

UNIVERSIDAD DE VALLADOLID
MÁSTER UNIVERSITARIO
Ingeniería Informática



TRABAJO FIN DE MÁSTER

Diseño de elemento de juego para aportar feedback basado en técnicas de estimación empleando árboles de decisión

Realizado por **María Robles del Blanco**



Universidad de Valladolid
7 de septiembre de 2022
Tutor: David Escudero Mancebo

Universidad de Valladolid



Máster universitario en Ingeniería Informática

D. David Escudero Mancebo, profesor del departamento de Informática, área de Ciencia de la Computación e Inteligencia Artificial.

Expone:

Que la alumna D. María Robles del Blanco, ha realizado el Trabajo final de Máster en Ingeniería Informática titulado "DISEÑO DE ELEMENTO DE JUEGO PARA APORTAR FEEDBACK BASADO EN TÉCNICAS DE ESTIMACIÓN EMPLEANDO ÁRBOLES DE DECISIÓN".

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Valladolid, 7 de septiembre de 2022

Vº. Bº. del Tutor:

D. David Escudero Mancebo

Agradecimientos

A mis compañeros de Máster, por hacer este camino mucho más agradable, sin vosotros este curso y sus clases presenciales no habrían sido lo mismo.

A mi tutor David Escudero, por su continuo esfuerzo y su enorme aportación de conocimiento durante el desarrollo de este trabajo.

Al grupo de Crossroads, en especial a Yania, David y Manuel, por haber confiado en mí para el desarrollo de la aplicación y por su gran dedicación.

Al grupo GEEDS de la UVa, en especial a Luis Miguel y José María, que han confiado en mí para el desarrollo de la aplicación, además de mostrar su disposición para realizar pruebas de las aplicaciones y aportar grandes opciones de mejora.

A la Fundación Española de Ciencia y Tecnología (FeCYT) y al proyecto código 16138 titulado ENCRUCIJADA-MUNDO: ECOHERRAMIENTAS LÚDICAS PARA LA TRANSICIÓN ENERGÉICA gracias a la cual he podido disfrutar de un contrato a tiempo parcial durante los meses de diciembre hasta junio, a la vez que realizaba el máster, y al proyecto RethinkingAction H2020 cod. 101037104 que financió el contrato durante el mes de julio. Gracias a ellos he descubierto el asombroso mundo del desarrollo iOS.

Resumen

Actualmente, el cambio climático es un problema de gran relevancia. Para concienciar a las generaciones más jóvenes de la importancia del cambio climático y de sus posibles soluciones para paliar sus efectos, existen juegos educativos como es el caso de Crossroads. Este videojuego permite que el usuario tome una serie de decisiones medioambientales y económicas y, como resultado, puede comprobar el efecto que causarían éstas a lo largo de los años en relación al aumento de la temperatura y del PIB. Las medidas políticas y decisiones que debe tomar el usuario en el juego son cuestiones complejas y aportar feedback que ayude a posicionarse al jugador puede ser ventajoso. Para ello, en este trabajo se ha desarrollado un estimador que genera la probabilidad de obtener un resultado aceptable económica y sosteniblemente dependiendo de las respuestas que haya seleccionado el usuario. Se ha utilizado una técnica de machine learning, en concreto árboles de decisión con el algoritmo C4.5, para entrenar un clasificador que permite estimar la probabilidad de llegar a una solución aceptable. Además, se ha comparado el resultado con un método empírico que se apoya en el teorema de Bayes para contrastar la bondad de las estimaciones hechas con el árbol de decisión. Los resultados muestran el potencial de esta técnica para enriquecer el juego con un recurso que permita ofrecer feedback en tiempo de juego a los usuarios de la aplicación.

Descriptores

Árboles de decisión, realimentación, juego educativo, cambio climático, aplicación iOS, probabilidad a posteriori

Abstract

Climate change is currently a very significant issue. To raise awareness among the younger generations of the importance of climate change and possible solutions to mitigate its effects, there are educational games such as Crossroads. This video game allows the user to make a series of environmental and economic decisions and, as a result, can check the effect that these would cause over the years in relation to the increase in temperature and GDP. The policy measures and decisions to be taken by the user in the game are complex and providing feedback to help position the player can be advantageous. For this purpose, in this project we have developed an estimator that calculates the probability of obtaining an economically and sustainably acceptable result depending on the responses selected by the user. A machine learning technique, specifically decision trees with the C4.5 algorithm, has been used to train a classifier to estimate the probability of achieving at an acceptable solution. In addition, the result has been compared with an empirical method that relies on Bayes' theorem to test the goodness of estimates made with the decision tree. The results show the potential of this technique to enrich the game with a resource to provide game-time feedback to the users of the application.

Keywords

Decision trees, feedback, educational game, climate change, iOS application, posterior probability

Índice general

Índice general	IV
Índice de figuras	VI
Índice de tablas	VII
1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
1.3. Juego Crossroads	3
1.4. Marco contextual	5
1.5. Estructura del documento	7
2. Estado del arte	8
2.1. Juegos educativos sobre cambio climático	8
2.2. Feedback en videojuegos	10
2.3. Inteligencia artificial en juegos	11
3. Metodología	13
3.1. Procedimiento	13
3.2. Predicción de probabilidades de éxito	18
3.3. Herramientas	21
3.4. Validación	23
4. Implementación en la aplicación móvil	26
4.1. API para el estimador	26
4.2. Modificaciones <i>front-end</i>	27
4.3. Modificaciones <i>back-end</i>	29
5. Resultados	31
5.1. Evaluación de ambas técnicas: mayor probabilidad	31

5.2. Caso de prueba de menor probabilidad	34
5.3. Probabilidades individuales	36
5.4. Información aportada por el árbol de decisión	38
5.5. Adaptación de la probabilidad del estimador	41
6. Discusión	43
6.1. Evaluación de ambas técnicas	43
6.2. Caso de prueba de menor probabilidad	44
6.3. Información aportada por el árbol de decisión	45
6.4. Estimación en base a los objetivos	46
6.5. Adaptación de la probabilidad del estimador	47
6.6. Utilidad real para los usuarios	48
7. Conclusiones y Líneas de trabajo futuras	50
Apéndices	53
Apéndice A Plan de Proyecto	54
A.1. Planificación temporal	54
A.2. Recursos	57
A.3. Viabilidad económica	58
Apéndice B Enlaces	60
Apéndice C Manual de instalación y despliegue	61
C.1. Proyecto estimación estado bueno	61
C.2. Aplicación móvil Crossroads (copia modificada)	63
C.3. Backend Crossroads (copia modificada)	63
Apéndice D Descripción del contenido del juego Crossroads	64
D.1. Preguntas	64
D.2. Ayudas	67
Bibliografía	80

Índice de figuras

1.1. Capturas de pantalla del juego Crossroads web	4
1.2. Caso de uso del juego Crossroads en la aplicación móvil iOS	5
3.3. Creación del conjunto de datos inicial	14
3.4. Distribución de los resultados de simulación sobre para la agrupación utilizada en el recomendador en porcentaje respecto del total. [48]	16
3.5. Esquema funcional del estimador bayesiano	24
3.6. Esquema funcional del estimador basado en árboles de decisión	24
4.7. Ejemplo de operación POST desde Postman para la estimación de estado bueno	27
4.8. Indicador gráfico	28
4.9. Resultado final del indicador gráfico en la aplicación móvil.	29
4.10. Operación añadida al <i>back-end</i> de Crossroads	30
5.11. Captura de pantalla de las preguntas con el indicador	33
5.12. Captura de pantalla de los resultados tras seleccionar la secuencia con mayor probabilidad	34
5.13. Captura de pantalla de los resultados tras seleccionar la secuencia con menor probabilidad	36
5.14. Captura de pantalla de los resultados tras seleccionar la secuencia con mayor probabilidad en cada pregunta	38
5.15. Gráfica de la función $f(x) = \gamma^{0.3}$	42
A.1. Diagrama de Gantt con la planificación de la estancia y el TFM	57

Índice de tablas

3.1. Atributos del conjunto de datos	17
5.2. Probabilidades generadas por el Teorema de Bayes. Caso de uso seleccionar secuencialmente la mayor probabilidad	32
5.3. Probabilidades generadas por el modelo con el algoritmo C4.5 de Weka. Caso de uso seleccionar secuencialmente la mayor probabilidad	32
5.4. Probabilidades individuales generadas por el modelo con el algoritmo C4.5. Caso de uso: seleccionar secuencialmente la menor probabilidad	35
5.5. Probabilidades individuales generadas por el Teorema de Bayes	37
5.6. Probabilidades individuales generadas por el modelo con el algoritmo C4.5	37
5.7. Probabilidades generadas por el modelo con el algoritmo C4.5 de Weka con función que procesa la salida. Caso de uso seleccionar secuencialmente la mayor probabilidad	42
A.1. Costes simulados	59

1: Introducción

La realimentación es un elemento fundamental en la educación. Conocer si las decisiones que se toman son o no adecuadas permite reconducir la conducta de la persona que se entrena. En este proyecto se pretende realizar un elemento de ayuda para un juego móvil educativo. El juego móvil en cuestión está enfocado a alumnos de la ESO o de Bachiller y trata temas complejos como el cambio climático. En él, el alumnado debe tomar decisiones sobre economía y medioambiente. Se pretende acompañar al juego de un elemento que guíe al alumno durante la partida para que conozca qué decisiones pueden ser más trascendentes y cuáles son más beneficiosas para el medioambiente y la economía.

1.1. Motivación

El cambio climático y la disponibilidad de recursos energéticos son problemas de gran importancia. Es necesario que las nuevas generaciones tomen conciencia de la importancia del problema y de la trascendencia de adoptar medidas políticas de alcance. La gamificación de temas complejos puede ayudar a que generaciones más jóvenes sean más conscientes de la problemática y de su importancia de una forma más amena.

Actualmente, existe una aplicación educativa llamada Crossroads 2.0 [11] cuyo fin es la concienciación sobre la necesidad de adoptar medidas políticas que frenen el impacto del cambio climático. El juego educativo consiste en una sucesión de toma de decisiones sobre políticas medioambientales y económicas tras las cuales se puede visualizar qué efectos producirían sobre el producto interno bruto (PIB) y la temperatura a lo largo de los años. La toma de decisiones se realiza contestando unas preguntas y a partir de las respuestas dadas, se grafica el resultado obtenido a lo largo de los años que muestra la evolución de la temperatura y el PIB.

En el juego se trabaja con medidas políticas que cuentan con cierta dificultad cognitiva. Es por ello que existe dificultad para que el alumnado llegue a conocer y comprender con claridad los efectos (tanto positivos como negativos) de las medidas políticas escogidas y cuáles podrían ser las decisiones más beneficiosas para el medioambiente y la economía. Uno de los principales problemas encontrados es la dificultad para que los alumnos saquen

conclusiones medioambientales y económicas al utilizar la aplicación educativa y que puedan identificar qué medidas políticas son más favorables. Se pretende que comprendan el efecto asociado a cada decisión que tomen, lo cual no es un proceso trivial ya que son cuestiones realmente complejas.

Las intervenciones de *feedback* o realimentación [33] se definen como aquellas acciones realizadas por agentes externos que proporcionan información sobre algún aspecto del desempeño de una tarea. Concretamente, en aplicaciones de juegos, el *feedback* generado para el usuario es una parte relevante del juego. El *feedback* suministra conocimiento sobre el juego y permite guiar al jugador provocando que se replantee la toma de ciertas decisiones. En este proyecto se quiere contribuir a la realimentación de información al jugador, lo que permitirá al usuario conocer qué decisiones son más beneficiosas y reconducir su partida para llegar a un estado favorable.

1.2. Objetivos

El objetivo general de este proyecto es generar un elemento de juego que guíe al usuario en la toma de decisiones durante el juego de Crossroads. Se acompañará la aplicación existente con un elemento sencillo de juego (un indicador numérico o gráfico) que muestre al usuario cuan lejos o cerca se encuentra de obtener un buen resultado, dependiendo de las opciones que haya escogido durante la partida. Para ello, se diseña un modelo que estime la probabilidad de conseguir un resultado aceptable con las opciones que haya seleccionado el usuario. Esto permite al usuario conocer qué opciones aumentan la probabilidad de obtener un resultado bueno y, por tanto, qué opciones son las más beneficiosas para el medioambiente y la economía.

Concretamente, los objetivos específicos para lograr este objetivo general son los presentados a continuación:

1. Revisión del estado del arte y trabajos previos relacionados para obtener conocimiento sobre el problema que se aborda.
2. Creación y evaluación de un sistema de predicción que estime la bondad de las opciones seleccionadas, indicando si éstas acercan al jugador a un resultado considerado aceptable o no.
3. Añadir el modelo de predicción creado en la aplicación móvil, utilizando para ello un elemento gráfico pertinente.

Cumpliendo estos objetivos específicos del problema, se consigue que el alumnado que utilice el juego de forma lúdica conozca cómo de lejos o cerca se encuentra de obtener un buen resultado, tanto medioambiental como económico. El resultado de este proyecto es un indicador en la aplicación móvil que, a lo largo del juego, irá modificando su escala para indicar al usuario si su situación se acerca a un resultado aceptable o no. De esta

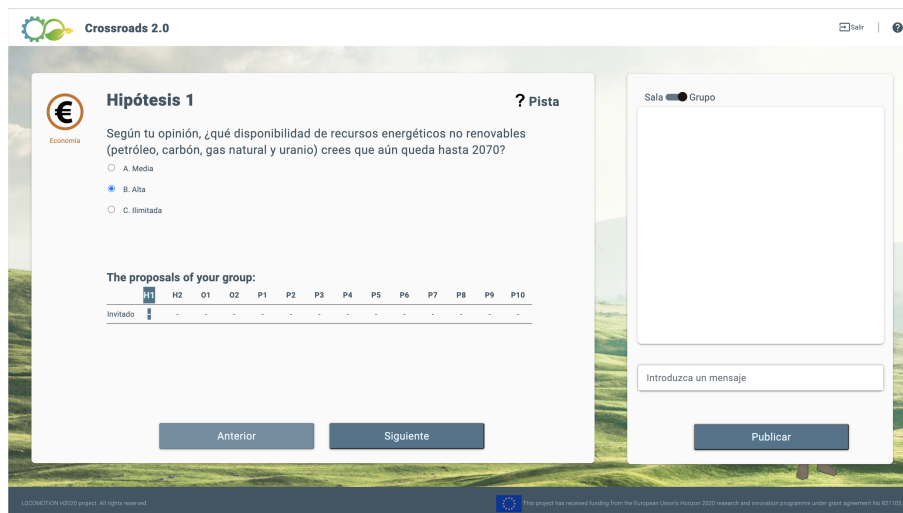
forma, el alumnado podrá comprender y asentar conocimientos de una manera amena sobre qué decisiones medioambientales y económicas son más perjudiciales.

1.3. Juego Crossroads

El **juego Crossroads 2.0** es el juego en el cual se añade el sistema de predicción de bondad de las opciones seleccionadas por el usuario. Este juego incorpora un cuestionario donde se van introduciendo determinados valores sobre economía y medioambiente. En concreto, los jugadores deben fijar unas hipótesis, elegir unos objetivos y una serie de medidas políticas. Este juego está disponible tanto para plataformas web como móvil. En la Figura 1.1 se muestran algunas capturas de pantalla de la aplicación web. En cuanto a plataformas móviles, se ha desarrollado para Android e iOS. En este caso, el objetivo específico de este proyecto es añadir el indicador que estime la bondad de las respuestas en la aplicación iOS, aunque como se añadirá una operación al *backend* de Crossroads 2.0, posteriormente sería sencillo implementarlo también en la aplicación web y Android. Sin embargo, en este proyecto TFM esta modificación sólo se realizará para la aplicación iOS.

A continuación, se presenta un caso de uso resumido de la aplicación móvil en la Figura 1.2 para poder comprender mejor el funcionamiento de ésta. Para ello se han realizado capturas de pantalla durante el desarrollo del juego. En la primera imagen 1.2a se presenta la primera pregunta del cuestionario que es la primera hipótesis que debe fijar el usuario. Posteriormente, se ha completado el formulario y en la imagen 1.2b se adjunta la captura de la última pregunta del formulario que corresponde a una medida política. En concreto, el juego cuenta con 3 hipótesis, 2 objetivos y 9 medidas políticas. No se ha mostrado ningún objetivo en este caso de uso por simplicidad. Una vez finalizado el cuestionario, se muestra la simulación de la evolución de temperatura y PIB hasta el año 2100 en base a los valores previos escogidos como se muestra en la imagen 1.2c, donde se representa en dos gráficas la evolución que tendría la temperatura y el PIB si se llevasen a cabo las medidas políticas acordadas y se cumpliesen las hipótesis seleccionadas. Un resultado aceptable sería cuando en estas gráficas de temperatura y PIB, la predicción de evolución representada fuera lo más equilibrada posible sin grandes variaciones a lo largo de los años ni crecimientos o decrecimientos extremos.

Como se ha comentado en la Sección 1.2, se desea implementar un sistema de predicción en la aplicación móvil Crossroads. Concretando el objetivo previamente introducido, se quiere implantar un sistema de feedback que ayude al usuario a asentar los conocimientos implícitos en las medidas políticas y los efectos que causan, así como que le permita reconducir la partida para obtener un resultado mejor y más beneficioso con el medioambiente y la economía. En este caso, un posible **elemento de feedback** es un indicador gráfico a través del cual se represente la probabilidad numérica de obtener un resultado aceptable dependiendo de las respuestas contestadas, entendiendo como resultado aceptable un estado que sea equilibrado en cuanto a medioambiente y temperatura. Más adelante se brindarán más detalles sobre el indicador gráfico desarrollado. Este elemento gráfico, irá variando dinámicamente entre valores de 0 a 100 según el usuario vaya contestando preguntas

(a) Pantalla *login*

(b) Pantalla pregunta (hipótesis 1)

Figura 1.1: Capturas de pantalla del juego Crossroads web

y según la probabilidad de conseguir acceder a un estado aceptable, acercándose a 100 cuando la probabilidad sea lo mayor posible, es decir, probabilidad igual a 1. Además, se modificará la escala de colores del indicador dependiendo del valor de la probabilidad para que sea más intuitivo para el usuario. Este indicador también podría ser sustituible por un elemento numérico simple que varíe de 0 a 1, representando la probabilidad numéricamente sin ningún material gráfico. Pero, dado el contexto de juego lúdico en el que se encuadra el proyecto, el elemento gráfico parece ser más adecuado.

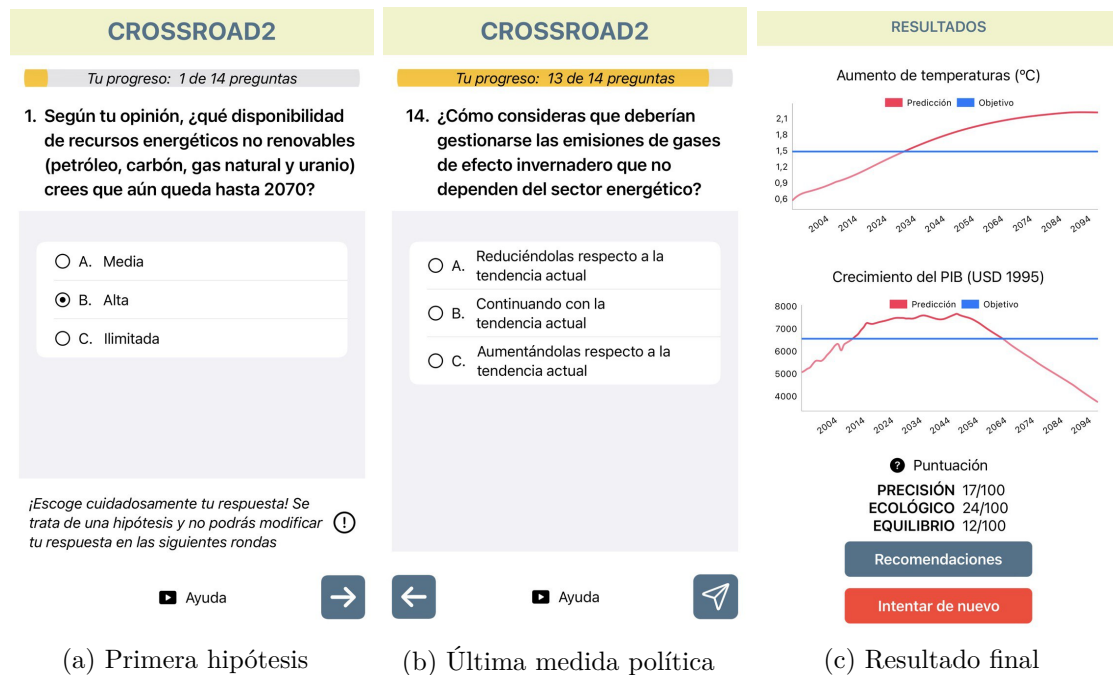


Figura 1.2: Caso de uso del juego Crossroads en la aplicación móvil iOS

1.4. Marco contextual

Este proyecto ha sido desarrollado en el marco del proyecto LOCOMOTION, financiado por la Unión Europea en el programa Horizon 2020 número de contrato 821105 [36] y especialmente en el marco del proyecto de la Fundación Española para la Ciencia y Tecnología (FECYT) 2020 código 16138 titulado ENCRUCIJADA-MUNDO: ECOHERRAMIENTAS LÚDICAS PARA LA TRANSICIÓN ENERGÉTICA [16]. Los modelos de simulación que se utilizan surgen del proyecto MEDEAS [38].

El proyecto se enmarca dentro de una línea de trabajos fin de grado y fin de máster tutelados por David Escudero y Yania Crespo, con apoyo de sus grupos de investigación respectivos y del GIR de la UVa GEEDS, grupo interdisciplinar liderado por Luis Miguel. El objetivo final es el desarrollo de herramientas gamificadas para la concienciación sobre el problema del cambio climático que se apoyen en conocimiento obtenido con simuladores basados en dinámica de sistemas.

Otros proyectos previos y relacionados directamente han sido:

- *Diseño de la interacción y desarrollo del back-end de Crossroads 2.0, un juego educativo para concienciar sobre el cambio climático*, realizado por Lucas González Calderón en el curso 2019/20 en la titulación de Grado en Ingeniería Informática.

- *Diseño de interfaz web para videojuego de trabajo cooperativo*, realizado por Elena Pastor en el curso 2019/20 en la titulación de Grado en Ingeniería en Diseño Industrial y Desarrollo de Producto.
- *Desarrollo del front-end y mejoras en el back-end de un juego didáctico multijugador de competición y consenso sobre el cambio climático*, realizado por Manuel Alda Peñafiel en el curso 2020/21 en la titulación de Grado en Ingeniería Informática.
- *Elaboración y ejecución de un plan de pruebas de usabilidad para juego educativo Crossroads 2.0.* realizado por María Galindo Álvarez en el curso 2020/21 en la titulación de Máster en Ingeniería Informática.
- *Uso de técnicas de Data Mining sobre series temporales obtenidas por simulación y aplicación de resultados en videojuego Crossroads* realizado por Adrián Manzano en el curso 2020/21 en la titulación de Máster en Ingeniería Informática.
- *Desarrollo de una aplicación Android para Crossroads 2.0, un juego educativo para concienciar sobre el cambio climático* realizado por David Crespo en el curso 2021/22 en la titulación de Grado en Ingeniería Informática.
- *Definición y automatización de pruebas de carga y escalabilidad para una aplicación web colaborativa* realizado por Manuel Alda Peñafiel en el curso 2021/22 en la titulación de Máster en Ingeniería Informática.

Este proyecto se beneficia del back-end desarrollado inicialmente por Lucas González y mejorado por Manuel Alda; se emplean datos procesados por Adrián Manzano para calcular el feedback; se ha colaborado con David Crespo y Manuel Alda para realizar una aplicación para un solo jugador en Android e iOS con back-end común.

El proyecto presente, por tanto, se construye sobre un juego desarrollado previamente que se enmarca en un proyecto europeo, LOCOMOTION [36]. El juego se llama Crossroads y su fin es la concienciación sobre los efectos que causan determinadas medidas políticas sobre el cambio climático. La aplicación lúdica existe ahora tanto para versión web como móvil (iOS y Android). El desarrollo de la aplicación móvil para iOS comenzó en diciembre de 2021 y se encuentra financiado por la Fundación Española de Ciencia y Tecnología (FECYT) [16]. Del desarrollo para iOS se encarga la alumna María Robles, por lo que el elemento de juego fruto de este proyecto se implantará en dicha plataforma.

Este proyecto se ha comenzado en la estancia de prácticas de la asignatura I+D+i del Máster en Ingeniería Informática y se ha finalizado en el TFM. La estancia de prácticas de la asignatura I+D+i se ha realizado en colaboración con el G.I.R. ECA-SIMM de la UVa, alguno de cuyos miembros forman parte del proyecto LOCOMOTION. La estancia consta de 190 horas durante las cuales se ha comenzado el desarrollo del trabajo, aunque no ha sido posible terminarlo en su totalidad por restricciones de tiempo como se proveyó. Sin embargo, este trabajo ha conformado la antesala para el desarrollo del TFM (Trabajo Fin de Máster), donde se ha continuado trabajando sobre el mismo tema para completar los objetivos propuestos.

