso de un estudiante, como los retrasos por reprobación, los compromisos laborales y las cargas curriculares con menos asignaturas de las recomendadas por el mapa curricular ideal (R. Berges, E. Mann, J. Zavala 2021).

En 2008, la Universidad del Caribe a través del Programa de Acompañamiento Estudiantil lanzó a producción la primera versión del SAES, un sistema auxiliar que permite gestionar los tutorados que son asignados entre profesores de tiempo completo y estudiantes que se encuentran en situación de riesgo académico. Desde el principio, este se diseñó para funcionar de manera reactiva, es decir, solo entra en acción cuando los estudiantes han alcanzado un estatus que puede conllevar una eventual deserción estudiantil. Es por ello que es importante, en medida de lo posible, atender las causas que conllevan caer en esta situación, las cuales ya se han descrito por el Programa Institucional de Desarrollo Académico de la Universidad del Caribe (Berges Portillo R. 2021). Modelado del riesgo de reprobación en estudiantes de la Universidad del Caribe. Proyecto Terminal. p.10) como:

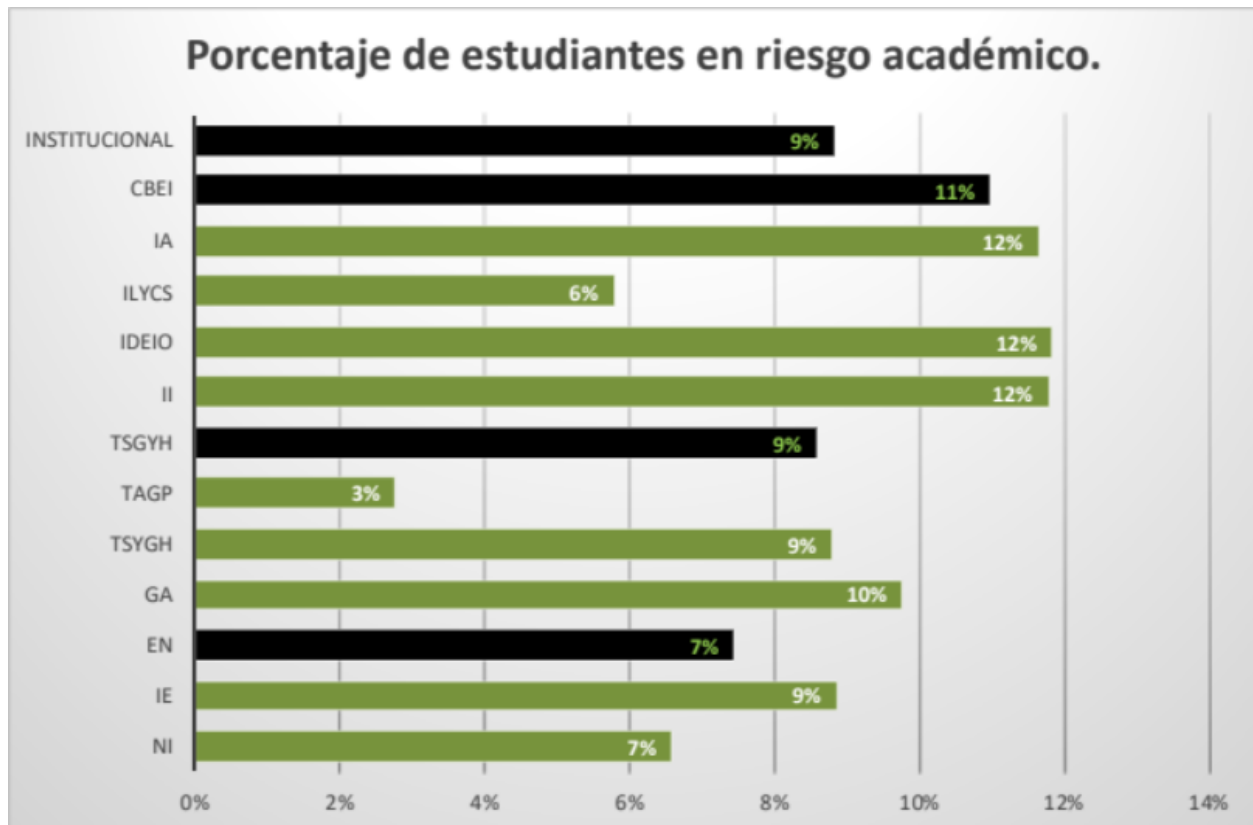
1. Ausentismo a clases por trabajo
2. Problemas socioeconómicos
3. Presentar una formación deficiente
4. Orientación profesional equivocada

1.2 Situación actual

Hasta el día de hoy, la Universidad del Caribe es una institución que desde su origen, ha ido creciendo en diferentes ámbitos que van desde su infraestructura hasta los programas educativos. Actualmente se imparten 9 licenciaturas y 9 posgrados, cuenta con un departamento de Educación Continua en donde se ofertan diplomados, cursos y talleres. De acuerdo a los indicadores actualizados hasta el mes de Agosto del 2021 por el Consejo Universitario de Acompañamiento Estudiantil (Zavala López E. (2022). Indicadores Agosto 2021(p. 2). Fuente de información. Universidad del Caribe. Cancún Q. Roo.), la población de estudiantes en el periodo de primavera fue de 3,281 estudiantes, de los cuales 290 (9%) recibieron tutoría individual porque se encontraban en riesgo académico. De esos 290 estudiantes, la población se distribuyó de la siguiente manera:

- 11% (102 de 928 alumnos) de la matrícula total en el departamento de CBeI.
- 9% (95 de 1105 alumnos) de la matrícula total en el departamento de TSGyH.
- 7% (93 de 1248 alumnos) de la matrícula total en el departamento de Economía y Negocios.

