



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO®

INSTITUTO TECNOLÓGICO DE CULIACÁN



RECONOCIMIENTO AUTOMÁTICO DE ATRIBUTOS DE PERSONALIDAD PARA ORIENTACIÓN VOCACIONAL

TESIS

PRESENTADA ANTE EL DEPARTAMENTO ACADÉMICO DE ESTUDIOS DE POSGRADO
DEL INSTITUTO TECNOLÓGICO DE CULIACÁN EN CUMPLIMIENTO PARCIAL DE LOS
REQUISITOS PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

POR:

ABEL ROBLES MONTOYA

INGENIERO EN SISTEMAS COMPUTACIONALES

DIRECTOR DE TESIS:

DRA. MARÍA LUCÍA BARRÓN ESTRADA

CULIACÁN, SINALOA

AGOSTO 2022



Instituto Tecnológico de Culiacán

Culiacán, Sin., 5 de Agosto del 2022

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN
OFICIO: DEPI/24260/08/2022

ASUNTO: **Autorización Impresión**

ABEL ROBLES MONTOYA
ESTUDIANTE DE LA MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
PRESENTE.

Por medio de la presente y en virtud de que ha completado los requisitos para el examen de grado de **Maestro en Ciencias de la Computación**, se concede autorización para la impresión de la tesis titulada: **“RECONOCIMIENTO AUTOMÁTICO DE ATRIBUTOS DE PERSONALIDAD PARA ORIENTACIÓN VOCACIONAL”** bajo la dirección del(a) **DRA. MARÍA LUCÍA BARRÓN ESTRADA.**

Sin otro particular reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

M.C. MARÍA ARACELY MARTÍNEZ AMAYA
JEFE(A) DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN



INSTITUTO TECNOLÓGICO DE CULIACÁN

DEPARTAMENTO DE DIVISIÓN DE
ESTUDIOS DE POSGRADO E INVESTIGACIÓN

C.c.p. archivo

MAMA/lucy *





"RECONOCIMIENTO AUTOMÁTICO DE ATRIBUTOS DE PERSONALIDAD PARA ORIENTACIÓN VOCACIONAL"

Tesis presentada por el(a):

ABEL ROBLES MONTOYA

Aprobada en contenido y estilo por:

DRA. MARÍA LUCÍA BARRÓN ESTRADA
Director de Tesis

DR. RAMÓN ZAFARAIN CABADA
Secretario

M.C. ROSALÍO ZAFARAIN CABADA
Vocal -1

DR. HÉCTOR RODRÍGUEZ RANGEL
Vocal -2

M.C. María Aracely Martínez Amaya
Jefe(a) de la División de Estudios de
Posgrado e Investigación



Dedicatoria

Dedico este trabajo de tesis a mis padres Claudia Yesmid Montoya M. y Abel Robles E. por haberme dado su apoyo incondicional en mis estudios durante todos estos años, por motivarme a cumplir todos mis objetivos y enfrentar nuevos retos. A mis hermanos Jesús Eduardo Robles M. y Sherley Castro M. por estar siempre presentes en mi vida. A toda mi familia, a quienes amo y han sido mi apoyo en todo momento y han creído en mí.

Agradecimientos

Agradezco a CONACYT por otorgarme el apoyo económico para que pudiera realizar mis estudios.

Al Tecnológico Nacional de México Campus Culiacán por permitirme estudiar un posgrado de calidad y proporcionarme el espacio para cursar la maestría y realizar mi trabajo de investigación.

A mi asesora de tesis la Dra. María Lucía Barrón Estrada por sus recomendaciones y apoyo para llevar a cabo este trabajo de tesis. También, al resto de mis profesores Dr. Héctor Rodríguez, Dr. Ricardo Quintero, Dr. Víctor González y Dr. Ramón Zatarain por compartir sus conocimientos y su experiencia.

Al Maestro en ciencias Héctor M. Cárdenas López por su apoyo en la realización de mi tesis.

Al departamento del posgrado por todo su apoyo en el proceso administrativo.

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Abel Robles Montoya

Culiacán, Sinaloa, México, 2022

Resumen

En México solo 6 de cada 10 alumnos de educación superior logran terminar sus estudios; aunque las causas más comunes son los problemas socioeconómicos y personales; también la insatisfacción académica y el bajo rendimiento escolar impactan en gran medida, esto sugiere una mala elección de carrera. La personalidad del estudiante se ha identificado como un factor clave que influye en la elección de carrera; ya que los estudiantes eligen entornos académicos acordes con su personalidad; aunque muchas veces, los jóvenes no siempre saben qué opciones son las más adecuadas para sus características en particular.

En este trabajo de investigación se presenta el desarrollo de un sistema de predicción de carreras profesionales utilizando como entrada un video del usuario; con el objetivo de ayudar a los estudiantes a tomar una mejor decisión. Para el desarrollo de este sistema fue necesario recopilar información de diferentes tipos como cuestionarios de personalidad e intereses profesionales, videos, así como información relativa a la formación académica y profesional de los participantes; con los que se generó un conjunto de datos (*dataset*) de personalidad e intereses profesionales. Este *dataset* se utilizó para entrenar modelos de aprendizaje máquina con el fin de predecir intereses profesionales a partir de video.

El sistema PersonApp se utilizó como base para el desarrollo del sistema para recopilar información, adecuándolo a las necesidades actuales, con éste se logró generar una base de datos de 128 cuestionarios de personalidad e intereses profesionales y 68 videos; los valores de cada atributo de los cuestionarios de personalidad e intereses muestran una distribución normal. En cuanto a los resultados de los modelos de predicción; en el caso del modelo de personalidad es de un error promedio de 0.18, mientras que para el modelo de intereses profesionales es de 0.16.

A partir de los datos analizados en los resultados, se puede ver un problema en los reconocedores al tratar de realizar las predicciones; problema originado debido a la distribución de los datos, la cual provoca que los modelos aprendan a minimizar el error arrojando valores cercanos a la media, lugar donde se encuentra el mayor volumen de datos. Estos problemas hacen al sistema de recomendación de carreras inexacto, por lo que se requiere seguir investigando para mejorar los modelos de aprendizaje.

Palabras clave

Abandono escolar

Aprendizaje máquina

Aprendizaje profundo

Conjunto de datos (*Dataset*)

HEXACO

Intereses profesionales

OCEAN

Orientación vocacional

Personalidad

Plataforma web

Pruebas estandarizadas de intereses

Pruebas estandarizadas de personalidad

Reconocimiento automático de intereses

Reconocimiento automático de personalidad

RIASEC

Sistema de predicción de intereses profesionales

Índice general

1. Introducción.....	1
1.1. Antecedentes	3
1.2. Descripción del problema	3
1.3. Justificación	4
1.4. Objetivos	5
1.4.1. Objetivo general	5
1.4.2. Objetivos específicos	5
1.5. Hipótesis	6
1.6. Organización de la tesis	6
2. Marco teórico.....	8
2.1. Abandono escolar	8
2.2. Orientación vocacional	9
2.2.1. Evaluación en Orientación Vocacional.....	9
2.2.2. Evaluación de los intereses	10
2.2.3. Modelo RIASEC.....	10
2.3. Evaluación de la personalidad	10
2.3.1. Modelo de los cinco grandes (Big-Five).....	11
2.3.2. Modelo HEXACO	12
2.4. Inteligencia artificial	12
2.4.1. Aprendizaje máquina.....	13
2.4.1.1. Aprendizaje supervisado	13
2.4.1.2. Aprendizaje no supervisado	13
2.4.1.3. Aprendizaje semi-supervisado	14
2.4.1.4. Aprendizaje por refuerzo	15
2.4.2. Aprendizaje profundo.....	16
3. Estado del arte	18
3.1. Pruebas estandarizadas de personalidad	18
3.2. Pruebas estandarizadas de intereses.....	19
3.3. Reconocimiento automático de personalidad	21
3.4. Reconocimiento automático de intereses profesionales	25

4.	Desarrollo del proyecto	26
4.1.	Propuesta de solución	26
4.2.	Metodología	26
4.3.	Análisis y selección de modelos de personalidad e intereses profesionales	27
4.4.	Implementación del sistema PersonApp 2.0	28
4.4.1.	PersonApp	28
4.4.2.	Requisitos funcionales PersonApp 2.0	28
4.4.3.	Requisitos de calidad	29
4.4.4.	Restricciones	30
4.4.5.	Actores	30
4.4.6.	Casos de uso	30
4.4.7.	Diagrama de contexto	32
4.4.8.	Arquitectura	33
4.4.9.	Implementación	33
4.5.	Recolección de datos y generación del dataset	38
4.6.	Diseño e implementación de modelos de predicción	39
4.6.1.	Modelo de predicción de personalidad	39
4.6.2.	Modelo de predicción de intereses profesionales	40
4.7.	Implementación del sistema de recomendaciones API-Interests	41
4.7.1.	Requisitos funcionales de API-Interests	41
4.7.2.	Requisitos de calidad	42
4.7.3.	Restricciones	42
4.7.4.	Diagrama de contexto	42
4.7.5.	Componentes del sistema	43
4.8.	Implementación del sistema de recomendación de carreras profesionales	45
4.8.1.	Requisitos funcionales	45
4.8.2.	Requisitos de calidad	46
4.8.3.	Restricciones	46
4.8.4.	Casos de uso	47
4.8.5.	Diagrama de contexto	47
4.8.6.	Arquitectura	48
4.8.7.	Implementación	49
4.9.	Lenguajes y herramientas de desarrollo	54

5. Resultados	56
5.1. Dataset de personalidad e intereses profesionales	56
5.2. Reconocedor de intereses profesionales	66
5.3. Reconocedor de personalidad	66
6. Conclusiones y trabajo futuro	68
6.1. Conclusiones	68
6.2. Aportaciones y limitaciones	69
6.3. Trabajo futuro	70
Referencias	71
Anexos	1
A. Tabla de correlaciones entre atributos	1
B. Casos de uso para el sistema PersonApp 2.0	2

Índice de figuras

Figura 1. Modelo triangular de consejo de orientación de estudios (Celdrán, 1990).	9
Figura 2. Ramas de la inteligencia artificial.....	12
Figura 3. Modelo estándar del aprendizaje por refuerzo (Patterson & Gibson, 2017).....	15
Figura 4. Arquitectura de una típica red neuronal artificial.	16
Figura 5. Pantalla de resultados de prueba de intereses.	21
Figura 6. Metodología de desarrollo del proyecto.	27
Figura 7. Diagrama de casos de uso.....	31
Figura 8. Diagrama de contexto de PersonApp 2.0.....	32
Figura 9. Vista lógica de la plataforma PersonApp 2.0.....	33
Figura 10. Cuestionario de personalidad.....	34
Figura 11. Cuestionario de intereses profesionales.	34
Figura 12. Aviso de privacidad de la plataforma.	35
Figura 13. Cuestionario para participantes que sean estudiantes.	36
Figura 14. Cuestionario para participantes graduados.	36
Figura 15. Menú principal de PersonApp 2.0.	36
Figura 16. Grabación de experimentos de video.	37
Figura 17. Mensaje de experimento completado.....	37
Figura 18. Correo de verificación.	38
Figura 19. Función para calcular el error de predicciones.	40
Figura 20. Función para calcular el porcentaje de predicciones dentro de un rango.	41
Figura 21. Diagrama de contexto de API-Interests.	43
Figura 22. Diagrama de componentes.....	44
Figura 23. Envío de petición a la API.	45
Figura 24. Diagrama de contexto del sistema CareerHelp.....	48
Figura 25. Arquitectura del sistema CareerHelp.	49
Figura 26. Menú principal de CareerHelp con opción de prueba en video.....	50
Figura 27. Menú principal de CareerHelp con opción de prueba estandarizada.....	50
Figura 28. Pantalla de espera de sugerencias de carreras.	51
Figura 29. Gráfica de valores RIASEC.....	52
Figura 30. Descripciones de cada atributo RIASEC.....	52
Figura 31. Lista de carreras recomendadas.	53
Figura 32. Descarga del reporte como archivo PDF.	54
Figura 33. Número de registros por área de conocimiento.	57
Figura 34. Cantidad de registros por sexo.....	57
Figura 35. Histograma del atributo escrupulosidad.....	59
Figura 36. Histograma del atributo apertura.	60
Figura 37. Histograma del atributo emocionalidad.	60
Figura 38. Histograma del atributo neuroticismo.....	61
Figura 39. Histograma del atributo amabilidad.....	61
Figura 40. Histograma del atributo honestidad.	62
Figura 41. Histograma del atributo extraversión.....	62
Figura 42. Histograma del atributo investigativo.....	63

Figura 43. Histograma del atributo social.	63
Figura 44. Histograma del atributo artístico.....	64
Figura 45. Histograma del atributo empresarial.....	64
Figura 46. Histograma del atributo convencional.	65
Figura 47. Histograma del atributo realista.....	65

Índice de tablas

Tabla 1. Requisitos funcionales para PersonApp 2.0.	29
Tabla 2. Requisitos de calidad para PersonApp 2.0.	29
Tabla 3. Actores del sistema de PersonApp 2.0.	30
Tabla 4. Casos de uso de PersonApp 2.0.	31
Tabla 5. Requisitos funcionales para API-Interests.	41
Tabla 6. Requisitos de calidad para API-Interests.	42
Tabla 7. Requisitos funcionales para CareerHelp.	46
Tabla 8. Requisitos de calidad para CareerHelp.	46
Tabla 9. Tabla de casos de uso de CareerHelp con su descripción.	47
Tabla 10. Estadística descriptiva por atributo.	58
Tabla 11. Porcentaje de datos por desviación estándar.	59
Tabla 12. Métricas de error por atributo RIASEC.	66
Tabla 13. Error absoluto medio por atributo OCEAN.	67

Capítulo 1

1. Introducción

El concepto de abandono escolar se usa para referirse a aquellos estudiantes que no finalizan sus estudios, ya sea por motivos personales, por reprobación o por vencimiento de la matrícula de sus materias (Linares & Hernández, 2017). Aunque en la Secretaría de Educación Pública (SEP) lo usan para referirse a aquellos alumnos que abandonaron los estudios sin tomar en cuenta los que reprobaban, se cambian de carrera, etc. En este trabajo se hará referencia a la primera definición siempre que se hable de abandono escolar.

De acuerdo con información de la SEP, en México solo 6 de cada 10 alumnos de educación superior (licenciatura universitaria y tecnológica) logran terminar sus estudios (2019-2020) (Secretaría de Educación Pública, 2020) . Para el 40% de los estudiantes que no terminan sus estudios universitarios se han identificado como las causas más comunes los problemas socioeconómicos y personales, pero también la insatisfacción académica y el bajo rendimiento escolar impactan en gran medida (Reyes, 2013), lo que sugiere una mala elección de carrera.

La personalidad del estudiante se ha identificado como un factor que influye en la elección de carrera. Según la teoría de Holland los estudiantes eligen entornos académicos acordes con su personalidad (Hansen, 2019), sin embargo, los jóvenes no siempre saben que opciones son las adecuadas para sus características en particular. Esto puede provocar diversas situaciones, por ejemplo: que no logre un rendimiento académico adecuado en esa carrera o bien que pierda el interés, por lo tanto, la orientación vocacional a través de herramientas como exámenes psicométricos pueden ayudar a los estudiantes a encontrar una carrera acorde con su personalidad y así disminuir el abandono escolar.

El análisis de la personalidad ha sido un tema ampliamente estudiado por múltiples disciplinas como la psicología, filosofía, y algunos otros (Sinisterra et al, 2009). Recientemente, la comunidad de visión artificial ha mostrado mucho interés en reconocer la

personalidad partiendo de información corporal de un individuo expresada a través del rostro, posturas, o comportamientos (Jacques et al., 2022). El análisis de la personalidad también puede ayudar en la sociedad en muchas áreas distintas, por ejemplo, en educación puede servir para identificar la estrategia correcta para enseñar a un alumno; al contratar personal se puede filtrar a las personas que tengan una personalidad afín al puesto requerido y, en orientación educativa se puede utilizar para ayudar a los jóvenes a decidir que carrera estudiar con el fin de evitar la deserción escolar.

La orientación vocacional es un conjunto de prácticas orientadas a la problemática vocacional que tiene el objetivo de brindar las herramientas necesarias para que cada sujeto tome la mejor elección a su alcance (Galilea, 2001). Actualmente la orientación vocacional se apoya de técnicas e instrumentos como entrevistas, cuestionarios o pruebas psicométricas, esto con el fin de ayudar a las personas a conocer más de sí mismas para elegir una profesión que les agrade y al mismo tiempo logren desempeñarse satisfactoriamente (Ramos & González, 2020). Algunos instrumentos vocacionales propuestos por muchos autores son la escala de Intereses Vocacionales de Kuder, el Test Explora y el Cuestionario de Áreas de Intereses Básicos Académico Profesionales; cada uno utiliza una metodología diferente para lograr identificar el área o profesión que mejor se adapta a cada persona (Alfaro-Barquero & Chinchilla-Brenes, 2019).

Dada la problemática de abandono escolar y considerando que la insatisfacción académica y el bajo rendimiento escolar están relacionados con una elección de carrera equivocada, en este trabajo de investigación se propone utilizar un sistema software que a través del reconocimiento de personalidad usando videos, sea capaz de predecir aquellas carreras profesionales más afines a cada estudiante. Para realizar este proceso se requieren dos modelos de inteligencia artificial para realizar las predicciones, el primero entrenado con videos grabados por los usuarios y cuestionarios de personalidad, de tal forma que el modelo reciba como entrada los videos y genere cómo salida los atributos de personalidad de ese usuario; el segundo modelo recibe los atributos de personalidad y responde con los intereses vocacionales, de esta forma al unir ambos modelos se logra predecir los intereses vocacionales a partir de los videos y se elimina la necesidad de que el usuario responda largos cuestionarios usados para orientación vocacional actualmente.

El principal problema que se presenta para lograr esto es que se requiere de una amplia base de datos de videos de usuarios, cuestionarios de personalidad, e intereses vocacionales, la cual contenga suficiente cantidad de registros y calidad de los datos para lograr que los modelos puedan aprender. Este trabajo de investigación presenta una propuesta para recolectar los datos necesarios para conformar el dataset requerido para los modelos de aprendizaje.

1.1. Antecedentes

Para la realización de este trabajo se utilizó como base la plataforma PersonApp desarrollada por Bátiz Beltrán (Bátiz-Beltrán, 2021) con el fin de crear un dataset de personalidad, utilizando cuestionarios de personalidad estandarizados y experimentos de video, para con este *dataset* entrenar modelos de aprendizaje máquina. El cuestionario de personalidad que aplica PersonApp es el IPIP-50. Para la grabación de los videos, el sistema describe tres experimentos donde el usuario debe hablar de un tema diferente para cada uno, y se graban 4 videos cortos de 15 segundos por experimento. Este sistema se desarrolló usando un modelo arquitectónico por capas sobre una arquitectura cliente-servidor y está diseñado para funcionar en navegadores web, siendo soportado por Chrome principalmente.