1.5. Estructura del documento

En adelante, este documento se estructura de la siguiente forma:

Capítulo 1. Introducción: Se describe la motivación del problema que se quiere abordar, los objetivos propuestos para ello y se detalla el contexto en el que se encuentra este proyecto.

Capítulo 2. Estado del arte: Se describe la situación actual sobre temas influyentes en este proyecto, como son los juegos educativos para el cambio climático, el *feedback* en videojuegos y el uso de técnicas de inteligencia artificial en juegos. Además de presentar el estado actual en este ámbito, se indica brevemente dónde se encuadra este proyecto.

Capítulo 3. Metodología: Se describen los datos de entrenamiento y se explican los procedimientos realizados para estimar la probabilidad de obtener un estado aceptable medioambientalmente. Se comentan las herramientas utilizadas y se describen una serie de casos de prueba con los que validar el modelo creado.

Capítulo 4. Implementación en la aplicación móvil: Se describen las modificaciones llevadas a cabo para integrar el modelo creado en la aplicación móvil de Crossroads.

Capítulo 5. Resultados: Se presentan los resultados obtenidos aportando evidencias sobre ambas técnicas utilizadas.

Capítulo 6. Discusión: Se interpretan los resultados obtenidos.

Capítulo 7. Conclusiones y líneas de trabajo futuras: Se describen las conclusiones del trabajo de forma general, así como las posibles líneas de desarrollo que presenta el proyecto.

Apéndice A. Plan de Proyecto: Se recoge la planificación inicial llevada a cabo para cumplir los objetivos propuestos, describiendo las tareas y su calendarización. Además, se enumeran los recursos utilizados en el proyecto y se presenta un estudio sobre la viabilidad económica del proyecto.

Apéndice B. Enlaces: Se adjuntan los enlaces de interés del proyecto, donde se puede encontrar todo el código utilizado en él.

Apéndice C. Manual de instalación y despliegue: Se recogen las instrucciones para realizar el proceso de instalación y despliegue del estimador realizado en la aplicación móvil Crossroads.

Apéndice D. Descripción del contenido del juego Crossroads: Se adjunta el contenido del juego Crossroads, incluyendo el texto de cada pregunta y respuesta junto con las ayudas que proporciona la aplicación.

2: Estado del arte

En este capítulo se recoge la información consultada sobre juegos educativos sobre cambio climático, referenciando algunos ejemplos encontrados, junto con información sobre *feedback* en videojuegos, explicando qué es y qué beneficios supone para los jugadores. Además, se explica la aplicación de la inteligencia artificial (IA) en juegos, describiendo el uso de técnicas de IA de forma general y en particular los árboles de decisión enfocados a la predicción de éxito.

2.1. Juegos educativos sobre cambio climático

El cambio climático es un problema ambiental existente que ya ha afectado a ecosistemas y se espera que tenga una evolución desfavorable, la cual suponga el aumento del calentamiento global y riesgos para la salud, dificultades para el acceso a alimentos o agua y perjuicios para el crecimiento económico [57]. Actualmente, el cambio climático es percibido internacionalmente como una de las principales amenazas globales [24]. Esta preocupación por dicha problemática ambiental ha crecido desde 2010 en España, ya que hoy en día la sociedad española es más consciente de la realidad del cambio climático y de que la crisis climática es cada vez más amenazante [7]. Sin embargo, la emergencia climática continúa creciendo y, por tanto, crece la importancia de que los jóvenes sean conscientes de lo que supone el cambio climático. Para ello, es necesario buscar una buena estrategia de comunicación a través de la cual se pueda llegar a los jóvenes y fomentar la concienciación y la acción social. Si se sitúa el foco en la actual población joven, también llamados “nativos digitales” [42], se ha de tener en cuenta que han crecido rodeados de ordenadores y videojuegos. Por este motivo parece lógico indicar que la mejor manera de comunicar el problema sea a través de un videojuego [13].

Actualmente, existen diversos videojuegos sobre el cambio climático. La mayoría se enfocan a la enseñanza en el aula. Por ejemplo, “La emergencia climática” [2] es un portal educativo desarrollado por amBientech, asesor educativo que desarrolla contenidos de ciencia, salud y medio ambiente online para educación secundaria [1]. “La emergencia climática” es uno de esos materiales educativos que proporciona amBientech e incorpora un cuestionario online donde se examina al usuario sobre cuestiones concretas de medioam-

biente y se aportan gráficos informativos sobre el planeta e información sobre los glaciares. Se trata de un cuestionario online para estudiantes con el fin de que consigan afianzar conceptos o conocer nuevos datos de una forma más amena en un contexto gamificado.

“Actúa con tu consumo” [22] es una página web de la mano de Greenpeace en la que se recogen una serie de guías para fomentar un consumo sostenible y respetuoso con el medio ambiente y los derechos humanos. Por ejemplo, la página web cuenta con: “Guía para calcular las emisiones de CO₂ en tus desplazamientos”, “Guía con consejos prácticos sobre la producción y consumo de papel en España” y “Guía general sobre el consumo sostenible”, entre otras. Este material está pensado para ser utilizado en un aula mediante una persona que lo dinamice, aunque también se puede utilizar de manera individual.

Por otra parte, “*My Green Energy Planet*” [39] es un juego español desarrollado por WWF España y Fundación AXA que se encuentra enfocado a los jóvenes. Aunque actualmente parece no proporcionar soporte, se trata de un juego que presenta un mundo virtual para que los participantes realicen acciones con el fin de fomentar el uso de energías renovables y saber tomar decisiones para una realidad ambientalmente sostenible y responsable. Los jugadores de este videojuego tienen la misión de gestionar recursos energéticos, mantener a salvo la comunidad de la contaminación y reaccionar ante alertas meteorológicas. Es un juego dinámico y participativo, con un ranking para conocer cómo están situados los participantes.

El diario *Financial Times* desarrolló un videojuego llamado “*The Climate Game*” [18]. En él, el jugador actúa como un líder mundial que intenta salvar el planeta de los peores efectos del cambio climático. Durante el trascurso del juego, se van realizando preguntas en las que se exponen situaciones y el jugador debe mantener el calentamiento global en 1,5°C, reduciendo las emisiones de dióxido de carbono. Al final de la partida, se le muestra al usuario la evolución del cambio climático que sus decisiones han acarreado.

En nuestro caso, Crossroads es un juego educativo enfocado al alumnado de la ESO y Bachiller. La plataforma web del juego tiene versión multijugador, en la que varios jugadores pueden formar parte de un mismo grupo y tomar las decisiones conjuntamente. Sin embargo, en las plataformas móviles, tanto Android como iOS, la versión es monojugador. En el juego Crossroads [11] los participantes se enfrentan a la toma de decisiones sobre estrategias de mitigación climática y conocerán las consecuencias sociales, económicas y ambientales de las decisiones tomadas a nivel global al final de la partida. A través del juego, los participantes descubren pronto la complejidad del problema y las dificultades para conseguir soluciones ideales. Para presentar el resultado que han obtenido los jugadores, se utiliza una simulación de lo que conllevaría aplicar las medidas seleccionadas de forma global sobre la temperatura. A distinción de los trabajos presentados anteriormente, también se presenta la simulación de lo que sucedería con la situación económica global, lo que permite tener en cuenta la evolución de las cuestiones medioambientales y las cuestiones económicas conjuntamente.

2.2. Feedback en videojuegos

El término *feedback* surge en la década de los años 1940, durante la Segunda Guerra Mundial [56]. Las Fuerzas Armadas de los Estados Unidos trabajaban en un proyecto para guiar a la artillería antiaérea de forma automática mediante el empleo del radar. Se pretendía predecir la trayectoria de los misiles atacantes y, de esta forma, recalcular la orientación de los disparos, utilizando para ello las diferencias entre las trayectorias prevista y real. Como resultado de este proyecto se introduce en la ciencia los conceptos de realimentación y de cantidad de información.

Comenzando con algunas definiciones, las intervenciones de *feedback* o realimentación [33] se definen como aquellas acciones realizadas por agentes externos que proporcionan información sobre algún aspecto del desempeño de una tarea. Mason y Bruning [37] definen *feedback* como cualquier mensaje generado en respuesta a la acción del alumno, cuyo resultado es ayudar a los alumnos a identificar errores y tomar conciencia de los conceptos así como motivar a un mayor aprendizaje. A su vez, John Hattie y Helen Timperley entienden el *feedback* como la información proporcionada por un agente (por ejemplo: profesor, libro, compañero) sobre aspectos de un desempeño o comprensión de una tarea [25]. Winne and Butler definen *feedback* como la información con la que el alumno puede confirmar, añadir, sobrescribir, ajustar o reestructurar la información en su memoria, ya sea conocimiento del dominio, conocimiento metacognitivo, creencias sobre uno mismo o sobre las tareas, o tácticas y estrategias cognitivas [26]. El *feedback* suministra conocimiento al usuario sobre el juego y le guía para tomar decisiones más beneficiosas. Esto implica que exista un flujo interactivo entre el alumno y el sistema [10], proveniente de información previa recopilada o generada por el alumno y que vuelve a él como salida después de haber sido procesada. Independientemente del dominio, la retroalimentación brinda a las personas la información necesaria para tener éxito en el juego [47]. Particularmente en los videojuegos sobre cambio climático, al tratar conceptos complejos como el medioambiente en el que influyen muchos factores, es muy importante el *feedback* para que guíe al usuario hacia un resultado exitoso.

Existen varios tipos de *feedback*. El *feedback* formativo [23] es aquel que ocurre en medio del juego y ofrece información que le permite al jugador cambiar su comportamiento o acción. En [17] se define *feedback* formativo como aquel que permite monitorear el aprendizaje proporcionando comentarios continuos que pueden ser utilizados por los alumnos para mejorar los conocimientos adquiridos. Por otra parte, existe el tipo de *feedback* sumativo [21], el cual ocurre al final de una actividad y generalmente no permite que el jugador cambie su comportamiento en función de la retroalimentación generada. En este proyecto se desea desarrollar *feedback* continuado durante el desarrollo del juego, actualizándolo con cada acción del jugador, para permitir a éste que conozca su situación y que pueda mejorarla. Por este motivo y con las definiciones anteriormente presentadas, el estilo de *feedback* que se desarrolla en este proyecto puede ser de tipo formativo.

2.3. Inteligencia artificial en juegos

Desde la década de 1950 hasta la actualidad, el desarrollo de juegos ha sido un objetivo de investigación importante para el dominio de la inteligencia artificial (IA) [20]. Al inicio, los investigadores se decantaron por los juegos de mesa tradicionales, como el ajedrez, las damas, el backgammon y el póker, consiguiendo finalmente la creación de programas capaces de competir contra oponentes humanos. En este ámbito, el pionero fue Arthur L. Samuel, profesor en la Universidad de Illinois, que tuvo la idea de crear un programa de juego de damas [19]. La investigación inicial se enfocó en juegos de “información perfecta”, como el ajedrez o las damas, lo que requería el desarrollo de soluciones dentro de un número grande de estados completamente especificados, con investigaciones posteriores que permitió extender el dominio de la aplicación al uso de juegos de “información imperfecta”, como el backgammon y el póker, donde se presenta la necesidad de incluir espacios de estados con estados ocultos, juegos probabilísticos y múltiples oponentes [20]. En este juego se emplea una búsqueda en el espacio de todos los posibles estados del juego para aportar una estimación de éxito en función de las decisiones que va tomando el jugador.

Dentro del campo de la inteligencia artificial existen las técnicas de *data mining* o minería de datos. *Data mining* [29] es el proceso para descubrir patrones y otra información valiosa existente en grandes conjuntos de datos. Se utilizan métodos estadísticos para identificar tendencias y otras relaciones significativas en grandes bases de datos [9]. La adopción de técnicas de minería de datos ha aumentado rápidamente en las últimas dos décadas debido a la evolución tecnológica del almacenamiento de datos y al crecimiento de big data. Se pueden distinguir dos tipos de modelos de *data mining*: descriptivos y predictivos [41]. Los modelos descriptivos suelen utilizar algoritmos de aprendizaje no supervisado, como clústering, para reproducir patrones que expliquen o generalicen la estructura intrínseca, las relaciones y las interconexiones de los datos. Por otra parte, los modelos predictivos aplican algoritmos de aprendizaje supervisado, como clasificación o regresión, para estimar valores desconocidos o futuros de variables dependientes basadas en las *features* de las variables independientes relacionadas. Los modelos predictivos suelen utilizar algoritmos como árboles de decisión o redes bayesianas [52]. En el juego Crossroads se emplea clústering para obtener una representación simplificada de las posibles entradas del juego. Una vez identificados los estados objetivos a los que el jugador debe llegar, el problema de estimación de éxito se convierte en un problema de clasificación.

Siguiendo con la técnicas de *data mining* de clasificación, se encuentran los árboles de decisión que son una herramienta que permite resumir la información de los datos de entrenamiento en forma de árbol. Esto se consigue con la aplicación de diferentes algoritmos de árboles de decisión, como *Iterative Dichotomiser 3* (ID3) [45], CHi squared Automatic Interaction Detector (CHAID) [31], *Classification and Regression Trees* (CART) [8] y C4.5 [46], utilizados para clasificación. Un árbol de decisión [49] se forma de arriba hacia abajo, partiendo de un nodo raíz. La construcción del árbol implica que los datos de entrada se dividan en subconjuntos que contengan instancias con valores similares (homogéneos). El algoritmo ID3 es el algoritmo base que se utiliza para realizar árboles de decisión y utiliza la entropía para calcular la homogeneidad de una muestra. En este caso concreto se

utilizará el algoritmo C4.5 [53], una mejora del algoritmo ID3 desarrollado por Quinlan. El algoritmo C4.5 construye un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente siguiendo la estrategia de profundidad primero (*depth-first*). Este algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la de mayor ganancia de información.

Las ventajas que conlleva aplicar árboles de decisión son las siguientes [46] [35]: permite analizar tanto variables continuas como discontinuas eficientemente y los resultados de forma gráfica o en reglas de expresiones pueden ser comprendidos fácilmente. Además, una de las principales ventajas que supone utilizar árboles de decisión es que se puede obtener información implícita del conjunto de datos de entrenamiento. Esto es especialmente interesante en este proyecto porque podemos obtener información sobre qué preguntas son más significativas e importantes para obtener un resultado bueno.

Desde su invención, se han utilizado árboles de decisión [35] en proyectos de diversos dominios, como comercio, marketing y medicina [12] [32]. En el ámbito de los videojuegos también se ha utilizado esta técnica de *data mining*, como en el trabajo [34] en el que se emplean árboles de decisión difusos para extraer las relaciones psicofisiológicas y realizar una clasificación del disfrute o diversión del jugador durante un juego de acción. En el presente trabajo se emplea la técnica de árboles de decisión para realizar la clasificación de alcanzar un buen resultado a partir de una serie de respuestas seleccionadas. En concreto, se utilizará el algoritmo C4.5 para realizar la clasificación de conseguir un resultado adecuado. Esta clasificación, junto con el error de predicción que genera, se convertirá en una estimación de probabilidad de obtener un resultado aceptable. De esta manera se lee la clasificación realizada como una probabilidad numérica. Además, para realizar esta estimación también se aplicará el teorema de bayes, pudiendo obtener un *baseline* con el que comparar el modelo de clasificación.

3: Metodología

En este capítulo se describen los datos utilizados para el entrenamiento del modelo, indicando su origen y formato. Se explican las dos técnicas seguidas para la estimación de probabilidad de éxito. Se comentan brevemente las herramientas tecnológicas utilizadas en la parte práctica de este proyecto. Por último, se muestra un esquema que ilustra el funcionamiento del modelo en el juego móvil y una serie de validaciones que se realizarán para comprobar si el funcionamiento del modelo en la aplicación móvil es el esperado.

3.1. Procedimiento

Conjunto de datos

El cuestionario que introduce el juego consta de 14 preguntas. Dos de ellas son objetivos y no se incluyen en los datos de entrada para el estimador de predicción. La finalidad de estos dos objetivos es establecer el valor de temperatura y PIB que se desea alcanzar. Estos valores son fijados por el alumno y le permiten comparar la evolución de temperatura y PIB contra sus objetivos propuestos, pero no influye en la clasificación que se pretende realizar en este proyecto. Para conocer más sobre las preguntas del juego y su contenido, en el Apéndice D se puede encontrar el texto de cada pregunta junto con sus posibles respuestas y también las ayudas del juego, un texto fijo asociado a cada pregunta con el fin de aportar más conocimiento sobre el tema que se trata.

Con respecto a las 12 preguntas restantes, 3 son hipótesis y 9, medidas políticas. A partir de estas cuestiones se calcula la curva que representa la evolución de temperatura y de PIB a lo largo de los años. Para conseguir la evolución hasta el año 2100 se utiliza el software Vensim. Vensim [58] es un software de simulación de sistemas que permite realizar representaciones de situaciones a lo largo del tiempo basándose en fórmulas matemáticas. Vensim, junto con Medea que se explicará a continuación, se han utilizado en Crossroads y en el proyecto europeo LOCOMOTION para simular el comportamiento de la temperatura y de la economía a lo largo de los años.

En la Figura 3.3 se presenta el proceso seguido en proyectos previos para la creación del conjunto de datos del que se parte en este trabajo. En primer lugar, se parte de un modelo (Medeas .MDL en el esquema 3.3). El modelo MEDEAS (*Modeling the Renewable Energy Transition in Europe*) [38] predice el comportamiento de ciertos indicadores socioeconómicos y medioambientales, como puede ser el PIB per cápita o el incremento de la temperatura media global. Este modelo recoge el conocimiento de cómo evolucionarían o se comportarían estos indicadores relacionados con el cambio climático y el consumo de recursos. A este modelo toma como entrada más de 5000 variables de entrada (**Variables de entrada** en el esquema 3.3) que son números reales. Estas variables de entrada, tras ser procesadas con el modelo, da lugar a diferentes posibles escenarios futuros que ha predicho el modelo dependiendo de las variables. Este procesamiento se realiza con el software de modelado de sistemas Vensim (**System Dynamics** en el esquema) que tomando como entrada el modelo Medeas junto con las variables de entrada, realiza repetidas ejecuciones de forma exhaustiva para obtener todos los posibles escenarios futuros donde se recoge el comportamiento que tendría cada indicador socioeconómico y ambiental. Esta salida que genera el software Vensim se ha representado como **Dataset** en el esquema 3.3. En relación al juego Crossroads, las diferentes opciones que presenta el formulario se representan con un conjunto de número naturales correspondiendo con las posibles respuestas (de la ‘a’ a la ‘e’) que se pueden contestar. A este conjunto de números naturales (**Formulario** en el esquema) se le aplica un proceso de *mapeo* con el cual para cada opción del formulario existe una variable de entrada que se define en el fichero **Variables de entrada**. Estas variables, como se ha comentado, se aplican al modelo Medeas utilizando Vensim y se obtiene la salida, nombrada como **Dataset** en el esquema funcional.

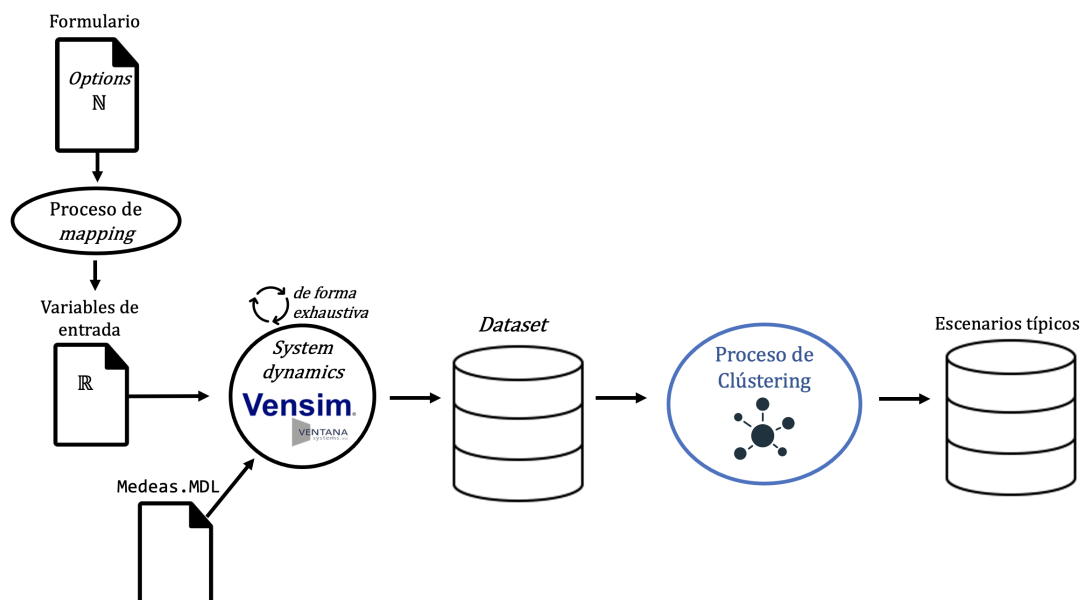


Figura 3.3: Creación del conjunto de datos inicial

Continuando con el esquema de la Figura 3.3, una vez obtenido el *Dataset* o salida que genera Vensim se ha aplicado la técnica de *clustering*. Se trata de un trabajo previo que se realizó con motivo de un proyecto fin de Máster [48], en el cual se agruparon las simulaciones en diferentes clústers dependiendo de la evolución de temperatura y PIB que generaban. Se ha realizado la técnica de *clustering* jerárquico aglomerativo en forma de árbol, denominado dendograma. Al aplicar *clustering* al conjunto de datos se consigue reconocer algunos patrones típicos de los datos y reducir la cantidad de elementos que componen el *dataset*. Con esta técnica de reducción de la dimensionalidad, se sustituye el conjunto de datos original por un conjunto alternativo de menor tamaño, pero representativo del original. Esta agrupación en clústers permitió asignar al *dataset* una clase cualitativa que identifica el clúster en el que se agrupa ese registro dependiendo de los patrones típicos reconocidos. Concretamente, al hablarse de datos que presentan una evolución se necesita la reducción de series temporales. Al finalizar este proceso de *clustering*, se puede sustituir cada resultado de la simulación por el clúster al que haya sido asignado. Esta agrupación se muestra en la Figura 3.4, donde se puede apreciar la distribución de los resultados de la simulación. Las columnas dividen los registros en función de la evolución de PIB y las filas, en función de temperatura. En concreto se poseen 8 clústers para PIB y 6 para la temperatura. En color azul se muestran los clústers que se consideran adecuados, cuya evolución es estable o no supera unos valores determinados. En gris se muestran aquellos que podrían ser registros admitidos con un ligero error. De esta forma, se puede comprobar la evolución media que puede tener la gráfica del PIB y de la temperatura en los resultado ideales (celdas azules). El conjunto de registros con simulaciones junto el clúster al que pertenecen, forman los **Escenarios típicos** que se muestran como resultado del proceso presentado en el esquema funcional en la Figura 3.3.

Los escenarios de temperatura adecuados son los patrones 0, 1 y 2 (Figura 3.4), ya que la temperatura se estabiliza en valores próximos a 2.5°C. Por otra parte, los patrones adecuados de PIB son 3, 4 y 5 donde el PIB no decrece. En los patrones 6 y 7 el PIB experimenta un crecimiento pero son escenarios irreales porque en ellos se asumen recursos ilimitados. Finalmente, los registros considerados adecuados representan menos del 10 % del total.

Como las evoluciones del medioambiente y economía se han calculado vía software, esto genera la posibilidad de obtener la simulación para todas las posibles combinaciones de parámetros de entrada. Es decir, para las 12 preguntas que influyen en ambas evoluciones, se tiene asociada la solución para cada combinación de esas 12 preguntas existentes con sus posibles respuestas.

Datos de entrenamiento

Los datos que se han utilizado para el proceso de entrenamiento del modelo en este proyecto, denominados como “Escenarios típicos” en el esquema de la Figura 3.3, es el resultado de trabajos previos realizados. Este *dataset* cuenta con 437.400 registros que almacenan todas las posibles combinaciones de respuestas para las preguntas. En concreto, los datos de los que se parte en este proyecto cuentan con los atributos presentados en la

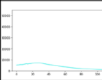
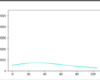
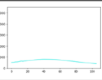
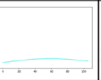
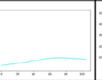
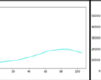
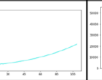
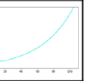

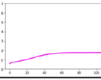
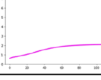
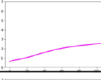
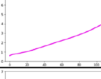
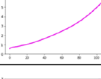
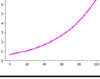
		0	1	2	3	4	5	6	7
									
0		0	2,6	1,1	4,4	0,7	0,1	0,2	0*
1		0	0,5	0	0,1	1,2	0*	1,2	0,4
2		10,3	5,78	3,60	15,90	1,5	0,1	0,1	0
3		31,5	3,7	0,7	11	0,5	0	0	0
4		0	0	0	0	0	0,8	0,6	1,2
5		0	0	0	0	0	0	0	0,5

Figura 3.4: Distribución de los resultados de simulación sobre para la agrupación utilizada en el recomendador en porcentaje respecto del total. [48]

Tabla 3.1. Los atributos que se han considerado han sido las 12 preguntas del cuestionario, la etiqueta del clúster al que pertenece el registro y la temperatura y PIB finales.