Figura 1 Gráfica que muestra el porcentaje de estudiantes en riesgo



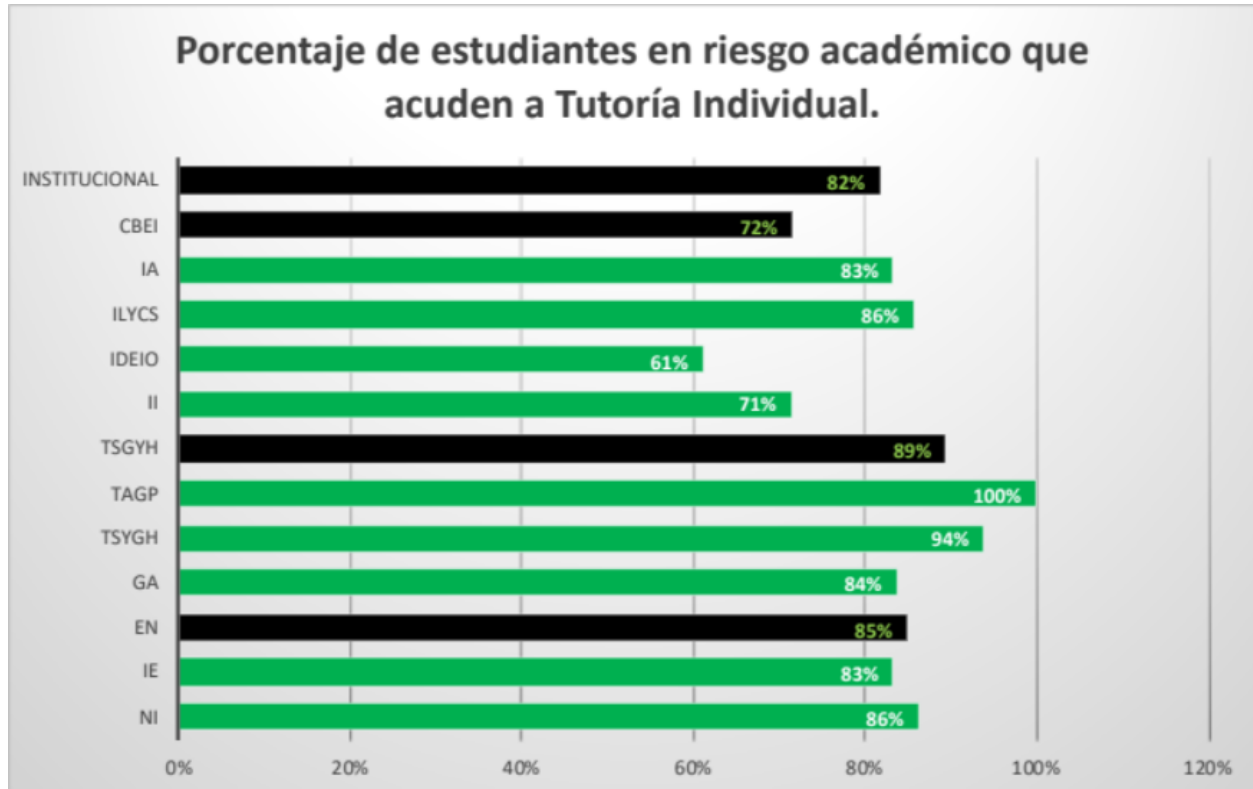
Tomado del consejo universitario de acompañamiento estudiantil - año 2022

Siendo TAGP la licenciatura que menos porcentaje de estudiantes que tiene en riesgo, mientras que IA, IDEIO e II, las licenciaturas que tiene el mayor porcentaje. La meta es mantener un porcentaje menor al 16% de los estudiantes en riesgo académico.

Por otra parte, se pretende que al menos el 60% de los estudiantes en riesgo asistan a las tutorías individuales. Según los datos proporcionados por el consejo estudiantil se manifiesta una participación del 72% para CBEI, 89% para TSGyH y 85% para EN, siendo IDEIO (61%) la

licenciatura que presenta la menor participación en tutorías individuales, mientras que de TAGP el 100% de los estudiantes acuden a estas tutorías.

Figura 2 Gráfica sobre la participación de los estudiantes en tutoría individual



Tomado del consejo universitario de acompañamiento estudiantil - año 2022

Para atender estos casos, el Programa de Acompañamiento Estudiantil, a través del SAES, integra un sistema con algunos recursos visuales básicos con la finalidad de brindar un panorama general del progreso del estudiante en tutorías y recibir avisos del tutor que se le haya sido asignado. Como se ha mencionado con anterioridad, este es reactivo, por lo que puede considerarse sistema remedial.

Recientemente, en un esfuerzo por promover el desarrollo de proyectos encaminados a resolver esta área de oportunidad, se han planteado propuestas como la de un proyecto de modelado de riesgo académico en el que se han entrenaron modelos de machine learning para clasificar si determinada carga curricular podría potencialmente desenlazar en la reprobación de una o más asignaturas. Esta fue la primera investigación formal realizada por estudiantes en la que se analizaron los datos del historial académico de egresados durante los periodos del 2016 al 2019, mismos que fueron proporcionados anonimizados con fines de investigación.

Hasta el día de hoy, no se han implementado soluciones tangibles (por ejemplo, una plataforma o sistema) donde se usen los modelos de machine learning y/o sistemas de recomendación en la Unicaribe por lo que aún hay oportunidad para nuevos proyectos que puedan satisfacer esta necesidad.