1.2. Descripción del problema

En la actualidad en las escuelas de nivel medio superior se incluyen en el currículo algunas materias relacionadas con la orientación vocacional, además los estudiantes participan en ferias donde universidades e instituciones de educación superior presentan sus ofertas educativas a los jóvenes que están próximos a graduarse de bachillerato. Estas medidas se implementan para ayudar a los estudiantes a tomar una mejor elección, sin embargo, el abandono escolar a causa de la incompatibilidad de los estudiantes y la carrera que eligen sigue siendo un problema en México como se observa en las cifras mostradas por la SEP (Secretaría de Educación Pública, 2020).

La personalidad de los estudiantes es un factor importante que se debe considerar para elegir una profesión, tanto como las destrezas y habilidades con las que cuenta. El principal

problema es que en los programas de orientación educativa que se implementan en las escuelas no se lleva un acompañamiento personalizado a cada estudiante ignorando muchas de sus características particulares, con esto se provoca desinformación y muchas dudas por parte de los estudiantes que aumentan sus posibilidades de tomar una mala decisión.

Por otra parte, existen cuestionarios de personalidad e intereses (Morales & Monzón 2014) los cuales son las herramientas más comunes que se encuentran en internet para recomendar carreras basándose en las respuestas de los usuarios a planteamientos de conducta, emociones o escenarios de trabajos, los problemas que presentan estas herramientas son: primero, algunas de ellas constan de una gran cantidad de preguntas, lo que implica dedicar hasta más de media hora a contestarlas, lo cual es mucho tiempo, esto puede hacer que algunas personas se les dificulte mantener la atención y sus respuestas no sean las más adecuadas; segundo, existe el riesgo que las personas no sean completamente honestas al contestar una pregunta, ya que existen aspectos propios percibidos como negativos que incluso se tiende a negarlos ante nosotros mismos, afectando así los resultados obtenidos en las pruebas.

1.3. Justificación

A pesar de los recursos que se destinan a la orientación educativa; tales como materias que se imparten en nivel medio superior o eventos de vinculación con universidades; el abandono escolar sigue siendo problema importante en México. El conocimiento de la personalidad e intereses profesionales son fundamentales para ayudar a los estudiantes a tomar una decisión informada. Las herramientas actuales que consideran la personalidad e intereses profesionales son muy útiles pero, presentan algunos inconvenientes, es por esto que una plataforma capaz de predecir y recomendar carreras profesionales a partir de video será de gran ayuda para combatir este problema, ya que permitirá obtener las recomendaciones de una manera más rápida, sin los problemas de la autopercepción que se presentan en las pruebas estandarizadas de orientación vocacional y será de fácil acceso para cualquier estudiante.

1.4. Objetivos

En esta sección se presenta el objetivo general y los objetivos específicos que se definieron para el desarrollo de este trabajo de investigación.

1.4.1. Objetivo general

Desarrollar un sistema inteligente que a través del análisis de video reconozca los atributos de personalidad de los estudiantes, utilizando para estas predicciones modelos de aprendizaje máquina, entrenados con un conjunto de datos de personalidad, intereses y experimentos de personalidad en video con la finalidad de recomendar una carrera profesional, similar a como lo hace una prueba estandarizada de orientación vocacional.

1.4.2. Objetivos específicos

Para llevar a cabo el objetivo general de esta investigación, se definieron 7 objetivos específicos que se listan a continuación:

- Seleccionar los modelos de personalidad e intereses profesionales aptos para aplicarse en conjunto y que sirvan para entrenar algoritmos de aprendizaje máquina.
- Implementar adaptaciones y nuevos módulos a la plataforma PersonApp para la aplicación de cuestionarios de personalidad e intereses profesionales.
- Recolectar datos de usuarios que respondan cuestionarios de personalidad, intereses profesionales y generen videos para conformar un dataset.
- Entrenar modelos de aprendizaje máquina que puedan predecir la personalidad a partir de video y los intereses profesionales a partir de la personalidad.
- Desarrollar un servicio en la nube que utilice los modelos para predecir los intereses a partir de los videos de los usuarios, interactuando con los clientes a través de una API de REST.
- Obtener una base de datos de carreras profesionales y los atributos recomendados para esas profesiones.
- Desarrollar en la plataforma PersonApp un apartado donde el usuario grabe un video y el sistema le muestre un listado con las carreras recomendadas.

1.5. Hipótesis

El modelo de predicción de personalidad mediante video podrá reconocer los atributos de personalidad con un margen de error absoluto medio menor o igual a 0.2 para cada atributo.

El modelo de predicción de intereses profesionales mediante personalidad podrá reconocer los atributos de intereses con un margen de error absoluto medio menor o igual a 0.1 para cada atributo.

1.6. Organización de la tesis

El documento de este trabajo de investigación está compuesto por 6 capítulos, en los cuales se documenta el desarrollo del proyecto, así como los resultados obtenidos. Los capítulos siguientes se describen a continuación:

En el capítulo 2 se presenta el marco teórico, en donde se describen los conocimientos y consideraciones teóricas en los que se fundamenta esta investigación. Se presentan temas relacionados con la evaluación de la personalidad y los intereses profesionales, así como algunos temas de las ciencias de la computación como el aprendizaje máquina.

El capítulo 3 muestra el estado del arte, donde se presentan algunos trabajos relacionados con la predicción de atributos de personalidad e intereses, también sobre las pruebas estandarizadas que se utilizan en el campo de la psicología para medir estos rasgos.

En el capítulo 4 se presenta a detalle el desarrollo del proyecto y la metodología diseñada para la implementación del sistema de recolección, las pruebas de personalidad e intereses profesionales, los modelos de predicción y el servicio de predicción de intereses profesionales.

El capítulo 5 muestra los resultados obtenidos al utilizar la plataforma de recolección de datos y los modelos de predicción, se presentan gráficos y tablas describiendo las cantidades y cualidades de los datos recolectados y también la precisión y error calculado en las predicciones de personalidad e intereses profesionales.

En el capítulo 6 se presentan las conclusiones de este trabajo, las aportaciones que se hicieron a esta área de investigación y el trabajo futuro.

Por último, se presentan las referencias utilizadas en este trabajo de investigación y los anexos incluidos en este trabajo.

En el anexo A se muestra una tabla de correlaciones entre los atributos de RIASEC y HEXACO.

En el anexo B se muestran los casos de uso completos para el sistema PersonApp 2.0.

Capítulo 2

2. Marco teórico

En este capítulo se describen los conceptos y fundamentos teóricos que respaldan la realización del trabajo de investigación. Se abordan los temas del abandono escolar, la orientación vocacional, los modelos para evaluar la personalidad y los intereses. También se revisan algunos conceptos sobre el campo de la inteligencia artificial relacionados con el aprendizaje máquina, y los tipos de aprendizaje.

2.1. Abandono escolar

El concepto de abandono escolar se refiere a aquellos alumnos que quedan fuera del sistema educativo por dejar de asistir a clase, y aunque ésta es una decisión personal de cada alumno, existen algunos factores o causas que más se repiten y exponen un problema más profundo.

Esto constituye un problema importante para los propios estudiantes, pero también para el sistema nacional de educación y la sociedad en general, ya que tiene una incidencia negativa en los procesos políticos, económicos, sociales y culturales del desarrollo del país. Algunas de las implicaciones más importantes son: a nivel social contribuye a la generación de pobreza aumentando el desempleo; en cuanto a las instituciones, disminuye los índices de eficiencia y calidad, así como los ingresos de las matrículas; por último, en el ámbito personal genera sentimientos de frustración y fracaso en los estudiantes, además de limitarles las ventajas de la educación en su desarrollo social (Linares & Hernández, 2017) .

En México las causas que más influyen para que los estudiantes dejen sus estudios son de tipo socioeconómico, de rendimiento académico o personales como enfermedades, maternidad, inseguridad, entre otras. Por otro lado, existen otros factores internos que causan a los alumnos incertidumbre sobre su horizonte profesional, llevándolos a tomar decisiones inadecuadas que muchas veces desembocan en deserción, estos factores tienen que ver con las aptitudes, valores, intereses y personalidad de los estudiantes, o incluso factores externos

como el entorno familiar o la desinformación sobre los perfiles ocupacionales y planes de estudio (Reyes, 2013; Linares & Hernández, 2017).

2.2. Orientación vocacional

En nuestra sociedad hoy en día es necesario conocerse a sí mismo, así como qué tipo de profesiones y oportunidades académicas son más adecuadas para cada persona según sus aptitudes y actitudes.

La orientación vocacional es el proceso que ayuda a los estudiantes a elegir una profesión, así como a prepararse y acceder a ésta posteriormente. El objetivo de este proceso es que los alumnos tengan conocimiento sobre si mismos y sus intereses, de las ofertas académicas, de los planes y programas de estudio y de las habilidades y competencias necesarias para llevar a cabo cada profesión, esto les permite a los estudiantes tomar decisiones considerando sus capacidades y gustos, para poder posicionarse correctamente en un contexto social-laboral (Vidal Ledo & Fernández Oliva, 2009).

2.2.1. Evaluación en Orientación Vocacional

La evaluación en la orientación vocacional tiene el propósito de identificar los intereses laborales de los alumnos, así como sus expectativas profesionales, su nivel de conocimientos y cómo se relaciona todo esto con la oferta laboral, tomando en cuenta también la motivación que puede tener la persona por una profesión en particular (Galilea, 2001).

Celdrán (Celdrán, 1990) propone un modelo triangular basado en el equilibrio entre las aptitudes, los intereses y el rendimiento escolar, este modelo refleja las probabilidades de éxito conforme se tenga una mayor consonancia entre las variables, ver Figura 1.

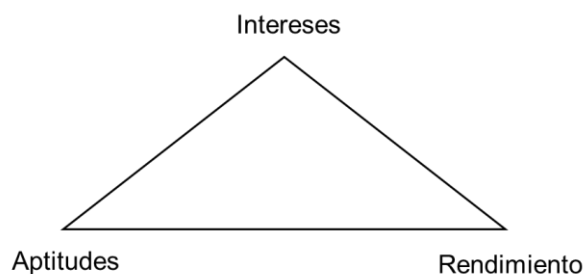


Figura 1. Modelo triangular de consejo de orientación de estudios (Celdrán, 1990).

En este modelo triangular, los intereses son el conjunto de metas deseadas por el estudiante, ya sea en el ámbito personal, académico, profesional o vocacional; las aptitudes son las capacidades que posee el alumno para ser entrenado o para adquirir una serie de habilidades y conocimientos necesarios; por otro lado, el rendimiento es el esfuerzo que realiza en sus actividades académicas medidas en las horas que le dedica al desarrollo de sus actividades académicas (Blanco & Frutos, 2001).

2.2.2. Evaluación de los intereses

El estudio sobre los intereses por parte de la psicología vocacional lleva más de 100 años y se ha utilizado en casos reales como en la primera guerra mundial; su importancia en la elección de trabajos fue reconocida por académicos desde principios del siglo pasado y rápidamente fue adoptada por la industria en Estados Unidos. En 1919, en el *Carnegie Institute of Technology*, Clarence S. Yoakum desarrolló el primer instrumento para medir los intereses de las personas, usando un enfoque de muestreo racional que pretende representar el dominio de los intereses (Hansen, 2019).

2.2.3. Modelo RIASEC

La teoría de Holland ha sido una de las más influyentes para el desarrollo de distintos cuestionarios para medir los intereses, esta teoría establece que es posible clasificar a cada persona en una de seis diferentes dimensiones: Realista, Investigativo, Artístico, Social, Emprendedor y Convencional, o bien en una combinación de estas, a este modelo se le conoce como RIASEC por las siglas de cada dimensión. Esta teoría, a través de los años, ha propiciado el desarrollo de muchos instrumentos como cuestionarios que sirven para medir los intereses de cada persona (Acosta-Amaya, 2018; Hansen, 2019).

2.3. Evaluación de la personalidad

La evaluación de la personalidad es un área de la psicología que se utiliza en muchas situaciones del mundo real, por ejemplo, para entender los síntomas de los pacientes o clientes de los psicólogos usando su personalidad, para descubrir las causas de un problema de aprendizaje, para seleccionar el mejor candidato para un puesto de trabajo, etc. En muchas

actividades de la vida, carrera profesional o incluso en el entorno social, la personalidad será evaluada de alguna manera, por esto es muy importante conocerla y saber cómo funciona.

Existen muchos instrumentos (Duane & Sydney, 2010) para evaluar la personalidad, los principales son los siguientes:

- Inventarios de auto-reporte u objetivos
- Técnicas proyectivas
- Entrevistas clínicas
- Procedimientos de evaluación de la conducta
- Procedimientos de muestreo de pensamientos y experiencias

Los inventarios de auto-reporte son técnicas donde las personas se describen a sí mismas contestando una serie de preguntas relativas a sus conductas y sentimientos, este tipo de instrumentos son los más confiables, ya que son menos subjetivos y cumplen con los principios de confiabilidad y validez (Duane & Sydney, 2010).

2.3.1. Modelo de los cinco grandes (Big-Five)

La teoría de los cinco grandes de la personalidad es una de las más reconocidas y estudiadas en la actualidad, tiene como objetivo buscar y explicar las estructuras básicas de la personalidad. Este modelo se forma de cinco dimensiones o rasgos mediante los cuales se intenta cuantificar las características de personalidad de los individuos (Martínez et al., 2011).

Los cinco factores de este modelo están descritos por el acrónimo OCEAN, cuyas siglas en inglés hacen referencia a las dimensiones: Apertura a la experiencia (*Openness to Experience*), Responsabilidad (*Conscientiousness*), Extroversión (*Extraversion*), Amabilidad (*Agreeableness*) y Neuroticismo (*Neuroticism*), cada factor es bipolar, esto indica que tiene su rasgo contrario y está compuesto de varios rasgos, cada uno de estos se utiliza para describir a las personas (Durupinar et al., 2011).

2.3.2. Modelo HEXACO

El modelo HEXACO es una propuesta alternativa a la estructura de los cinco grandes, este modelo alternativo ha acumulado evidencia considerable a su favor y consiste en seis dimensiones en lugar de cinco, tres de estas se corresponden muy estrechamente a las de los cinco grandes. El acrónimo HEXACO significa por sus siglas en inglés: Honestidad o Humildad (*Honesty-Humility*), Emocionalidad (*Emotionality*), Extroversión (*eXtraversion*), Amabilidad (*Agreeableness*), Responsabilidad (*Conscientiousness*), Apertura a la experiencia (*Openness to Experience*) (Ashton & Lee, 2007).

2.4. Inteligencia artificial

El término inteligencia se refiere a la capacidad de aprender y comprender, para solucionar problemas o tomar decisiones (Real Academia Española, 2022). La inteligencia artificial, después de varios eventos clave en su historia, se ha convertido en una ciencia o disciplina de estudio cuyo propósito es construir máquinas capaces de realizar acciones que requerirían inteligencia si fueran realizadas por humanos (Negnevitsky, 2005).

El campo de estudio de la inteligencia artificial es muy amplio y se compone de varias ramas, como se puede ver en la Figura 2 el aprendizaje máquina es una rama de la inteligencia artificial, que a su vez contiene dentro al aprendizaje profundo (Oracle México, 2022).

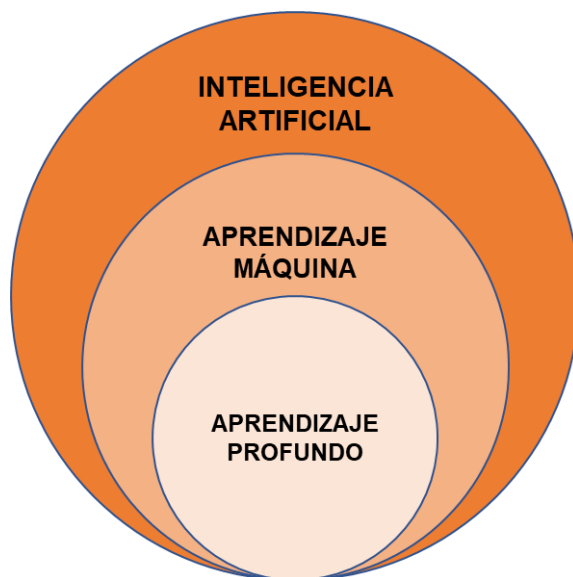


Figura 2. Ramas de la inteligencia artificial.

2.4.1. Aprendizaje máquina

El aprendizaje máquina es una rama de la inteligencia artificial, el cual consiste en un mecanismo adaptativo que permite a las computadoras aprender a partir de la experiencia, en lugar de ser explícitamente programadas, aprenden por analogía utilizando un conjunto de ejemplos que presentan diferentes casos o valores, dicho de una forma más técnica se trata de extraer información y representarla en algún tipo de modelo. La capacidad de aprender permite a los sistemas mejorar su eficiencia con el tiempo, es por esto por lo que el aprendizaje máquina es la base de los sistemas adaptativos. Las implementaciones más populares del aprendizaje máquina son los algoritmos genéticos y las redes neuronales artificiales (Patterson & Gibson, 2017; Negnevitsky, 2005).

2.4.1.1. Aprendizaje supervisado

Este tipo de aprendizaje requiere una gran cantidad ejemplos con sus respectivas entradas y salidas deseadas. El objetivo de estos algoritmos es aprender los patrones en la información y construir un conjunto de reglas para predecir futuras entradas (Swamynathan, 2017).

El aprendizaje supervisado se puede dividir en dos diferentes tipos de problemas: clasificación y regresión:

- **Regresión:** Se utiliza para entender la relación que existe entre las variables independientes y las variables dependientes, o, dicho de otro modo, entender las relaciones entre los valores de entrada y los de salida. Algunos de los algoritmos de regresión más populares son la regresión lineal, logística y polinomial.
- **Clasificación:** Se utiliza un algoritmo para asignar los datos de prueba en categorías específicas. Identifica las entradas del conjunto de datos y trata de predecir como estas entradas deberían ser etiquetadas o definidas según sus características y de acuerdo con el resto de los datos (IBM Cloud Education (B), 2020).

2.4.1.2. Aprendizaje no supervisado

El aprendizaje no supervisado se usa cuando no se cuenta con la etiqueta deseada para cada entrada, el objetivo es estudiar los patrones en el *dataset* para identificar similitudes que

pueden decir como agrupar cada información. Esta habilidad para descubrir similitudes y diferencias en la información hace que sean ideales para el análisis exploratorio de datos, la segmentación de clientes y el reconocimiento de imágenes (Swamynathan, 2017).