Atributo	Descripción	Valores
H1	Valor discreto que almacena la respuesta asociada a la hipótesis 1: “Disponibilidad de recursos no renovables”	{a, b y c}
H2	Valor discreto que almacena la respuesta asociada a la hipótesis 2: “Impacto del cambio climático en la economía”	{a, b y c}
H3	Valor discreto que almacena la respuesta asociada a la hipótesis 3: “Evolución de la población mundial”	{a, b, c, d y e}
M1	Valor discreto que almacena la respuesta asociada a la medida política 1: “Evolución del PIB per cápita”	{a, b, c, d y e}
M2	Valor discreto que almacena la respuesta asociada a la medida política 2: “Programa de bosques”	{a y b}
M3	Valor discreto que almacena la respuesta asociada a la medida política 3: “Evolución de la energía nuclear”	{a, b y c}
M4	Valor discreto que almacena la respuesta asociada a la medida política 4: “Generación de energía renovable”	{a, b y c}
M5	Valor discreto que almacena la respuesta asociada a la medida política 5: “Producción de biocombustibles líquidos”	{a, b y c}
M6	Valor discreto que almacena la respuesta asociada a la medida política 6: “Transformación del transporte terrestre”	{a y b}
M7	Valor discreto que almacena la respuesta asociada a la medida política 7: “Eficiencia energética”	{a, b y c}
M8	Valor discreto que almacena la respuesta asociada a la medida política 8: “Reciclado de minerales”	{a y b}
M9	Valor discreto que almacena la respuesta asociada a la medida política 9: “Emisiones de gases de efecto invernadero”	{a, b y c}
cluster_label	Valor discreto que almacena el clúster al que pertenece el registro dependiendo de la evolución de temperatura y PIB. Hay 26 clústers diferentes.	{0..26}
finalTemp	Valor continuo que indica el valor de temperatura final obtenida en la evolución, es decir, el valor de temperatura simulado para el año 2100 que es el último año de la simulación. La temperatura se encuentra en grados celsius.	Min: 1.41 Max: 7.72
finalPIB	Valor continuo que indica el valor de PIB final de la evolución en USD, es decir, el PIB obtenido para el año 2100 que es el último año de la evolución.	Min: 275.80 Max: 55787.80

Tabla 3.1: Atributos del conjunto de datos

A partir de la agrupación de los registros en clústers, se han escogido aquellos registros que están agrupados en los clústers que se consideran aceptables, dependiendo de la evolución de PIB y temperatura. Los clústers seleccionados son los que se marcan en azul en la Figura 3.4 y en el atributo `cluster_label` toman los siguientes valores: 03, 04, 13, 14 y 15, correspondiendo el primer dígito al patrón de temperatura en el que se encuentran (0 ó 1) y el segundo, al patrón de PIB (3, 4 ó 5) [48]. Además, se ha realizado un filtrado a mayores de los registros que se encontraban en estos 5 clústers. Se han filtrado aquellos registros cuya temperatura final (atributo `finalTemp`) sea mayor de 1.75°C, ya que en el objetivo de París se propusieron estabilizar la temperatura en 1.5°C. Con respecto al PIB, se han filtrado aquellos registros cuyo valor final de PIB (atributo `finalPIB`) se encuentre en el rango de 6000 a 7500 USD, ya que se consideran valores económicamente aceptables.

Para clasificar los registros considerados aceptables, se ha añadido un atributo nuevo al conjunto de datos denominado `solucion`. Sus valores posibles son `solucion_aceptable` y `solucion_no_aceptable`. Aquellos registros que se encuentran en los clústers con evolución adecuada y, además, cumplen con los rangos finales establecidos, se han almacenado como registros con `solucion_aceptable`. Todos los demás registros que no cumplen alguna de estas condiciones se han etiquetado como `solucion_no_aceptable`. En concreto, de los 437.400 registros que contiene el conjunto de datos, 1.623 se han etiquetado como `solucion_aceptable`. En resumen, este nuevo atributo (“*booleano*”) se corresponde con la clasificación que se utilizará posteriormente para predecir y en ella se distinguen aquellos registros con una solución aceptable o no, dependiendo de la simulación de PIB y temperatura que generan.

3.2. Predicción de probabilidades de éxito

Se presenta a continuación los métodos empleados para estimar la probabilidad de obtener un resultado aceptable sabiendo las opciones que selecciona el usuario. Se quiere computar $P(H|E)$, que denota la probabilidad de E condicionada a H donde H denota un estado aceptable y E es un subconjunto de opciones seleccionadas por el usuario, por ejemplo: $E = \{P1='a', P2='b', P3='c', P4=NA, P5=NA, \dots, P12=NA\}$ siendo P_i la opción elegida en la pregunta ‘i’. Cabe destacar que en este ejemplo se ha querido representar a partir de la pregunta 4 que el usuario aún no ha contestado esas preguntas. Se han marcado como valores no disponibles (N.A.) las preguntas restantes, para representar que los *missing values* pueden formar parte del conjunto de respuestas. El hecho de soportar *missing values* es un aspecto realmente importante en el contexto de juego en el que nos encontramos, ya que actualmente en el juego el usuario puede ir contestando preguntas sin ningún orden y sin ninguna restricción. Por ello, el modelo de predicción debe ser tolerante con las preguntas que el usuario aún no haya contestado, es decir, el modelo debe soportar *missing values*. Esta característica de valores aún no disponibles ha supuesto una condición en las herramientas utilizadas, ya que la biblioteca `scikit-learn` que se pretendió utilizar en un primer momento es incompatible con la existencia de valores aún no disponibles en el conjunto de datos [51]. Tras haber estudiado los datos de entrada y conocer la importancia de *missing values* en este conjunto de datos, se decidió utilizar

Weka, herramienta que sí soporta la existencia de valores aún no disponibles y se describe en la Sección 3.3.

Estimación empírica

Para una primera aproximación de computar la probabilidad de obtener un resultado aceptable sabiendo las opciones que selecciona el usuario se ha aplicado un método empírico, el teorema de Bayes. Este teorema permite conocer esta probabilidad condicionada y su desarrollo conformará un *baseline* con el que se compare posteriormente el modelo de *machine learning*.

El teorema de Bayes [61] de la probabilidad condicional postula que si se parte de una hipótesis H y una evidencia E que se basa en esa hipótesis, entonces

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \quad (3.1)$$

Aplicando este teorema, se utiliza $P(E)$ para denotar la probabilidad de las opciones escogidas por el usuario entre todas las opciones posibles. Como se ha comentado anteriormente, E indica un subconjunto de las opciones que haya seleccionado el usuario, por ejemplo $P1='a'$, $P2='b'$, $P3='c'$ siendo P_i la opción elegida en la pregunta 'i'. $P(H)$ denota la probabilidad de conseguir un resultado cuyo estado sea aceptable, o lo que es lo mismo, equilibrado en cuanto a la evolución de temperatura y PIB. La probabilidad condicionada $P(E|H)$ es la probabilidad de la selección de unas opciones concretas sabiendo que el estado que se alcanza es aceptable. Estas tres probabilidades se calculan como se explica a continuación, teniendo en cuenta que parte de un conjunto de datos en el cual se encuentran todas las posibles combinaciones de respuestas de las 12 preguntas con las que cuenta el juego y la clasificación creada que indica si se genera una solución aceptable o no.

- $P(H)$ se puede estimar empíricamente como el resultado de la división entre el número de registros con estado aceptable sobre el número total de registros, con todas las combinaciones posibles de opciones. Por tanto, sería: $P(H) = n_b/n_t$, donde:
 - n_b : número de registros clasificados con estado aceptable
 - n_t : número de registros totales, es decir, el número de posibles combinaciones existentes de las opciones
- $P(E)$ se puede estimar empíricamente como la probabilidad de que se seleccionen esas opciones determinadas. Es decir, $P(E) = n_o/n_t$, donde:
 - $n_o = \prod C_i$: cardinalidad de opciones no contestadas, siendo $i \in \text{opciones_no_contestadas}$ y C_i la cardinalidad de cada pregunta i .
 - n_t : número de registros totales

- $P(E|H)$ se puede estimar empíricamente como el número de registros con estado bueno que contienen las opciones seleccionadas sobre el número de registros buenos. Por lo tanto, quedaría: $P(E|H) = n_{bo}/n_b$ donde:
 - n_{bo} : número de registros con estado bueno que contienen las opciones seleccionadas
 - n_b : número de registros totales clasificados con estado bueno

Asumiendo las probabilidades y el cálculo de cada una de ellas comentado, se puede obtener $P(H|E)$ que denota la probabilidad de alcanzar un estado aceptable dada una selección de opciones como se ha mostrado en la Ecuación 3.1. Si se sustituye y simplifica, quedaría

$$P(H|E) = \frac{\frac{n_{bo}}{n_b} \cdot \frac{n_b}{n_t}}{\frac{n_o}{n_t}} = \frac{n_{bo}}{n_o} \quad (3.2)$$

Por tanto, la probabilidad de alcanzar un estado aceptable dado un conjunto de opciones seleccionadas se calcularía a partir de la división entre el número de registros con estado aceptable que contienen esas opciones entre el número de registros que contienen esas opciones.

Se asume que el cálculo de las estimaciones generadas por el teorema de Bayes es muy preciso porque el *dataset* contiene una colección de valores exhaustiva del número de posibles entradas, es decir, todas las combinaciones posibles de respuestas. Esto no es posible en otros casos porque la ejecución de tantas iteraciones del modelo con Vensim conlleva mucho tiempo de procesamiento, por lo que se suele realizar un muestreo con un conjunto de datos de menor tamaño pero representativo. Por este motivo, no es posible en la mayoría de casos contar con un conjunto de datos tan exhaustivo, por lo que probar un estimador basado en *machine learning* parece conveniente.

Estimación con árbol de decisión

Para crear un sistema de predicción que indique si se puede obtener un resultado aceptable o no a partir de una serie de valores de entrada, se pueden utilizar árboles de decisión. Los árboles de decisión [27] es un algoritmo de *machine learning* supervisado. Los árboles se construyen a partir de los diferentes valores de parámetros y destacan en ellos dos elementos: los nodos de decisión y las hojas. Las hojas del árbol es el resultado final, que en este caso concreto sería la etiqueta de clasificación (solución aceptable o solución no aceptable). En los nodos de decisión se dividen los datos dependiendo de los valores de las *features* del conjunto de datos.

Los árboles de decisión se pueden interpretar como estimadores de probabilidad [15], se conecta la estimación de probabilidad con la minimización del error de predicción. Siendo E un subconjunto de opciones escogidas por el usuario y H un estado aceptable,

las opciones escogidas en E tiene probabilidad $P(H|E)$ de ser un miembro de la clase H . Un árbol de decisión (denotado por θ) puede considerarse como un estimador de la verdadera probabilidad. En este caso, la probabilidad estimada depende del conjunto de opciones de E y del árbol de decisión θ , por lo que se denota por $P(H|E, \theta)$. Normalmente, $P(H|E, \theta) \neq P(H|E)$ a menos que el verdadero modelo objetivo sea el propio árbol de decisión θ o se haya verificado exhaustivamente para cada E . Suponiendo que n es el número de ejemplos en un nodo hoja de un árbol que clasifica E y n_y es el número de ejemplos de n que tienen etiqueta de clase H , la probabilidad estimada de que E sea etiquetado de la clase H es

$$P(H|E, \theta) = \frac{n_y}{n}$$

De esta forma, se puede obtener la probabilidad numérica extraída del conocimiento que almacena el árbol de decisión.

Utilizando la API de Weka desde Python, se ha creado un árbol de decisión con el algoritmo C4.5 que permite clasificar los datos en solución aceptable o no dependiendo de las respuestas dadas por el usuario. Es decir, el clasificador predice si se puede conseguir un resultado aceptable o un resultado no aceptable a partir de una secuencia de respuestas. Con esta implementación el usuario posee libertad de navegar por el formulario contestando preguntas en el orden que considere y el modelo creado con Weka soporta valores faltantes o *missing values* como parámetros, que serán aquellas preguntas que aún no han sido contestadas.

El modelo devuelve como salida la clasificación junto con un porcentaje de precisión de esa predicción. Cuanto mayor sea ese número, más seguridad habrá de que el registro pertenezca a la clase predicha. En este caso, se ha realizado un procesamiento a la salida generada por el modelo para que devuelva únicamente la probabilidad de que se consiga un estado aceptable. Por tanto, el modelo creado te devuelve la probabilidad de que dado un conjunto de respuestas se obtenga un estado aceptable.

3.3. Herramientas

A continuación se explican las herramientas utilizadas para el desarrollo del proyecto. Las tecnologías utilizadas para las cuestiones de gestión del proyecto se mencionan en el Apéndice A (Sección A.2).

Python

Para el desarrollo del proyecto, el lenguaje principal utilizado ha sido **Python** [44]. Es un lenguaje de alto nivel multiparadigma, ya que soporta orientación a objetos, programación imperativa y funcional. En este caso, se han utilizado varias bibliotecas disponibles para leer, explorar, limpiar y transformar datos, como **numpy** y **pandas**.

Para procesar ficheros utilizando el lenguaje Python se ha utilizado **Anaconda** [3], una distribución libre que permite realizar tareas de ciencia de datos y *machine learning* con Python. Esta herramienta simplifica el despliegue y la administración de las bibliotecas. Concretamente, se ha utilizado para desplegar de forma sencilla **Jupyter notebook** [30], un entorno informático interactivo basado en la web. Un documento de Jupyter Notebook es un documento JSON que contiene una lista ordenada de celdas de entrada/salida que pueden contener código o texto, entre otras. En este proyecto, los cuadernos Jupyter se han utilizado para crear, editar y ejecutar fácilmente código Python para el desarrollo del proyecto.

Además de utilizar Python para el estudio y transformación de datos, se ha utilizado para implementar la funcionalidad del teorema de Bayes (Sección 3.2). Se ha realizado una función en Python con la cual se calcula la probabilidad de llegar a un estado aceptable partiendo de una selección de opciones concretas. Esta función utiliza la base de datos con los registros de las evoluciones y la clasificación de solución aceptable. El principal motivo por el que se ha realizado una funcionalidad del teorema *ad hoc* y no se ha utilizado ninguna proporcionada por terceros es la posibilidad de tener opciones no observadas, es decir, por si alguna de las 12 preguntas aún no ha sido contestada. Esta solución particular permite computar la probabilidad sin tener que haber contestado todas las preguntas, es decir, teniendo *missing values* como parámetros de la función, que como se ha comentado previamente al inicio de la Sección 3.2, es una restricción derivada del contexto del juego. Además, el usuario puede contestar las preguntas en el orden que considere sin que esto influya en la funcionalidad implementada.

Weka

Weka [61] es una colección de algoritmos de *machine learning* utilizados para minería de datos. Posee herramientas que permite la preparación, clasificación, regresión, agrupación, minería de reglas de asociación y visualización de datos. Se trata de un software de código abierto, escrito en Java. Posee una interfaz gráfica para crear modelos y leer datos de forma sencilla, así como una interfaz de línea de comandos a través de la cual se pueden crear modelos y guardar con un solo comando y diferentes *flags*. Además, se puede importar como biblioteca en programas de Java o Python.

En concreto, en este proyecto se ha utilizado en Python. Se ha utilizado la biblioteca `python-weka-wrapper` [43] que se trata de una API de Weka para Python que permite ejecutar los algoritmos de Weka en ficheros Python. A través de la API de Weka se puede acceder a todas las operaciones que posee Weka desde código Python. La principal razón por la que se ha utilizado Weka y no otra biblioteca de Python para *machine learning*, como puede ser `sklearn`, es la posibilidad de tratar los valores faltantes o *missing values* y su importancia en el proyecto. En el contexto de este trabajo, es de gran relevancia poder manejar *missing values*, ya que el usuario puede tener aún preguntas pendientes sin contestar. Los algoritmos de árboles de decisión o Naive Bayes que posee `sklearn` no permiten trabajar con conjuntos de datos que posean valores faltantes, por lo que se ha

tenido que optar por trabajar con Weka y crear una implementación propia del teorema de Bayes utilizando Python.

En concreto, se ha utilizado la biblioteca `python-weka-wrapper` para la creación y evaluación del árbol de decisión. Se ha utilizado el algoritmo J48 de Weka, que implementa el algoritmo C4.5 de árboles de decisión.

Framework Flask

Flask [40] es un framework escrito en Python que permite realizar aplicaciones web en pocas líneas de código. En este proyecto, se ha utilizado Flask para realizar la API que permita añadir al backend de la aplicación Crossroads 2.0 actual la nueva operación de estimación de bondad sobre las opciones. De esta forma, se ha creado fácilmente una API para que desde la aplicación móvil se pueda acceder a la probabilidad que devuelve el clasificador generado.

Xcode, Swift y SwiftUI

Xcode [6] es un entorno de desarrollo integrado para sistemas operativos macOS, con las herramientas necesarias para desarrollar aplicaciones para sistemas operativos de la familia Apple. En concreto, en este trabajo se ha utilizado Xcode para desarrollar la aplicación del juego Crossroads, cuyo desarrollo y mantenimiento ha continuado activo durante el desarrollo de este trabajo fin de Máster. La aplicación se ha desarrollado en **Swift** [4], lenguaje de programación creado por Apple para el desarrollo de aplicaciones para iOS y macOS en 2014. La interfaz gráfica de la aplicación móvil Crossroads se ha desarrollado con **SwiftUI** [5], framework que permite la creación de vistas utilizando código en Swift.

3.4. Validación

A continuación se presentan esquemas gráficos para describir cómo funcionarían los estimadores integrados en la aplicación móvil. Ambos estimadores se han creado con los datos de entrenamiento y la clasificación comentados en la Sección 3.1.

En la Figura 3.5 se muestra el esquema funcional del estimador realizado con el teorema de Bayes. En esta figura se muestra el flujo de información que se generaría al utilizar el estimador integrado en la aplicación. Este flujo comienza cuando el usuario selecciona opciones de las preguntas del juego Crossroads a través de la aplicación móvil. Un posible subconjunto de opciones seleccionadas podría ser: {1: 'a', 2: 'b', 3: 'b', 4: ?, ..., 12: ?} donde se almacena la opción seleccionada para cada una de las 12 preguntas indicando con un signo de interrogación aquellas preguntas que no han sido contestadas aún, *missing values*. Este conjunto de opciones es la entrada del estimador bayesiano (“Input: opciones” en el esquema). Para la previa creación del estimador se han utilizado los “Escenarios típicos” cuya procedencia se describe en la Sección 3.1. Una vez que se le envía al estimador bayesiano el subconjunto de opciones seleccionadas, es necesario realizar

consultas a la base de datos “Escenarios típicos” para computar la probabilidad. La salida que genera el estimador es un número entre 0 y 1 que representa la probabilidad de obtener un resultado bueno habiendo seleccionado ese conjunto de opciones. Esta probabilidad se muestra en un indicador gráfico en la aplicación móvil y permite guiar al usuario durante el desarrollo de la partida.

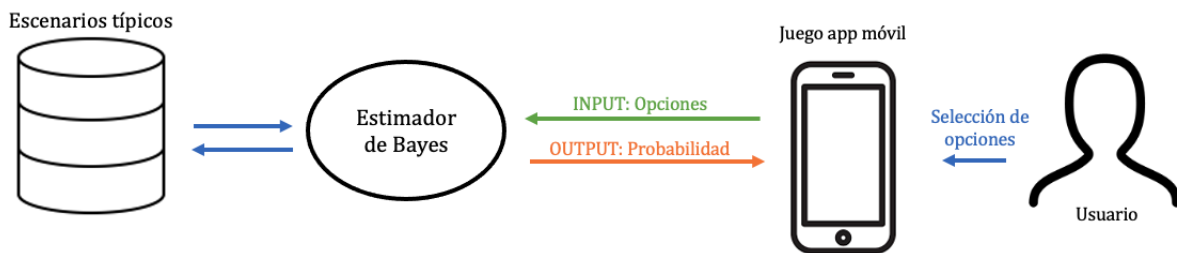


Figura 3.5: Esquema funcional del estimador bayesiano

De forma análoga, en la Figura 3.6 se presenta el esquema que indica el funcionamiento del estimador basado en un árbol de decisión. Este esquema aunque recibe la misma entrada y produce la salida en el mismo formato que el bayesiano (Figura 3.5), posee ciertas diferencias. De nuevo, el flujo de información comienza cuando selecciona las opciones de las diferentes preguntas a través del juego móvil de Crossroads. Estas opciones son la entrada que recibe el estimador basado en árbol de decisión. Para la creación de este estimador basado en el algoritmo C4.5, se han utilizado para entrenamiento la base de datos de “Escenarios típicos”. A diferencia del estimador bayesiano, el modelo clasificador no necesita acceder de nuevo a los datos de entrenamiento cuando realiza el cómputo de la probabilidad. Tras realizar el cálculo, el estimador devuelve la probabilidad de obtener un buen resultado a partir de las preguntas seleccionadas. Este número entre 0 y 1 se muestra en la aplicación móvil a través de un elemento gráfico para poder conducir al usuario a un buen escenario durante el juego.

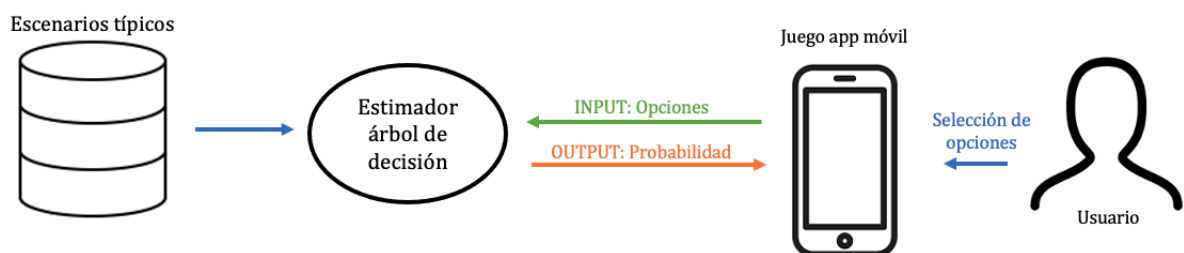


Figura 3.6: Esquema funcional del estimador basado en árboles de decisión

Para representar la probabilidad generada con el estimador basado en árbol de decisión en la aplicación móvil se ha creado el indicador gráfico que se describe en la Sección 4.2.

Para el proceso de validación que determine el correcto funcionamiento del modelo creado y su implantación en el juego, se han elaborado una serie de casos de prueba para constatar resultados. Los casos de prueba son los descritos a continuación:

- Elegir de forma secuencial la opción que aporte mayor probabilidad en cada pregunta. Se espera obtener un buen resultado evolutivo tanto medioambiental como económico.
- Elegir de forma secuencial la opción que aporte menor probabilidad en cada pregunta. Se espera obtener un mal resultado evolutivo, es decir, o se que se incremente enormemente la temperatura global o se desplome el PIB, o ambas.

Los resultados obtenidos tras realizar estas pruebas se recogen en las Secciones [5.1](#) y [5.2](#).

4: Implementación en la aplicación móvil

En este capítulo se explican las modificaciones llevadas a cabo para conseguir implementar en la aplicación móvil el indicador gráfico, describiendo para ello las modificaciones que han sido necesarias tanto en la parte de *back-end* como de *front-end*. Los repositorios donde se encuentra cada proyecto de los que se comentan en este capítulo se encuentran en el Apéndice B.

4.1. API para el estimador

Una vez que se ha creado el modelo final basado en el algoritmo C4.5 que clasifica si una secuencia de respuestas conduce a un estado bueno o no, se ha realizado una función en el lenguaje Python. A continuación se explica brevemente esta función creada, comentando la entrada que toma, la salida que genera y el procedimiento que lleva a cabo:

- **Input:** diccionario con las respuestas contestadas, donde cada **key** del diccionario es el índice de la pregunta (de 1 a 12) y el **value** es la respuesta (de 'a' a 'e'). Si alguna pregunta de 1 a 12 no se encuentra en el diccionario, significa que es una pregunta aún no contestada.

Ejemplo: {1: 'b', 2: 'b', 5: 'd'}

- **Output:** número en el rango de 0 a 1 que indica la probabilidad de obtener un resultado aceptable con las respuestas dadas.

Ejemplo: 0.03

- **Procedimiento:** para obtener una estimación a partir de una serie de respuestas se utiliza el modelo basado en el algoritmo C4.5 creado con Weka. Con la entrada que recibe como parámetro se crea una instancia que clasifica el modelo, el cual se lee desde Python utilizando la biblioteca `python-weka-wrapper`. Tras obtener la clasificación, se tiene como resultado la probabilidad de conseguir un estado bueno, que es la salida de la función.

Para acceder a esta función es necesario crear una interfaz de programación de aplicaciones (API por sus siglas en inglés). Para ello se ha utilizado el framework `flask` que permite la creación de una API en pocas líneas de código. Esta API se compone de una operación HTTP POST a través de la cual se accede a la función que calcula la estimación anteriormente descrita. Su funcionamiento se muestra en la Figura 4.7, donde se ha incluido una captura de pantalla de la aplicación Postman a través de la cual se ha realizado la llamada a la API implementada.