1.3 Problema

El Programa de Acompañamiento Estudiantil (**PAE**) mantiene un enfoque reactivo a la hora de proporcionar asesoramiento curricular. Desafortunadamente, este proceso entra en marcha una vez que las fechas de altas y bajas han concluido, ya que es cuando se asegura que los estudiantes mantendrán una selección de asignaturas definitiva. Como es de esperar, este proceso lleva tiempo, ya que además se necesita realizar una búsqueda de estos estudiantes en riesgo y asignarles un tutor, misma que se realiza de forma manual, lo cual es ineficiente en términos del tiempo de respuesta, por lo tanto, el proceso no es el adecuado para realmente garantizar que estos estudiantes puedan superar esa situación satisfactoriamente.

2. MARCO TEÓRICO

2.1. Web scraping

Es una técnica que consiste en la extracción y/o recuperación de datos de la World Wide Web (WWW) para posteriormente almacenarlos en un archivo, sistema o base de datos para su posterior análisis. Debido a que constantemente se genera una enorme cantidad de datos heterogéneos en la Internet, el web-scraping es ampliamente reconocido como una técnica eficiente y poderosa para recolectar grandes cantidades de datos. Esta técnica nos ayudará a obtener los datos de los estudiantes de la Universidad del Caribe que usen SURE haciendo uso de sus credenciales para entrar al SIGMAA, SIPP y SASS con la finalidad recolectar el historial académico, datos personales, horarios académico, situación actual, oferta escolar, de prácticas y servicio social, entre otros, para alimentar al dashboard y al sistema de recomendación.

El lenguaje de programación más utilizado para realizar web-scraping es Python, y lo usaremos principalmente porque tiene varias bibliotecas y una documentación extensa.

2.2 Selenium

Framework específico para aplicaciones web que facilita la automatización de pruebas funcionales. Selenium controla remotamente las instancias de los navegadores para emular las (posibles) interacciones de los usuarios, con el objetivo de validar las funciones de la aplicación.

De igual manera Selenium permite simular las tareas más habituales, como podrían ser: completar campos de texto, seleccionar opciones de los desplegables, interactuar con la interfaz de usuario, etc. Las funcionalidades que Selenium nos puede aportar en este proyecto son: hacer click en determinados enlaces, movimientos de ratón, ejecución de Javascript arbitraria,

obtención de datos del sitio web (títulos, cabeceras, etc) y acceso al código HTML que cola página complementandolo con técnicas de web-scraping.

Selenium dispone de varias herramientas con funcionamientos y objetivos distintos que permiten realizar las pruebas dependiendo del entorno y el proyecto para el que se realicen. Para este proyecto se usarán los siguientes componentes:

2.2.1 Selenium Web Driver

El Webdriver actúa como un enlace entre el código y el navegador, se encarga de recibir los comandos que enviará posteriormente al navegador para ejecutar las acciones que se necesitan para interactuar con la página web.

Esto se realiza a través de un driver específico que varía en función de en qué navegador quieras ejecutar las pruebas. Entre los lenguajes de programación que se encuentran soportados están:

- Java
- C#
- Python
- Ruby
- PHP
- JavaScript

2.3 Selenium Grid

Uno de los componentes de la suite de Selenium que permite diseñar pruebas automatizadas para aplicaciones web en orientado a la orquestación de pruebas en diversas plataformas, sistemas operativos y navegadores web. Selenium Grid es una extensión de Selenium RC para la ejecución de pruebas en diversos servidores en paralelo, por lo cual reduce el tiempo de ejecución y el costo, ya que permite la ejecución de las pruebas al mismo tiempo en varios navegadores y en diversos sistemas operativos. Selenium Grid cuenta con dos componentes: Selenium Hub y Remote Control.

2.4 Shiny

El paquete Shiny es un paquete de R con el cual se pueden diseñar interfaces gráficas de forma rápida para el análisis y visualización de datos ingresados por R (Shiny Team, s/f). Mediante este paquete, se pueden realizar funciones como el ingreso de datos o interacción dinámica por parte del usuario. El funcionamiento de Shiny consiste, en esencia, de dos grandes categorías de componentes: la categoría de componentes reactivos, los cuales son elementos que originan una reacción en el resto de la aplicación (por ejemplo, presionar un botón); y la categoría de componentes reactantes, los cuales cambian su presentación o funcionalidad luego de recibir un evento de un componente reactivo (RStudio Team, 2017). Estos dos tipos de componentes son de utilidad para la generación de páginas web dinámicas (ideales para el proyecto), donde el cambio del estado de ciertos elementos dentro de una página web no requiere que la página web se refresque o que se navegue a una página distinta dentro de un sitio, ya que basta con generar un evento reactivo para que el contenido se actualice automáticamente.

En este proyecto se hará uso de Shiny para desarrollar la aplicación web en donde se mostrarán estadísticas, el progreso del estudiante e implementar la parte gráfica del sistema de recomendación.

2.4.1 Biblioteca bs4Dash

El paquete bs4Dash es un complemento a la biblioteca de Shiny basada en su predecesor shinydashboard, cambia el aspecto visual de los componentes de una aplicación web de Shiny para hacerlos más atractivos y modernos, permite la opción de elegir entre distintos temas (paletas de colores) que una aplicación Shiny utiliza para sus componentes y añade nuevos elementos web preconstruidos y prediseñados con la finalidad de lograr una apariencia visual uniforme y consistente (RStudio Team, 2014).

2.4.2 Biblioteca ggplot2

Los gráficos que se pueden generar a través del lenguaje R tienen poca capacidad de configuración y no son muy atractivos visualmente, sin embargo, están presentes para una visualización rápida. Para la generación de gráficos con apariencia moderna y altamente configurable en la aplicación, se planea usar el paquete ggplot2. Éste permite generar un gran número de distintos tipos de gráficos estadísticos, así como configurar muchos aspectos de su apariencia.