El aprendizaje no supervisado se utiliza principalmente para tres tareas, el agrupamiento, la asociación y la reducción de dimensionalidad:

- **Agrupamiento:** El agrupamiento es una técnica de minado de datos donde se agrupa los datos sin etiquetar según sus características en común, los algoritmos de agrupamiento se pueden clasificar en algunos tipos específicamente exclusivos, superpuestos, jerárquicos y probabilísticos.
- **Asociación:** La asociación es un método basado en reglas que sirve para buscar la relación entre las variables de un dataset. Un ejemplo muy común de este tipo de métodos es el análisis de las compras en una tienda, ya que permite conocer mejor las relaciones entre los diferentes productos.
- **Reducción de dimensionalidad:** La reducción de la dimensionalidad es una técnica que se utiliza cuando el número de características o dimensiones en un dataset es demasiado grande, lo que hace este método es reducir el número de entradas a un tamaño que sea más manejable, sin comprometer demasiado la integridad del *dataset* (IBM Cloud Education (C), 2020).

2.4.1.3. Aprendizaje semi-supervisado

El aprendizaje semi-supervisado es un paradigma de aprendizaje que estudia como las computadoras aprenden utilizando datos tanto etiquetados como sin etiquetar. Tradicionalmente, el aprendizaje ha sido estudiado desde el paradigma supervisado con todos los datos etiquetados y desde el paradigma no supervisado con todos los datos sin etiquetar. El objetivo del aprendizaje semi-supervisado es comprender como combinar los datos etiquetados y no etiquetados para cambiar el comportamiento del aprendizaje y crear así algoritmos que usen esta combinación de datos como una ventaja. Este tipo de aprendizaje es de gran interés en el aprendizaje máquina porque puede utilizar datos no etiquetados para mejorar las tareas del aprendizaje supervisado cuando hay pocos datos etiquetados (Goldberg, 2009).

2.4.1.4. Aprendizaje por refuerzo

El aprendizaje por refuerzo es un área del aprendizaje máquina que aborda la resolución de problemas enfrentando a un agente que aprende un comportamiento por prueba y error, interactuando con un entorno dinámico. El objetivo es entrenar a un agente para que aprenda una solución lo más cercana posible a la solución óptima. Se utilizan principalmente dos estrategias para resolver un problema utilizando el aprendizaje por refuerzo, la primera es buscar en el espacio de solución aquella que mejor se desarrolle en el entorno, este enfoque se utiliza en los algoritmos genéticos, la segunda es usar técnicas de estadística, métodos dinámicos de programación para estimar la ganancia al tomar ciertas opciones del espacio de solución.

En la Figura 3 se observa el modelo estándar del aprendizaje por refuerzo, para cada iteración el agente sigue una estrategia de prueba y error interactuando con el entorno, su comportamiento se va modificando según las recompensas que reciba del entorno, el agente observa el entorno y toma acciones dentro del mismo a través del actuador el cual puede cambiar el estado del entorno, en cada paso el agente recibe una entrada observacional reflejando el estado del entorno (Patterson & Gibson, 2017).

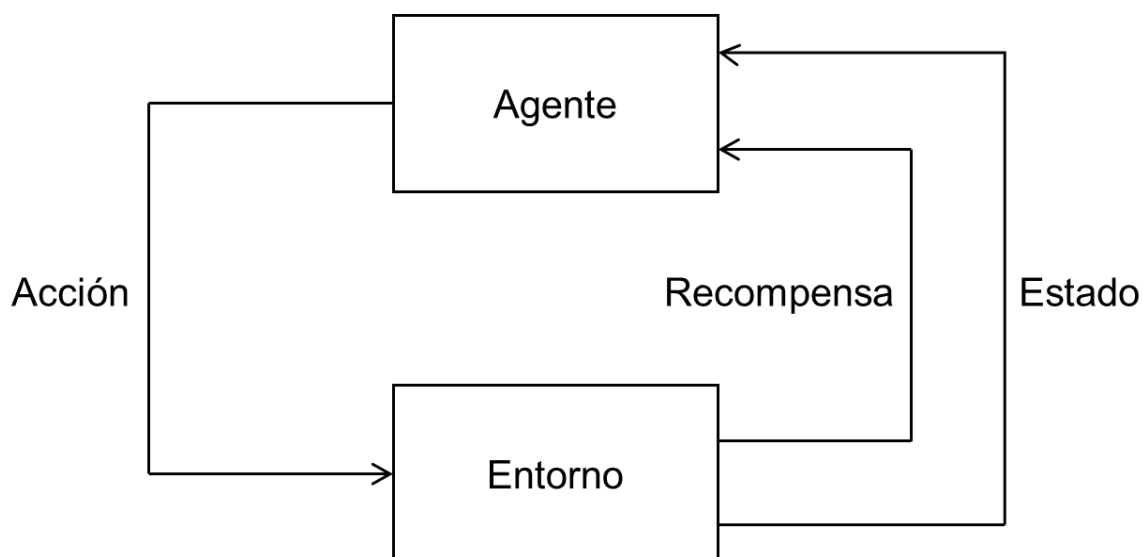


Figura 3. Modelo estándar del aprendizaje por refuerzo (Patterson & Gibson, 2017).

2.4.2. Aprendizaje profundo

El aprendizaje profundo ha sido difícil de definir para muchos, ya que ha ido cambiando de forma en la última década. Una definición muy concreta especifica que el aprendizaje profundo involucra redes neuronales con más de dos capas. Para entender esto es necesario conocer más a fondo las redes neuronales y su funcionamiento; las redes neuronales forman parte del aprendizaje máquina y pueden ser definidas como un modelo de razonamiento basado en el cerebro humano.

Una red neuronal artificial consiste en un número simple y altamente interconectado de procesadores, también llamados neuronas, las cuales son análogas a las neuronas del cerebro humano, estas neuronas están unidas por conexiones que cuentan con un peso, cada neurona recibe un número de señales de entrada, pero solo puede producir una señal de salida. Las redes neuronales aprenden a través de repetidos ajustes que se realizan a los pesos de cada conexión, el cual representa la importancia de esta neurona.

En la Figura 4 se puede ver la arquitectura típica de una red neuronal, al lado izquierdo se tienen los datos o señales de entrada, que son los valores que la red neuronal recibe para ser entrenada, estos pasan por la capa de entrada y se distribuyen a cada neurona de la capa oculta, por último, pasan por la capa de salida para ser procesadas las señales o valores de salida de la red (Patterson & Gibson, 2017; Negnevitsky, 2005).

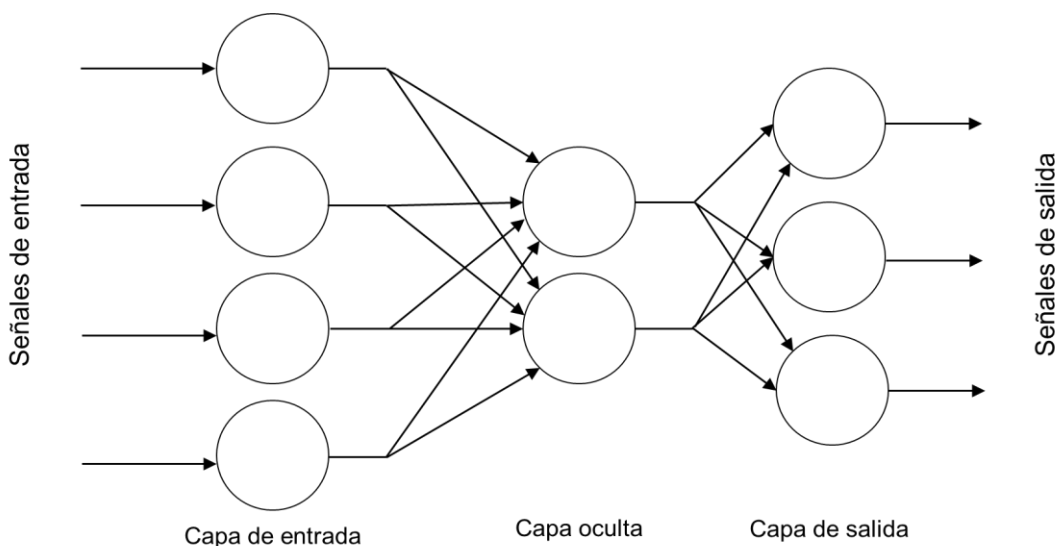


Figura 4. Arquitectura de una típica red neuronal artificial (Negnevitsky, 2005).

Este tipo de redes neuronales se diferencian de los modelos utilizados en los años ochenta debido a la evolución en ciertas facetas, a pesar de que siguen siendo conceptualmente los mismos, algunas características que han traído la evolución de estos modelos son:

- Mayor número de neuronas.
- Maneras más complejas de conectar las capas y neuronas.
- Un gran aumento en el poder de cómputo disponible para los entrenamientos.
- Extracción automática de características (Patterson & Gibson, 2017).

Uno de los tipos más populares y versátiles de redes neuronales son las redes neuronales convolucionales, estas redes proveen una forma más escalable para la clasificación de imágenes y las tareas de reconocimiento de objetos, aunque una de sus desventajas es que son computacionalmente muy demandantes, debido al gran número de parámetros que manejan, requiriendo procesadores gráficos para realizar los entrenamientos de las redes. Estas redes tienen tres tipos principales de capas:

- Capa de convolución.
- Capa de agrupación.
- Capa completamente conectada.

La capa convolucional es la primera capa de la red convolucional, si bien las capas convolucionales pueden ir seguidas de capas convolucionales adicionales o capas de agrupación, la capa completamente conectada es la capa final. Con cada capa, CNN aumenta su complejidad, identificando partes más grandes de la imagen. Las capas anteriores se centran en características simples, como colores y bordes. A medida que los datos de la imagen avanzan a través de las capas de CNN, comienza a reconocer características o formas más grandes del objeto hasta que finalmente determina el objeto deseado (IBM Cloud Education (A), 2020).

Capítulo 3

3. Estado del arte

En este capítulo se presentan algunas investigaciones relacionadas con este proyecto de tesis, estas investigaciones pertenecen a las áreas de la psicología y computación, específicamente se revisaron trabajos relacionados con la personalidad, los intereses profesionales, el aprendizaje máquina y el reconocimiento de personalidad e intereses mediante inteligencia artificial.

3.1. Pruebas estandarizadas de personalidad

En el artículo presentado por Cruz (Cruz, 2011) se establece la vigencia y aplicabilidad del cuestionario de personalidad Eysenck en sus diferentes versiones, para esto se realizó una revisión sistemática, encontrando 26 artículos que cumplieran con los criterios establecidos, se encontró un uso constante del cuestionario de personalidad, así como de identificación de rasgos de diferentes poblaciones. En el trabajo se concluye que el cuestionario de personalidad Eysenck se ha utilizado en casi todo el mundo de forma constante, demostrando un avance importante en las investigaciones que utilizan este instrumento.

En su trabajo Lee y Ashton (Lee & Ashton, 2018) presentan las propiedades psicométricas del instrumento HEXACO-100, se examinaron las respuestas de más de diez mil autoinformes realizados de manera virtual y casi tres mil informes de observadores, como resultado se observaron tres cosas: primero, cada uno de los seis factores del modelo están definidos por sus facetas consecuente, segundo, las escalas de los factores mostraron una baja correlación lo que los hace más independientes, donde solo la honestidad, humildad y simpatía tuvieron una correlación absoluta por debajo de .20, tercero, los factores y las escalas mostraron fuertes convergencias de las correlaciones. Como parte de las conclusiones, los autores recomiendan el uso del instrumento HEXACO-100 en entornos de investigación, siempre que se desee obtener mediciones de la personalidad.

En la investigación realizada por Gow y colaboradores (Gow et al., 2005) se examinó la estructura del cuestionario de personalidad IPIP-50, para esto utilizaron una muestra de novecientos adultos, para cada caso se comparó con el NEO-FFI (*NEO Five Factor Inventory*) y EPQ-R (*Eysenck Personality Questionnaire-Revised Short Form*), los resultados muestran que todos los factores están fuertemente correlacionados con los de NEO-FFI, donde las correlaciones obtenidas fueron de 0.85 para el caso de IPIP y 0.84 para EPQ-R. Como conclusión se dice que las escalas del cuestionario IPIP tienen una buena consistencia interna y se relacionan fuertemente con las escalas de dos de los cuestionarios principales.

(Johnson, 2014) presentó el desarrollo de una versión de 120 preguntas del cuestionario de personalidad IPIP-NEO, así como de pruebas realizadas con más de veinte mil registros, con esta muestra se realizaron las pruebas de fiabilidad y consistencia correspondientes. Como conclusión se indica que las propiedades psicométricas del instrumento IPIP-NEO 120-item son comparables favorablemente a la versión completa del cuestionario.

3.2. Pruebas estandarizadas de intereses

En el estudio realizado por (Hansen & Neuman, 1999) se examinó la validez de los criterios del Cuestionario de Intereses y Habilidades Campbell (CISS por su nombre en inglés *Campbell Interest and Skill Survey*), para predecir carreras universitarias comparándolo con el Inventario de Interés Fuerte (SII por su nombre en inglés *Strong Interest Inventory*). Se utilizó una muestra de 62 mujeres y 66 hombres, algunos datos recolectados incluyen los registros de puntaje de cada estudiante en la escala CISS y la escala de carrera SII, así como información sobre la carrera elegida; se observó una fuerte concordancia entre las carreras elegidas y el componente de interés en el perfil del estudiante, también una correlación de excelente a moderada entre las ocupaciones elegidas y el componente de preferencia de la escala ocupacional de la CISS para el 69.3% de las mujeres y el 75.8% de los hombres.

(Naylor & Kidd, 2000) presentan en su investigación un estudio sobre la validez predictiva del instrumento Inventario de Evaluación de Carreras (CAI por su nombre en inglés *Career Assessment Inventory*), Esta prueba se aplicó a 160 estudiantes de secundaria, el objetivo era predecir la escala del rasgo investigativo y comparar tres años después los valores obtenidos con las carreras elegidas por estas personas. Los resultados mostraron que la correlación entre

el puntaje en la escala de investigativo y la opción de carreras de ciencia y tecnología comparada con otras opciones fue de 0.46, lo que valida positivamente las capacidades predictivas del CAI.

En el trabajo realizado por Flores y colaboradores (Flores et al., 2006) se revisó la validez y consistencia del instrumento SII (Strong Interest Inventory) y SCI (Skills Confidence Inventory), los cuales están basados en la teoría de Holland para evaluar las seis dimensiones de intereses propuestas, para esto se usó una muestra de 487 estudiantes mexicanos y los resultados fueron comparados con muestras normativas de los correspondientes cuestionarios. El ajuste del modelo para ambas medidas fue deficiente en comparación con los resultados obtenidos con muestras de referencia, además los resultados apoyan la validación para las puntuaciones del código de Holland, se encontraron también diferencias de género en ambos modelos.

La relación que existe entre los individuos y su ambiente de trabajo ha sido investigado tanto de forma teórica como empírica en las últimas décadas, en el estudio realizado por Ehrhart y Makransky (Ehrhart & Makransky, 2007) se presentan dos antecedentes, las preferencias profesionales y personales, los autores plantean la hipótesis de que la preferencia ocupacional y personalidad basada en cinco factores predicen la idoneidad laboral, así como también características como la aptitud entre individuos y puestos de trabajo. Los resultados indican que las preferencias ocupacionales son mejores predictores de los trabajos ideales y ocupaciones que la personalidad.

Uno de los instrumentos más reconocidos para evaluar los intereses es el *O*NET Interest Profiler*, en (Rounds et al., 2010) se revisan sus características y la evidencia que lo apoya. O*NET es una herramienta de exploración de carreras diseñada para orientación profesional, planificación y exploración de carreras, este artefacto mide seis tipos de intereses ocupacionales: Realista (R), Investigativo (I), Artístico (A), social (S), Emprendedor (E) y Convencional (C), los cuales se le conoce como el modelo de RIASEC propuesto por Holland.

Esta prueba se realiza en línea desde la plataforma desarrollada por el departamento del trabajo de los Estados Unidos y es gratuita. Además de poder realizar la prueba esta

plataforma cuenta con una base de datos de profesiones las cuales están clasificadas según los intereses que vayan acorde. La prueba tiene dos versiones, la versión completa consta de 60 preguntas y la versión reducida *O*NET Mini-IP* de 30, al terminar cualquiera de las pruebas los usuarios pueden ver una gráfica con sus resultados, así como una descripción de cada rasgo, ver Figura 5.

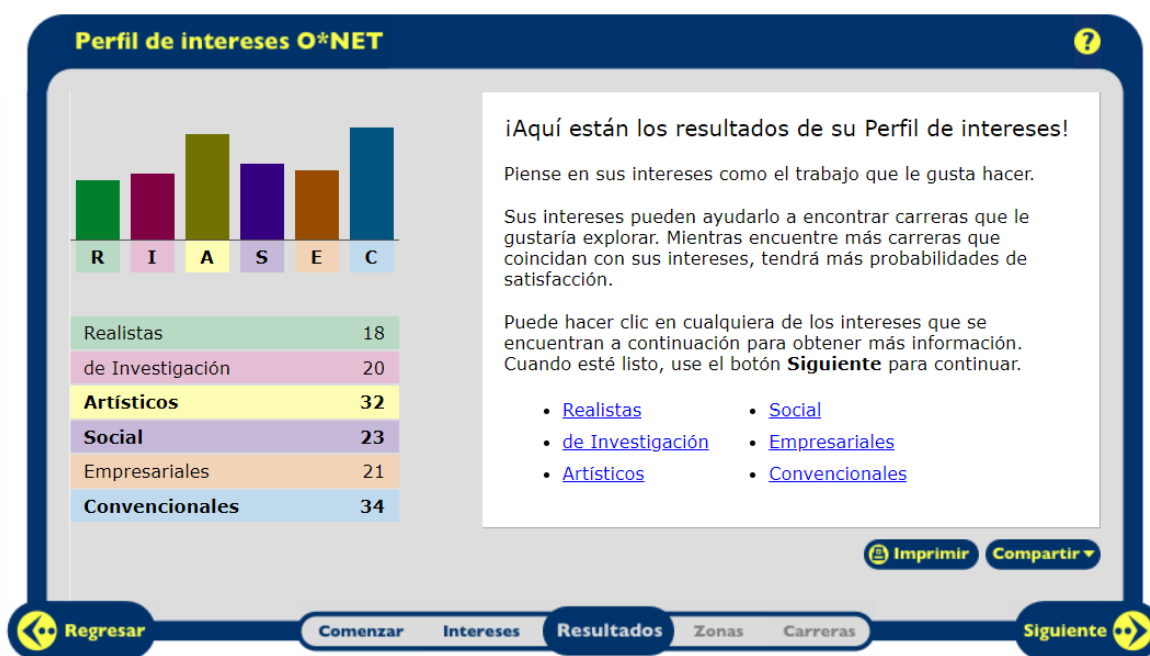


Figura 5. Pantalla de resultados de prueba de intereses.