Figura 4.7: Ejemplo de operación POST desde Postman para la estimación de estado bueno

Para utilizar la función creada a través de la API, es necesario enviar una operación POST a la ruta `/estadoBueno`, añadiendo las respuestas seleccionadas en formato JSON como *body* de la petición, como se muestra en la parte superior de la captura de pantalla de la Figura 4.7. En la parte inferior de esa misma figura se muestra la salida obtenida en formato JSON, donde en la etiqueta denominada `estimation` contiene la probabilidad de conseguir un estado bueno.

4.2. Modificaciones *front-end*

Para añadir el elemento gráfico a la aplicación móvil iOS ha sido necesario crear una copia del proyecto principal donde se está desarrollando. Para ello se ha copiado el proyecto que se encuentra alojado en <https://gitlab.inf.uva.es/marrobl/crossroad2>.

Para representar la probabilidad generada con el estimador basado en árbol de decisión en la aplicación móvil, se ha creado el indicador gráfico que se muestra en la Figura 4.8. En esta figura se muestran 5 capturas de pantalla de 5 estados del indicador con probabilidades diferentes. Para representar una probabilidad numérica de forma agradable visualmente, debido al contexto de juego en el que se encuadra el proyecto, se ha definido

este elemento gráfico antes que un simple indicador numérico, ya que puede ser más atractivo visualmente para el alumnado.

Este elemento está compuesto por una semicircunferencia gris fija, a la cual se superpone un sector que aumenta o reduce su ángulo en función del número de probabilidad. En el centro de la figura se indica numéricamente la probabilidad escalada en el rango de 0 a 100. A la vez que se aumenta o disminuye la distancia que ocupa el sector también se modifica el color dependiendo del rango numérico en el que se encuentre la probabilidad, para que sea un elemento más intuitivo para el usuario. Con respecto a la gama de colores, en la Figura 4.8a se muestra una probabilidad baja del 16% por lo que el color que predomina es el rojo advirtiendo al usuario que se encuentra en un posible estado de alerta o peligro. Sin embargo, a medida que la probabilidad aumenta, el color rojo asociado a un resultado negativo o perjudicial desaparece y, a su vez, aparece el verde, indicando al usuario que se acerca a un resultado más beneficioso.

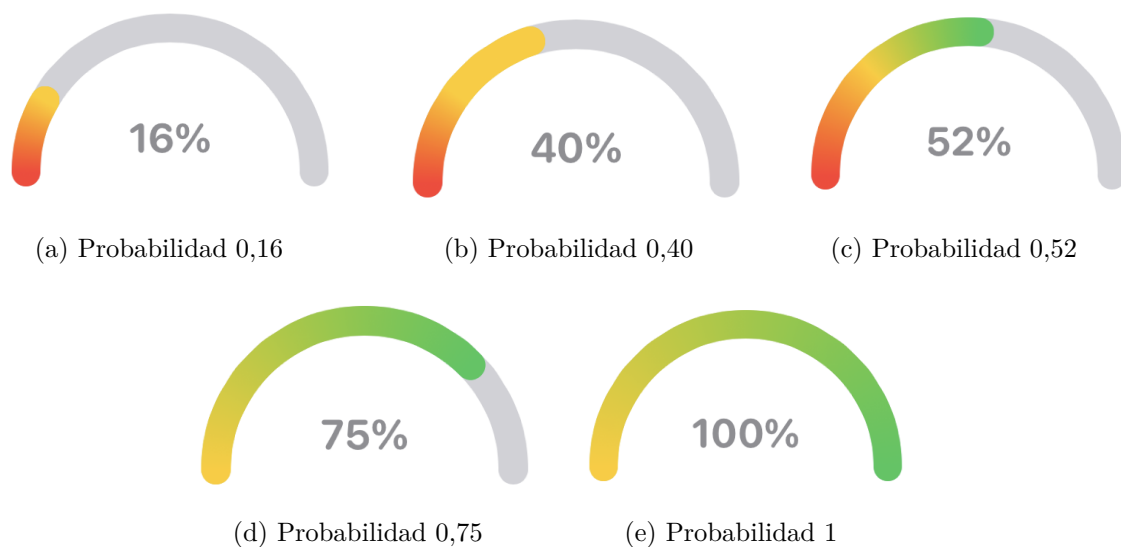


Figura 4.8: Indicador gráfico

Este ítem gráfico se ha creado desde cero utilizando para ello Swift y SwiftUI y se ha añadido a la copia del proyecto de iOS de Crossroads. Se ha creado una vista que contiene este elemento y a la cual se le envía por parámetro la probabilidad a mostrar. Tras haber incluido este ítem en la aplicación de Crossroads, el resultado obtenido es el presentado en la Figura 4.9. Es una captura del simulador durante la partida, en la que se muestra que se han contestado 6 preguntas y la probabilidad de obtener un resultado buen que muestra el indicador es del 16%. Este indicador varía dinámicamente cuando el usuario contesta una nueva pregunta o modifica alguna contestada previamente.

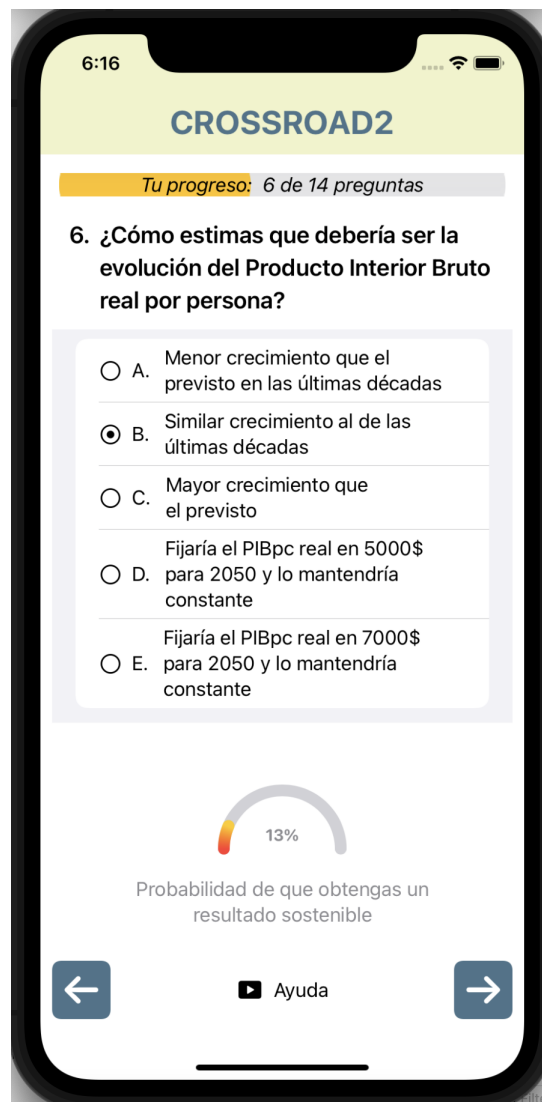


Figura 4.9: Resultado final del indicador gráfico en la aplicación móvil.

4.3. Modificaciones *back-end*

El *back-end* de Crossroads está desarrollado en lenguaje Java. Como el estimador de este proyecto está implementado en lenguaje Python, ha sido necesario crear la API en Python para que desde el *back-end* de Java se pueda acceder a la función Python.

Para ello, se ha duplicado el *back-end* de Crossroads y se ha añadido el código necesario para crear una nueva operación **POST**. En concreto, los detalles de la operación añadida se muestran en la Figura 4.10 que contiene una captura de pantalla de la documentación de Swagger, donde se encuentran todas las operaciones que proporciona la API.

Esta operación recibe por parámetro un diccionario con un conjunto de preguntas y respuestas y realiza a su vez una petición **POST** a la API de Python, la cual devuelve la

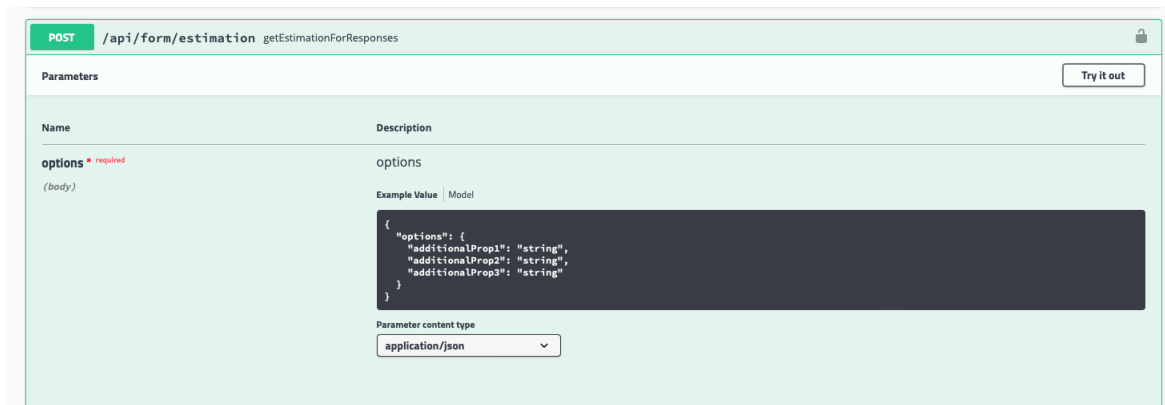


Figura 4.10: Operación añadida al *back-end* de Crossroads

estimación. La operación añadida al *back-end* Java devuelve a su vez la estimación que ha recibido. Esta función añadida al *back-end* en Java es la que se utiliza en la interfaz de iOS para mostrar la probabilidad de obtener un estado bueno. Es decir, la operación implementada en Java actúa de intermediario para enviar la información desde la aplicación Crossroads a la función de Python y a la inversa. De esta forma, se ha conseguido añadir código Python al *back-end* existente desarrollado en Java, logrando que se muestre en la aplicación móvil la estimación calculada por el modelo basado en el algoritmo C4.5.

5: Resultados

Tras haber desarrollado la función que calcula la probabilidad de obtener un resultado exitoso basada en el teorema de Bayes y, por otra parte, un modelo que utiliza el algoritmo C4.5, se presentan a continuación las evaluaciones realizadas con ambos métodos y se muestran los resultados obtenidos. En este capítulo se recogen los siguientes casos de uso: estimar las probabilidades de todas las preguntas escogiendo secuencialmente la que presente mayor probabilidad, escoger secuencialmente la que presenta menor probabilidad, se han calculado las probabilidades individuales de cada opción posible y se presenta la información implícita en el conjunto de datos aportada por el árbol de decisión.

5.1. Evaluación de ambas técnicas: mayor probabilidad

Tanto el estimador basado en Bayes como el árbol de decisión toman la misma entrada: una secuencia de las preguntas seleccionadas por el usuario, pudiendo haber alguna pregunta que aún no se haya contestado. Asimismo, ambos generan una salida numérica entre 0 y 1, que indica la probabilidad de conseguir un resultado aceptable.

En la primera evaluación realizada con los dos estimadores se ha simulado el caso de uso en el cual el usuario escoge respuestas de forma secuencial y selecciona las opciones que le aporten mayor probabilidad. Se han calculado las probabilidades que generan las opciones de cada pregunta del cuestionario. Como se quiere simular el caso de uso en el que el usuario escoge de forma secuencial cada pregunta y selecciona la mayor probabilidad, en todas las preguntas se han añadido las preguntas anteriores con las respuestas de mayor probabilidad. Por ejemplo, en la cuarta pregunta (M1) se ha añadido a la entrada la opción de las preguntas anteriores con más probabilidad: $H1='b'$, $H2='b'$, $H3='d'$. Es decir, en este caso de uso el usuario iría escogiendo cada pregunta en el orden en el que se le ofrece.

El resultado obtenido tras realizar el cómputo de las probabilidades con el estimador bayesiano se presenta en la Tabla 5.2. Por otra parte, en la Tabla 5.3 se presenta la salida generada por el estimador basado en árboles de decisión. En ambas tablas se muestra la misma estructura: por columnas se encuentran las preguntas del cuestionario, en total son

12, de las cuales las 3 primeras son hipótesis (H_i) y las siguientes son 9 medidas políticas (M_i). Por filas se encuentran las diferentes respuestas que puede haber, de la ‘a’ a la ‘e’. En cada celda se muestra la probabilidad de lograr un resultado bueno seleccionado esa opción en esa pregunta, incluyendo en la entrada también las preguntas anteriores con mayor probabilidad. Las celdas sombreadas en gris significan que esa pregunta no tiene esa respuesta.

Preguntas												
	H1	H2	H3	M1	M2	M3	M4	M5	M6	M7	M8	M9
a	0.00402	0.00198	0.00093	0.0	0.02984	0.20988	0.15741	0.25	0.33333	0.33333	0.33333	1.0
b	0.00405	0.00613	0.0	0.12088	0.21193	0.20988	0.22222	0.30556	0.27778	0.33333	0.33333	0.0
c	0.00307	0.00403	0.0	0.0		0.21605	0.26852	0.25		0.33333		0.0
d			0.02418	0.0								
e			0.00556	0.0								

Tabla 5.2: Probabilidades generadas por el Teorema de Bayes. Caso de uso seleccionar secuencialmente la mayor probabilidad

Preguntas												
	H1	H2	H3	M1	M2	M3	M4	M5	M6	M7	M8	M9
a	0.00396	0.00192	0.00093	0.0	0.03326	0.21708	0.15123	0.2644	0.3251	0.33333	0.33333	1.0
b	0.00406	0.00627	0.0	0.12517	0.21708	0.21708	0.21091	0.30453	0.28395	0.30864	0.33333	0.0
c	0.00312	0.00397	0.0	0.0004		0.21708	0.28909	0.29835		0.33333		0.0
d			0.02511	0.0								
e			0.00532	0.0								

Tabla 5.3: Probabilidades generadas por el modelo con el algoritmo C4.5 de Weka. Caso de uso seleccionar secuencialmente la mayor probabilidad

Las probabilidades generadas por el estimador basado en el algoritmo C4.5 son las que se han utilizado para animar el elemento de juego (Figura 4.8). Por tanto, las cifras mostradas en la Tabla 5.3 son las que se muestran en el indicador del juego cuando el usuario escoge las respuestas. Estas cifras se van modificando en el indicador gráfico dinámicamente cada vez que el usuario seleccione una opción diferente. Es decir, cuando el usuario selecciona una nueva opción o cambia alguna de las que ya había seleccionado previamente, la estimación se recalcula y se actualiza el indicador gráfico con la nueva probabilidad. En la Figura 5.11 se muestra una captura de pantalla del desarrollo del juego Crossroads móvil. En esta partida se ha replicado el caso de uso en el que el jugador escoge la opción cuya probabilidad es mayor. Concretamente, en la captura de pantalla se muestra la medida política 5 (M5) con sus 3 opciones posibles. En esta pregunta se ha escogido la última opción (‘c’) ya que es la que mayor probabilidad aporta. Esta probabilidad se puede visualizar con el indicador de la parte inferior de la pantalla, donde se muestra que la probabilidad de obtener un buen resultado es 0.29 (29%), correspondiendo con la cifra que se muestra en la columna M5 y fila ‘c’ de la Tabla 5.3: 0.29835. Para mostrar la

probabilidad en el indicador gráfico, se ha redondeado a dos decimales el resultado que devuelve el estimador y se ha representado en forma de porcentaje para el usuario.



Figura 5.11: Captura de pantalla de las preguntas con el indicador

Siguiendo con el caso de prueba, la secuencia obtenida a partir de ambas tablas con mayor probabilidad es: H1: b, H2: b, H3: d, M1: b, M2: b, M3: c, M4: c, M5: b, M6: a. A partir de la medida política 6, las últimas preguntas tienen la misma probabilidad en sus opciones. Por tanto, para realizar esta prueba, en las medidas políticas restantes se ha escogido la opción que se ha considerado más adecuada dependiendo del tema tratado

en cada pregunta. Esta secuencia se ha probado en la aplicación móvil para ilustrar la evolución climática y económica que supone esta toma de decisiones. Las gráficas obtenidas se presentan en la Figura 5.12. Estas gráficas se muestran al finalizar una ronda en el juego Crossroads. En ellas también se añaden los objetivos que ha seleccionado el usuario en las dos preguntas del cuestionario dedicadas a ello. En este caso, se ha elegido como objetivo de temperatura 1.5°C , mientras que para el PIB se ha establecido el objetivo de 6500 USD. Estos objetivos tienen la finalidad de que el usuario pueda comparar la evolución que ha obtenido con el objetivo que se ha propuesto, pero no influyen en la estimación realizada con el árbol de decisión.

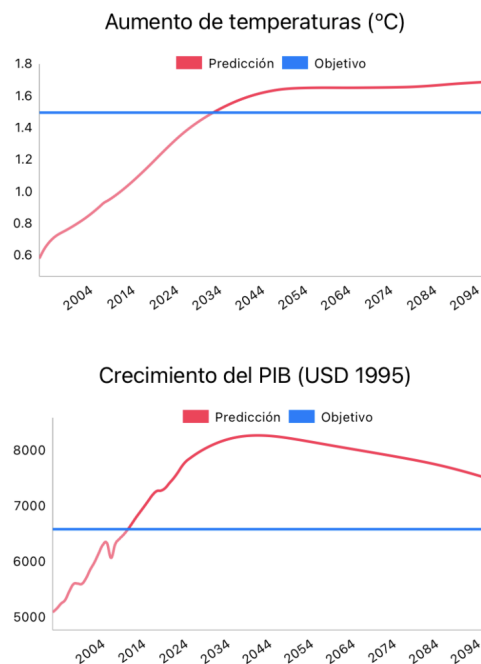


Figura 5.12: Captura de pantalla de los resultados tras seleccionar la secuencia con mayor probabilidad

5.2. Caso de prueba de menor probabilidad

A continuación, se ha realizado el caso de uso de escoger la secuencia menos probable con el modelo de clasificación creado. En este caso de uso se selecciona secuencialmente la opción menos probable de cada pregunta.

Las probabilidades obtenidas por el modelo basado en el algoritmo C4.5 se presentan en la Tabla 5.4, donde por columnas se presentan las diferentes preguntas, en total 12, y en las filas, las diferentes opciones con las que cuenta cada pregunta. Las celdas muestran la probabilidad obtenida para esa opción (fila) en esa pregunta (columna). Las celdas coloreadas en gris indica que no existe dicha opción para la pregunta. Como se ha comentado anteriormente, en este caso de uso se ha realizado la selección de preguntas de

manera secuencial, por ejemplo en la pregunta M1 se han seleccionado las opciones menor probables de las preguntas anteriores (H1, H2 y H3). Es decir, la selección de opciones se ha ido acumulando a lo largo de las 12 preguntas.

	Preguntas											
	H1	H2	H3	M1	M2	M3	M4	M5	M6	M7	M8	M9
a	0.00396	0.00178	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
b	0.00406	0.00514	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
c	0.00312	0.00243	0.0	0.0		0.0	0.0	0.0		0.0		0.0
d			0.00886	0.0								
e			0.00003	0.0								

Tabla 5.4: Probabilidades individuales generadas por el modelo con el algoritmo C4.5. Caso de uso: seleccionar secuencialmente la menor probabilidad

Como se muestra en la Tabla 5.4, a partir de la pregunta M1 o medida política 1, la probabilidad obtenida es 0, es decir, que en ningún caso se podrá obtener un resultado final considerado aceptable. La secuencia de menor probabilidad obtenida con el modelo creado ha sido la siguiente: H1: c, H2: a, H3: a y en las demás preguntas se ha escogido una opción aleatoria ya que aportan todas la misma probabilidad, 0. Aplicando esta secuencia obtenida en el formulario de la aplicación móvil Crossroads, se obtienen la siguiente evolución de temperatura y PIB presentadas en la Figura 5.13 que se obtiene al finalizar el formulario.

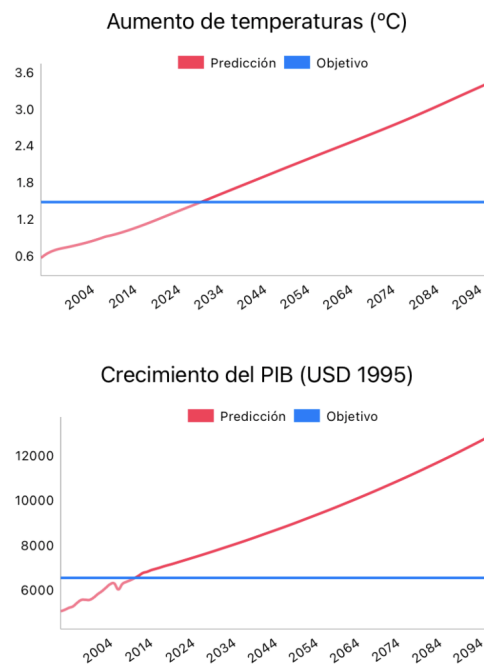


Figura 5.13: Captura de pantalla de los resultados tras seleccionar la secuencia con menor probabilidad

5.3. Probabilidades individuales

La siguiente prueba que se ha realizado ha sido el cómputo de cada opción independientemente. Se ha calculado la probabilidad que tiene cada opción de cada pregunta de forma individual, sin seleccionar más de una pregunta a la vez. En la Tabla 5.5 se presentan las probabilidades individuales de cada posible respuesta obtenidas con el teorema de Bayes. En la Tabla 5.6 se han almacenado las probabilidades individuales computadas con el modelo de clasificación creado. En ambas tablas cada columna contiene una de las 12 preguntas del cuestionario y en las filas se muestran las posibles respuestas. En cada celda se muestra la salida generada por el estimador si se selecciona esa pregunta (columna) con esa respuesta (fila). Las celdas grises representan que no existe en el juego ese índice de respuesta en esa pregunta. Para conocer con profundidad a qué hace referencia cada pregunta y qué respuestas posee, en el Apéndice D se ha recogido el texto asociado a cada pregunta y respuestas que se muestran en el juego Crossroads.

Preguntas												
	H1	H2	H3	M1	M2	M3	M4	M5	M6	M7	M8	M9
a	0.00402	0.00184	0.00079	0.0	0.00042	0.00369	0.00206	0.00293	0.00285	0.0006	0.00354	0.00749
b	0.00405	0.0058	0.0	0.00966	0.00701	0.0037	0.00316	0.00401	0.00457	0.00163	0.00388	0.00316
c	0.00307	0.0035	0.0	0.00583		0.00374	0.00591	0.00419		0.00891		0.00049
d			0.01484	0.0								
e			0.00293	0.00306								

Tabla 5.5: Probabilidades individuales generadas por el Teorema de Bayes

Preguntas												
	H1	H2	H3	M1	M2	M3	M4	M5	M6	M7	M8	M9
a	0.00396	0.00184	0.00079	0.0	0.00042	0.00371	0.00208	0.003	0.00287	0.0006	0.00361	0.00748
b	0.00406	0.00582	0.0	0.00966	0.00701	0.00371	0.00313	0.00398	0.00455	0.00163	0.00381	0.00316
c	0.00312	0.00347	0.0	0.00583		0.00371	0.00592	0.00415		0.00891		0.00049
d			0.01484	0.0								
e			0.00293	0.00306								

Tabla 5.6: Probabilidades individuales generadas por el modelo con el algoritmo C4.5

En ambas tablas, si nos fijamos en las opciones de cada pregunta con más probabilidad, la secuencia obtenida es la siguiente: H1: b, H2: b, H3: d, M1: b, M2: b, M3: c, M4: c, M5: c, M6: b, M7: c, M8: b, M9: a. En la Figura 5.14 se muestra las gráficas obtenidas en la aplicación móvil Crossroads tras ejecutar dicha secuencia. Aunque en la medida política 3 (M3) del modelo clasificador las probabilidades eran iguales para todas las opciones, se ha escogido la opción con mayor probabilidad de la tabla del teorema de Bayes. La secuencia obtenida está clasificada como una secuencia no aceptable, ya que se consigue un resultado en cuanto a PIB bastante elevado.