2.5. FastApi

Una API (Interfaz de programación de aplicaciones) es una conexión entre programas de computadora. Las API se pueden implementar en Python utilizando varios marcos, algunos de los cuales incluyen Django, Flask o FastAPI.

FastAPI es un framework web moderno y rápido (alto rendimiento) para construir APIs con Python. Se lanzó en 2019, más tarde que los marcos populares como Django (2006) y Flask (2011), y ganó popularidad porque es rápido y de alto rendimiento. En comparación con Flask, FastAPI tiene los siguientes beneficios:

- Asyncio para concurrencia: Llamado con *async* y *await* palabras clave.
- Pydantic para la validación de datos: aplica el esquema y detecta el tipo de datos en tiempo de ejecución.
- Interfaz de usuario de Swagger para la generación automatizada de documentación: habilita la prueba de puntos finales de API con la extensión `/docs` . La documentación también promueve la comprensión. Una alternativa a la interfaz de usuario de Swagger es Redoc, a la que se puede acceder con la extensión `/redoc` .
- Características de seguridad y autenticación.
- Curva de aprendizaje inferior en comparación con las otras bibliotecas que se han mencionado con anterioridad.

2.6. Azure - infraestructura como servicio (Azure Ubuntu VM)

La infraestructura como servicio (IaaS) se refiere a los servicios en línea, a las funcionalidades informáticas que son proporcionadas por el proveedor, están pueden ser servidores, redes y almacenamiento a través de Internet. Es decir, dentro de la infraestructura de un proveedor de servicios la empresa u organización compradora de este servicio utilizan sus propias plataformas y aplicaciones.

En este proyecto, se usará una máquina virtual de Ubuntu ya que se necesita instalar software personalizado para gestionar el despliegue de la aplicación web hecha con Shiny y R, para ejecutar bots de web scraping usando Selenium e instalar bibliotecas y recursos adicionales que en PaaS no serían posibles de gestionar.

2.6.1 Microsoft Azure para SURE

Microsoft Azure es un conjunto en constante expansión de servicios en la nube que aporta a nuestro proyecto. Nos otorga la libertad de crear, administrar e implementar aplicaciones en una red mundial enorme con las herramientas que proporciona y marcos favoritos. Microsoft Azure es una plataforma tanto para los que están iniciando en el uso de estas tecnologías como para los que ya tienen años de experiencia en el desarrollo de aplicaciones basadas en la nube.

2.7. Azure - Plataforma como servicio (CosmoDB)

En la plataforma como servicio (PaaS) la organización no gestiona y tampoco controla la infraestructura, servidores, sistemas operativos o almacenamiento, más tiene todo el control sobre todas las aplicaciones que trae la plataforma. Esta plataforma es un ambiente de desarrollo e implementación completo en la nube, pero tienen recursos que permiten entregar todo, desde aplicaciones sencillas basadas en la nube hasta aplicaciones empresariales complejas habilitadas para la nube.

CosmoDB es una base de datos NoSQL completamente administrada para el desarrollo de aplicaciones modernas. Ofrece disponibilidad SLA y seguridad de grado empresarial (Azure, 2022).

En este proyecto se necesitará usar bases de datos para el almacenamiento de los datos, es por ello que se recurre a esta plataforma ya que ofrece alta disponibilidad y control total sobre la instancia.

2.8. APP Web (HTTP, HTML)

2.8.1 HTTP

El Protocolo de transferencia de hipertexto, o HTTP, es un protocolo de mensajería que permite a los navegadores web comunicarse con los servidores web (donde se almacenan los sitios web). Este protocolo está diseñado para transferir lo que llamamos hipertextos, páginas web o páginas HTML: textos complejos con enlaces, figuras, formularios, botones y objetos incrustados como animaciones o reproductores de música.

2.8.2 HTML

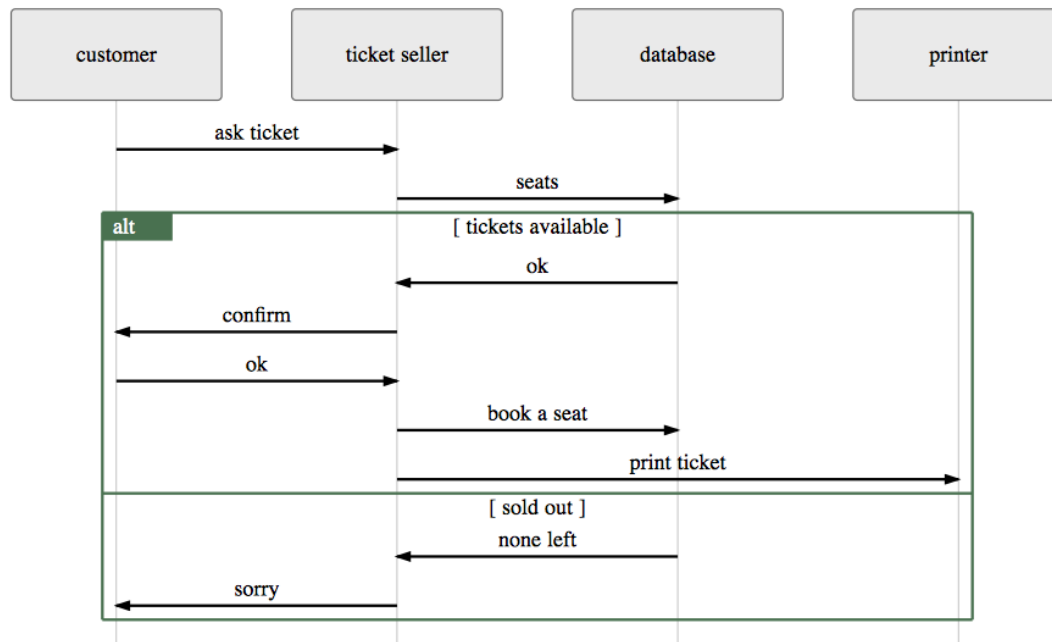
HTML, siglas de HyperText Markup Language (Lenguaje de Marcado de Hipertexto), es el lenguaje de marcado predominante para la elaboración de páginas web. Es usado para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos tales como imágenes. HTML se escribe en forma de "etiquetas", rodeadas por corchetes angulares (<,>).