3.3. Reconocimiento automático de personalidad

La evaluación de la personalidad es algo muy importante para evaluar si una persona es apta para desarrollar cierto trabajo, normalmente esta evaluación se realiza con base en el indicador de tipo Myers-Briggs, pero debido a que se considera demasiado rígido y dicotómico, los autores optaron por HEXACO para medir los rasgos de personalidad basados en datos de redes sociales como interacciones y comentarios. Se realizó un preprocesado de los datos eliminando palabras vacías o irrelevantes (*stopwords*) e interpretando emoticones, luego, extrajeron los rasgos usando TF-IDF, posteriormente se realiza una clasificación mediante un algoritmo de bosque aleatorio (*Random Forest*). Los resultados arrojan una precisión de hasta el 90%, lo que indica que la evaluación de la personalidad se puede

predecir de una manera adecuada, esto en función de los comentarios e interacciones que una persona realiza en redes sociales. Como conclusiones se utilizaron diccionarios y bibliotecas artificiales que logran tener un mayor grado de precisión que las bibliotecas literarias; también se encontró que el uso de algoritmos de optimización como *Particle Swarm* puede servir para encontrar fácilmente los valores óptimos para el modelo de *Random Forest* (Ramadhan et al., 2021).

La personalidad ha jugado un papel importante en la computación en los últimos años, es por esto por lo que el reconocimiento de la personalidad ha ganado mucho interés recientemente, en este trabajo (Aslan & Güdükbay, 2019) sugieren una nueva forma de reconocer los cinco grandes rasgos de personalidad a partir de video.

La personalidad y las emociones influyen en actividades de nuestro día a día como el habla, las expresiones faciales, los movimientos corporales y los factores de lenguaje y estos son afectados por elementos ambientales. Se desarrolló un sistema capaz de reconocer la personalidad basado en elementos como los rasgos faciales, factores ambientales y el audio. Se emplearon diferentes modelos de redes neuronales pre entrenadas como ResNet y VGGish para extraer características de alto nivel; primero entrenaron las subredes por separado y después se realizó una optimización de los modelos. Para el entrenamiento se usó el dataset ChaLearn First Impressions V2 con estas configuraciones y modelos se obtuvo la precisión media más alta comparada con el estado del arte. Como conclusiones se propusieron una nueva metodología para identificar rasgos de personalidad aparente de videos, mostrando como resultados que el enfoque multimodal es mejor que el enfoque unidireccional (Aslan & Güdükbay, 2019).

La apariencia puede afectar las interacciones sociales, lo que a su vez afecta el desarrollo de la personalidad, se han encontrado amplias investigaciones que demuestran que la morfología facial y las señales sociales brindan información sobre la personalidad y el comportamiento de las personas. (Xu et al., 2021) proponen un nuevo enfoque para predecir los rasgos de personalidad basado en los cinco grandes, utilizando imágenes faciales fijas, usando un dataset construido con 13, 347 imágenes faciales y características de personalidad. Luego se utilizó el aprendizaje profundo para entrenar una red neuronal. Los resultados muestran que se puede predecir la personalidad con más del 70 por ciento de precisión, las características

de las redes neuronales con aprendizaje profundo son mejores comparadas con los manuales; por último, se encontraron ciertas diferencias en las características personales de los estudiantes universitarios de diferentes antecedentes educativos (Xu et al., 2021).

El reconocimiento de la personalidad puede usarse para el aprendizaje en las escuelas, debido a la pandemia de COVID las escuelas se vieron forzadas a cambiar las aulas por plataformas de aprendizaje en línea. En este estudio los autores (Pakinee & Puritat, 2021) introducen el concepto de gamificación aplicado al E-Learning. Poniendo un enfoque en mejorar la participación de diferentes tipos de personalidad de los estudiantes universitarios; la gamificación se desarrolló implementando las ventajas y desventajas de cada juego para mejorar el desempeño de cada estudiante. Para esto se realizaron tres evaluaciones, en la primera se prueba si el alumno alcanzó los conocimientos sobre el tema, el segundo se trata de monitorear la actividad de los estudiantes en la plataforma y la tercera se trata de evaluar las experiencias de usuario mediante entrevistas. Según la investigación, elegir un elemento del juego en función de un rasgo de personalidad no necesariamente mejora el conocimiento; sino que demuestra un mejor compromiso en el curso. Como conclusión, en este trabajo los autores realizaron una plataforma digital para el aprendizaje de las materias de un curso, se compararon dos grupos; donde el primero utilizó la plataforma y el segundo no, se observó en el experimento que la plataforma ayuda a interesar a los estudiantes en los temas de la materia; sin embargo, no se logra aumentar el rendimiento académico ni la retención de los conocimientos. Se concluyó que estas plataformas pueden ayudar en el aprendizaje a distancia, pero para mejorar el aprendizaje es importante la presencia del profesor también (Pakinee & Puritat, 2021).

En la investigación realizada por Golbeck y colaboradores (Golbeck et al., 2011), se presenta un trabajo donde las redes sociales aparecen como un lugar donde las personas pueden presentarse al mundo, mostrando muchos detalles de su información personal. Se ha empezado a entender cómo se puede usar esta información para mejorar la experiencia del usuario a través de interfases. En esta investigación se centran en la personalidad de los usuarios, la cual se ha demostrado que ha sido útil para predecir la satisfacción laboral, profesional, éxito en las relaciones románticas e incluso las preferencias por diferentes tipos de interfases; hasta ahora es necesario el uso de pruebas de personalidad para medir la

personalidad de los usuarios, lo cual es muy impráctico de usar en redes sociales. Se presenta un método para medir la personalidad usando información pública de perfiles de Facebook, esto se logró utilizando diferentes métodos de aprendizaje máquina y el modelo de personalidad de los cinco grandes. Como conclusiones se recolectaron pruebas de personalidad de varios participantes y a través de una API se obtuvo la información pública de sus perfiles, utilizando el dataset generado a partir de esta información se entrenaron dos modelos de aprendizaje máquina, logrando una precisión del 11% de su valor real.

En el trabajo realizado por Yeye y colaboradores (Yeye et al., 2020) se habla sobre cómo la personalidad afecta el estilo de vida y el trabajo individual de las personas, así como en el desarrollo de la carrera profesional. Tradicionalmente los investigadores utilizan escalas para medir la personalidad; sin embargo, en la tarea de la psicología por evaluar a las personas como parte de un proceso de obtención de un trabajo, siempre existe la posibilidad de obtener respuestas dirigidas con el fin de influenciar los resultados. En los años recientes con el desarrollo del aprendizaje máquina, se han creado nuevas técnicas para el reconocimiento de la personalidad; en este trabajo los autores utilizan videos y cuestionarios de personalidad de los participantes para entrenar modelos capaces de predecir los rasgos de personalidad, para esto se utilizó el modelo de los cinco grandes, obteniendo como resultado una correlación entre las predicciones y los valores de los cuestionarios, que va en un rango de 0.32 a 0.44 para todas las dimensiones, con un valor p menor a 0.05, lo que da una manera no intrusiva de reconocer la personalidad. Como conclusión se utilizaron videos de presentación y cuestionarios de personalidad para entrenar modelos de aprendizaje máquina, con los cuales se puede predecir los rasgos de personalidad de los cinco grandes, estando cerca de obtener una correlación media para cada uno de los rasgos de personalidad.

(William & Badholia, 2015) revisan en su investigación diversos algoritmos de clasificación basados en el teorema de bayes, árboles, funciones y leyes comúnmente usadas para esta tarea. En esta investigación realizaron una comparación entre varios tipos de algoritmos de clasificación, se compararon algoritmos de clasificación usados para la identificación y predicción de la personalidad, utilizando un dataset de valores de HEXACO. El principal objetivo fue probar y examinar los algoritmos de clasificación y predicción para encontrar los que tengan un mejor rendimiento en el dataset de personalidad. Como conclusiones se

encontró el algoritmo con menor error absoluto medio en la clasificación y predicción, valor que comúnmente está asociado con una mayor precisión, la regresión lógica y el algoritmo de Naïve Bayes son los que mostraron mejores resultados con el menor ratio de error, el tiempo también fue un parámetro al evaluar la eficacia de los modelos.

3.4. Reconocimiento automático de intereses profesionales

Elegir una profesión es una decisión muy importante que todo estudiante o profesional debe tomar en su vida, y es de vital importancia conocer los propios intereses, esto se puede hacer realizando alguna prueba de intereses; ya sea en línea o en alguna institución. En la investigación realizada por (Pavithran & Ashraf, 2020) se propone el uso del aprendizaje máquina, para analizar información extraída de las redes sociales Facebook y Twitter, para predecir los intereses de una persona; para esto se usaron técnicas de minado de texto y utilizaron la base de datos de carreras de O*NET. La metodología consistió en extraer textos de Twitter, preprocesar las palabras y luego etiquetarlas en alguna dimensión de RIASEC, para posteriormente hacer el recuento. Como conclusiones argumentan que es posible predecir los intereses vocacionales de una persona usando textos extraídos de redes sociales.

Los métodos tradicionales para medir el interés tienden a capturar características comunes, pero pueden ser inexactos al capturar relaciones específicas dentro de una muestra dada. En la investigación realizada por Song (Song et al., 2022) se propone el uso de aprendizaje máquina ya que puede explicar de manera efectiva la compleja relación entre los intereses y la elección de carrera. El objetivo fue mejorar la precisión de las predicciones de la elección de trabajo, basado en acciones mediante la aplicación de aprendizaje automático; para esto se usó una muestra de 81,267 participantes empleados y desempleados; se comparó la precisión predictiva del método tradicional de aplicación de cuestionarios de intereses con un nuevo método de aprendizaje automático; mejorado para predecir los intereses y las aspiraciones profesionales. Los resultados muestran que, en comparación con el método tradicional, el método de aprendizaje automático proporciona una mayor precisión general para predecir ambos tipos de elecciones profesionales.

Capítulo 4

4. Desarrollo del proyecto

En este capítulo se presenta una propuesta para coadyuvar en la solución del problema de deserción escolar, la cual intenta sustituir las pruebas estandarizadas de intereses vocacionales que son largas y tediosas por el reconocimiento de personalidad a través de video y la predicción de intereses vocacionales. Se presenta la metodología implementada, la cual consta de cinco pasos, cada uno descrito en una sección independiente; además se describen todas las tecnologías utilizadas durante la investigación.

4.1. Propuesta de solución

Dado los factores que se analizaron en la descripción del problema sobre la deserción escolar en México, en donde se habla de la personalidad como uno de los factores más determinantes para la elección de una carrera profesional. Se propone una solución, la cual consiste en utilizar las investigaciones actuales en torno al aprendizaje máquina y el reconocimiento automático de personalidad para el desarrollo de un sistema que ayude en esta problemática.

El sistema que se propone se desarrolla recolectando pruebas de personalidad e intereses para generar un *dataset*, el cual se usará para entrenar modelos de aprendizaje máquina que puedan predecir los intereses. Estos modelos se implementarían en un sistema que grabe videos de los usuarios, realice las predicciones. Posteriormente se genera una lista de carreras recomendadas con base en los intereses del usuario, ayudando así a los jóvenes a tomar una decisión que favorezca a sus habilidades y características individuales.

4.2. Metodología

En este capítulo se presenta la metodología diseñada para la implementación de la propuesta de solución; esta metodología consiste en cinco pasos como se muestra en la Figura 6. Metodología de desarrollo del proyecto.

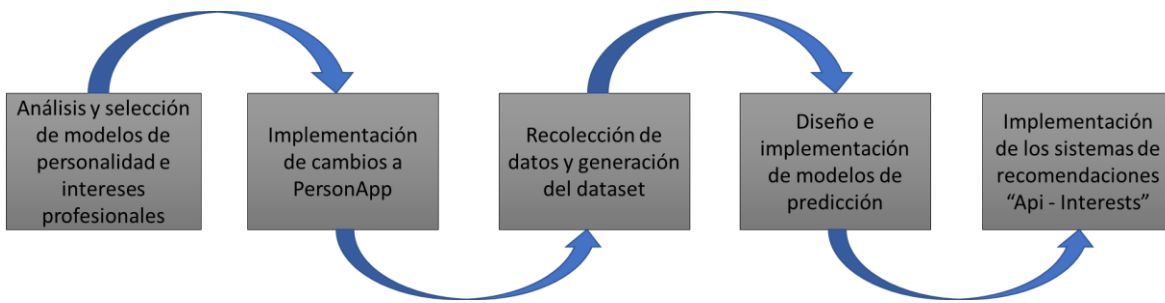


Figura 6. Metodología de desarrollo del proyecto.

En el primer paso se seleccionaron los modelos y herramientas que se usaron, en los pasos dos y tres se implementó el sistema de recolección y se generó la base de datos, en los últimos dos se entrenaron los modelos de aprendizaje y se implementó el sistema de recomendaciones. Cada una de las etapas de la metodología se detallan en las secciones siguientes.

4.3. Análisis y selección de modelos de personalidad e intereses profesionales

En las secciones 3.1 y 3.2 se presentan los antecedentes sobre los modelos de personalidad e intereses, encontrando que es imprescindible seleccionar modelos que tengan una estrecha relación entre sí. Los modelos de aprendizaje máquina solo pueden encontrar relaciones entre las variables si estas existen, por esto a pesar de que el modelo OCEAN sea el más reconocido e investigado se seleccionó a HEXACO. Este presenta relaciones más fuertes con los atributos de intereses que se utilizan comúnmente, los cuales son representados en el modelo RIASEC propuesto por Holland.

Otra de las razones por las que RIASEC fue considerado para esta investigación es la base de datos y cuestionario de intereses que proporciona el departamento del trabajo de estados unidos, información con la cual es posible realizar las pruebas de intereses profesionales. También permite interpretar los resultados y buscar en la base de datos las carreras más afines de acuerdo con la clasificación predefinida en la base de datos para cada valor de RIASEC.

4.4. Implementación del sistema PersonApp 2.0

En esta sección se documenta el desarrollo del sistema de recolección de datos, así como las herramientas y tecnologías utilizadas en el proyecto.

4.4.1. PersonApp

PersonApp (Bátiz-Beltrán, 2021) es una aplicación web desarrollada para recabar información sobre la personalidad (modelo OCEAN) aplicando cuestionarios y experimentos en video, con el propósito de generar un *dataset*. En PersonApp se implementó el cuestionario de personalidad IPIP-50; mientras que, para la grabación de los videos, el sistema describe tres experimentos donde el usuario debe hablar de un tema diferente para cada uno y se graban 4 videos cortos de 15 segundos por experimento.

El sistema PersonApp se desarrolló usando un modelo arquitectónico por capas sobre una arquitectura cliente – servidor.

La aplicación PersonApp se usó como base para desarrollar una nueva versión llamada PersonApp 2.0 donde se incorporaron los nuevos requerimientos establecidos para generar la base de datos de videos, personalidad e intereses vocacionales. Estos requerimientos se describen en la sección 4.4.2.

4.4.2. Requisitos funcionales PersonApp 2.0

Los requerimientos funcionales describen los principales procesos que debe efectuar un sistema. Para definir los requisitos funcionales se tomaron en cuenta los objetivos generales y específicos mencionados en las secciones 1.4.1 y 1.4.2 respectivamente.

En esta sección se muestran los requisitos funcionales para el sistema de recolección de datos de PersonApp 2.0. En la Tabla 1 se muestran los requisitos funcionales ordenados según su importancia, en la columna de prioridad se utilizan los valores de ALTA, MEDIA y BAJA para representar la importancia de cada requisito.

Tabla 1. Requisitos funcionales para PersonApp 2.0.

Requisito	Descripción	Prioridad
RF-1	El sistema permitirá a los usuarios realizar una prueba vocacional estandarizada.	ALTA
RF-2	El sistema permitirá a los usuarios realizar una prueba de personalidad estandarizada.	ALTA
RF-3	El sistema mostrará un aviso de privacidad de los datos.	ALTA
RF-4	El sistema deberá almacenar información académica y profesional.	MEDIA
RF-5	El sistema deberá tener menú donde los usuarios pueden ver su avance.	MEDIA
RF-6	El sistema mostrará el avance del tiempo durante la grabación de video.	BAJA
RF-7	El sistema mostrará elementos visuales para que el usuario distinga cuando terminó de grabar un experimento.	BAJA
RF-8	El sistema deberá tener una opción para detener la grabación cuando sea necesario y poder empezar desde el inicio.	BAJA
RF-9	El sistema validará el correo de los usuarios.	BAJA

4.4.3. Requisitos de calidad

Los requerimientos de calidad describen propiedades generales del sistema, estos se obtuvieron en base a los de la versión anterior y tomando en cuenta las nuevas funcionalidades requeridas. Se clasificaron los requisitos de calidad en rendimiento, portabilidad y usabilidad. En la Tabla 2 se muestran los requisitos de calidad ordenados por la prioridad en escala ALTA, MEDIA y BAJA.

Tabla 2. Requisitos de calidad para PersonApp 2.0.

Requisito	Atributo	Descripción	Prioridad
RC-1	Rendimiento	El sistema deberá guardar los datos en no más de 4 segundos.	ALTA
RC-2	Rendimiento	El sistema responderá a cada petición en no más de 4 segundos.	ALTA
RC-3	Portabilidad	El sistema funcionará en distintos tamaños de pantallas.	MEDIA
RC-4	Usabilidad	El sistema deberá ser intuitivo y fácil de usar.	BAJA
RC-5	Usabilidad	El sistema contará con una interfaz gráfica amigable para el usuario.	BAJA

4.4.4. Restricciones

Para el correcto funcionamiento del sistema y todos sus módulos se deberán tomar en cuenta las siguientes restricciones:

- El usuario deberá contar con cámara y micrófono que sean compatibles.
- El usuario deberá autorizar el uso de la cámara y micrófono de su dispositivo.
- El dispositivo deberá contar con acceso a internet con una velocidad mínima de un megabyte por segundo.
- El sistema funcionará en el navegador Chrome, ya sea en versión móvil o de escritorio.

4.4.5. Actores

Los actores son todas aquellas entidades que interactúan de alguna manera con el sistema. En la Tabla 3 se muestra una lista de los actores principales definidos en este sistema.

Tabla 3. Actores del sistema de PersonApp 2.0.

Actor	Descripción
Participante	Es todo usuario que se registra en el sistema con el fin de contestar las pruebas y grabar los videos.