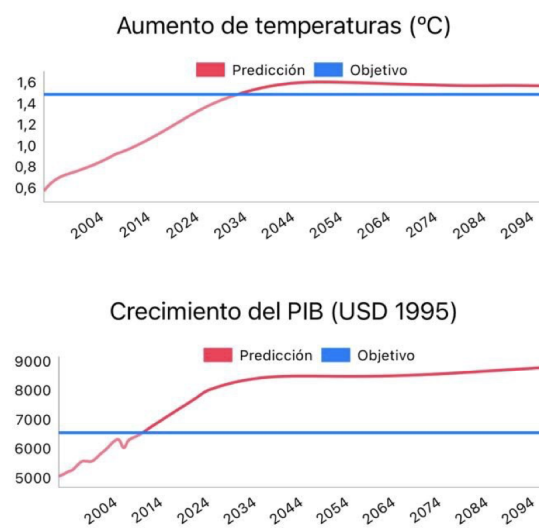


Figura 5.14: Captura de pantalla de los resultados tras seleccionar la secuencia con mayor probabilidad en cada pregunta

5.4. Información aportada por el árbol de decisión

Los árboles de decisión tienen una ventaja sobre la estimación bayesiana que tiene que ver con la posibilidad de ofrecer información sobre el problema que se está utilizando. Por un lado, puede utilizarse representaciones arbóreas que dividen las ramas en función de la importancia de las variables de análisis. Por otro lado, las divisiones se basan en métricas de información mutua que aportan también información interpretable sobre la importancia de cada variable. En el Fragmento de código 5.1 se presenta la estructura del árbol de decisión generado con Weka. En este árbol se han reducido el número de hojas para que pudiera ser más abordable su interpretación. El árbol utilizado en el estimador final posee un mayor número de hojas por nodo. Éste en concreto posee 24 hojas y un tamaño total de 34, por lo que se encuentra formado por 10 nodos de decisión.

```

1  J48 pruned tree
2  -----
3
4  m2 = a: no-acceptable (218700.0/91.0)
5  m2 = b
6  |   h3 = a: no-acceptable (43740.0/69.0)
7  |   h3 = b: no-acceptable (43740.0)
8  |   h3 = c: no-acceptable (43740.0)
9  |   h3 = d
10 |   |   m1 = a: no-acceptable (8748.0)
11 |   |   m1 = b
12 |   |   |   h2 = a: no-acceptable (2916.0)

```

```

13 | | | | h2 = b
14 | | | | | m9 = a
15 | | | | | m7 = a: no-acceptable (324.0/57.0)
16 | | | | | m7 = b: no-acceptable (324.0/110.0)
17 | | | | | m7 = c: acceptable (324.0/158.0)
18 | | | | | m9 = b
19 | | | | | m7 = a: no-acceptable (324.0)
20 | | | | | m7 = b: no-acceptable (324.0/42.0)
21 | | | | | m7 = c: acceptable (324.0/104.0)
22 | | | | | m9 = c: no-acceptable (972.0)
23 | | | | h2 = c: no-acceptable (2916.0)
24 | | | m1 = c: no-acceptable (8748.0/347.0)
25 | | | m1 = d: no-acceptable (8748.0)
26 | | | m1 = e
27 | | | | h2 = a
28 | | | | | m9 = a
29 | | | | | m7 = a: no-acceptable (324.0)
30 | | | | | m7 = b: no-acceptable (324.0/36.0)
31 | | | | | m7 = c: acceptable (324.0/95.0)
32 | | | | | m9 = b: no-acceptable (972.0)
33 | | | | | m9 = c: no-acceptable (972.0)
34 | | | | h2 = b: no-acceptable (2916.0)
35 | | | | h2 = c: no-acceptable (2916.0)
36 | | h3 = e: no-acceptable (43740.0/256.0)
37
38 Number of Leaves : 24
39
40 Size of the tree : 34

```

Fragmento de código 5.1. Árbol de decisión simple

Para continuar estudiando la información que contiene el árbol y aquellas preguntas consideradas más determinantes, se ha evaluado la ganancia de información que almacenan las *features* del modelo, es decir, las preguntas. Para ello, Weka proporciona la utilidad ***InfoGainAttributeEval*** [60], que evalúa el valor de un atributo midiendo la ganancia de información con respecto a la clase, es decir, mide cómo cada atributo contribuye a disminuir la entropía general. Un atributo que aporta un buen nivel de información para la clase que se desea predecir, reduce al máximo la entropía. En el Fragmento de código 5.2 se presenta los valores obtenidos de ganancia de información de las *features* del modelo.

```

1 | Ranked attributes:
2 | 0.005404172 h3
3 | 0.003267021 m1
4 | 0.002593283 m7
5 | 0.002562067 m2
6 | 0.001814089 m9

```

```

7 | 0.000524542 h2
8 | 0.000496102 m4
9 | 0.000146286 m6
10 | 0.00006304 m5
11 | 0.000041914 h1
12 | 0.000005739 m8
13 | 0.000000116 m3

```

Fragmento de código 5.2. Evaluación de *Info Gain*

La que presenta mayor ganancia de información es la hipótesis 3 (h3) la cual trata sobre la evolución de la población mundial, seguida de la medida política 1 (m1) que es referida a la evolución del Producto Interior Bruto real por persona. Al contrario, la que menor ganancia de información aporta es la medida política 3 o m3 y esta pregunta recoge la información de la evolución sobre la potencia de energía nuclear instalada a nivel mundial.

Por otra parte, la ganancia de información [55] tiene una característica no deseada y es que favorece a aquellas variables que cuentan con un gran número de valores y estas variables altamente ramificadas dividan los datos en subconjuntos con valores de entropía bajos. Esto se soluciona al utilizar el ratio de ganancia. El ratio de ganancia se calcula dividiendo la ganancia de información entre la información intrínseca, definida como la entropía de las proporciones del subconjunto de datos. Es decir, este ratio intenta disminuir el sesgo que introduce la ganancia de información en aquellas variables se encuentran más ramificadas.

En el Fragmento de código 5.3 se presenta el ratio de ganancia obtenido para cada atributo del conjunto de datos de entrada. Para ello, se ha utilizado el evaluador que proporciona Weka *GainRatioAttributeEval* [59], el cual se suele emplear en tareas de selección de *features*. En la salida que se presenta en el Fragmento de código 5.3, en cada fila se encuentra el valor del ratio de ganancia obtenido y el atributo al que hace referencia, estos atributos son las diferentes preguntas del juego.

```

1 | Ranked attributes:
2 | 0.0025620665 m2
3 | 0.0023274504 h3
4 | 0.0016361794 m7
5 | 0.0014070293 m1
6 | 0.0011445626 m9
7 | 0.0003309491 h2
8 | 0.0003130055 m4
9 | 0.0001462864 m6
10 | 0.000039774 m5
11 | 0.0000264447 h1
12 | 0.000005739 m8
13 | 0.0000000733 m3

```

Fragmento de código 5.3. Evaluación de *Gain Ratio*

La pregunta más significativa con la que se ha obtenido un mayor ratio de ganancia es la medida política 2 (m2) la cual cuestiona la importancia de llevar a cabo un programa mundial de forestación y reforestación de bosques. Se encuentra seguida de la hipótesis 3 (h3) que trata sobre la evolución de los niveles de población mundial. En último lugar, con el menor ratio de ganancia obtenido y por tanto, la pregunta que aporta menor información al modelo creado, se encuentra la medida política 3 (m3) en la cual se pregunta sobre la evolución de energía nuclear a nivel mundial.

5.5. Adaptación de la probabilidad del estimador

En el caso de uso en el cual el usuario selecciona las respuestas secuencialmente, como se muestra en la Tabla 5.3, las probabilidades no aumentan de una pregunta a otra paulatinamente. Por ejemplo, en la medida política 8 o M8 la probabilidad de cualquiera de las dos opciones con las que cuenta es 0.33, mientras que en la siguiente pregunta (M9) en la opción a, la probabilidad aumenta hasta llegar a 1. Atiendo a las probabilidades desde la pregunta H1 hasta la penúltima pregunta M8 (de 0.00406 en H1 a 0.33 en M8), el aumento que se produce es menor que el que se produce entre las dos últimas preguntas (de 0.33 en M8 a 1.0 en M9). En la primera pregunta, la hipótesis 1, con opción b se cuenta con una probabilidad de 0.00406, mientras que al llegar a la medida política 8 en cualquiera de las opciones la probabilidad ha ascendido hasta 0.33. El repentino aumento sucede en la última pregunta, medida política 9, donde se alcanza una probabilidad 1.0 al escoger la opción 'a'.

El hecho de que se produzca un aumento mucho mayor en la probabilidad de estas últimas opciones comparándolo con el aumento que se produce durante las primeras preguntas, tiene como consecuencia que, al añadir el estimador en la aplicación móvil con el elemento gráfico, el indicador aumente de una manera brusca al pasar de la penúltima pregunta con una probabilidad del 33 % a la última que muestra 100 %. Por este motivo, se quiere realizar un procesamiento a la probabilidad que genera el estimador, para que se redistribuir las probabilidades y conseguir que no haya un aumento tan brusco entre las últimas preguntas. Para ello, se ha empleado una función típica utilizada para añadir contraste a las imágenes [50]. La función utilizada es:

$$f(x) = x^\gamma; \quad \text{siendo } \gamma = 0.3$$

El valor de γ se ha estimado de forma empírica y la función resultante es la presentada en la Figura 5.15.

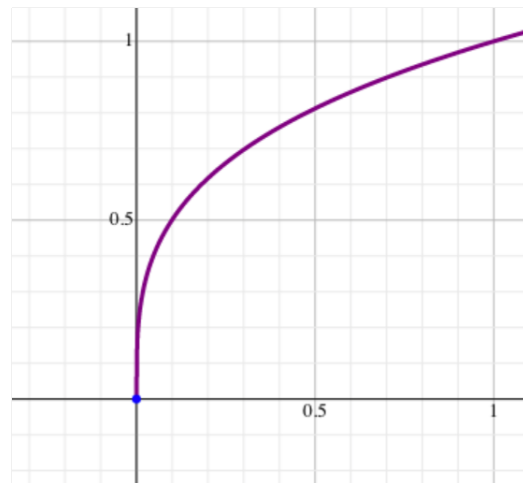


Figura 5.15: Gráfica de la función $f(x) = \gamma^{0.3}$

Tras aplicar la función a la probabilidad que genera el modelo basado en el algoritmo C4.5, se ha vuelto a probar el caso de uso de escoger secuencialmente la opción con mayor probabilidad en cada pregunta como en la anterior Sección 5.1. La Tabla 5.7 presenta las probabilidades obtenidas tras aplicar esta nueva función. La secuencia escogida es la misma que la que se presenta en la Sección 5.1: H1: b, H2: b, H3: d, M1: b, M2: b, M3: c, M4: c, M5: b, M6: a. A partir de la medida política 6, en las medidas políticas restantes se ha escogido la opción que se ha considerado más adecuada dependiendo del tema tratado en cada pregunta.

	Preguntas											
	H1	H2	H3	M1	M2	M3	M4	M5	M6	M7	M8	M9
a	0.19024	0.15317	0.12302	0.0	0.36024	0.63239	0.56741	0.67093	0.71385	0.71922	0.71922	1.0
b	0.19161	0.21838	0.0	0.53611	0.63239	0.63239	0.62694	0.69998	0.68545	0.70281	0.71922	0.0
c	0.17705	0.19044	0.0	0.09564		0.63239	0.68915	0.6957		0.71922		0.0
d			0.33111	0.0								
e			0.20781	0.0								

Tabla 5.7: Probabilidades generadas por el modelo con el algoritmo C4.5 de Weka con función que procesa la salida. Caso de uso seleccionar secuencialmente la mayor probabilidad

6: Discusión

En este capítulo se interpretan los resultados numéricos y las observaciones reportadas en el capítulo anterior.

6.1. Evaluación de ambas técnicas

En la Sección 5.1 se han presentado las probabilidades obtenidas por ambos estimadores, bayesiano y algoritmo C4.5, en un caso de uso concreto. Atendiendo a los resultados numéricos obtenidos que se han recopilado en las Tablas 5.2 y 5.3, las probabilidades mostradas para cada pregunta son bastante similares y se encuentran escasas diferencias entre las probabilidades presentadas en ambas tablas. También se puede apreciar como la probabilidad de obtener un resultado bueno aumenta a medida que se van seleccionando las preguntas con mayor probabilidad.

Si bien es cierto que en la pregunta medida política 3 (M3) el modelo basado en el algoritmo C4.5 asocia la misma probabilidad a todas las opciones (Tabla 5.3 columna M3), mientras que el teorema de Bayes asocia la mayor probabilidad a la última opción (Tabla 5.2 columna M3 fila 'c'). Esto está relacionado con la evaluación de este atributo (M3) y su importancia en el modelo, porque esta pregunta es la que menor ratio de ganancia tiene del conjunto de datos (Fragmento de código 5.3 última línea). Por tanto, al ser menos informativa para el modelo, puede ocurrir que la decisión que tome el usuario en esa pregunta no influya en la clasificación final que haga el árbol de decisión, por lo que asocia la misma probabilidad a cada una de las opciones de esta pregunta M3. Cabe destacar que en las demás preguntas, tanto el estimador basado en el teorema de Bayes como el basado en el árbol de decisión, puntúan con la mayor probabilidad a las mismas opciones. Por ejemplo, en la columna H1, que corresponde a la hipótesis 1, tanto en el teorema de Bayes (Tabla 5.2) como en el clasificador creado (Tabla 5.3), la fila 'b' presenta la mayor probabilidad en ambas tablas. Entonces, comparando ambos estimadores se puede observar que el resultado del algoritmo C4.5 escogido para el desarrollo del modelo es bastante similar al resultado obtenido con el estimador bayesiano, por lo que se puede deducir que el resultado generado con el modelo es correcto.

En la Figura 5.12 se muestra la evolución climática y económica tras haber seleccionado la secuencia con mayor probabilidad obtenida. Se ha conseguido una estabilidad de la temperatura a lo largo de los años, estableciéndose un aumento menor de 2°C dentro de 100 años. En cuanto al PIB, hay crecimiento hasta el año 2034 aproximadamente, a partir del cual el PIB disminuye ligeramente hasta llegar a un valor aceptable de 7500 USD aproximadamente. Entonces, el estudio en el que el usuario se vaya guiando por la mayor probabilidad en cada respuesta tiene como resultado un escenario evolutivo sostenible en cuanto a economía y temperatura. Por tanto, se puede determinar que el indicador gráfico que se ha añadido al juego y que utiliza el estimador basado en el algoritmo C4.5 funciona correctamente y conduce al usuario hacia buenos escenarios evolutivos responsables con el incremento de la temperatura y conteniendo la subida del PIB.

6.2. Caso de prueba de menor probabilidad

Continuando con los casos de prueba realizados, en la Sección 5.2 se presenta el resultado obtenido tras escoger la menor probabilidad con el modelo basado en el algoritmo C4.5. Como se puede apreciar en las probabilidades obtenidas en la Tabla 5.4, se obtiene a partir de la hipótesis 3 una probabilidad nula de obtener un resultado aceptable. Para ello en la hipótesis 1 (H1) se ha seleccionado la respuesta 'c' (Tabla 5.4 columna H1 y fila 'c') que es la que menor probabilidad presenta con 0.00312. En concreto, esta pregunta trata sobre la disponibilidad de recursos energéticos no renovable que el usuario cree que queda hasta 2070 y la respuesta con menor probabilidad indica que una disponibilidad de recursos ilimitada. Continuando con la hipótesis 2, que trata del impacto que estima el usuario que producirá el cambio climático derivado de la actividad humana, se ha escogido la respuesta 'a' que presenta la menor probabilidad con 0.0178 (Tabla 5.4 columna H2 y fila 'a') y corresponde a un impacto bajo. Entonces, tras estos resultados, si el usuario supone una cantidad ilimitada de recursos no renovables y que va a haber un impacto mínimo a consecuencia del cambio climático, las posibilidades de obtener un resultado aceptable se ven reducidas enormemente.

Continuando con este ejemplo, en la Figura 5.13 se expone la gráfica tanto de incremento de temperatura como de crecimiento de PIB obtenido después de haber aplicado la secuencia que conlleva a un resultado no aceptable. Atendiendo a la temperatura, se obtiene un incremento de 3.6°C , lo que supone un aumento enorme incompatible con la vida. Por otro lado, el PIB también aumenta linealmente. Aunque el aumento del PIB no es perjudicial para la vida, el aumento de la temperatura es preocupante, ya que un incremento de 3° de temperatura es un valor demasiado elevado y presenta un aumento lineal a lo largo de los años, por lo que no se consigue una estabilidad final de la temperatura. Es un resultado preocupante y se encuentra clasificado como no aceptable.

6.3. Información aportada por el árbol de decisión

En la Sección 5.4 se ha presentado un árbol de decisión simplificado. Además, se han mostrado los resultados obtenidos tras aplicar a los atributos el evaluador de ganancia de información y ratio de ganancia que proporciona Weka.

Atendiendo al árbol de decisión (Fragmento de código 5.1 líneas 4 y 5) el primer nodo de decisión que se encuentra en él es la medida política 2 (m2). Esta medida política trata llevar a cabo o no un programa mundial de forestación y reforestación de bosques. Cuenta con dos respuestas: la opción 'a' recoge que no se lleve a cabo un programa de forestación y reforestación ya que implicaría una gran reducción de las superficies cultivables que son necesarias para la producción de alimentos y energía; y la opción 'b' que expone que sí se lleven a cabo programas de ese tipo para compensar las emisiones de CO2 y ayudar a frenar el aumento de la temperatura media del planeta. En el árbol simplificado, todas las opciones que contienen la primera opción de no implantar un programa de forestación, están consideradas como resultado no aceptable. Por tanto, la selección de un programa de forestación y reforestación es una respuesta crucial para poder obtener un buen resultado durante el juego.

Una vez seleccionada la opción 'b' en la medida política 2, el siguiente nodo de decisión que se encuentra es la hipótesis 3 (h3) referida a los niveles de evolución de población mundial. La única opción de este árbol simplificado que alberga la posibilidad de conseguir resultados aceptables a partir de esta pregunta es la opción 'd', que corresponde a establecer un límite poblacional de 5.000 millones de personas para 2050. Por tanto, tras este conocimiento obtenido a partir del árbol de decisión, destaca que para obtener un resultado sostenible se deba reducir la población alrededor de 3.000 millones por debajo de las cifras actuales [62].

La importancia denotada por la medida política 2 y la hipótesis 3 para el modelo se puede comprobar también en el Fragmento de código 5.3 (líneas 2 y 3), que son las dos preguntas con mayor ratio de ganancia, 0.0025 para la medida política 2 y 0.0023 para la hipótesis 3. Esto es lógico ya que el algoritmo C4.5 que utiliza el árbol de decisión se encarga de dividir los nodos de decisión en base a la evaluación del ratio de ganancia.

Comparando los resultados obtenidos al utilizar la ganancia de información y la ratio de ganancia, el ranking de preguntas de ambos no corresponde. Esto se debe al problema de la métrica ganancia de información, la cual favorece a las variables que cuentan con un mayor número de valores y están más ramificadas, como es el caso de h3 que cuenta con 5 opciones posibles y está clasificada en primer lugar como la que aporta mayor ganancia de información. Sin embargo, evaluando la ratio de ganancia, la hipótesis 3 aparece en segundo lugar y la medida política 2 (que cuenta únicamente con 2 opciones posibles) asciende a la más significativa para el árbol de decisión.

6.4. Estimación en base a los objetivos

En la Sección 5.3 se ha computado la probabilidad de lograr un estado bueno probando cada opción de manera individual. En esta evaluación no se ha incluido ninguna secuencia de respuestas en la entrada de los estimadores, sino que se ha estimado cada opción independientemente. Atendiendo a los resultados mostrados en las Tablas 5.5 y 5.6, las probabilidades generadas por el estimador bayesiano y la salida obtenida del algoritmo C4.5 son bastante similares, en algunos casos iguales como por ejemplo para la opción 'b' de la medida política 1 donde ambos estimadores generan una probabilidad de 0.00966 (columna M1 fila 'b' en Tablas 5.5 y 5.6).

En la medida política 3 (m3), el árbol de decisión estima la misma probabilidad para todas las opciones posibles (Tabla 5.6 columna M3), mientras que el teorema de Bayes estima diferentes probabilidades para las 3 opciones (Tabla 5.5 columna M3). Esto se debe a que, como se verá en detalle más adelante, esta pregunta para el modelo creado tiene poca significación, siendo la que menos información aporta al modelo.

En la Figura 5.14 se muestra el resultado del juego móvil tras haber seleccionado en cada pregunta la respuesta que mayor probabilidad tiene individualmente. Con respecto a la gráfica de la temperatura, se obtiene una estabilidad climática a lo largo de los años, cuyo valor medio se establece en 1.6°C y es un escenario bastante favorable. Sin embargo, el PIB se dispara los primeros años, pero se consigue estabilizar alcanzando finalmente un valor medio de 9000 USD. Por este valor tan elevado de PIB obtenido finalmente, este escenario que se muestra en la figura está clasificado en el *dataset* como estado no aceptable.

Al haber escogido la opción que aporta mayor probabilidad de cada pregunta independientemente, el resultado que se obtiene si se considera sostenible ambientalmente, pero para los filtros aplicados al conjunto de datos de entrenamiento los valores de PIB obtenidos son demasiados elevados. Un elevado incremento del PIB no es un problema para la vida, sin embargo, en el conjunto de datos utilizado para entrenar el modelo se ha establecido que el PIB se mantenga entre valores de 6000 y 7500 USD. Para aceptar también otro tipo de escenarios donde se modifique la definición de estado aceptable y, por ejemplo se pueda ser más tolerante con la subida del PIB y/o más permisivo con subidas de temperatura, se deberían realizar ciertas modificaciones sobre el modelo, comentadas a continuación.

El juego Crossroads incluye 14 preguntas en total. Para la predicción de conseguir un resultado aceptable sólo influyen las medidas políticas y las hipótesis que son 12 preguntas del total. Las dos preguntas restantes corresponden a objetivos que marca el usuario para establecerse límites los cuales pretende alcanzar. Estos objetivos no influyen en la evolución que se muestra del clima y la temperatura, simplemente el valor que se haya propuesto como objetivo se grafica una línea horizontal junto a la evolución de temperatura y de PIB, para poder conocer cómo de cerca o lejos se encuentra la evolución que ha conseguido con respecto al objetivo que ha marcado el usuario.

El indicador gráfico creado, en vez de estimar la probabilidad de obtener un resultado final sostenible, podría estimar el acercamiento hacia los objetivos que se ha propuesto el usuario. En ese caso, cuando el usuario se plantee los objetivos de PIB y temperatura las hipótesis y medidas políticas que conteste determinarán la probabilidad de conseguir esos objetivos propuestos. Esta probabilidad calculada en base a los objetivos guiará al usuario para conseguirlos, independientemente de si esos objetivos albergan una situación climática y económica sostenible. Por ejemplo, como objetivo el usuario se puede proponer un aumento de temperatura por debajo de 3°C , lo que significa que la temperatura media del planeta podría ascender hasta 3°C por encima del valor actual, lo que supondría un valor demasiado elevado e insostenible. Aún así, es un objetivo posible que el usuario puede seleccionar.

Para llevar este cambio a cabo, habría que modificar el modelo que se ha creado. La clasificación utilizada para entrenar el modelo cambiaría, ya que en este caso la clase que se quiere predecir estaría basada en los objetivos y no en un escenario considerado aceptable. La nueva clasificación estaría determinada por el conjunto de objetivos, en concreto 3 objetivos para la temperatura y 5 para el PIB, el mismo número que posibles opciones hay. Con los datos actuales de los que se ha partido para realizar el modelo basado en el algoritmo C4.5 se podrían reutilizar, ya que contiene los clústers con la agrupación de las diferentes evoluciones que ocasionan las respuestas junto con la temperatura y PIB final que se alcanza. Quizás, la mejor aproximación sea realizar un árbol de decisión con cada combinación de objetivos posibles que puede marcar el usuario. De esta manera, dependiendo de los objetivos que marque el usuario se utilizará un árbol de decisión u otro para que se pueda estimar la probabilidad de acercarse a los objetivos propuestos.

6.5. Adaptación de la probabilidad del estimador

En la Sección 5.5 se muestra el resultado obtenido tras haber aplicado una función que redistribuye la probabilidad generada por el modelo de predicción creado. En esa sección se ha utilizado la misma secuencia de opciones que la presentada en la Sección 5.1 con el fin de comparar ambos resultados obtenidos y conocer si el aumento de la probabilidad a lo largo de las preguntas es más paulatino.

En la Tabla 5.3 se muestran las probabilidades obtenidas con el modelo basado en el algoritmo C4.5, seleccionando la secuencia que aporta mayor probabilidad. Por otra parte, en la Tabla 5.7 se recogen las probabilidades obtenidas del modelo basado en el algoritmo C4.5 tras aplicar la función matemática a la salida, utilizando la misma secuencia de mayor probabilidad. Comparando ambos resultados, se puede comprobar cómo las probabilidades obtenidas en las primeras preguntas son más significativas cuando se aplica la función que sin ella. Por ejemplo, en la columna H1 y la fila 'b' de la Tabla 5.7 se muestra la probabilidad de la opción 'b' en la pregunta H1 tras haber aplicado la función potencial y es 0.19161, mientras que en esa misma columna y fila de la Tabla 5.3, es decir, sin aplicar la función matemática, se obtiene una probabilidad de 0.00406. En este caso, se ha conseguido solucionar la problemática de que el usuario comience a jugar y en la primera

pregunta obtenga probabilidades casi imperceptibles de 0.00406. En su lugar obtiene una probabilidad de 0.19161 que en el elemento gráfico se mostrará como 19 % y es un valor más motivador para el jugador.

En la Tabla 5.3 en las dos últimas preguntas (columnas M8 y M9 fila a) se obtiene una probabilidad de 0.33 para la penúltima pregunta y se convierte en una probabilidad de 1.0 en la última pregunta. Hay un aumento del 0.67 en la probabilidad entre esas dos últimas preguntas, un aumento bastante drástico si se compara con la probabilidad obtenida hasta la penúltima pregunta M8 que es 0.33. Aplicando la función potencial a la salida del modelo clasificador, en la penúltima pregunta M8 se obtiene una probabilidad de 0.72 y en la última, una probabilidad de 1 (Tabla 5.7 columnas M8 y M9 fila a). Por tanto, se ha logrado por una parte que tras haber contestado a la mayoría de las preguntas el jugador obtenga un 0.72 de probabilidad frente a un 0.33 que obtenía sin aplicar la función y además se ha disminuido el aumento de probabilidad que tenía lugar entre la penúltima y última preguntas. De esta forma, en el elemento gráfico en la aplicación móvil se consigue que el incremento de la probabilidad que se muestra entre preguntas es mucho más sutil, con ligeros cambios entre preguntas y no existen drásticos aumentos sobre todo al final de las preguntas.