El lenguaje HTML es un estándar reconocido en todo el mundo y cuyas normas define un organismo sin ánimo de lucro llamado World Wide Web Consortium, más conocido como W3C. Como se trata de un estándar reconocido por todas las empresas relacionadas con el mundo de Internet, una misma página HTML se visualiza de forma muy similar en cualquier navegador de cualquier sistema operativo. El propio W3C define el lenguaje HTML como "un lenguaje reconocido universalmente y que permite publicar información de forma global". Por convención, los archivos de formato HTML usan la extensión .htm o .html.

2.9. diagramas de sirena (mermaid diagrams)

Mermaid es una herramienta de gráficos y diagramas basada en JavaScript que utiliza definiciones de texto inspiradas en Markdown y un renderizador para crear y modificar diagramas complejos. El propósito principal de Mermaid es ayudar a que la documentación se ponga al día con el desarrollo, aunque en este proyecto en particular se integrará como una herramienta de visualización para representar el progreso en el mapa curricular de cada estudiante. Se puede integrar fácilmente al proyecto a través de bibliotecas de visualización de datos como DiagrammeR en R.

Figura 3 Diagrama de secuencias de sirena



Tomado de http://rich-iannone.github.io/DiagrammeR/graphviz_and_mermaid.html

2.10. Base de Datos NoSQL (mongoDB)

Existen prototipos diseñados para el manejo de bases de datos no relacionales, uno de ellos es bigtable, un sistema de distribución y almacenamiento que permite gestionar datos en una escala muy grande, diseñado en 2003 por Google con el fin de escalar peta bytes de datos de miles de máquinas. Bigtable no es compatible con modelos completos de bases de datos relacionales, es por eso que permite un control dinámico a los clientes que deciden utilizarlo.

2.10.1 Mongo DB

Usa un formato propio para guardar datos (BSON), con el cual no es necesario definir la estructura de la base ni los tipos datos; soporte de índices y referencias entre documentos,

lenguaje de consultas, optimiza la interfaz con los desarrolladores lo que hace que trabajar con ella sea muy simple, etc.

MongoDB es compatible con las metodologías de desarrollo actuales, como por ejemplo el método Agile, que permite a los desarrolladores realizar iteraciones de forma rápida y continua sobre el modelo de datos. Además permite a los desarrolladores almacenar con facilidad datos polimórficos, así como datos semi-estructurados y estructurados, en un almacén de datos individual.

3. ESTADO DEL ARTE

3.1 Using Association Rules for Course Recommendation

El trabajo presentado por Bendakir y Aïmeur (2016), tiene como objetivo orientar a los estudiantes al elegir los cursos adecuados para completar sus estudios académicos.

El sistema denominado RARE, parte de dos puntos importantes para la obtención de la variable de respuesta. En primer lugar mediante la técnica de minería de datos realiza la búsqueda de reglas de asociación y clasificación de acuerdo con los datos históricos de cursos realizados por antiguos estudiantes. En segundo lugar, el sistema ofrece a los estudiantes actuales calificar las recomendaciones, con la finalidad de poder beneficiarse realizando una mejora de las reglas de asociación.

Los autores mencionan que con el paso del tiempo la percepción del curso suele ser diferente para los estudiantes. Es por ello que es importante que las reglas no solo estén basadas en

técnicas de asociación para datos históricos, resaltando lo importante que es reforzar los criterios mediante las calificaciones actuales emitidas por los estudiantes.

3.2 Recommendation System for Continuing Education Courses

Los autores Paoli, Mariotti y Blattner (2003), exploran la educación continua y los retos que puede presentar continuar con el proceso de educación continua en el sistema SUPSI (Scuola Universitaria della Svizzera Italiana) acorde a las necesidades del usuario.

Son exploradas tres diferentes métodos para realizar la herramienta que es capaz de recomendar cursos que sean del interés para el usuario. El primero es G-Rank recomienda al usuario elementos más gustados en el público general. El segundo Collaborative Filtering recomienda al usuario elementos más gustados de usuarios similares. El tercero B-Rank recomienda cursos similares a los usuarios. Por último, Content-based recomienda cursos etiquetados con las preferencias del usuario.

Este estudio muestra que un sistema híbrido es óptimo para realizar las predicciones de los cursos de interés para el usuario. El emplear un sistema híbrido proporciona la ventaja de aumentar la precisión, exactitud y el F1 de aceptación de los cursos.

3.3 Courses Recommendation Algorithm Based On Performance Prediction In E-Learning

El trabajo presentado por Koffi, Ouattara, Mambe, Oumtanaga y Assohoun (2021), profundiza en los algoritmos mediante los cuales los sistemas de recomendación cumplen con su propósito. El algoritmo tiene como propósito el recomendar o desrecomendar la carga de un curso.

La metodología que emplean para dar solución a este problema en particular es la exploración de los diferentes modelos de aprendizaje, aprendizaje profundo y K-NN. Los alumnos que formaron parte de la muestra fueron evaluados para conocer los modelos de aprendizaje que ellos mantienen, de esta manera se les pudo catalogar y determinar si eran aptos o no para estudiar un curso en específico.