4.4.6. Casos de uso

Con base en los requisitos funcionales y las restricciones del sistema, se establecieron cuatro casos de uso para describir los procesos más importantes que el sistema proporciona a cada uno de los actores involucrados.

En la Tabla 4 se muestra la lista de los casos de uso para el sistema PersonApp 2.0, y en el anexo B se presentan los casos de uso desarrollados, donde se detalla cada caso de uso, los actores involucrados, las precondiciones, el escenario principal y las postcondiciones.

En la Figura 7 se presenta el diagrama de casos de uso donde se muestran las interacciones de los actores con el sistema PersonApp 2.0.

Tabla 4. Casos de uso de PersonApp 2.0.

Caso de uso	Nombre	Descripción
CU-01	Presentar prueba estandarizada.	El participante entra al sistema y contesta un cuestionario.
CU-02	Registrar información académica y profesional.	El participante entra al sistema y contesta varias preguntas sobre su información académica o profesional.
CU-03	Visualizar pruebas completadas.	El participante entra al sistema y puede visualizar las pruebas que ha completado.
CU-04	Grabar video.	El participante entra al sistema y se graba realizando diversos experimentos.

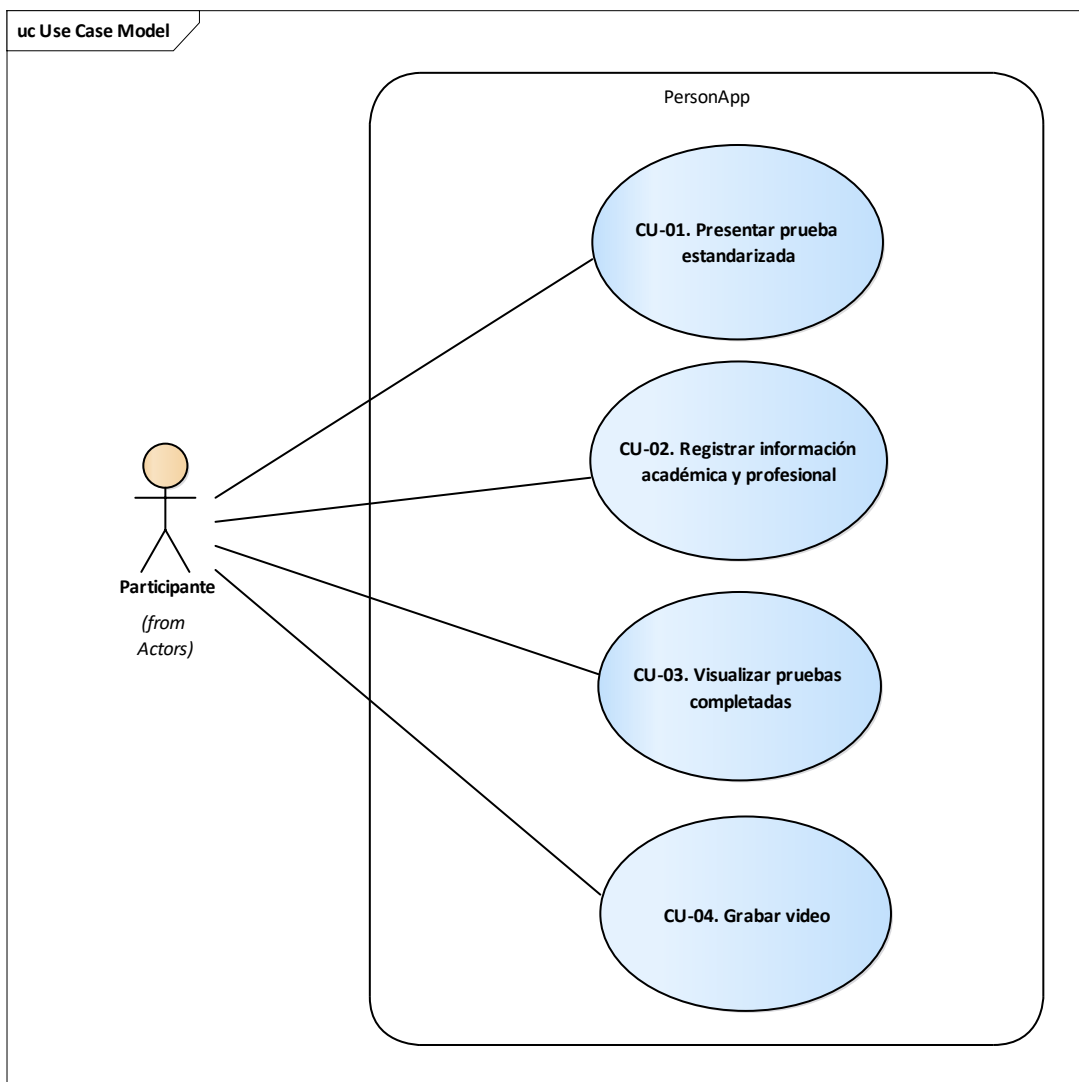


Figura 7. Diagrama de casos de uso.

4.4.7. Diagrama de contexto

En esta sección se muestra el diagrama de contexto, en donde se definen los límites del sistema, así como las interacciones con entidades externas. En la Figura 8 se muestra en la parte superior a los actores que utilizan el sistema, los cuales interactúan con el sistema PersonApp 2.0; en la parte inferior se encuentran las entidades de las cuales depende el sistema, en este caso el servidor de autenticación que permite a los usuarios ingresar al sistema, la base de datos que almacena toda la información generada en la plataforma y el servicio de predicción de intereses el cual recibe peticiones con los videos de los usuarios y realiza predicciones de intereses.

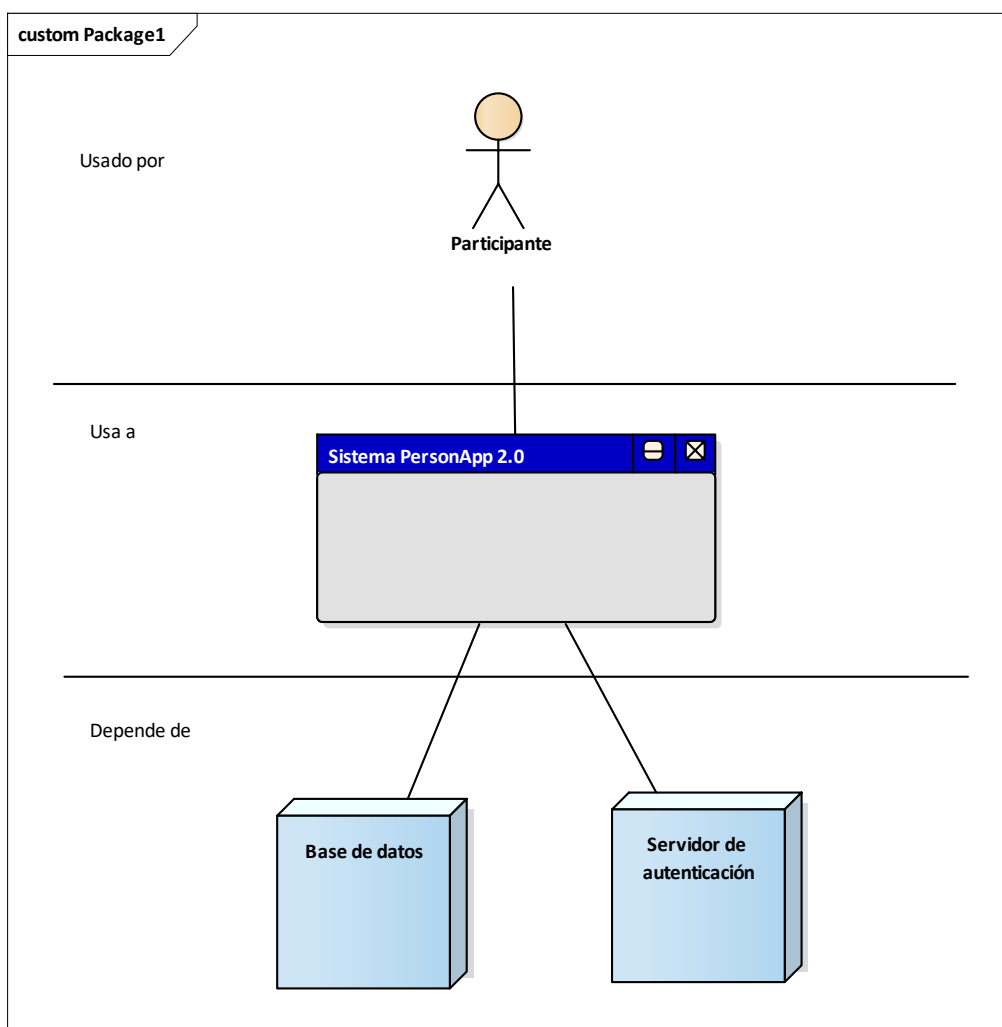


Figura 8. Diagrama de contexto de PersonApp 2.0.

4.4.8. Arquitectura

La arquitectura del sistema PersonApp original se compone de una arquitectura estilo cliente-servidor con un modelo arquitectónico de 3 capas, las cuales se definen como la capa de presentación, capa lógica de la aplicación y la capa de datos.

En la Figura 9 se muestra el diagrama de la vista lógica de la aplicación, los componentes de color azul fueron los que se modificaron para esta nueva versión de PersonApp, los grises se reutilizaron sin modificaciones.

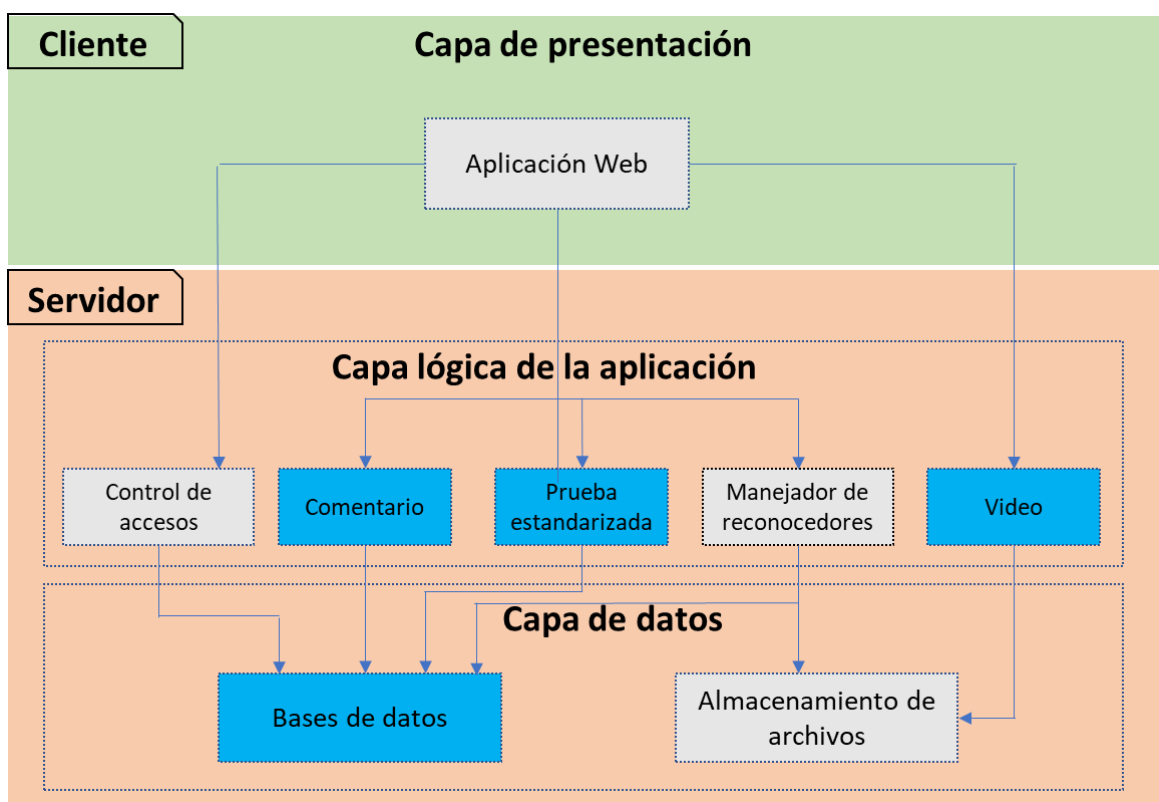
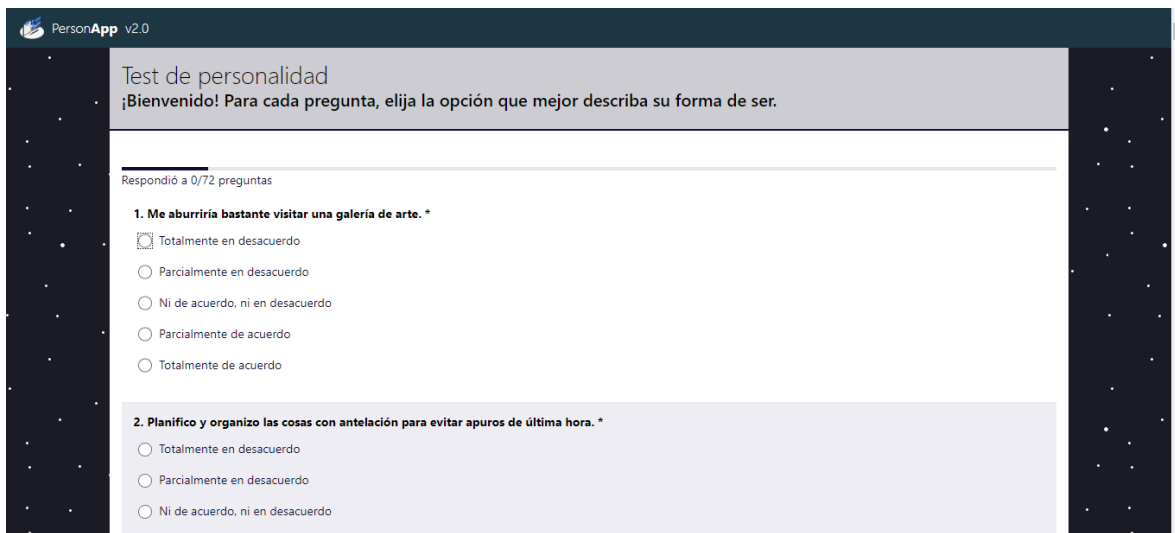


Figura 9. Vista lógica de la plataforma PersonApp 2.0.

4.4.9. Implementación

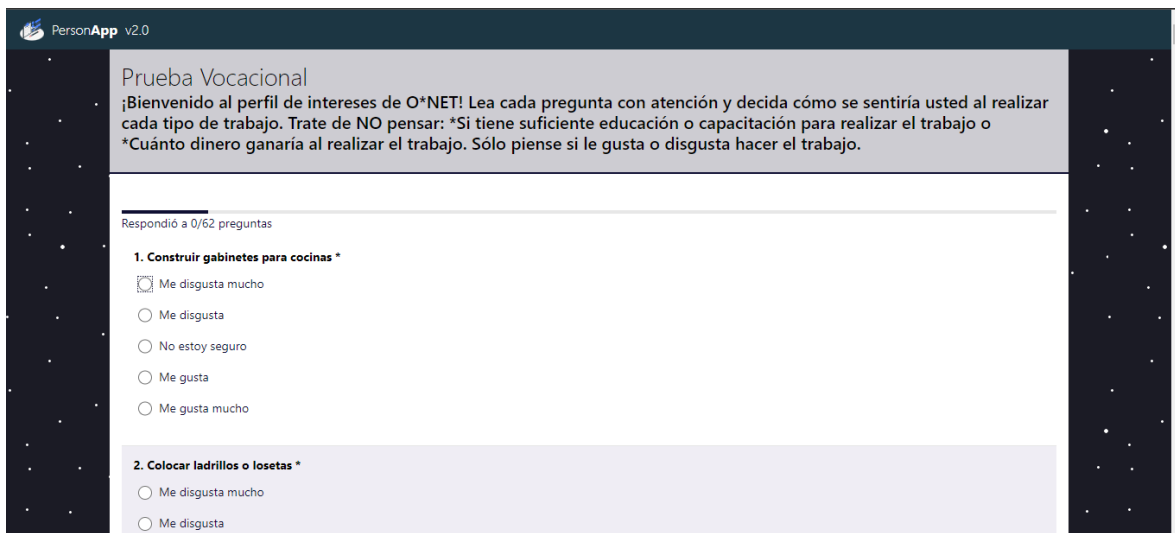
La implementación se realizó usando Google Firebase para el hospedaje, la autenticación y el almacenamiento de datos, para la interfaz de usuario se utilizó HTML, CSS, JavaScript junto con la librería de React.

Para los requisitos funcionales RF-1 y RF-2 se agregaron pruebas de personalidad e intereses, en cada cuestionario los usuarios tienen que escoger una opción para cada pregunta, el sistema simultáneamente lleva un contador del tiempo y al finalizar la prueba se verifica que todas las preguntas hayan sido contestadas, ver Figura 10. Cuestionario de personalidad. y Figura 11. Cuestionario de intereses profesionales. Para esto se utilizó la herramienta surveyjs.io, donde se capturó cada una de las preguntas del cuestionario, se descargó un archivo JSON con las preguntas y las opciones, por último usando librería survey-react se elaboraron los cuestionarios.



The screenshot shows a web application interface for a personality test. At the top, it says 'PersonApp v2.0' and 'Test de personalidad'. Below that, a welcome message reads: '¡Bienvenido! Para cada pregunta, elija la opción que mejor describa su forma de ser.' A progress bar indicates 'Respondió a 0/72 preguntas'. The first question is: '1. Me aburriría bastante visitar una galería de arte. *'. It has five radio button options: 'Totalmente en desacuerdo' (selected), 'Parcialmente en desacuerdo', 'Ni de acuerdo, ni en desacuerdo', 'Parcialmente de acuerdo', and 'Totalmente de acuerdo'. The second question is: '2. Planifico y organizo las cosas con antelación para evitar apuros de última hora. *'. It has three radio button options: 'Totalmente en desacuerdo', 'Parcialmente en desacuerdo', and 'Ni de acuerdo, ni en desacuerdo'.

Figura 10. Cuestionario de personalidad.



The screenshot shows a web application interface for a vocational test. At the top, it says 'PersonApp v2.0' and 'Prueba Vocacional'. Below that, a welcome message reads: '¡Bienvenido al perfil de intereses de O*NET! Lea cada pregunta con atención y decida cómo se sentiría usted al realizar cada tipo de trabajo. Trate de NO pensar: *Si tiene suficiente educación o capacitación para realizar el trabajo o *Cuánto dinero ganaría al realizar el trabajo. Sólo piense si le gusta o disgusta hacer el trabajo.' A progress bar indicates 'Respondió a 0/62 preguntas'. The first question is: '1. Construir gabinetes para cocinas *'. It has five radio button options: 'Me disgusta mucho' (selected), 'Me disgusta', 'No estoy seguro', 'Me gusta', and 'Me gusta mucho'. The second question is: '2. Colocar ladrillos o losetas *'. It has two radio button options: 'Me disgusta mucho' and 'Me disgusta'.