6.6. Utilidad real para los usuarios

La aplicación Crossroads está enfocada al alumnado para que conozcan los principales conceptos sobre la sostenibilidad. Guiarles hacia un estado final sostenible se puede lograr con el estimador de probabilidad desarrollado. Sin embargo, se puede poner en duda la utilidad del sistema desarrollado, ya que al final sólo aporta un dato numérico al usuario. No obstante, existen unas ayudas particulares de cada pregunta que favorecen que el usuario comprenda mejor los conceptos que se quieren impartir en cada pregunta y podría completar la explicación del dato numérico que aparece en el estimador creado, pudiendo añadir cierta información sobre porqué algunas opciones aumentan esa probabilidad o, por el contrario, la disminuyen. Para revelar si el elemento gráfico desarrollado es útil o no para los usuarios, se podría llevar a cabo en un futuro una prueba con usuarios reales donde se pregunte sobre cómo les ha ayudado el elemento para obtener un resultado sostenible y si les ha servido para reconducir su partida hacia un mejor resultado.

En la aplicación web, Crossroads es multijugador para que los alumnos de una misma clase puedan dividirse en grupos y competir entre ellos. Para ello, un moderador debe crear la partida y los grupos. Este moderador suele ser el profesor y puede comprobar el estado de cada grupo en todo momento. Otra posible utilidad de mostrar la probabilidad de obtener un resultado sostenible podría ser en la aplicación web. En ese caso, al moderador se le mostraría la probabilidad de cada grupo y esto le permitirá ver de una forma sencilla e intuitiva el estado de la toma de decisiones de cada grupo y si éstas conllevan a un estado final aceptable. De esta forma, los alumnos que se encuentren jugando a Crossroads web no verían el indicador gráfico y éste solo estaría disponible para el profesor/moderador

que le permite obtener una visión general del desarrollo de las partidas en los diferentes grupos.

Para implementar esta posible solución no supondría una gran carga de trabajo en un principio. En este proyecto se ha desarrollado una operación en el *backend* de Crossroads a la que se le envía por parámetro las respuestas y devuelve la estimación de obtener un estado bueno, computando esta probabilidad a partir del modelo basado en el algoritmo C4.5 creado. Crossroads posee un único *backend* que se utiliza tanto para la aplicación web como para las aplicaciones móviles. Por tanto, para añadir esta operación a la aplicación de Crossroads web haría falta crear el indicador gráfico (Figura 4.8) con Angular y conectarlo con la operación del *backend* ya creada. Una vez creado el elemento gráfico y conectado con el *backend*, se computaría la probabilidad de obtener un resultado aceptable cada vez que un grupo conteste una nueva respuesta y, de esta forma, el moderador tendría una visión general del desarrollo de cada uno de sus grupos dentro del juego.

7: Conclusiones y Líneas de trabajo futuras

El cambio climático es un gran desafío actual y en concreto la transición hacia una sociedad con bajas emisiones de carbono. En este contexto se encuentra el trabajo llevado a cabo, que cuenta y se establece sobre varios trabajos consolidados previamente que se encuadran dentro de la misma temática y el mismo proyecto europeo. Este trabajo se encuadra dentro del proyecto europeo Locomotion H2020 y se ha desarrollado como una función que se anexiona al juego Crossroads. Este juego expone ciertas situaciones climatológicas y económicas donde el usuario o jugador deberá ir escogiendo aquellas que considere mejores opciones para poder obtener un resultado sostenible. En este trabajo se ha realizado un estimador que calcula la probabilidad de obtener un buen resultado dependiendo de las opciones que vaya seleccionando el jugador a lo largo del desarrollo del juego. Se entiende como un buen resultado aquel en el que la evolución de la temperatura a lo largo de los años no presente un incremento incompatible con la vida y, por otra parte, la evolución de la economía no decrezca bruscamente. Entonces, a partir de las respuestas que marque el usuario se ha conseguido calcular una estimación de la probabilidad de obtener un buen resultado.

Para crear el estimador, en primer lugar se han leído y estudiado aquellos trabajos previos importantes que aportan algún conocimiento para el problema que se pretendía resolver. Una vez preprocesados y transformados los datos de entrada, se han utilizado árboles de decisión para estimar las probabilidades de éxito. La clasificación que predice el árbol de decisión se convierte en una probabilidad a través de utilizar el error de predicción que devuelve. Esta metodología se ha comparado con un método empírico, el teorema de Bayes cuya aplicación ha sido posible porque se dispone de todas las posibles combinaciones de preguntas y respuestas junto con el resultado que genera cada una. Tras obtener los resultados de las dos metodologías utilizadas, las probabilidades obtenidas son muy similares, por lo que el uso de árboles de decisión está correctamente empleado. Algunas de las ventajas que se obtienen al aplicar el uso de árboles de decisión frente al uso de teorema de Bayes es la información que recogen éstos encontrada implícitamente en el conjunto de datos. Al haber aplicado el algoritmo de árboles de decisión se han

podido distinguir las preguntas más significativas para determinar si un conjunto de respuestas es aceptable o no. Además, a diferencia de la aplicación del teorema de Bayes, los árboles de decisión no necesitan acceder al conjunto de datos inicial en cada estimación de probabilidad que se realice, ya que con la información que almacena el modelo creado con los diferentes valores de cada atributo es suficiente.

También se ha creado un indicador gráfico que muestra un dato numérico y se anima dinámicamente cuando este dato se modifica. Este elemento gráfico, creado *ad-hoc*, es el encargado de mostrar la probabilidad que genera el estimador e indicarle al usuario cuán lejos o cerca se encuentra de obtener un resultado aceptable. Para realizar la conexión entre el estimador creado basado en árboles de decisión, concretamente en el algoritmo C4.5, y el elemento gráfico para mostrar la probabilidad en el juego de manera gráfica y amena para el jugador, ha sido necesario crear un API para poder obtener la probabilidad numérica desde la capa *frontend* de la aplicación y mostrarla en el indicador gráfico.

Finalmente, observando los resultados obtenidos se puede comprobar cómo tomando como guía la probabilidad que genera el estimador creado y seleccionando aquellas preguntas que poseen la mayor probabilidad, se obtiene un resultado con una evolución sostenible, es decir, un resultado en el que la evolución de la temperatura se consigue estabilizar a lo largo de los años sin superar un aumento de 2°C y, a su vez, la evolución PIB no experimenta un decrecimiento. Además, a partir de la información almacenada en el árbol de decisión se han conocido más detalles sobre el conjunto de datos inicial, como qué preguntas son las más significantes para obtener un buen resultado y cuáles las que menos. Se ha logrado crear un estimador que conduce al usuario a un buen resultado final y le informa sobre qué opciones le aportan una mayor probabilidad para obtener un buen resultado.

Como aspectos a mejorar de este trabajo y posibles líneas futuras, se encuentra la necesidad de realizar un test de usabilidad con usuarios reales donde prueben la aplicación Crossroads con el estimador de resultado aceptable. Al realizar estos test de usabilidad, que pueden ser unas encuestas a los usuarios, se podrá obtener información sobre si el elemento gráfico y el estimador de probabilidad les ha ayudado a asentar conocimientos sobre la influencia de determinadas políticas en el medioambiente y en la economía, si les ha ayudado a conducir su partida hacia un mejor resultado y si les ha servido para mejorar su puntuación. En cuanto al modelo basado en árboles de decisión, se ha obtenido una precisión bastante elevada, sin embargo, otra posible mejora sería realizar un estudio y una búsqueda con cambios en determinados hiperparámetros del modelo que permitiesen aumentar la precisión del modelo y así conseguir que la probabilidad que se obtenga sea más acertada.

Otra posible línea de trabajo futura sería mejorar el proceso de clustering realizado sobre los datos de los que se parte en este trabajo. El proceso de clustering realizado en un trabajo previo se encarga de almacenar la evolución de temperatura y economía en diferentes grupos teniendo en cuenta el último valor obtenido en la secuencia, es decir, el valor obtenido para el último año de la evolución. Por este motivo, sería más aconsejable refinar este proceso de clústering y conseguir que también se dividan los datos dependiendo

de la evolución que tiene la secuencia a lo largo de los años y no sólo con el último dato. De esta forma, se mejoraría la clasificación de resultados buenos del conjunto de datos y se podrían obtener unos mejores resultados del estimador, obviando aquellas secuencias cuya evolución no es buena pero su valor final sí lo es.

Apéndices

Apéndice A

Plan de Proyecto

En este apéndice se expone el plan de proyecto donde se desarrolla la planificación del mismo junto con la división del trabajo en tareas para cumplir los objetivos presentados en la Sección 1.2. Además, se incluye la metodología y los recursos utilizados.

A.1. Planificación temporal

Este proyecto se comenzó durante la estancia de prácticas de la asignatura de I+D+i del Máster. El objetivo de esta estancia era comenzar y realizar avances en el proyecto que luego se terminarán en el Trabajo Fin de Máster. Por tanto, las tareas realizadas durante la estancia de I+D+i en el grupo ECA-SIMM de la UVa se añadirán a esta planificación debido a que son la base de la que se ha partido en este TFM.

Para el desarrollo de este proyecto se ha utilizado la metodología de investigación en la ingeniería [54]. Este método permite iterar sobre diferentes fases para obtener resultados cada vez más adecuados que resuelvan el problema planteado. En concreto, se ha adaptado la metodología al problema que se expone en este proyecto y se han obtenido las siguientes etapas de división del método:

- Observación de soluciones existentes
- Propuesta y desarrollo de soluciones para el problema concreto
- Medición y análisis de la solución obtenida

A continuación se exponen las tareas acordadas para el desarrollo del proyecto basadas en la metodología de investigación.

Tareas

Para poder lograr los objetivos propuestos en el proyecto, se han identificado una serie de tareas cuyo fin es solventar el problema planteado. En este proyecto las tareas que se han establecido son las mostradas a continuación. Cada tarea tiene asociado un identificador cuyo primer dígito determina la fase en la que se va a realizar. En este proyecto se han identificado 6 fases que se realizarán secuencialmente. Además, se ha añadido a cada tarea el resultado que se obtiene de la misma, para comprender mejor la función de cada una y la salida esperada.

- **Tarea 1.1.** Realizar un estudio sobre la información previa disponible, como trabajos previos relacionados que permitan conocer el problema y crear un estado del arte.
 - Resultado esperado: conocimiento sobre el problema del que se parte y trabajos previos, identificando los factores relevantes a tener en cuenta. Constituirá la base de conocimiento del proyecto.
- **Tarea 1.2.** Realizar un estudio exhaustivo sobre la tecnología que mejor se adapte al problema a resolver.
 - Resultado esperado: estudio que determina cuál es la mejor tecnología o herramienta para desarrollar el trabajo basado en artículos y otros trabajos.
- **Tarea 2.** Realizar una exploración de los datos de entrada, aplicando operaciones de limpieza y transformación si fuera necesario.
 - Resultado esperado: fichero de datos de entrada pre-procesado. Se modificarán los registros que tengan valores erróneos o nulos. Además, se añadirá la clasificación para cada registro indicando si genera un resultado aceptable o no.
- **Tarea 3.1.** Aplicación del teorema de Bayes con los datos de entrada, para que genere la probabilidad de llegar a un estado final aceptable dadas las opciones seleccionadas por el usuario. De esta manera, se obtendrá un *baseline* con el cual se pueda comparar el clasificador realizado con *machine learning*.
 - Resultado esperado: función que devuelve la probabilidad de alcanzar un estado aceptable conociendo las respuestas seleccionadas por el usuario.
- **Tarea 3.2.** Creación de un estimador que genere la probabilidad de que el resultado sea aceptable a partir de las respuestas escogidas.
 - Resultado esperado: modelo que clasifica las opciones escogidas dependiendo de si derivarán a un resultado aceptable, generando la probabilidad de que se obtenga un estado aceptable. El modelo debe trabajar con *missing values* como posibles valores de entrada, es decir, para los estados que se quieran predecir se puede incluir opciones aún no contestadas, por lo que estos serían valores faltantes. Asimismo, las diferentes preguntas, no tienen porque ser contestadas

en orden secuencial ya que en la aplicación móvil real el usuario puede avanzar y retroceder entre las diferentes preguntas sin haber contestado necesariamente las preguntas previas.

- **Tarea 4.1.** Comparación entre el modelo creado con *machine learning* y la aplicación del teorema de Bayes.
 - Resultado esperado: estudio comparativo sobre los resultados que generan ambas técnicas.
- **Tarea 4.2.** Búsqueda del mejor modelo de *machine learning* para poder obtener la mayor precisión posible.
 - Resultado esperado: clasificador final que determina con un alto nivel de precisión si las opciones seleccionadas conducen a un estado aceptable.
- **Tarea 5.** Búsqueda de información relacionada con juegos educativos para conocer qué elemento de juego es más adecuado para representar la probabilidad que devuelve el clasificador, seguido de la realización de pruebas con usuarios reales para poner en práctica el conocimiento obtenido y afinar qué elemento de juego es mejor y más adecuado para la problemática presentada.
 - Resultado esperado: decisión por un elemento de juego, por ejemplo: un círculo que muestre el avance de probabilidad o una ruleta que actúe como indicador que posea una animación cuando se modifique la probabilidad de alcanzar un estado beneficioso.
- **Tarea 6.1** Creación de una API para acceder al modelo creado y añadirla al backend existente de la aplicación Crossroads 2.0.
 - Resultado esperado: API con una operación que contenga el conocimiento del modelo clasificador creado y devuelva la probabilidad numérica de obtener un estado aceptable dadas unas opciones determinadas.
- **Tarea 6.2** Integración del sistema de feedback en la aplicación móvil.
 - Resultado esperado: aplicación móvil con el sistema de feedback que facilita al usuario comprender qué medidas políticas sobre medioambiente y economía le conducen a un estado equilibrado.

Planificación inicial

En la Figura A.1 se expone el diagrama de Gantt con la planificación que ha sido considerada para el desarrollo del proyecto. Como la estancia en prácticas ha sido la antesala para el comienzo de este Trabajo Fin de Máster, se ha incluido la planificación de la estancia en este esquema. El trabajo realizado durante la estancia dual tiene una duración de 190 horas, mientras que para el TFM se estiman 150 horas de trabajo lo que

corresponde a 6 ECTS. En la planificación presentada, desde el 9 de mayo hasta el 13 de junio se han realizado las prácticas en el G.I.R y a partir de esa fecha, comienza la planificación para el TFM.

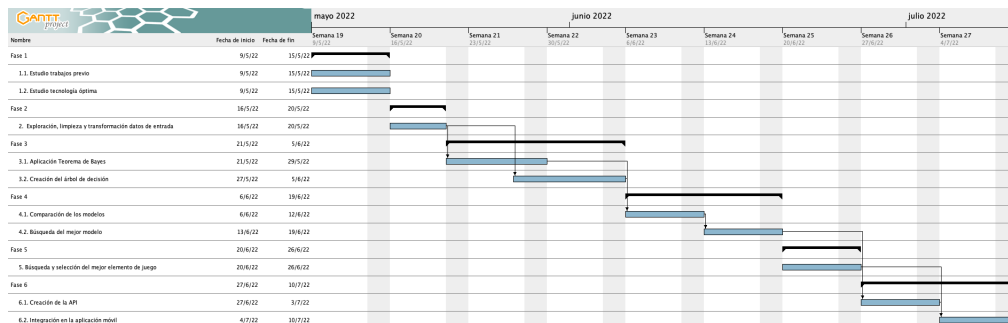


Figura A.1: Diagrama de Gantt con la planificación de la estancia y el TFM

A.2. Recursos

Recursos humanos

El desarrollo de este trabajo se ha realizado individualmente por la alumna. Se ha dedicado 340 horas en total, teniendo en cuenta tanto la duración de la estancia y el TFM. Para llevar a cabo el seguimiento del proyecto, se han realizado reuniones diarias con el tutor del trabajo, David Escudero Mancebo, donde se ha comprobado el desarrollo de las tareas, el cumplimiento de los objetivos y se han propuesto soluciones para aquellos problemas que han surgido. Por otra parte, para el seguimiento del desarrollo de la aplicación iOS, se ha realizado una reunión semanal, en concreto los viernes, con el equipo de Crossroads formado por 4 integrantes. En estas reuniones de seguimiento con el equipo Crossroads se han comentado los aspectos acerca del desarrollo de la aplicación móvil.

Recursos hardware y software

El recurso *hardware* utilizado para elaborar el trabajo es el ordenador portátil de la alumna. Es un MacBook Pro (2019) que cuenta con las siguientes características técnicas:

- **Sistema operativo** macOS Monterey (versión 12.4)
- **Procesador** 2,6 GHz Intel Core i7 de 6 núcleos
- **Memoria** 16 GB 2667 MHz DDR4
- **Gráficos** AMD Radeon Pro 5300M 4 GB Intel UHD Graphics 630 1536 MB

Los diferentes recursos *software* utilizados en el desarrollo del proyecto se presentan a continuación:

- **Google Scholar, IEEE Xplore, ResearchGate, Elsevier y AlmenaUVa** donde se ha buscado la información necesaria para el desarrollo del trabajo y la realización del estado del arte.
- **Office 365**, en concreto *Microsoft Word* empleado para tomar notas y visualizar documentos, *Microsoft PowerPoint* para realizar los esquemas funcionales de las Figuras 3.5 y 3.6 y *Microsoft Excel* donde se han creado hojas de cálculo para recoger los datos de las probabilidades obtenidas.
- **Latex en Overleaf** para el desarrollo de la memoria.
- **Adobe Reader** para la visualización de documentos PDF.
- **GanttProject** para el desarrollo del diagrama de Gantt presentado (Figura A.1).
- **Microsoft Teams y Webex** para realizar las reuniones *online* con el tutor y el equipo Crossroads.
- **GitLab** donde se ha almacenado los archivos relacionados con el proyecto.
- **Python, Jupyter Notebook, Weka, Flask, Xcode, Swift y SwiftUI** utilizado para la parte práctica de este proyecto, como se ha explicado en la Sección 3.3.

A.3. Viabilidad económica

Para calcular el presupuesto de un proyecto hay que tener en cuenta costes de plantilla, gastos generales y cargos por uso de suministros [28]. Los costes de plantilla incluyen los sueldos de los desarrolladores y la contribución por trabajador a la Seguridad Social. Se estima que el salario de un ingeniero de software con conocimientos sobre ciencia de datos sea 30.000 € brutos anuales con una jornada laboral de 1.800 horas/año. Si se adiciona al salario un 30 % correspondiente a los gastos sociales, el coste que supone este trabajador para la empresa sería 39.000. Conociendo estos datos, el coste por hora para este proyecto serían 21,67 €/hora según la estimación anterior. Por tanto, teniendo en cuenta la duración de trabajo se compone del TFM junto con la estancia de I+D+i, las horas empleadas en el mismo ascienden a 340 horas. Con la estimación realizada, el coste de plantilla relativo a este proyecto sería 7367,80 €.

Los gastos generales incluyen costes como el alquiler de las oficinas, la luz y el agua. Normalmente, estos costes se calculan aplicando un porcentaje de los costes directos de plantilla. Por tanto, para calcular los gastos generales del proyecto se ha estimado que sean un 30 % de los costes de plantilla dando lugar a 2.210,34 € de gastos generales.

El ordenador donde se ha desarrollado el proyecto aporta un coste al proyecto. Para calcularlo es necesario calcular la depreciación del mismo a lo largo del tiempo teniendo en cuenta la duración del proyecto. Esto se puede calcular a partir de su valor inicial o precio

de adquisición, su vida útil y su valor residual o el valor estimado cuando su vida útil termine [14]. La cuota de depreciación anual sería la presentada en la siguiente fórmula:

$$Cuota\ de\ depreciación\ anual = \frac{valor\ inicial - valor\ residual}{vida\ útil\ en\ años} \quad (A.1)$$

Teniendo en cuenta que el valor de adquisición del ordenador fueron 2.699 €, la vida útil estimada es de 5 años y el valor residual podría ser 100 €, se obtiene una cuota de depreciación anual presentada en la Fórmula A.2.

$$Cuota\ de\ depreciación\ anual = \frac{2.699 - 100}{5} = 519,8 \quad (A.2)$$

Si 519,8€ es la cuota de depreciación anual, para este proyecto que ha durado alrededor de 4 meses sería 173,26 €.

En el coste del proyecto también se deben incluir costes asociados a las herramientas utilizadas, como licencias de *software*. En este proyecto todas las tecnologías utilizadas no necesitan ninguna licencia de pago para su uso o la licencia está proporcionada por la Universidad de Valladolid, como en el caso de Office 365. Por lo que el coste asociado de licencias de software para el proyecto es 0 €.

Finalmente, en la Tabla A.1 se recogen los costes anteriormente calculados y el coste final estimado del proyecto que asciende a 9.751,40 €.

Costes de Plantilla	7.367,80 €
Gastos Generales	2.210,34 €
Material	173,26 €
Licencias software	0 €
Total	9.751,40 €

Tabla A.1: Costes simulados

Apéndice B

Enlaces

En este apéndice se recogen todos los enlaces de los repositorios donde se encuentran los trabajos de este proyecto.

Copia de la aplicación iOS Crossroads: <https://gitlab.inf.uva.es/marrobl/tfm-crossroadsapp>

Copia del backend de Crossroads: <https://gitlab.inf.uva.es/marrobl/tfm-crossroadsbackend>

API del estimador de estado bueno: <https://gitlab.inf.uva.es/marrobl/estimadorestadobueno>

Ficheros de preprocesamiento de datos: <https://gitlab.inf.uva.es/marrobl/tfm-procesamientodatos>

Apéndice C

Manual de instalación y despliegue

En este apéndice se explica cómo descargar e instalar el proyecto del estimador realizado. Para ello, se deben descargar los siguientes proyectos:

- Proyecto que contiene el modelo C4.5 y la API para realizar la estimación de un resultado aceptable.
- Proyecto de la aplicación móvil
- Proyecto del *backend* de Crossroads

A continuación, se explican los pasos a seguir para la instalación y despliegue de estos proyectos. Todos los repositorios mencionados a continuación corresponden a los enlaces adjuntos en el Apéndice B.

C.1. Proyecto estimación estado bueno

En un repositorio se ha almacenado el modelo C4.5 que se utiliza para realizar la estimación y la función Python que utiliza dicho modelo y calcula la estimación. Además, en ese repositorio también se cuenta con un fichero que contiene la API realizada con Flask para utilizar externamente el estimador.

Manual de instalación y despliegue

Se puede descargar el código fuente clonando el repositorio ejecutando el comando que aparece en el Fragmento de Código C.1.

```
1 | git clone https://gitlab.inf.uva.es/marrobl/  
   |   estimadorestadobueno.git/
```

Fragmento de código C.1. Clonar proyecto del estimador

Una vez clonado, se habrá creado un directorio `estimadorestadobueno` que contiene toda la lógica del estimador y su API. Para acceder a él se utiliza el comando de terminal `cd`. Cuando nos situamos dentro del directorio, el siguiente paso es crear un entorno virtual para crear un entorno de Python aislado para gestionar las dependencias de diferentes proyectos. Para ello se debe instalar la biblioteca `virtualenv`, crear el entorno virtual y activarlo, como se muestra en el siguiente Fragmento de Código C.2.

```
1 pip3 install virtualenv // Descargar biblioteca
2 virtualenv . // Crear entorno virtual
3 source bin/activate // Activar entorno
```

Fragmento de código C.2. Descargar, crear y activar entorno virtual

Ahora se debe instalar los paquetes necesarios para la ejecución del código. Los paquetes que se necesitan son `Flask` y `python-weka-wrapper3`.

```
1 Python3 -m pip install Flask
2 pip3 install python-weka-wrapper3
```

Fragmento de código C.3. Descargar paquetes necesarios

Una vez instaladas las dependencias ya se puede desplegar la API, ejecutando para ello el comando:

```
1 flask run
```

Fragmento de código C.4. Desplegar API

Cuando se ha desplegado, la API se encontrará escuchando en la máquina local en el puerto 5000.

Contenido

El repositorio que contiene la API para el estimador de estado bueno cuenta con los siguientes archivos:

- Archivo `app.py`: contiene código de la API.
- Archivo `estimador.py`: contiene la función que realiza el cálculo de la estimación de estado bueno utilizando el modelo basado en el algoritmo C4.5.
- Directorio `model`: almacena el modelo basado en el algoritmo C4.5 que se utiliza para predecir.

C.2. Aplicación móvil Crossroads (copia modificada)

Para añadir el ítem gráfico (Figura 4.8) que muestre la estimación de estado bueno en la aplicación móvil de Crossroads, se ha realizado una copia del proyecto principal de iOS. Para poder instalar el proyecto en la máquina local, se debe clonar el repositorio que contiene el proyecto con el comando que se indica en el Fragmento de Código C.5.

```
1 | git clone https://gitlab.inf.uva.es/marrobl/tfm-crossroadsapp.git
```

Fragmento de código C.5. Clonar proyecto copia de la app iOS Crossroads

Una vez descargado se habrá creado un nuevo directorio llamado `crossroad2` que contiene todo el proyecto iOS. Para abrir el proyecto es necesario contar con el sistema operativo `macOS` y la versión 13.4 de Xcode o superior. El siguiente paso será abrir desde Xcode el proyecto que se ha descargado. Una vez hecho esto, ya se puede acceder a todo el código del frontend iOS. En concreto, la vista creada para el elemento gráfico se encuentra en `Crossroad2/views/utils/SpinnerView.swift`.