La evaluación de resultados de predicción se realiza a partir del error cuadrático medio y la matriz de confusión, en el caso del aprendizaje profundo se obtuvieron los siguientes valores; exactitud 97.28%, precisión 97.31%, sensibilidad 99.98% y F1 score 98.51%. La matriz de confusión evaluada en los valores de K-NN presenta los siguientes valores ; exactitud 97.03%, precisión 97.20%, sensibilidad 99.57% y F1 score 98.37%. El aprendizaje profundo mantiene una menor tasa de error sobre K-NN.

3.4 Predicción del rendimiento en una asignatura empleando la regresión logística ordinal

Los autores Jobany J. Heredia R., Aida G. Rodríguez H., José A. Vilalta A. (2014), Tiene como punto de partida identificar el rendimiento que los estudiantes presentan en la asignatura Modelos Probabilísticos de los Procesos(MPP). Empleando un modelo de regresión logística ordinal se espera evaluar el perfil del estudiante con la finalidad de evaluar el rendimiento que este tendrá en la asignatura. La regresión logística ordinal evalúa dos estados, bien el estudiante tendrá un rendimiento positivo o tendrá un rendimiento negativo.

En total se exploraron los datos de 274 estudiantes pertenecientes a dos cursos académicos distintos, se agruparon las materias semejantes en cuanto a sus contenidos a partir de este agrupamiento se obtuvo un promedio de las calificaciones, de tal manera se obtuvieron las principales variables predictoras. Los datos de los estudiantes se separaron por curso académico y período de clases en el cual se comprendía(2010- 2011 / 2011-2012).

El modelo se evaluó en un total de 103 estudiantes, de los cuales hubo 82 aciertos correspondiente a un total del 80% de precisión, y 21 errores correspondiente a 20% de precisión. Este modelo es comparado con distintos modelos empleados para realizar una tarea similar, el modelo de regresión logística ordinal se encuentra semejante en precisión a un modelo que emplea redes neuronales y superior a un modelo que emplea árboles de clasificación.

4. DESCRIPCIÓN DEL PROYECTO

Los estudiantes de la Universidad del Caribe tienen la necesidad de una herramienta que les oriente a la hora de diseñar sus cargas curriculares para el semestre, se requiere de una herramienta auxiliar que permita a los estudiantes, entre otras cosas, obtener recomendaciones de asignaturas basadas en sus necesidades particulares como, horarios de disponibilidad, su historial académico y un conjunto de reglas predefinidas como la seriación de asignaturas y la promoción del cierre de los ciclos escolares con la finalidad de disminuir la tasa de reprobación estudiantil derivado de una mala selección de materias que puedan conllevar a la reprobación, y subsecuentemente, la deserción escolar.

4.1 Propuesta

Se propone desarrollar una sistema completo, nuevo y complementario, que integre una plataforma web con elementos visuales propios de un dashboard para cuantificar el progreso, un sistema de recomendación curricular para diseñar cargas académicas adecuadas con un bajo riesgo de reprobación y una infraestructura de datos sólida que permita obtener la información de los estudiantes mediante la orquestación de tareas automatizadas y web-scraping. Estará pensada para hacer uso de los sistemas existentes como el SIGMAA, SIPP y SASS, por lo tanto no será un sustituto sino una forma de aprovechar la información que nos proporciona la universidad en sus sistemas principales con la finalidad de proporcionar apoyo a la toma de decisiones para la carga curricular y así, reducir la tasa de reprobación y deserción estudiantil.

4.2 Justificación

En la actualidad, la Universidad del Caribe mantiene un enfoque reactivo a la hora de ofrecer acompañamiento estudiantil, esto sucede cuando el estudiante entra en una situación de riesgo por reprobación, tiempo de permanencia, o bien, por revalidación. Por lo tanto, esperamos que SURE logre reducir los casos críticos en donde se requiere de acompañamiento estudiantil a través de una nueva herramienta de apoyo a la toma de decisiones inicialmente para estudiantes del programa Ingeniería en datos e inteligencia organizacional. Si atacamos el problema directamente desde un enfoque preventivo, es decir, antes que el estudiante cargue asignaturas, es probable que la tasa de reprobación disminuya.

4.3 Objetivos

4.3.1 Objetivo General

Proveer a los estudiantes una herramienta de toma de decisiones para visualizar, obtener recomendaciones y diseñar cargas curriculares adecuadas haciendo uso de la información de los sistemas existentes (SIGMAA, SIPP y SASS).

4.3.2 Objetivos Específicos

- Obtener y almacenar la información de los sistemas informáticos de la Universidad Del Caribe (SIGMAA, SISPP, SAES) en nuestra base de datos a través de la orquestación de pruebas automatizadas y web-scraping para estudiantes que usen SURE.
- Diseñar e implementar una aplicación web en Shiny con recursos visuales propios de un dashboard y un sistema de recomendación curricular integrado.
- Reevaluar y ajustar a las necesidades actuales los modelos de machine learning existentes con la finalidad de diseñar cargas académicas preliminares adecuadas para los estudiantes.
- Implementar un sistema de recomendación de cargas curriculares de acuerdo a reglas como la seriación de asignaturas, priorizar el cierre de ciclos y modelos de machine learning.