Figura 11. Cuestionario de intereses profesionales.

En la Figura 12 se muestra el aviso de privacidad que aparece en la pantalla sobre un modal cuando el usuario ingresa al sistema. Este mensaje tiene como objetivo informar al usuario que toda la información recabada será solo para uso interno de este trabajo de investigación y que ésta no será utilizada ni publicada sin el consentimiento de los usuarios, esto de acuerdo con el requisito RF-3.

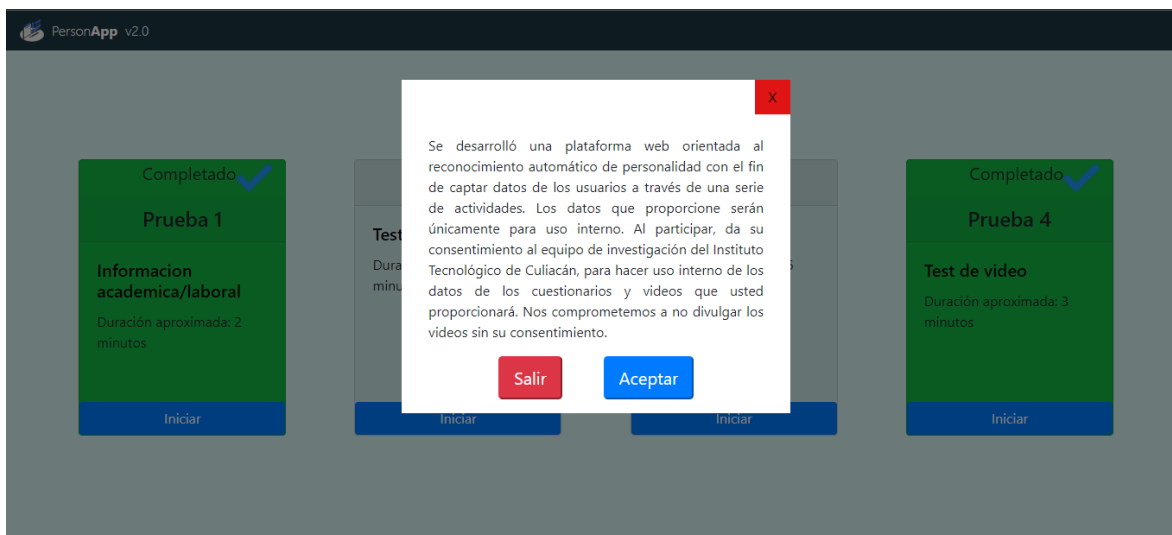


Figura 12. Aviso de privacidad de la plataforma.

En la Figura 13 y Figura 14 se muestran las preguntas que se realizan a los usuarios relacionadas con su información académica o profesional, esto según si el usuario es estudiante o ya se graduó, cumpliendo el requisito RF-4.

Figura 13. Cuestionario para participantes que sean estudiantes.

Figura 14. Cuestionario para participantes graduados.

Para que el usuario tenga una mejor experiencia se agregó un menú en donde los usuarios pueden ver que pruebas ya terminaron y cuales aún siguen sin contestar, coloreando de color verde y dibujando una palomita sobre las pruebas ya contestadas, esto atendiendo el requisito RF-5, ver Figura 15.

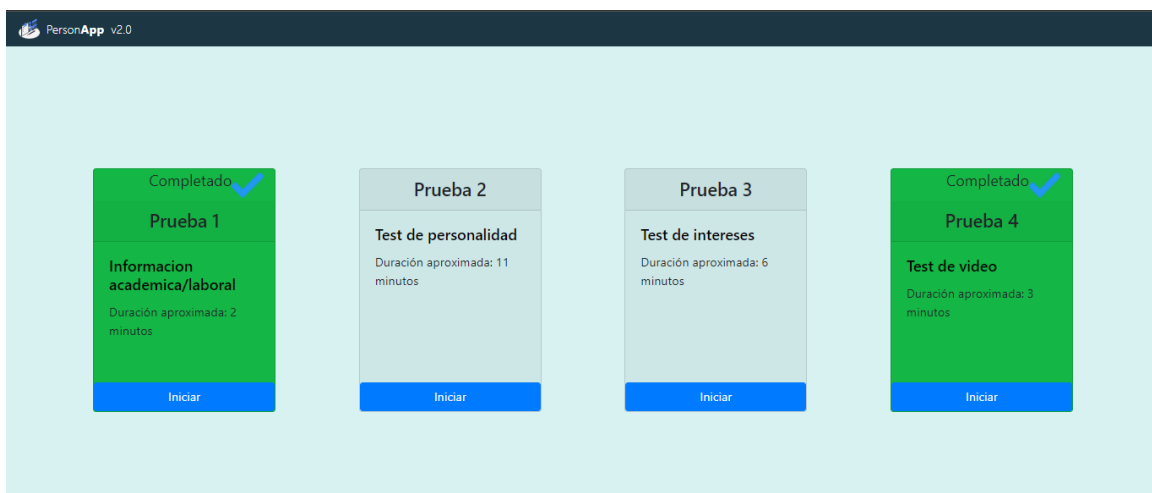


Figura 15. Menú principal de PersonApp 2.0.

Con el fin que el usuario tenga más control sobre la grabación de los videos se agregó un contador y un botón para poder detener y reiniciar la grabación, esto según los requisitos RF-6 y RF-8, ver Figura 16.

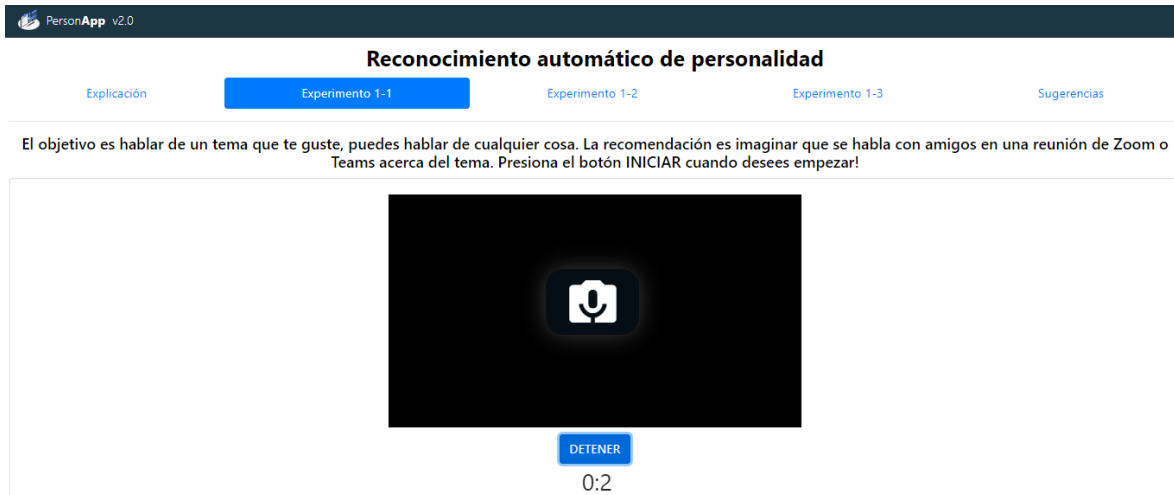


Figura 16. Grabación de experimentos de video.

Ayudando a la experiencia de usuario y siguiendo el requisito RF-07, al terminar de grabar un experimento aparece la frase “¡Gracias por tu participación!” y se cambia el número del experimento por la frase “Completado” subrayado de color verde, ver Figura 17.

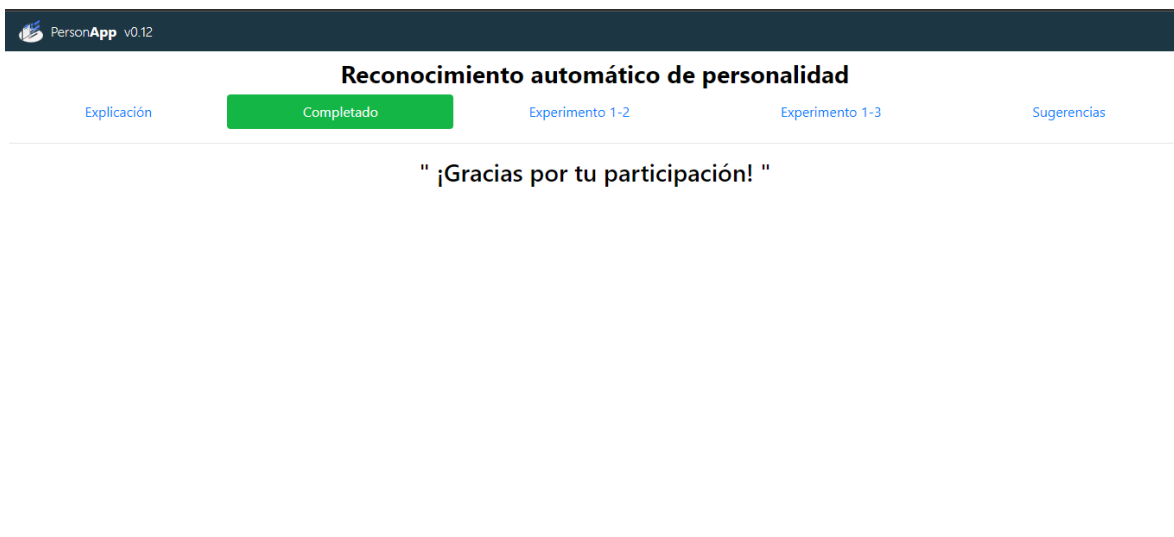


Figura 17. Mensaje de experimento completado.

Para la verificación del correo electrónico como se describe en el requisito RF-9, se utilizó el servicio de Firebase, el cual envía un correo electrónico, con un enlace para verificar que el propietario del correo es quien se registró, ver Figura 18.

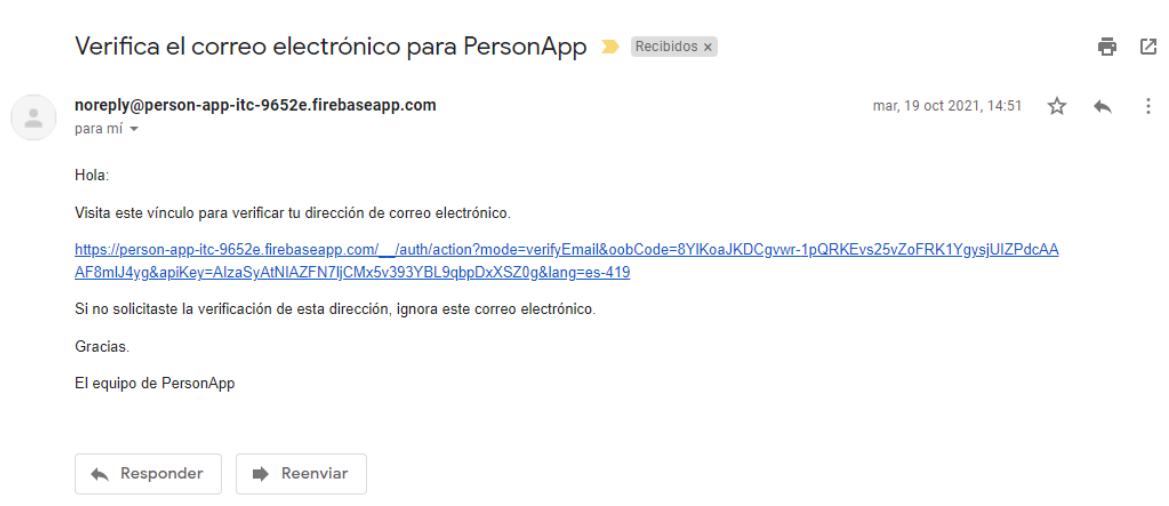


Figura 18. Correo de verificación.

4.5. Recolección de datos y generación del dataset

La aplicación se distribuyó entre un grupo de 30 personas de diferentes áreas de conocimiento para verificar el funcionamiento en diferentes plataformas. Posteriormente, se seleccionaron algunas personas de diferentes áreas (mayormente estudiantes y profesionales) para que respondieran las pruebas, esto con el fin de que la base de datos tenga una buena proporción de personas con características similares en cuanto a personalidad. Esto es de suma importancia para que sea posible la predicción de los intereses.

En la primera etapa de recolección de datos, los participantes fueron seleccionados entre compañeros, amigos y familiares, en esta fase se empezó por recabar la mayor cantidad de datos posibles sin prestar tanta atención a que se mantenga balanceado el número de registros por área de conocimiento. En una segunda etapa, estudiantes de nivel licenciatura del Instituto Tecnológico de Culiacán se incorporaron al proyecto para implementar un proceso enfocado a la recolección de datos con el fin de incrementar la cantidad de registros, en esta iteración, se buscaron personas de las áreas donde menos registros había en la BD.

Para que los datos puedan servir para entrenar a las redes neuronales artificiales, estos tienen que representar adecuadamente los intereses y la personalidad de los usuarios, de no ser así ninguna red neuronal ni ningún otro tipo de técnica será capaz de encontrar la relación entre estas variables. Para que la base de datos pueda servir para el propósito de este trabajo, es de suma importancia descartar aquellos datos inconsistentes los cuales representan respuestas aleatorias.

Para identificar estos casos se implementaron dos mecanismos:

- El primero es el tiempo que el usuario tarda en contestar las preguntas, se descartaron datos atípicos y se graficó una curva normal para obtener un tiempo mínimo, con el cual se considera que el usuario contestó las preguntas a conciencia en las pruebas de personalidad e intereses fueron 210 y 120 segundos respectivamente.
- El segundo son las preguntas de control, el usuario es instruido a contestar cierta opción en particular, por ejemplo, se le pide al usuario que escoja la opción “Me disgusta mucho” y si contesta cualquier otra opción se considera que contestó al azar el cuestionario.

4.6. Diseño e implementación de modelos de predicción

En esta sección se presentan los dos modelos de aprendizaje máquina desarrollados para la predicción de personalidad e intereses profesionales; así como las librerías y algoritmos utilizados.

4.6.1. Modelo de predicción de personalidad

En esta sección se describe el modelo de red neuronal utilizado para la predicción de la personalidad a través de video, este fue propuesto por (Zatarain-Cabada et al., 2022) y consiste en una red neuronal convolucional continua.

Esta red recibe vectores de $200 \times 200 \times 3$, estas son las dimensiones correspondientes a cada imagen que pasará por la red, las imágenes pasan por una serie de capas alternadas de convolución2D, BatchNormalization, MaxPooling2D, Flatten y varias capas densas al final. Para esta implementación se usó la función de activación tangente y *mean average error*

como función de pérdida. Esta red tiene un total de 53,664,453 de parámetros, de los cuales 53,662,917 pueden ser entrenados y 1,536 no.

4.6.2. Modelo de predicción de intereses profesionales

El módulo de predicción de intereses es el encargado de recibir la personalidad del usuario y hacer predicciones sobre los valores RIASEC, para esto se utiliza una red neuronal elaborada a partir funciones obtenidas de las librerías de Tensorflow y Keras, la estructura de la red neuronal se describe a continuación.

La red neuronal consiste en un modelo secuencial, donde la capa de entrada recibe siete parámetros de entrada correspondientes al modelo de personalidad, posteriormente se intercalan cuatro capas densas de 128 neuronas seguidas de una capa de Dropout del 50% y la capa de salida cuenta con una sola neurona, esta es la predicción del valor de RIASEC que se intenta predecir según el conjunto de datos, ya que esta red solo funciona para predecir un atributo a la vez es necesario entrenar seis modelos, uno para cada valor de RIASEC. Los hiperparámetros utilizados fueron una función de activación *softplus* y una función de pérdida *mean_squared_logarithmic_error*, se escogieron estas funciones después de varias pruebas con diferentes combinaciones.

Para evaluar la eficacia del modelo se desarrollaron dos funciones que se describen a continuación, la función calcular error recibe un vector que contiene los valores predichos y otro que contiene los valores reales, para cada par de valores se obtiene la diferencia y se guarda en un nuevo vector, por último, la función regresa el error absoluto medio por cada atributo de RIASEC, ver Figura 19.

```
1 def calcular_error(predictions_values, real_values):
2     error = []
3     for prediccion, real in zip(predictions_values , real_values):
4         error.append(abs(prediccion - real))
5     return np.sum(error, axis=0)/len(real_values)
```

Figura 19. Función para calcular el error de predicciones.

La función `get_num_errors` recibe los valores predichos, los valores reales y el error que se quiere calcular, esta función retorna el porcentaje de datos donde el error obtenido es menor o igual al error recibido por parámetro, Figura 20.

```
def get_num_errors(predictions_values, real_values, error):
    correctos = 0
    errors = []
    for prediccion, real in zip(predictions_values, real_values):
        errors.append(abs(prediccion - real))
    for error_arr in errors:
        for i in range(len(error_arr)):
            if error_arr[i] <= error:
                correctos += 1
    return correctos/len(predictions_values)*100
```

Figura 20. Función para calcular el porcentaje de predicciones dentro de un rango.

4.7. Implementación del sistema de recomendaciones API-Interests

En esta sección se describe la implementación del servicio de predicciones de intereses API-Interests, el cual es un sistema externo a la plataforma de PersonApp 2.0, este se desarrolló con la intención de que este sea un módulo independiente y pueda comunicarse con cualquier otro sistema para ofrecer el servicio de predicción de intereses.

4.7.1. Requisitos funcionales de API-Interests

En esta sección se presentan los requisitos funcionales del sistema de recomendaciones API-Interests, en la Tabla 5 se muestra el identificador del requisito y su descripción.

Tabla 5. Requisitos funcionales para API-Interests.

Requisito	Descripción
RF-01	El sistema podrá recibir peticiones desde cualquier dominio.
RF-02	El sistema podrá recibir el identificador de los usuarios en formato Json.
RF-03	El sistema podrá descargar los videos de los usuarios de la base de datos.
RF-04	El sistema podrá procesar los videos y obtener un número de imágenes predefinidas.
RF-05	El sistema podrá hacer predicciones de intereses con base en las imágenes obtenidas.

4.7.2. Requisitos de calidad

A continuación, en la Tabla 6 se muestran los requisitos de calidad con una descripción y prioridad.

Tabla 6. Requisitos de calidad para API-Interests.