C.3. Backend Crossroads (copia modificada)

Para añadir la operación necesaria en este TFM se ha realizado una copia del *backend* original de Crossroads. Esta copia se puede descargar clonando el repositorio como se muestra en el siguiente Fragmento de Código C.6:

```
1 | git clone https://gitlab.inf.uva.es/marrobl/tfm-crossroadsbackend.git
```

Fragmento de código C.6. Clonar proyecto copia del backend de Crossroads

Para instalar las dependencias, crear la base de datos y desplegar el proyecto, se deben seguir los mismos pasos que se indican en la Sección “Manual de despliegue e instalación” del Trabajo Fin de Grado “Desarrollo del front-end y mejoras en el back-end de un juego didáctico multijugador de competición y consenso sobre el cambio climático”¹.

¹Disponible en: <https://uvadoc.uva.es/handle/10324/50085>

Apéndice *D*

Descripción del contenido del juego Crossroads

En este apéndice se añade el contenido del juego Crossroads, enseñando el texto que se muestra en la aplicación en cada pregunta y respuesta del mismo. Además, se muestra el texto de ayuda para el usuario que acompaña a cada pregunta.

D.1. Preguntas

- **Hipótesis 1.** Según tu opinión, ¿qué disponibilidad de recursos energéticos no renovables (petróleo, carbón, gas natural y uranio) crees que aún queda hasta 2070?
 - a. Media
 - b. Alta
 - c. Ilimitada

- **Hipótesis 2.** ¿Qué nivel de impacto sobre la economía mundial estimas que producirá el cambio climático derivado de la actividad humana?
 - a. Bajo
 - b. Medio
 - c. Alto

- **Hipótesis 3.** ¿Cómo piensas que convendría que evolucionaran los niveles de población mundial?
 - a. Disminuyendo respecto del previsto por las tendencias históricas
 - b. Continuando las tendencias históricas de crecimiento
 - c. Aumentando respecto de lo previsto en las tendencias históricas

- d. Limitándose a 5.000 millones de personas para 2050
- e. Limitándose a 7.000 millones de personas para 2050
- **Objetivo 1.** ¿En qué valor fijarías el límite del incremento de la temperatura media del planeta?
 - a. Por debajo de 1,5 °C
 - b. Por debajo de 2 °C
 - c. Por debajo de 3 °C
- **Objetivo 2.** De acuerdo con el índice de desarrollo humano (IDH), ¿hacia qué nivel de bienestar consideras que debería tender la población mundial?
 - a. Muy alto
 - b. Alto
 - c. Medio
 - d. Bajo
- **Medida política 1.** ¿Cómo estimas que debería ser la evolución del Producto Interior Bruto real por persona?
 - a. Menor crecimiento que el previsto en las últimas décadas
 - b. Similar crecimiento al de las últimas décadas
 - c. Mayor crecimiento que el previsto
 - d. Fijaría el PIBpc real en 5000\$ para 2050 y lo mantendría constante
 - e. Fijaría el PIBpc real en 7000\$ para 2050 y lo mantendría constante
- **Medida política 2.** ¿Crees que sería importante llevar a cabo un programa mundial de forestación y reforestación de bosques?
 - a. No, pues esto implicaría una gran reducción de las superficies cultivables que son necesarias para la producción de alimentos y energía
 - b. Sí, ya que es una forma de compensar las emisiones de CO₂ y ayudar a frenar el aumento de la temperatura media del planeta
- **Medida política 3.** ¿Cómo consideras que debería ser la evolución de la potencia instalada de energía nuclear a nivel mundial?
 - a. Debería disminuirse
 - b. Debería mantenerse
 - c. Debería aumentarse
- **Medida política 4.** ¿Cómo piensas que debería ser la evolución de la capacidad de generación de energía renovable para la obtención de electricidad y calor?

- a. Hacia un menor crecimiento que la tendencia histórica
 - b. Hacia una continuación de la tendencia histórica de crecimiento
 - c. Hacia un mayor crecimiento que la tendencia histórica
- **Medida política 5.** ¿Cómo consideras que debería ser la producción de biocombustibles líquidos (bioetanol y biodiésel)?
- a. Debería estabilizarse en el nivel actual de producción
 - b. Debería continuarse con la tendencia histórica de crecimiento en la producción
 - c. Debería aumentarse la producción respecto de la tendencia histórica
- **Medida política 6.** ¿Crees que se debería promover una transformación del transporte terrestre hacia un mayor uso de vehículos eléctricos e híbridos?
- a. No, deberíamos continuar con el nivel actual de uso de los derivados del petróleo
 - b. Sí, deberíamos transitar hacia un mayor uso de vehículos alternativos
- **Medida política 7.** ¿Cómo piensas que debería ser la inversión en investigación relacionada con la eficiencia energética?
- a. La inversión debería ser menor respecto de la tendencia histórica porque mantener los niveles actuales implicaría un gasto inasumible en recursos humanos y económicos
 - b. La inversión al menos debería tratar de mantener los valores máximos de mejora alcanzada en muchas tecnologías
 - c. La inversión debería ser mayor respecto de la tendencia histórica debido a que aún pueden alcanzarse mejoras en la eficiencia energética en algunas tecnologías
- **Medida política 8.** ¿Aumentarías el porcentaje de reciclado de minerales necesarios para las tecnologías vinculadas a las energías renovables?
- a. No, mantendría constantes los porcentajes actuales
 - b. Sí, elevaría los porcentajes actuales
- **Medida política 9.** ¿Cómo consideras que deberían gestionarse las emisiones de gases de efecto invernadero que no dependen del sector energético?
- a. Reduciéndolas respecto a la tendencia actual
 - b. Continuando con la tendencia actual
 - c. Aumentándolas respecto a la tendencia actual

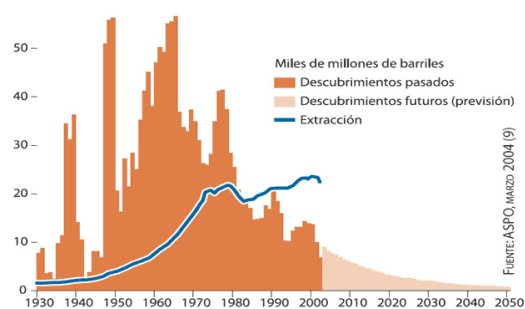
D.2. Ayudas

A continuación se adjunta el texto de las ayudas asociadas a cada pregunta. Estos textos se muestran durante el transcurso del juego con el fin de ayudar y guiar al jugador a comprender mejor el problema que se está planteando en cada pregunta. En las ayudas de cada pregunta se muestra un texto acompañado de imágenes que contiene una descripción más detallada del ámbito o contexto que trata la pregunta e intenta explicar la situación actual para que el usuario pueda tomar decisiones más acertadas. Estas ayudas han sido desarrolladas por integrantes del grupo GEEDS de la UVa para añadir al juego de Crossroads de plataforma web como móvil.

■ Ayuda Hipótesis 1.

Los recursos no renovables son aquellos elementos dispuestos en la naturaleza en cantidades finitas y limitadas. Estos recursos no renovables se utilizan, principalmente, como fuentes de energía; de forma que, atendiendo a estos materiales, se puede hacer una distinción entre fuentes de energía no renovables convencionales y fuentes de energía no renovables no convencionales. El primer grupo engloba a aquellos recursos que provienen de los grandes depósitos de carbono que, generados por la descomposición de materia orgánica, se encuentran atrapados en multitud de cuencas sedimentarias. Estos elementos se conocen como combustibles fósiles e incluyen el petróleo, el carbón y el gas natural. Por otro lado, la segunda categoría alude, especialmente, a la energía nuclear, que es aquella que se genera a partir de una serie de procesos químicos que consisten en la desintegración de los átomos de minerales como el uranio o el plutonio.

Desde mediados del siglo pasado, numerosos estudios vienen advirtiendo de la posibilidad de alcanzar el pico del petróleo, es decir, el momento de máxima extracción de crudo, a partir del cual se espera que su producción decrezca irremediablemente.



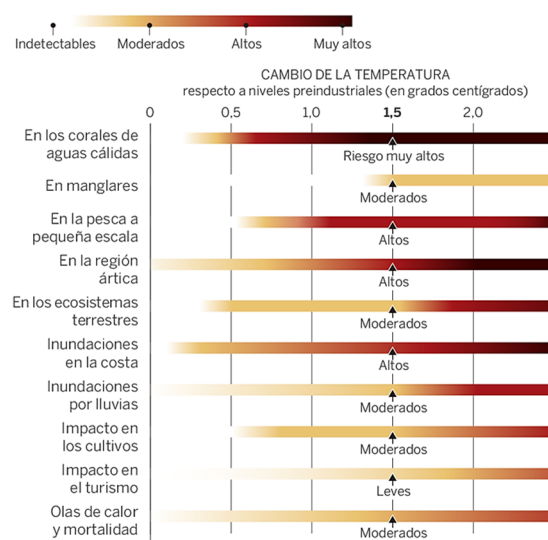
E igualmente podemos afirmar de los demás recursos energéticos no renovables y que no pueden ser restituidos en un breve lapso de temporal.

No obstante, hablamos de estimaciones, las cuales nos impelen a tomar medidas sobre su extracción y consumo, dada nuestra dependencia total de estos recursos

para la obtención de electricidad y calor. Pues no será lo mismo suponer una elevada disponibilidad que una más baja, dada que afectación que ello supondría para todo el mundo, si bien en desigual medida.

■ **Ayuda Hipótesis 2.**

El sistema económico mantiene una relación bidireccional con el cambio climático: por un lado, contribuye a su avance a través de las emisiones de gases de efecto invernadero y, por el otro, recibe los impactos de tales alteraciones del clima. Aunque existe gran incertidumbre con respecto a cuál podría ser la magnitud de los impactos climáticos sobre la economía, sí que hay cierto consenso sobre cuáles podrían ser algunos de estos efectos. Entre ellos encontramos fenómenos de todo tipo: problemas en la salud y reducción de la productividad laboral debido a mayores y más frecuentes olas de calor, destrucción de infraestructuras, viviendas y maquinaria debido a un incremento de incendios e inundaciones, falta de acceso a agua y comida por el aumento de sequías, y un largo etcétera que podemos ilustrar a través del siguiente esquema realizado por el Panel Intergubernamental contra el Cambio Climático (IPCC, SR15).

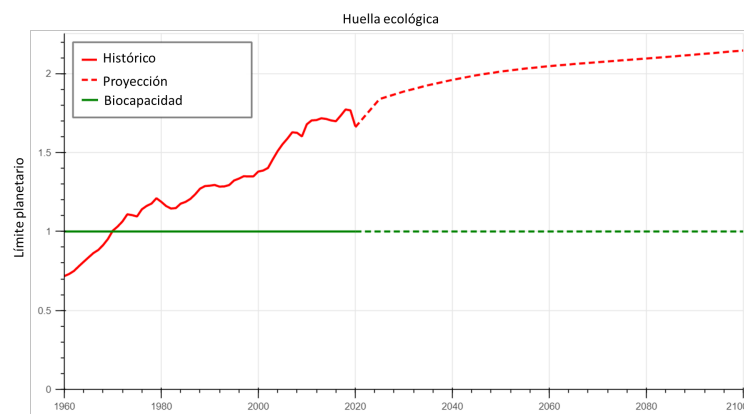


Dadas las altas interrelaciones que existen entre sectores económicos y regiones en un sistema globalizado como el actual, pequeñas disrupciones locales como puede ser la paralización de algún sector estratégico podrían desembocar en impactos de mayor escala e incluso en grandes crisis globales. Si bien, el impacto final dependerá de la magnitud de los fenómenos climáticos extremos, de la exposición a ellos y de la capacidad para adaptarnos.

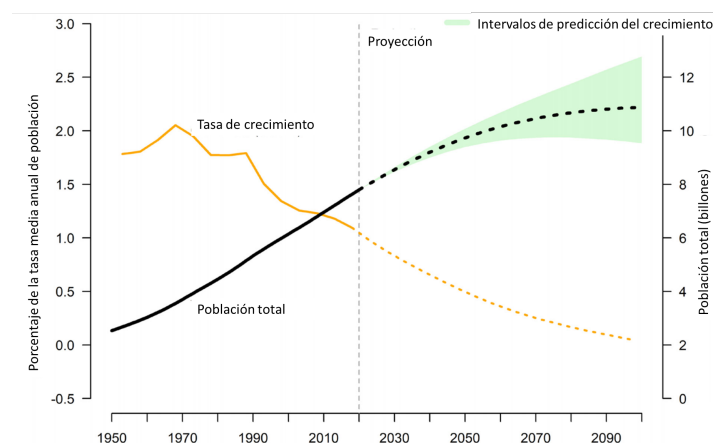
Por eso, aun reconociendo el problema, no es menos cierto que la incertidumbre sobre el cálculo de estos daños es alta. Aun así, no podemos omitir los riesgos en nuestras plantificaciones presentes y futuras.

■ Ayuda Hipótesis 3.

El aumento mundial de la población ha ido paralelo a la capacidad de aprovechamiento de los recursos naturales, especialmente los energéticos. Ha sido la progresiva disponibilidad de uso que nos ha proporcionado la energía fósil para gran parte de nuestras actividades lo que ha permitido una gran expansión por todo el mundo, con el consiguiente sobreconsumo que ello ha supuesto. Solemos referirnos a esto con la expresión "huella ecológica".



Desde comienzos de los años setenta del pasado siglo XX se tiene constancia de haber pasado los límites ecosistémicos del planeta. Restituírnos dentro de los límites ecosistémicos, lo que en ocasiones se denomina retorno a la sostenibilidad socioambiental.º retorno hacia sociedades ecológicamente sostenibles requiere reducir nuestro consumo de los distintos bienes ecosistémicos.



Ya desde épocas pretéritas se ha tenido como una máxima válida aquello de "nada en demasíaz los teóricos de la política tenían consciencia de que el número de la población no debía sobrepasar las posibilidades de mantenerse con los recursos del

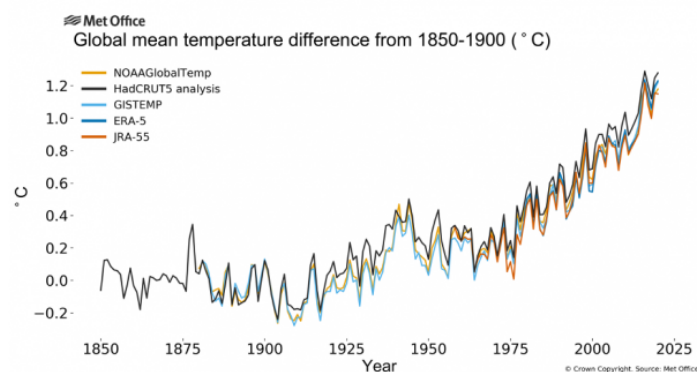
entorno próximo, y aunque en nuestros días este entorno parece haberse ensanchado debido a la globalización de nuestros procesos comerciales, la lógica de antaño no ha variado, dado que vivimos en un mundo limitado y en progresivo deterioro.

■ Ayuda Objetivo 1.

El Acuerdo de París establece como objetivo "mantener el aumento de la temperatura media mundial muy por debajo de 2 °C con respecto a los niveles preindustriales, y proseguir los esfuerzos para limitar ese aumento de la temperatura a 1,5 °C con respecto a los niveles preindustriales, reconociendo que ello reduciría considerablemente los riesgos y los efectos del cambio climático".

En la actualidad, ya la temperatura ya se ha incrementado más de 1°C desde el periodo comprendido entre 1850 y 1900, que es el que normalmente utiliza el Panel Intergubernamental de Expertos sobre el Cambio Climático (IPCC) como referencia de la etapa anterior a la primera revolución industrial.

Este organismo estima que, si las tendencias de emisiones de gases de invernadero continúan creciendo según tendencias históricas, alcanzaremos un incremento de temperatura de 1,5°C aproximadamente en 2040.

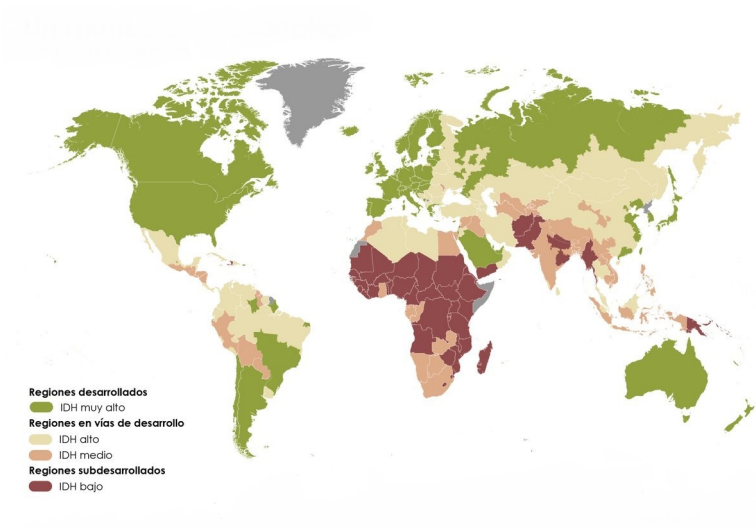


Esta elevación de la temperatura medida de la superficie de la Tierra supondría graves consecuencias para muchos ecosistemas y la vida de millones de personas, ya sea de manera directa o indirecta, que se agravará en la medida en que este incremento de la temperatura sea mayor. Pero, de nuevo, hablamos de riesgos y la pregunta que cabe anteponer es si estos deben determinar la toma de decisiones políticas que afectarán a nuestro modo de vida.

■ Ayuda Objetivo 2.

El IDH se construye teniendo en cuenta tres aspectos fundamentales de la vida del ser humano. La riqueza, medida con el producto nacional bruto per cápita en paridad de poder adquisitivo (es decir, teniendo en cuenta los diferentes niveles de precios de los distintos países), la salud, medida con la esperanza de vida al nacer, y la educación, con los años de escolarización. Como resultado, el IDH toma para

cada país un valor entre cero y uno que nos permite clasificarlo en cuatro niveles de desarrollo: muy alto (más de 0,8), alto (entre 0,7 y 0,8), medio (entre 0,55 y 0,7) y bajo (menos de 0,55).

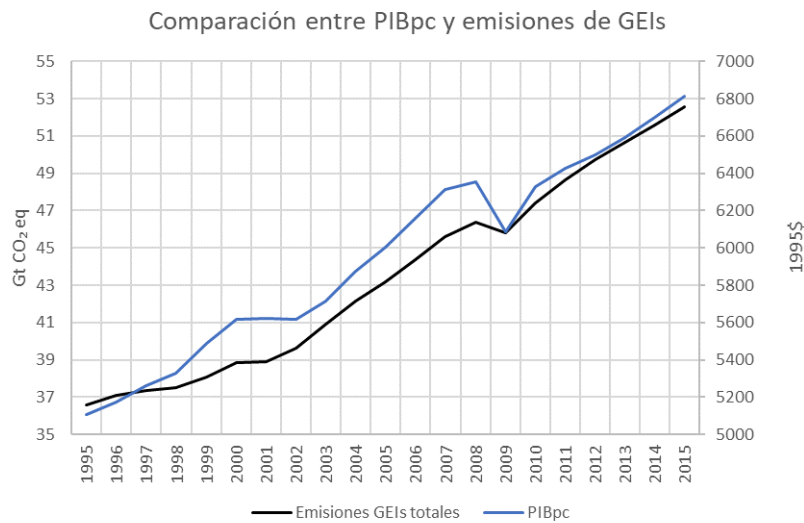


Un resultado medianamente bueno de IDH exige que estén cubiertas las necesidades básicas (alimentación, cobijo, salud, educación, etc.) de la población de un país. Para ello, además de instituciones que garanticen la satisfacción universal de esas necesidades, será necesario que se alcancen unos mínimos niveles de riqueza (usualmente medido en términos de PIB), lo cual conllevará también un cierto gasto de recursos e impactos medioambientales y climáticos.

■ Ayuda Medida política 1.

El producto interior bruto por persona (PIBpc) se utiliza como medida del nivel productivo de una economía. Se calcula haciendo la división del valor de los bienes y servicios producidos en un año entre la población del país.

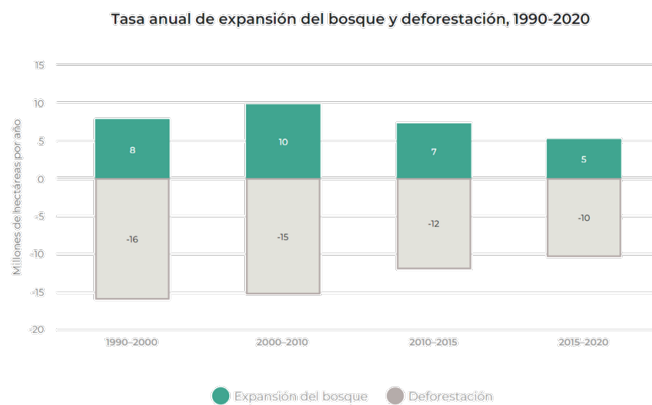
Un mayor PIBpc suele implicar mayores niveles de consumo de la población, aunque puedan distribirse estos de forma desigual. A su vez, los mayores niveles de consumo implican más gasto de recursos y mayor generación de residuos, entre ellos la liberación de gases de efecto invernadero (GEIs).



■ Ayuda Medida política 2.

Aunque sea en términos utilitarios, nuestros entornos naturales nos proporcionan toda suerte de provechos, incluidos la capacidad que tienen algunos ecosistemas para albergar de manera natural parte de nuestros residuos, entre ellos el dióxido de carbono (CO₂) que se encuentra en proporciones crecientes en el aire.

A esta absorción de CO₂ contribuyen, principalmente, la vida vegetal marina y los bosques a través del proceso de la fotosíntesis. Sin embargo, la actividad humana está provocando la muerte de los primeros y pérdida de grandes masas forestales.



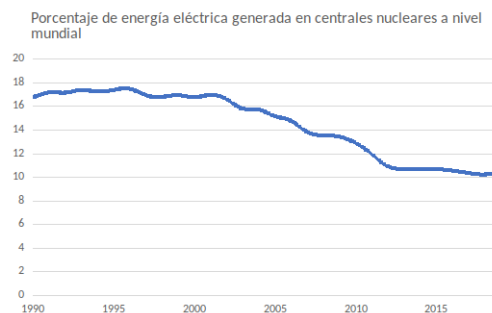
Estas pérdidas, gráficamente representas por la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), además comportan una grave disminución de diversidad biológica, con todas las desventajas que ello implica para el ser humano.

Pero aunque se podría pensar que medidas de corrección apropiadas serían aquellas que fomentaran la forestación y reforestación, lo cierto es que ello conflictuaría con las necesidades crecientes para otros usos de la tierra: por ejemplo para satisfacer la demanda de alimento tanto para los seres humanos como para el ganado cuya carne consumimos en grandes proporciones; pero también para despachar las progresivas demandas de biocombustibles, posibilitar la apertura a nuevos espacios urbanizables e incluso de lugares para implantar alguno de los sistemas de energía renovable que ayudarían a reducir nuestra dependencia de los recursos energéticos fósiles.

■ Ayuda Medida política 3.

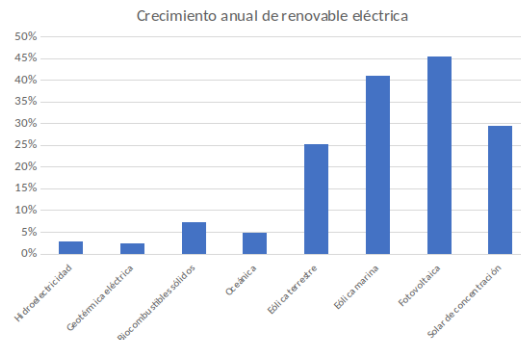
Las centrales nucleares producen energía eléctrica a partir de la fisión nuclear de determinados elementos radiactivos. Tienen tres inconvenientes principales: los riesgos asociados a fallos operativos que puedan dar lugar a emisiones de material radioactivo, la generación de residuos radiactivos que deben ser aislados durante miles de años y la escasa cantidad de uranio en la corteza terrestre para generar este tipo de energía.

En el plano de la viabilidad económica, las instalaciones nucleares necesitan de una inversión inicial muy fuerte, pero tienen unos costes operativos bajos (es decir, una vez construidas, las centrales nucleares permiten producir electricidad de forma barata). No obstante, a esos costes deberían añadirse los de la gestión de los residuos y los de los riesgos asociados a un desastre nuclear, que, aunque improbable, puede generar grandes daños. Si bien, no son esos los motivo ni de que el porcentaje de energía eléctrica generada mediante centrales nucleares sea relativamente pequeño con respecto al total, ni de que este haya ido disminuyendo con el paso del tiempo.