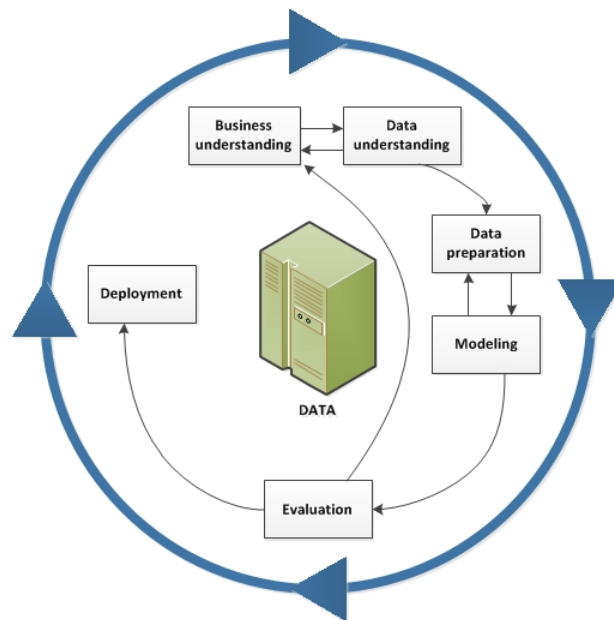
4.4 Hipótesis

Un sistema de recomendación curricular con el uso de reglas y modelos de aprendizaje automático disminuirá el porcentaje de estudiantes en situación de riesgo académico por debajo del 9% en IDeIO.

4.5 Metodología

En la elaboración de este proyecto se emplearán dos metodologías específicas con la intención de seguir los procedimientos para el alcance de los objetivos planteados. En primer lugar, se comenzará trabajando con la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) para la reevaluación de los modelos de machine learning existentes, ya que es necesario evaluar las nuevas necesidades del negocio y ajustarse a ellos haciendo las adecuaciones necesarias. Las fases de la metodología CRISP-DM como se describen en la figura 4.

Figura 4 Ciclo de vida de un proyecto de minería de datos



Tomado de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Por otra parte, para el desarrollo del software en donde se implementarán los modelos, que esencialmente comprende al despliegue (la última etapa de CRISP-DM), se optará por la metodología de **Extreme Programming** ya que el equipo de trabajo es reducido y necesitamos de una alta flexibilidad y tolerancia al cambio debido al reducido tiempo para finalizar el proyecto.

5. PLAN DE TRABAJO

El plan de trabajo contiene las actividades a realizar durante este periodo. Las actividades son clasificadas de acuerdo a la metodología de desarrollo del sistema y se les da la prioridad que necesiten. En el anexo 1 [\[*\]](#) se puede encontrar a detalle el plan de trabajo.

Figura 5 Plan de trabajo para desarrollo de sistema SURE

Tareas	Responsable	Fecha de inicio	Fecha final	Días	Estado	Junio				Julio			
Entendimiento del negocio													
Reunión con los interesados	Kenneth Díaz	2022-06-08	2022-06-15	7	Completado								
Determinar tecnologías para infraestructura de datos, visualización de información y análisis de datos	Kenneth Díaz	2022-06-15	2022-06-30	15	Completado								
Diseñar wireframes y mockups de sistema final	Kenneth Díaz	2022-06-15	2022-07-15	30	Completado								
Determinar fuentes de información	Kenneth Díaz	2022-06-30	2022-07-15	15	Completado								
Entendimiento de los datos													
Análisis exploratorio de calificaciones 2016 - 2021	Kenneth Díaz	2022-09-09	2022-09-16	7	En progreso								
Preparación de los datos					En progreso								
Limpiar, ordenar y adecuar datos para reentrenar modelos y probarlos	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-09-16	2022-09-23		En progreso								
Evaluación													
Recuperar, rediseñar y reentrenar	Saúl Durán, Juan												

Tomado de elaboración propia

6. RIESGOS

Figura 6 Tabla de riesgos con acciones preventivas para Proyecto Terminal SURE

Descripción de riesgos	Nivel de riesgo	Acciones preventivas
<ul style="list-style-type: none"> Que la universidad modifique drásticamente los actuales sistemas donde se obtiene la información 	Muy alto	<ul style="list-style-type: none"> Orquestar tareas de actualización de información periódicamente Almacenar información en bases de datos independientes
<ul style="list-style-type: none"> Servidores de la universidad fuera de línea por mantenimiento o desastres naturales 	Alto	<ul style="list-style-type: none"> Ofrecer la información más actualizada hasta que los sistemas se restablezcan
<ul style="list-style-type: none"> Dificultades técnicas para completar el sistema a tiempo dada la magnitud del proyecto 	Alto	<ul style="list-style-type: none"> Priorizar y delimitar el proyecto solo a las partes críticas donde se requieren información para el sistema de recomendación
<ul style="list-style-type: none"> Desconfianza estudiantil ya que se maneja información confidencial como contraseñas, información personal, etc. 	Medio	<ul style="list-style-type: none"> Garantizar que la información procesada será usada para efectos como se marca en el artículo 49, numeral I, capítulo único "Disposiciones generales" del reglamento escolar vigente.

Tomado de elaboración propia

7. REFERENCIAS

- Anaya, E. (2020, 20 mayo). ¿Qué es Selenium y para qué sirve? - Inmediatum. Inmediatum - Empresa Top en Software Especializado. Recuperado 6 de septiembre de 2022, de <https://inmediatum.com/blog/piensa-digital/que-es-selenium-y-para-que-sirve/>
- B. Paoli, F. Mariotti, M. Blattner, M. Gay, D. Bianchi and M. Rossi, 2013. Recommendation System for Continuing Education Courses [PDF] Manno: Scuola Universitaria della Svizzera Italiana, Disponible en: https://dlwqtxts1xzle7.cloudfront.net/70693929/Recommendation_System_for_Continuing_Edu20210930-16115-1uyptsk-with-cover-page-v2.pdf
- Barragan Charry A. (2013). Implementación de una base de datos NoSQL para la generación de una matriz O/D. (Trabajo de Grado). Universidad Catolica De Colombia Facultad De Ingeniería Programa De Ingeniería De Sistemas Bogotá D.C.
- Corral Espinosa J. (2021). Análisis y desarrollo de aplicación web para el análisis estadístico de datos de encuestas.(Trabajo de titulación, Universidad del Azuay Facultad de Ciencias de la Administración Escuela de Ingeniería de Sistemas y Telemática). Recuperado de <https://dspace.uazuay.edu.ec/bitstream/datos/11154/1/16692.pdf>