Requisito	Atributo	Descripción	Prioridad
RC-1	Rendimiento	El sistema deberá responder una petición en un tiempo máximo de dos minutos.	ALTA
RC-2	Seguridad	El sistema mantendrá privados los correos que se le envían.	ALTA

4.7.3. Restricciones

Para el correcto funcionamiento del sistema y todos sus módulos se deberán tomar en cuenta las siguientes restricciones:

- El sistema funcionará en la versión 20 de Ubuntu.
- El sistema deberá contar con una conexión estable a internet.
- El sistema deberá contar con al menos 2 GB de memoria RAM disponible.
- El sistema deberá contar con al menos 1.5 GB de espacio en disco disponible.

4.7.4. Diagrama de contexto

En esta sección se presenta el diagrama de contexto del sistema, como se puede ver en la Figura 21 el usuario interactúa con el sistema, quien a su vez depende de la base de datos de Firebase, ya que es de la base de datos donde almacenan y obtienen los videos del usuario en cuestión.

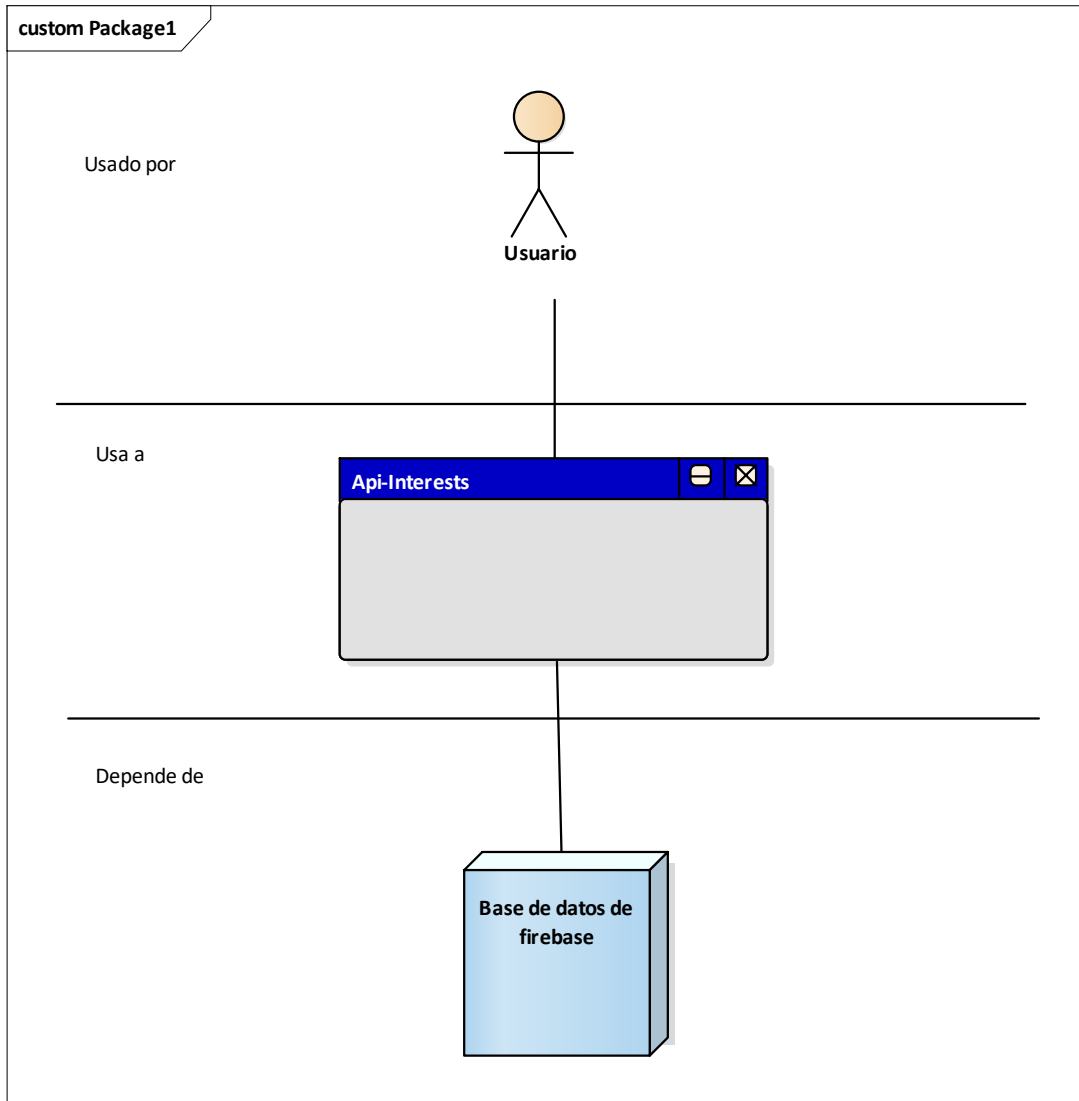


Figura 21. Diagrama de contexto de API-Interests.

4.7.5. Componentes del sistema

API-Interests fue desarrollado utilizando el lenguaje de programación Python con el framework Flask, así como varias librerías para el aprendizaje máquina. En la Figura 22 se muestra el diagrama de componentes donde se puede entender mejor cómo el sistema PersonApp 2.0 interactúa con el servicio de predicciones.

El sistema provee la interfaz email, la cual es un identificador único para cada usuario, el servicio recibe este email y lo utiliza para consultar la base de datos de Firebase; en esta consulta el servicio descarga todos los videos pertenecientes al usuario actual; los videos

pasan por el reconocedor de personalidad que utiliza un modelo de aprendizaje profundo para predecir los valores de HEXACO, posteriormente estos valores se pasan por el modelo de predicción de intereses, enviando las predicciones de RIASEC al cliente. PersonApp 2.0, como se describió en la sección anterior, cuenta con los mecanismos para interpretar los valores de RIASEC y mostrar al usuario las carreras más adecuadas de acuerdo con estos valores.

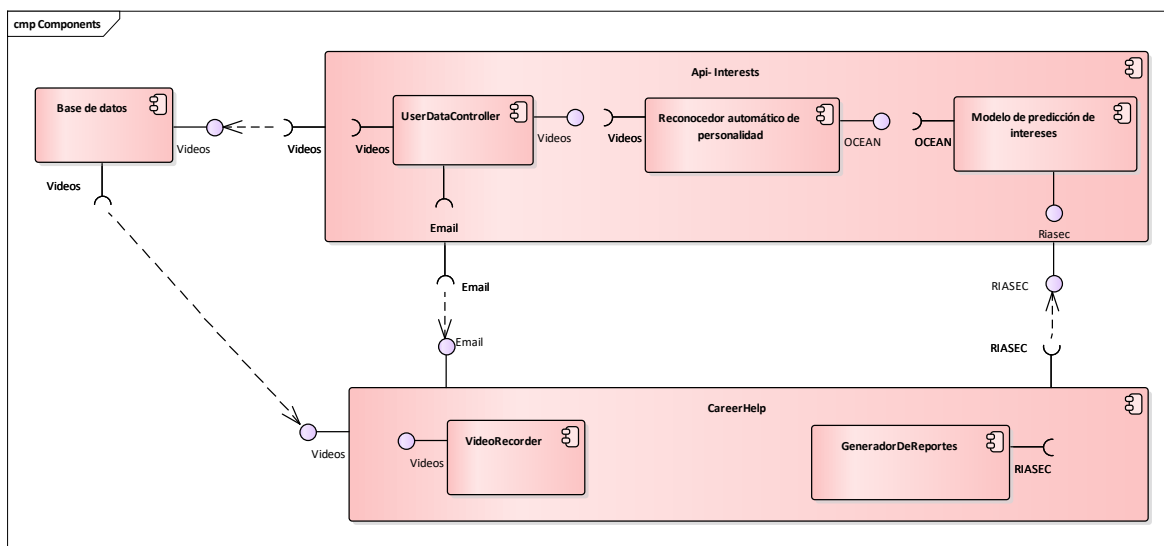


Figura 22. Diagrama de componentes.

El servicio está alojado en la plataforma de *Amazon Web Services* (AWS por su nombre en inglés), esta proporciona una máquina virtual de forma gratuita, el sistema operativo seleccionado para esta máquina virtual fue Ubuntu ya que es el más usado y hay más documentación sobre los procesos que se requieren implementar. Se utilizó la máquina virtual como un servidor para alojar API-Interests, también se configuraron las reglas de seguridad para que la API pueda recibir peticiones de cualquier dominio.

En la Figura 23 se muestra la interacción de la API funcionando en línea, recibe el correo del usuario en formato JSON y responde con las predicciones de RIASEC también en el mismo formato.

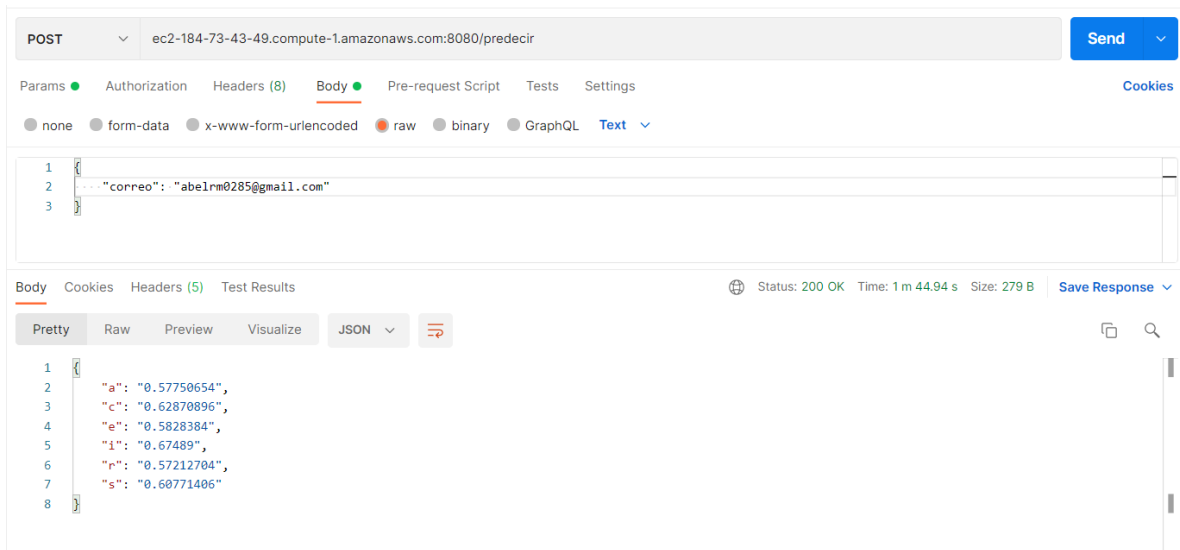


Figura 23. Envío de petición a la API.

4.8. Implementación del sistema de recomendación de carreras profesionales

En esta sección se describe la implementación del sistema de recomendación de carreras profesionales llamado CareerHelp. Este sistema permite a los usuarios realizar una prueba vocacional, eligiendo entre dos opciones, el clásico, que se conforma de 60 preguntas de opción múltiple o el de video, el cual consiste en grabar un clip de un minuto hablando de un tema de interés. Al finalizar cualquiera de los dos se le muestra al usuario un reporte de los resultados con una lista de carreras sugeridas.

4.8.1. Requisitos funcionales

En esta sección se muestran los requisitos funcionales para el sistema de recomendación de carreras profesionales, en la Tabla 7 se muestra el identificador del requisito, así como una descripción.

Tabla 7. Requisitos funcionales para CareerHelp.

Requisito	Descripción
RF-1	El sistema deberá permitir el registro de usuarios mediante correo electrónico.
RF-2	El sistema permitirá al usuario recuperar su contraseña.
RF-3	El sistema deberá autenticar los correos que se intenten registrar.
RF-4	El sistema deberá permitir a los usuarios realizar una prueba de orientación vocacional estandarizada.
RF-5	El sistema permitirá la grabación de una prueba en video sobre una temática libre.
RF-6	El sistema mostrará al usuario una gráfica con los valores que obtuvo el usuario en la prueba vocacional.
RF-7	El sistema mostrará al usuario una lista de carreras recomendadas.
RF-8	El sistema permitirá la descarga de los resultados en formato PDF.

4.8.2. Requisitos de calidad

Se clasificaron los requisitos de calidad en rendimiento, portabilidad y usabilidad. En la Tabla 8 se muestran los requisitos de calidad ordenados por la prioridad en escala ALTA, MEDIA y BAJA.

Tabla 8. Requisitos de calidad para CareerHelp.

Requisito	Atributo	Descripción	Prioridad
RC-1	Rendimiento	El sistema deberá guardar los datos en no más de 5 segundos.	ALTA
RC-2	Rendimiento	El sistema responderá a cada petición en no más de 2 minutos.	ALTA
RC-3	Portabilidad	El sistema funcionará en distintos tamaños de pantallas.	MEDIA
RC-4	Usabilidad	El sistema contará con una interfaz gráfica amigable para el usuario.	BAJA

4.8.3. Restricciones

Para el correcto funcionamiento del sistema se deberán tomar en cuenta las siguientes restricciones:

- El sistema deberá contar con una conexión estable a internet.
- El sistema deberá contar con al menos 2 GB de memoria RAM disponible.

- El dispositivo deberá contar con cámara y micrófonos compatibles.
- El usuario deberá autorizar el uso de la cámara y micrófono de su dispositivo.
- El dispositivo deberá contar con acceso a internet con una velocidad mínima de un megabyte.
- El sistema funcionará en el navegador Chrome ya sea en versión móvil o de escritorio.

4.8.4. Casos de uso

Con base a los requisitos y restricciones se elaboraron los casos de uso, los cuales describen las interacciones entre los usuarios y el sistema. En la Tabla 9 se muestra una lista de los casos de uso para este sistema.

Tabla 9. Tabla de casos de uso de CareerHelp con su descripción.

Caso de uso	Título	Descripción
CU-01	Registrarse en el sistema.	El usuario se puede registrar mediante su correo electrónico.
CU-02	Iniciar sesión en el sistema.	El usuario se puede iniciar sesión en el sistema.
CU-03	Presentar prueba vocacional estandarizada.	El usuario entra al sistema y contesta un cuestionario.
CU-04	Presentar prueba vocacional por video.	El usuario entra al sistema y se graba realizando un experimento.
CU-05	Obtener recomendación de carrera.	El usuario entra al sistema y después de realizar la prueba vocacional le aparece un reporte de sus resultados.

4.8.5. Diagrama de contexto

En esta sección se presenta el diagrama de contexto del sistema, como se puede ver en la Figura 24 se muestra las relaciones entre el sistema CarreHelp con los usuarios y servidores o servicios de los que depende para su funcionamiento.

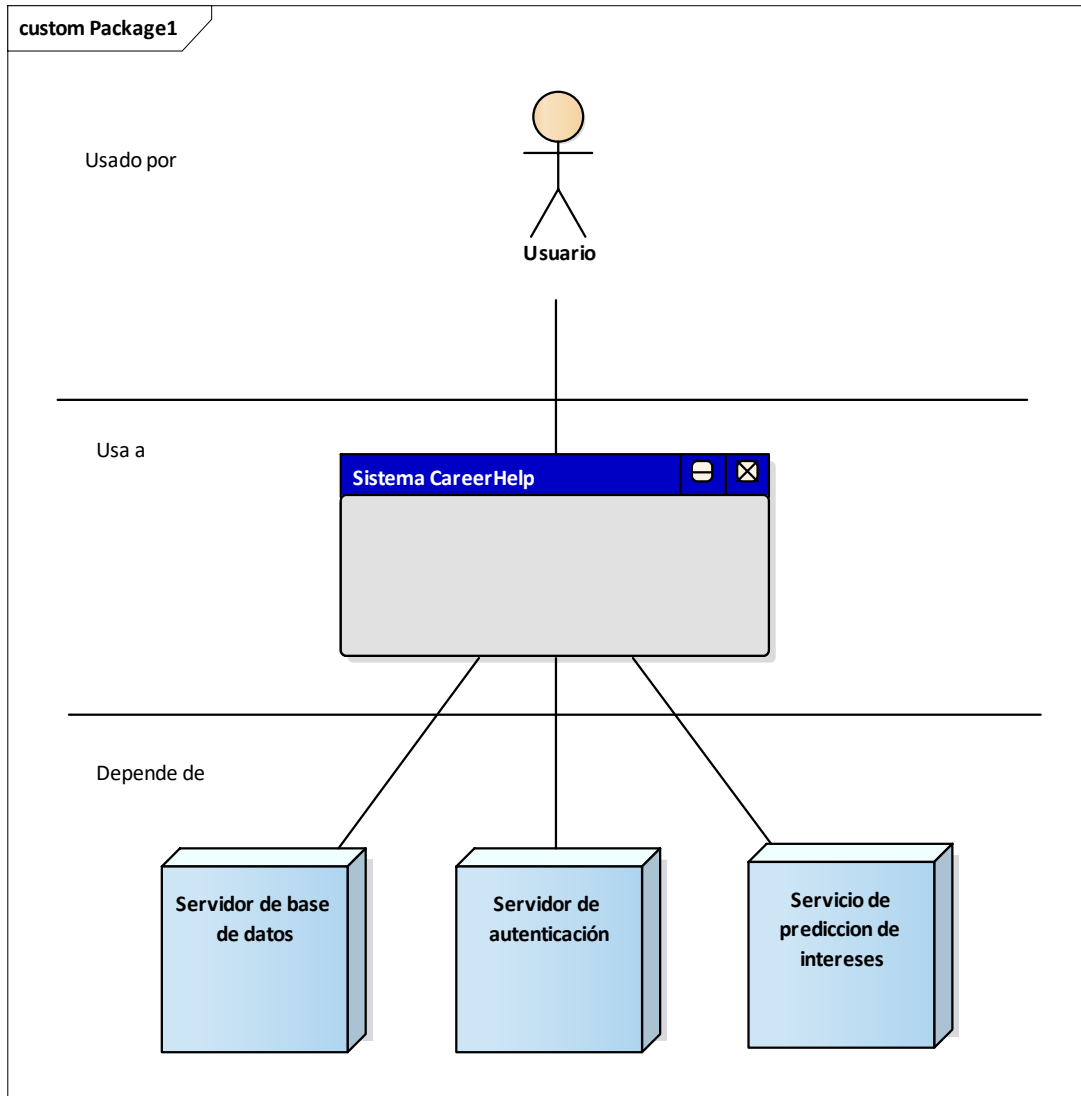


Figura 24. Diagrama de contexto del sistema CareerHelp.