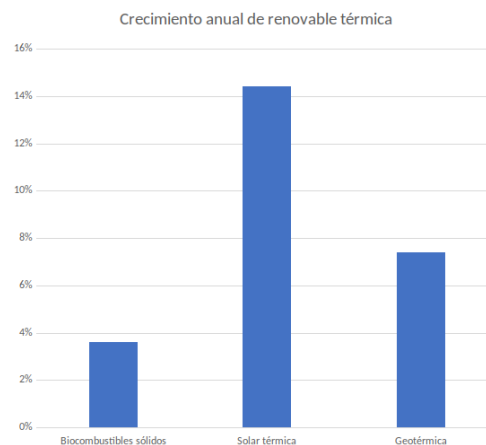


■ Ayuda Medida política 4.

La producción de energía renovable se divide entre tecnologías de generación eléctrica y tecnologías de generación térmica. Entre las primeras se encuentran la hidroeléctrica, la eólica, la solar fotovoltaica, la solar termoeléctrica, la geotérmica eléctrica, los biocombustibles y la oceánica.



Mientras que entre las tecnologías de generación de energía renovable térmica se encuentran los biocombustibles sólidos (carbón vegetal, leña, residuos agrícolas y forestales), la energía solar térmica y la energía geotérmica.



Todas estas tecnologías renovables implican tanto un mayor consumo de materiales escasos en la corteza terrestre, que por tanto pueden suponer una limitación en la instalación de estas tecnologías, como una ocupación de suelos que compite de forma importante con otros usos de las tierras debido a que las tecnologías renovables adolecen de una escasa densidad energética.

■ **Ayuda Medida política 5.**

Los biocombustibles constituyen una mezcla de diversas sustancias procedentes de la biomasa, es decir, de la acumulación de un amplio conjunto de materia orgánica que se vuelve susceptible de ser utilizada como fuente de energía alternativa a los combustibles fósiles.

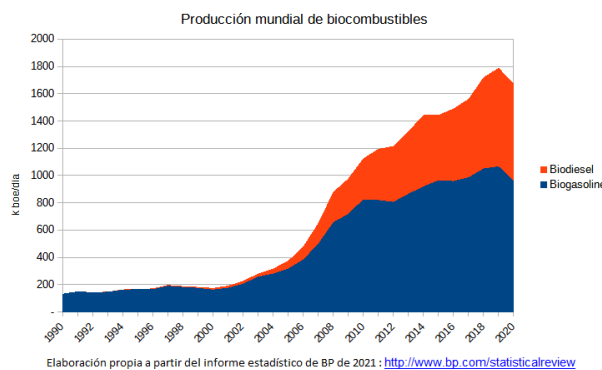
Entre los biocombustibles más frecuentes destacan el bioetanol (que se origina mediante la fermentación alcohólica de los azúcares concentrados en determinadas especies vegetales) o el biodiesel (que se genera a partir de aceites vegetales).

La principal ventaja de su utilización radica en que, durante su combustión, las emisiones netas de dióxido de carbono (CO₂) se consideran nulas.

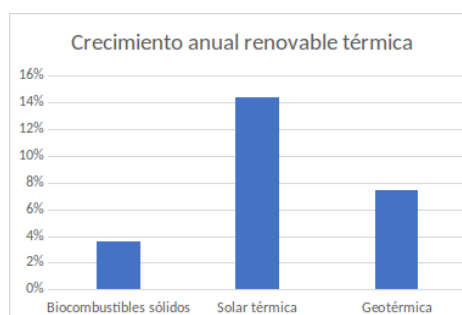
Esto no quiere decir que no se libere dióxido de carbono, sino que esta liberación se equilibra con el que se absorbe a lo largo de todo su proceso de producción.

Por el contrario, entre sus inconvenientes se halla el aumento de las emisiones de óxidos de nitrógeno (N_xO_y) a causa del uso de fertilizantes nitrogenados en los campos de cultivo.

Así, mientras numerosas hectáreas de terreno son expoliadas con el fin de aumentar la producción de los biocombustibles, se debate sobre si se cumple entonces con el objetivo que se persigue, o sea, con la reducción de los gases de efecto invernadero que repercuten notablemente sobre el proceso de calentamiento global.



Con todo, según la Agencia Internacional de Energía (AIE), para el año 2018, los biocombustibles proporcionaron el 93 % de toda la energía renovable dedicada al transporte, un 3,97 %, y es que, la producción mundial de biocombustibles duplicó el crecimiento del año previo para alcanzar un récord de 154 mil millones de litros.



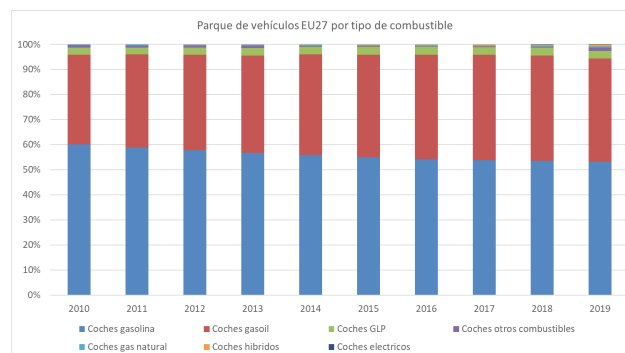
Pero todas estas tecnologías renovables implican tanto un mayor consumo de materiales escasos en la corteza terrestre (lo que, por tanto, puede suponer una limitación en la instalación de este tipo de tecnologías), como una ocupación de suelos que compite

de forma importante con otros usos de las tierras, debido a que las tecnologías renovables adolecen de una escasa densidad energética.

■ Ayuda Medida política 6.

El parque de vehículos terrestres es mayoritariamente consumidor de productos derivados del petróleo, tanto gasolinas como gasóleos. Si bien existen combustibles alternativos a estos últimos que es posible utilizar en los motores de los vehículos terrestres, como la electricidad, el gas o los biocombustibles, que en la actualidad representan un porcentaje muy pequeño del total de vehículos.

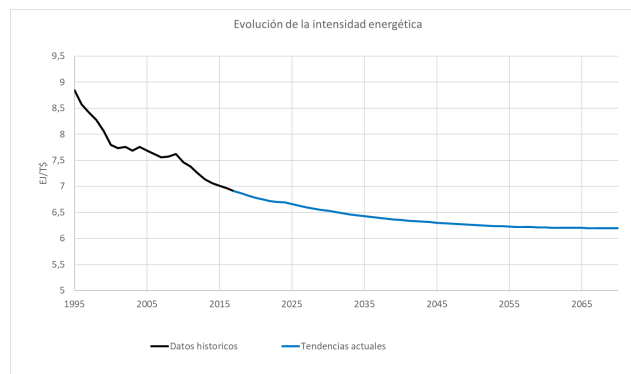
Si decidiéramos seguir esta tendencia, continuaría el aumento de las emisiones de gases de efecto invernadero (GEIs) y el progresivo agotamiento de recursos fósiles. Si, por el contrario, decidiéramos intentar la transición hacia un mayor uso de vehículos que utilicen combustibles alternativos a los derivados del petróleo, ello implicaría una gran dependencia de minerales escasos pero que son necesarios para las baterías de los vehículos eléctricos e híbridos. Se suma a ello el requerimiento de una mayor necesidad de tierras para esta producción, lo que también implicaría un aumento de las emisiones de otros gases de efecto invernadero (principalmente el óxido nitroso).



■ Ayuda Medida política 7.

La intensidad energética relaciona la producción económica con el consumo energético. Así una menor intensidad implica un menor consumo energético necesario para obtener la misma producción económica.

Al respecto, la tendencia histórica reciente es de una constante mejora en la intensidad energética, sin embargo, este ritmo de mejora es cada año más bajo, debido a que, en general, la mayoría de las tecnologías ya han alcanzado o están muy cerca de los límites teóricos de eficiencia y cada vez es más costoso y complejo obtener un incremento significativo de la intensidad energética. Lo cual no obsta para que se intente, si así se desea, aun a sabiendas de que ello implicaría además una notable inversión económica y humana dedicada a la investigación para tratar de llevar a cabo ese propósito de incrementar la intensidad energética.



■ **Ayuda Medida política 8.**

Las diversas tecnologías usadas en los distintos sistemas de producción de energías renovables utilizan materiales muy escasos en la corteza terrestre, como el neodimio (utilizado en los generadores eólicos), el telurio y el indio utilizados en las tecnologías fotovoltaicas) o el litio, el cobalto y el níquel (utilizados en los distintos sistemas para la acumulación eléctrica). Pero estos son solo algunos ejemplos del uso de esos minerales necesarios para la transición energética. Pero, ocurre que, debido a su escasez, la mayoría de estos minerales, incluso si solamente se mantiene el ritmo actual de implantación de tecnologías renovables, llegarán en unas pocas décadas a su agotamiento tanto del total de sus reservas como sus recursos. Por ello, se plantea modificar las tasas de reciclado de los más importantes de ellos, muchos de los cuales presentan tasas de reciclaje inferiores al 1

Mineral	Tasas de reciclado actuales
Aluminio (Al)	56%
Cadmio (Cd)	15%
Cromo (Cr)	90%
Cobre (Cu)	48%
Galio (Ga)	1%
Indio (In)	1%
Hierro (Fe)	71%
Litio (Li)	1%
Magnesio (Mg)	39%
Manganeso (Mn)	53%
Molibdeno (Mo)	30%
Níquel (Ni)	60%
Plomo (Pb)	74%
Plata (Ag)	64%
Estaño (Sn)	75%
Telurio (Te)	1%
Titanio (Ti)	91%
Vanadio (V)	1%
Zinc (Zn)	40%

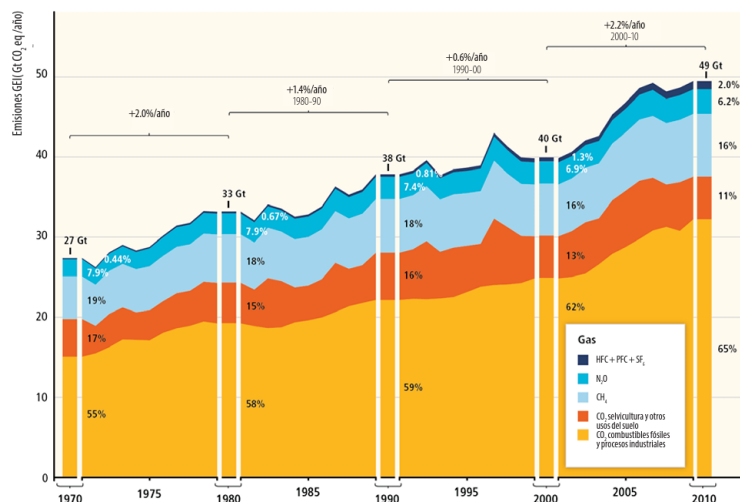
El problema añadido es que, por un lado, el reciclaje nunca es completo (siempre hay pérdidas durante el proceso) y, por otro lado, este tipo de acciones -de poder realizarse- comportaría un sobre gasto económico y energético.

■ **Ayuda Medida política 9.**

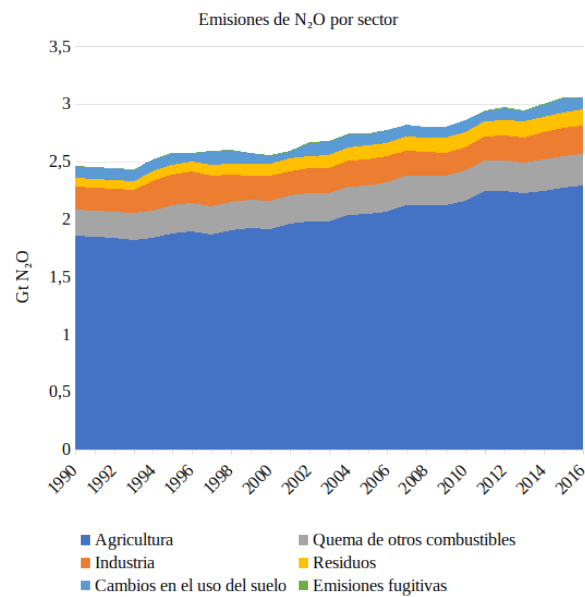
Mucho se oye hablar de los intentos de reducir las emisiones de dióxido de carbono (CO₂). Pero este no es el único gas de efecto invernadero. La Convención Marco de las Naciones Unidas contra el Cambio Climático (CMNNCC) refiere también otros gases de efecto invernadero de larga vida (GEILV), como el metano (CH₄) y el óxido nitroso (N₂O), cuyo aumento ilustra la siguiente gráfica del Grupo Intergubernamental de Expertos sobre el Cambio Climático:

Pero tratar de reducir estos otros gases requiere que, además de moderar nuestra dependencia de las energías fósiles, disminuyamos otras tantas actividades económico-productivas.

Así, por ejemplo, para el caso del metano (CH₄) convendría que redujéramos el sector agrícola-ganadero.



Evitar mayores porcentajes de liberación de óxido nitroso (N₂O) también nos obligaría a una considerable reducción del sector agrícola.



Se entiende, por tanto, que la exigencia de cambio en nuestras pautas de producción y consumo sería enorme si quisiéramos reducir la acumulación de gases de efecto invernadero (GEIs) en la atmósfera y con ello evitar mayores incrementos de la temperatura media global, ayudando a mitigar los posibles impactos sobre nuestras vidas y hacer que la adaptación a las nuevas condiciones climáticas sean lo menos drásticas posibles, si es que ese es nuestro propósito para las generaciones presentes y aún para las venideras.

Bibliografía

- [1] AMBIENTECH. ambientech. <https://ambientech.org/>, 2022. [Accedido 08-julio-2022].
- [2] AMBIENTECH. La emergencia climática. <https://ambientech.org/itinerariosad/emergencia-climatica/indice.html>, 2022. [Accedido 08-julio-2022].
- [3] ANACONDA INC. Anaconda. <https://www.anaconda.com/>. [Accedido 01-junio-2022].
- [4] APPLE INC. Swift - apple developer. <https://developer.apple.com/swift/>. [Accedido 30-mayo-2022].
- [5] APPLE INC. Swiftui tutorials - apple developer. <https://developer.apple.com/tutorials/swiftui>. [Accedido 30-mayo-2022].
- [6] APPLE INC. Xcode. <https://apps.apple.com/es/app/xcode/id497799835?mt=12>. [Accedido 30-mayo-2022].
- [7] ARTURO BUSTELO RIAL, PATRICIA NOVOA ESTEVEZ, ABRAHAM GONZÁLEZ MARTÍNEZ, BÁRBARA SANTANA RODRÍGUEZ, JUAN MANUEL ROMERO IGLESIAS, MARIAN QUIÑONES ÁLVAREZ, PABLO ANGEL MEIRA CARTEA, MÓNICA ARTO BLANCO, MIGUEL PARDELLAS SANTIAGO . La sociedad española ante el cambio climático. percepción y comportamientos en la población. https://accesoesee.idearainvestigacion.com/Informe_sociedad_espa%C3%B1ola_CC_2020.pdf, 2021. [Accedido 07-julio-2022].
- [8] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., AND STONE, C. J. *Classification and regression trees*. Routledge, 2017.
- [9] BRIGHTERION. A closer look at ai: data mining. <https://brighterion.com/data-mining/>, 2017. [Accedido 13-julio-2022].
- [10] BURGOS, D., VAN NIMWEGEN, C., VAN OOSTENDORP, H., AND KOPER, R. Game-based learning and the role of feedback. a case study. *Citeseer* (2007).

- [11] CAPELLÁN-PÉREZ, I., ÁLVAREZ-ANTELO, D., AND MIGUEL, L. J. Global sustainability crossroads: A participatory simulation game to educate in the energy and sustainability challenges of the 21st century. *Sustainability* 11, 13 (2019), 3672.
- [12] CHO, Y. H., KIM, J. K., AND KIM, S. H. A personalized recommender system based on web usage mining and decision tree induction. *Expert systems with Applications* 23, 3 (2002), 329–342.
- [13] DOUGLAS, B. D., AND BRAUER, M. Gamification to prevent climate change: A review of games and apps for sustainability. *Current Opinion in Psychology* 42 (2021), 89–94.
- [14] ESTEFANÍA MAC. Fórmulas de contabilidad básica: ¿cómo calcular la depreciación de una computadora? <https://www.cuidatudinero.com/13074026/como-calcular-la-amortizacion-de-una-computadora-portatil>, 2019. [Accedido 20-julio-2022].
- [15] FAN, W., GREENGRASS, E., MCCLOSKEY, J., YU, P. S., AND DRAMMEY, K. Effective estimation of posterior probabilities: Explaining the accuracy of randomized decision tree approaches. In *Fifth IEEE International Conference on Data Mining (ICDM'05)* (2005), IEEE, pp. 8–pp.
- [16] FECYT. Fundación española de ciencia y tecnología. <https://www.fecyt.es/>. [Accedido 23-abril-2022].
- [17] FEDERATION UNIVERSITY. Types of feedback. <https://federation.edu.au/staff/learning-and-teaching/teaching-practice/feedback/types-of-feedback>, 2022. [Accedido 28-julio-2022].
- [18] FINANCIAL TIMES. The climate game. <https://ig.ft.com/climate-game/>, 2022. [Accedido 08-julio-2022].
- [19] FÜRNKRANZ, J. Machine learning in games: A survey. *Machines that learn to play games* (2001), 11–59.
- [20] GALWAY, L., CHARLES, D., AND BLACK, M. Machine learning in digital games: a survey. *Artificial Intelligence Review* 29, 2 (2008), 123–161.
- [21] GOMEZ ESCODA, A. El feedback sumativo y el feedback formativo en las clases de español en cuatro escuelas de småland.: Percepciones y preferencias de los alumnos., 2016.
- [22] GREENPEACE. Actúa con tu consumo. <http://archivo-es.greenpeace.org/espana/es/Que-puedes-hacer-tu/Actua-con-tu-consumo/>, 2017. [Accedido 08-julio-2022].
- [23] GRESALFI, M. S., AND BARNES, J. Designing feedback in an immersive videogame: supporting student mathematical engagement. *Educational Technology Research and Development* 64, 1 (2016), 65–86.

- [24] GUTERRES, A. Threat to global security more complex, probably higher than during cold war, secretary-general warns munich security conference. <https://www.un.org/press/en/2022/sgsm21146.doc.htm#:~:text=Extremism%20and%20terrorism%20flourish%20where,and%20the%20COVID%2D19%20pandemic.>, 2022. [Accedido 07-julio-2022].
- [25] HATTIE, J., AND TIMPERLEY, H. The power of feedback. *Review of educational research* 77, 1 (2007), 81.
- [26] HATTIE, J., AND TIMPERLEY, H. The power of feedback. *Review of educational research* 77, 1 (2007), 82.
- [27] HIRAN, K. K., JAIN, R. K., LAKHWANI, K., AND DOSHI, R. *Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples (English Edition)*. BPB Publications, 2021.
- [28] HUGHES, B., AND COTTERELL, M. *Software project management*. Mc Graw Hill Education, 2009.
- [29] IBM CLOUD EDUCATION. Data mining. <https://www.ibm.com/cloud/learn/data-mining>, 2021. [Accedido 13-julio-2022].
- [30] JUPYTER. Jupyter notebook. <https://jupyter.org/>, 2022. [Accedido 26-junio-2022].
- [31] KASS, G. V. An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29, 2 (1980), 119–127.
- [32] KHAING, K. T. Enhanced features ranking and selection using recursive feature elimination (rfe) and k-nearest neighbor algorithms in support vector machine for intrusion detection system. *International Journal of Network and Mobile Technologies* 1, 1 (2010), 1832–6758.
- [33] KLUGER, A. N., AND DENISI, A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin* 119, 2 (1996), 254.
- [34] LEVILLAIN, F., ORERO, J. O., RIFQI, M., AND BOUCHON-MEUNIER, B. Characterizing player’s experience from physiological signals using fuzzy decision trees. In *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games* (2010), IEEE, pp. 75–82.
- [35] LIN, C. F., YEH, Y.-C., HUNG, Y. H., AND CHANG, R. I. Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education* 68 (2013), 199–210.

- [36] LOCOMOTION. Locomotion h2020. <https://www.locomotion-h2020.eu/>. [Accedido 23-abril-2022].
- [37] MASON, B. J., AND BRUNING, R. Providing feedback in computer-based instruction: What the research tells us. *Retrieved February 15 (2001)*, 2007.
- [38] MEDEAS. Medeas. modeling the renewable energy transition in europe. <https://www.medeas.eu/>, 2022. [Accedido 28-julio-2022].
- [39] MINISTERIO PARA LA TRANSICIÓN ECOLÓGICA Y EL RETO DEMOGRÁFICO. My green energy planer. <https://www.miteco.gob.es/es/ceneam/recursos/pag-web/green-energy-planet.aspx/>, 2016. [Accedido 08-julio-2022].
- [40] PALLETS PROJECTS. Flask. <https://flask.palletsprojects.com/en/2.1.x/>. [Accedido 30-mayo-2022].
- [41] PEÑA-AYALA, A. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications* 41, 4 (2014), 1432–1462.
- [42] PRENSKY, M. Digital natives, digital immigrants. *Gifted*, 135 (2005), 29–31.
- [43] PYTHON SOFTWARE FOUNDATION. Python3 wrapper for the weka machine learning workbench. <https://pypi.org/project/python-weka-wrapper3/>. [Accedido 30-mayo-2022].
- [44] PYTHON SOFTWARE FOUNDATION. Welcome to python.org. <https://www.python.org/>. [Accedido 30-mayo-2022].
- [45] QUINLAN, J. R. Induction of decision trees. *Machine learning* 1, 1 (1986), 81–106.
- [46] QUINLAN, J. R. Improved use of continuous attributes in c4. 5. *Journal of artificial intelligence research* 4 (1996), 77–90.
- [47] ROGERS, R. The motivational pull of video game feedback, rules, and social interaction: Another self-determination theory approach. *Computers in Human Behavior* 73 (2017), 446–450.
- [48] SANTOS, A. M. Uso de técnicas de data mining sobre series temporales obtenidas por simulación y aplicación de resultados en videojuego crossroads. <https://uvadoc.uva.es/handle/10324/50430>, 2021. [Accedido 13-junio-2022].
- [49] SAYAD, S. Árbol de decisión - clasificación. https://www.saedsayad.com/decision_tree.htm, 2022. [Accedido 04-julio-2022].
- [50] SCHETTINI, R., GASPARINI, F., CORCHS, S., MARINI, F., CAPRA, A., AND CASTORINA, A. Contrast image correction method. *Journal of Electronic Imaging* 19, 2 (2010), 023005.

- [51] SCIKIT-LEARN DEVELOPERS. Imputation of missing values. <https://scikit-learn.org/stable/modules/impute.html#imputation-of-missing-values>, 2022. [Accedido 28-julio-2022].
- [52] ŠVÁBENSKÝ, V., VYKOPAL, J., AND ČELEDA, P. Toward an automated feedback system in educational cybersecurity games. *Rigorous thesis. Masaryk University, Faculty of Informatics, Brno* (2019).
- [53] TAKEYAS, B. L. Algoritmo c4.5. [http://www.itnuevolaredo.edu.mx/takeyas/apuntes/Inteligencia%20Artificial/Apuntes/tareas_alumnos/C4.5/C4.5\(2005-II-B\).pdf](http://www.itnuevolaredo.edu.mx/takeyas/apuntes/Inteligencia%20Artificial/Apuntes/tareas_alumnos/C4.5/C4.5(2005-II-B).pdf), 2005. [Accedido 04-julio-2022].
- [54] TICHY, W. F., HABERMANN, N., AND PRECHELT, L. Summary of the dagstuhl workshop on future directions in software engineering: February 17–21, 1992, schloß dagstuhl. *ACM SIGSOFT Software Engineering Notes* 18, 1 (1993), 35–48.
- [55] TUNG.M.PHUNG. Information gain, gain ratio and gini index. <https://tungphung.com/information-gain-gain-ratio-and-gini-index/>, 2020. [Accedido 07-junio-2022].
- [56] UNIVERSIDAD DE SANTO TOMÁS COLOMBIA. La cibernética, el arte del timonel. https://web.archive.org/web/20131104091944/http://pwp.etb.net.co/ferqui/lec_ciber_arte_timonel.htm, 2013. [Accedido 26-junio-2022].
- [57] VALÉRIE MASSON-DELMOTTE, HANS-OTTO PÖRTNER, JIM SKEA, PANMAO ZHAI, DEBRA ROBERTS AND PRIYADARSHI R. SHUKLA. Global warming of 1.5°C. an ipcc special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, 2019.
- [58] VENTANA SYSTEMS. Vensim software. <https://vensim.com/vensim-software/>. [Accedido 23-abril-2022].
- [59] WEKA SOURCEFORGE.IO. Gainratioattributeeval class. <https://weka.sourceforge.io/doc.dev/weka/attributeSelection/GainRatioAttributeEval.html>. [Accedido 06-junio-2022].
- [60] WEKA SOURCEFORGE.IO. Infogainattributeeval class. <https://weka.sourceforge.io/doc.dev/weka/attributeSelection/InfoGainAttributeEval.html>. [Accedido 06-junio-2022].
- [61] WITTEN, I. H., AND FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, 2000.
- [62] WORLDOMETER. Población mundial actual. <https://www.worldometers.info/es/poblacion-mundial/>, 2022. [Accedido 17-julio-2022].