D. Dangui. A. Sylvain, L. Koffi, N. Ouattara, D Moïse, S. Oumtanaga and A. ADJE , 2021.

Courses Recommendation Algorithm Based On Performance Prediction In E-Learning

[PDF] Abiyán: Universidad Félix Houphouët-Boigny, Disponible en:

https://d1wqtxts1xzle7.cloudfront.net/70693929/Recommendation_System_for_Continuing_Edu20210930-16115-1uyptsk-with-cover-page-v2.pdf [Consultado el 22 Agosto 2022]

IBM, 2021. Conceptos básicos de ayuda de CRISP-DM, Recuperado 10 de septiembre de 2022

Disponible en:

<https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Jobany J. Heredia R. Aida G. Rodríguez H., José A. Vilalta A, 2014. Predicción del rendimiento

en una asignatura empleando la regresión logística ordinal[PDF] La habana: Instituto

Superior Politécnico “José Antonio Echeverría” Disponible en:

https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07052014000100009

[Consultado el 1 septiembre 2022]

N. Bendakir and E. Aïmeur , 2006. Using Association Rules for Course Recommendation

[PDF] Montréal: e, Université de Montréal ´ Pavillon Andre-Aisenstad, Disponible en:

<https://www.aaai.org/Papers/Workshops/2006/WS-06-05/WS06-05-005.pdf> [Consultado el 22 Agosto 2022]

Pérez Pérez V. (s.f.). Desarrollo de un sitio web para un colegio. (Proyecto final de la carrera).

Universidad Politécnica de Valencia.

R. Berges, E. Mann, J. Zavala, 2021. Modelado de riesgo de reprobación en estudiantes de la

Universidad del Caribe. [PDF] Cancún: Universidad del Caribe. Disponible en:

https://drive.google.com/file/d/1O9n4lVO-uV3wn_2ZowcH07ZMNhdhBIJy/view?usp=s

haring [Consultado el 17 Agosto 2022]

Wong, Kay Jan. (2021). FastAPI. towards data science. Recuperado 6 de septiembre de 2022, de

<https://fastapi.tiangolo.com/es/>

Zhao Bo.(2017).Web Scraping.doi:10.1007/978-3-319-32001-4_483-1. Recuperado 6 de

septiembre de 2022

¿IaaS? ¿SaaS? ¿PaaS? Lo que debes saber de Azure | Ibermatica365. (2022, 5 abril). Recuperado

6 de septiembre de 2022, de

<https://www.ibermatica365.com/iaas-saas-paas-guia-para-entender-azure/>

8. ANEXOS

Anexo 1 - PLAN DE TRABAJO

Tareas	Responsable	Fecha de inicio	Fecha final	Días	Estado
Entendimiento del negocio					
Reunión con los interesados	Kenneth Díaz	2022-06-08	2022-06-15	7	Completado
Determinar tecnologías para infraestructura de datos, visualización de información y análisis de datos	Kenneth Díaz	2022-06-15	2022-06-30	15	Completado
Diseñar wireframes y mockups de sistema final	Kenneth Díaz	2022-06-15	2022-07-15	30	Completado
Determinar fuentes de información	Kenneth Díaz	2022-06-30	2022-07-15	15	Completado
Entendimiento de los datos					
Análisis exploratorio de calificaciones 2016 - 2021	Kenneth Díaz	2022-09-09	2022-09-16	7	En progreso
Preparación de los datos					En progreso
Limpiar, ordenar y adecuar datos para reentrenar modelos y probarlos	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-09-16	2022-09-23		En progreso
Evaluación					
Recuperar, rediseñar y reentrenar	Saúl Durán, Juan				

Recuperar, rediseñar y reentrenar modelos existentes realizados el <i>modelado de riesgo de reprobación</i>	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-09-23	2022-09-30	7	Sin empezar
Diseñar un nuevo modelo basado en reglas y modelos de riesgo de reprobación	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-09-23	2022-10-08	15	Sin empezar
Despliegue					
Codificación del frontend con shiny	Kenneth Díaz	2022-07-13	2022-10-08	87	En progreso
Construcción de API Resful	Kenneth Díaz	2022-08-18	2022-10-08	51	En progreso
Codificación del backend con shiny	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-08-04	2022-10-21	78	En progreso
Implementación del modelo en la aplicación web	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-10-01	2022-10-08	7	Sin empezar

Lanzamiento final					
Diseñar infraestructura transaccional	Kenneth Díaz	2022-07-07	2022-07-24	17	Completado
Diseño e implemetación de bots para web-scraping y data pipelines	Kenneth Díaz	2022-07-07	2022-07-24	17	Completado
Limpieza de los datos obtenida de plataformas existentes	Kenneth Díaz	2022-07-24	2022-08-18	25	Completado
Despliegue en servidor web para testing	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-11-01	2022-11-15	14	Sin empezar
Liberación de prueba beta para estudiantes	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-11-01	2022-11-15	14	Sin empezar
Despliegue final para producción	Saúl Durán, Juan Antonio Yam, Kenneth Díaz	2022-10-01	2022-11-30	60	Sin empezar