4.8.6. Arquitectura

La arquitectura se compone de un modelo de capas sobre una arquitectura cliente servidor, en la Figura 25 se muestra el diagrama de la arquitectura del sistema, del lado del cliente se encuentra la capa de presentación compuesta por la aplicación web. Del lado del servidor se encuentra la capa lógica y la capa de datos, la capa lógica tiene los componentes lógicos necesarios para cumplir las funcionalidades del sistema y se comunica con la capa de presentación y la de datos, la capa de datos contiene el almacenamiento de los archivos de video y la base de datos.

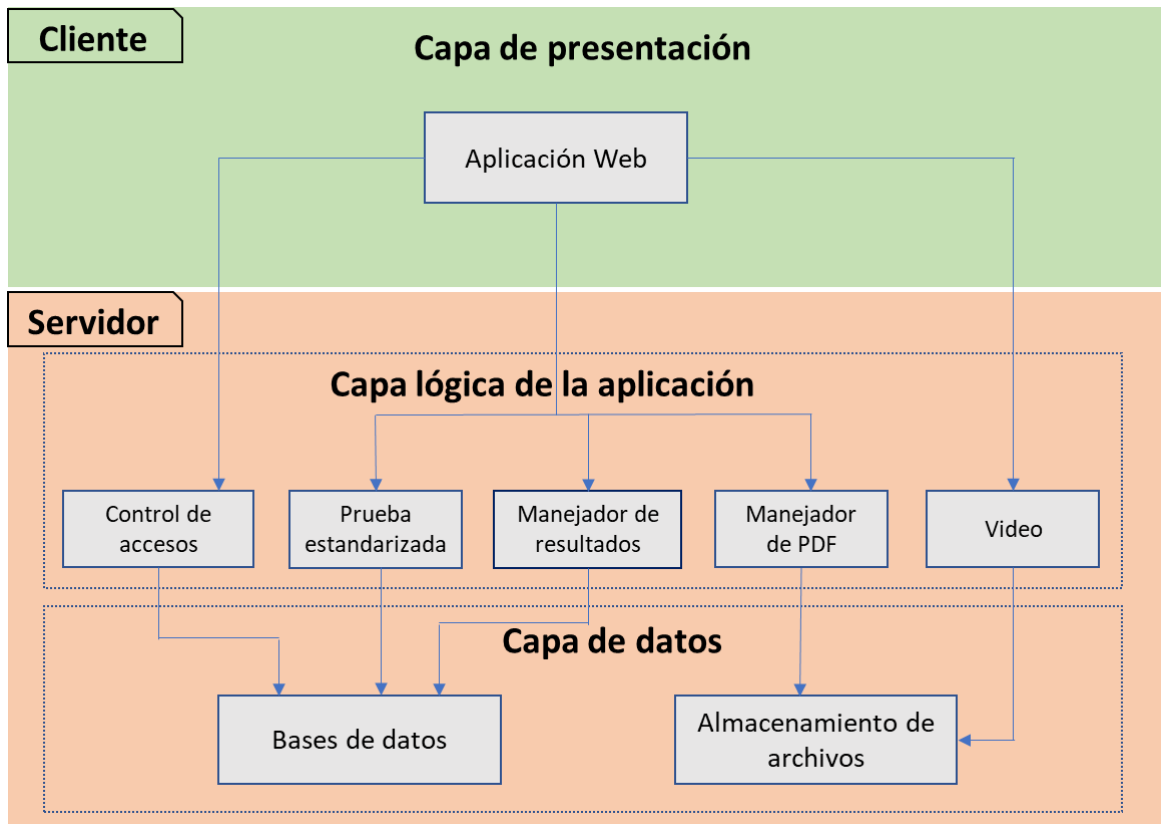


Figura 25. Arquitectura del sistema CareerHelp.

4.8.7. Implementación

La implementación se realizó utilizando los servicios que ofrece Google Firebase para el hospedaje, la autenticación y el almacenamiento de datos. Para el desarrollo de la aplicación web se utilizó HTML, CSS y JavaScript junto con la librería React.

Para la implementación de este sistema se utilizaron los componentes existentes del sistema PersonApp, los cuales tienen que ver con el manejo de los usuarios, la aplicación de pruebas estandarizadas y grabación de video. Estos componentes ayudaron a cumplir con los requisitos funcionales RF-1, RF-2, RF-3, RF-4 y RF-5.

Se desarrolló el menú principal del sistema utilizando las tecnologías web anteriormente mencionadas, en este menú se agregaron dos opciones cada una con una imagen, título, descripción y un botón para iniciar la prueba. La distribución de los elementos se realizó mediante el uso de CSS *grid layout*, dividiendo la pantalla en dos filas y una columna para

mostrar un elemento en cada fila, también se utilizaron *CSS media queries* para lograr un diseño responsive y así pueda utilizarse en todo tipo de pantallas.

En la Figura 26 y Figura 27 se puede ver el menú principal, la primera opción corresponde a la prueba de video, donde el sistema realizara predicciones sobre los intereses profesionales, en segunda opción se trata de la prueba estandarizada que consta de un cuestionario de 62 preguntas.

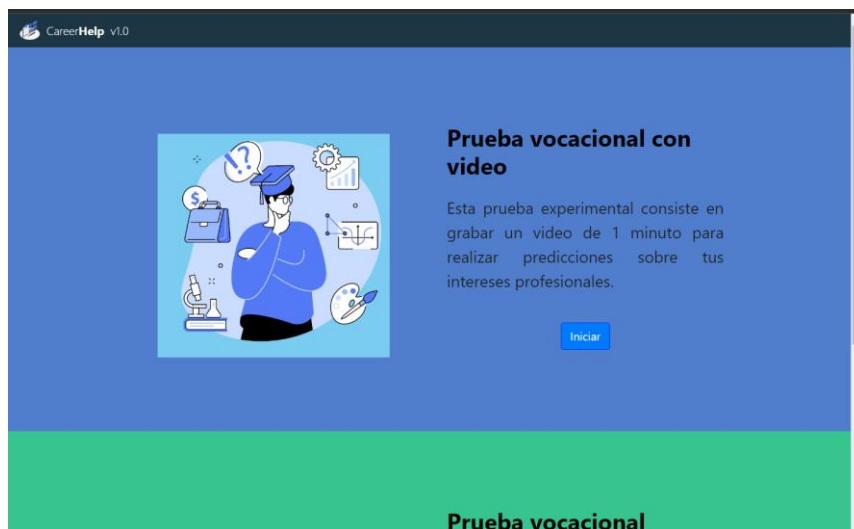


Figura 26. Menú principal de CareerHelp con opción de prueba en video.

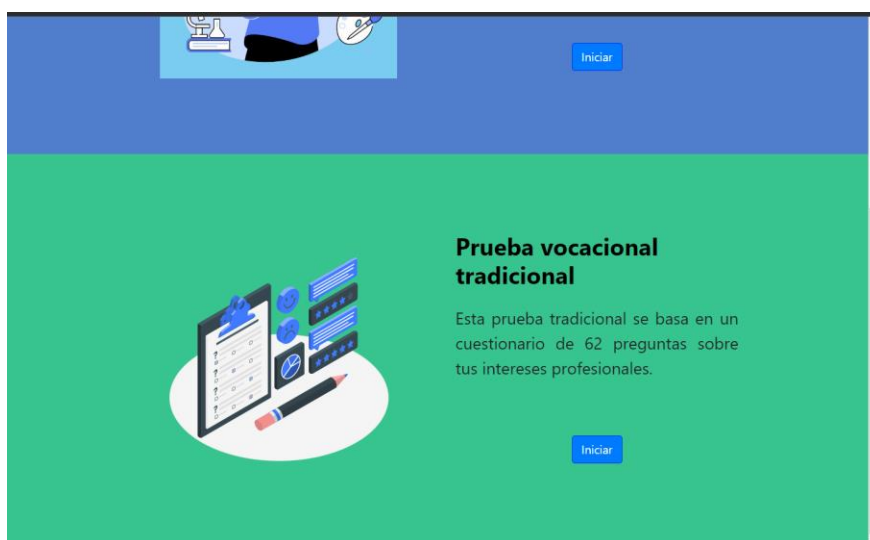


Figura 27. Menú principal de CareerHelp con opción de prueba estandarizada.

Al finalizar la prueba vocacional con video el usuario es dirigido a una pantalla de espera como se muestra en la Figura 28, en esta pantalla se envía una petición a través de API-REST al servicio de predicción de intereses, enviando el correo de usuario en un formato JSON. Al usuario se le pide esperar un máximo de 2 minutos, cuando el servidor responde con los valores correspondientes a cada atributo de RIASEC se redirige al usuario a la página de resultados.

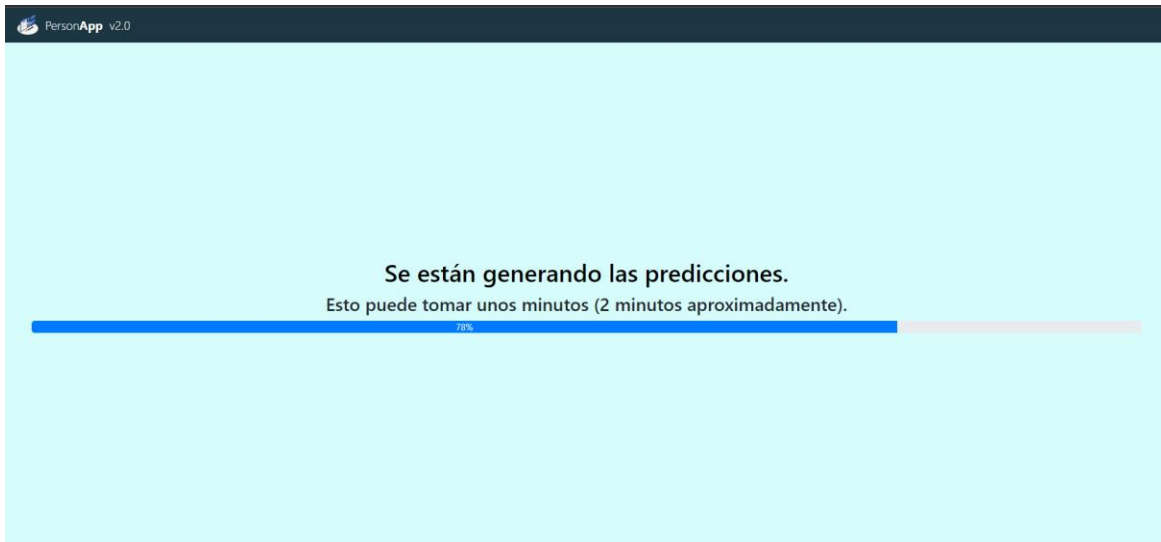


Figura 28. Pantalla de espera de sugerencias de carreras.

Para la implementación del requisito RF-6 se utilizó el componente Bar de la librería react-chartjs-2, este permite crear gráficos de barras personalizables. En la Figura 29 se observa la gráfica que muestra el puntaje obtenido por cada atributo, en la parte superior se listan los dos atributos con mayor puntaje, el eje de las ordenadas representa el valor que puede estar en el rango de cero a uno, el eje de abscisas muestra el nombre de cada atributo.

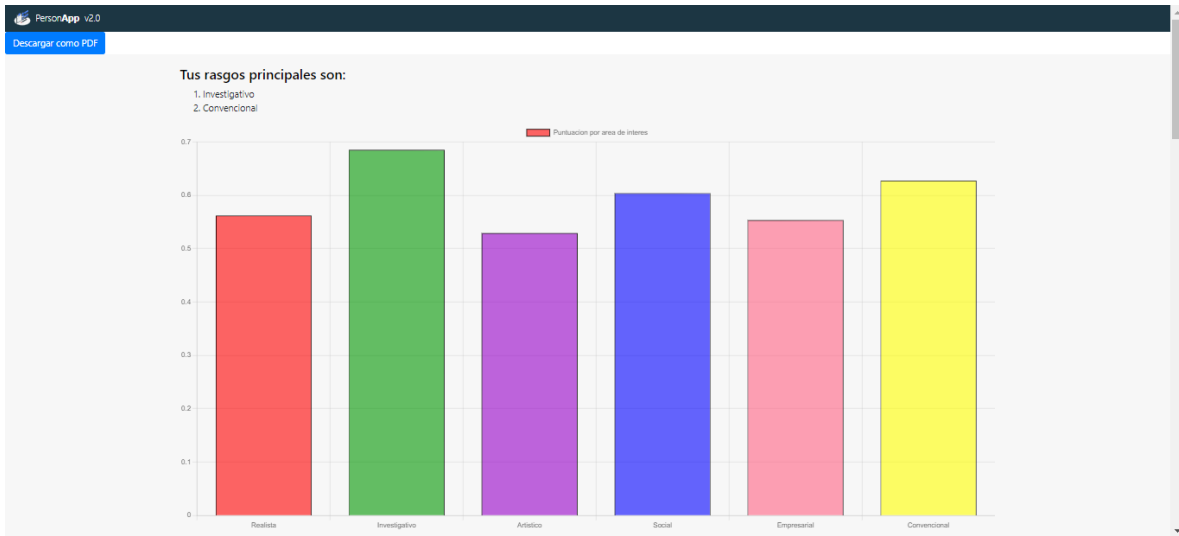


Figura 29. Gráfica de valores RIASEC.

Después de la gráfica se le muestra al usuario una lista de características generales que van con cada atributo, ver Figura 30. Al igual que otras interfaces se diseñó con un *grid layout* para que el número de filas y columnas cambie de acuerdo con el tamaño de pantalla.



Figura 30. Descripciones de cada atributo RIASEC.

Para cumplir el requisito RF-7 se utilizó la versión en español de la base de datos O*NET 4.0, la cual fue desarrollada por el departamento del trabajo de los estados unidos y está compuesta por información sobre carreras profesionales, intereses, habilidades etc. Se descargó la información necesaria convirtiéndola en archivos Json, para después ser manipulada a través de JavaScript.

En la Figura 31 se muestra la lista que se genera según los intereses de cada usuario, en esta lista aparecen las carreras cuyos rasgos de RIASEC mejor se adecuen a los del usuario en cuestión, al seleccionar una carrera de la lista aparece un modal donde se muestra una descripción de la carrera.

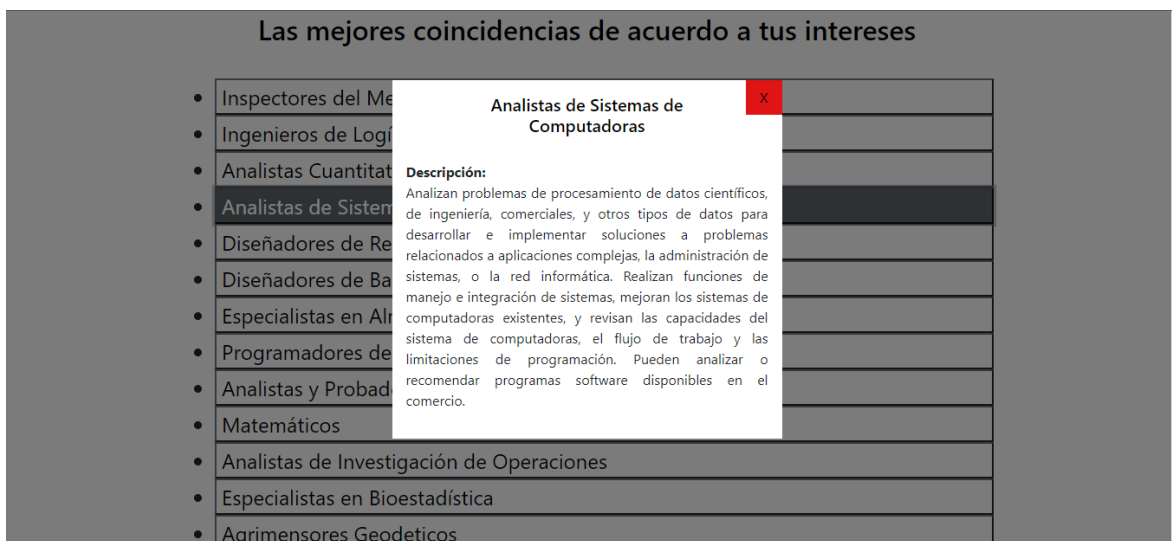


Figura 31. Lista de carreras recomendadas.

En la Figura 32 se muestra la opción para descargar el reporte como un archivo PDF, esto según el requisito RF-8, para implementar esta función se utilizó la librería react-to-print, la cual permite agregar el contenido pasándole la referencia a los componentes de React y se usa como disparador (*trigger*) el evento onClick de un botón.

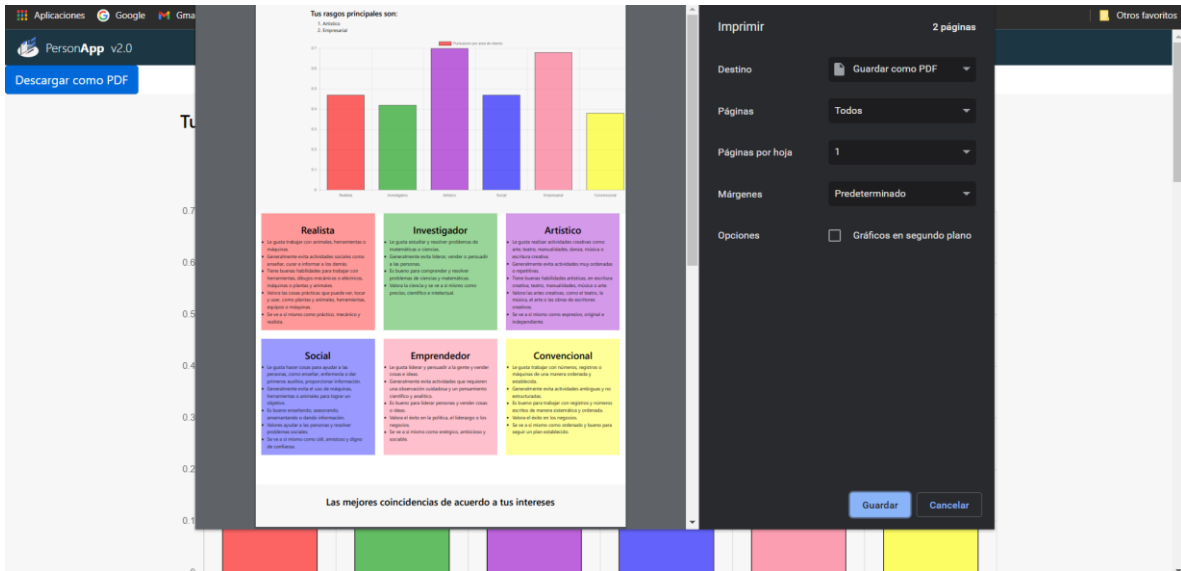


Figura 32. Descarga del reporte como archivo PDF.

4.9. Lenguajes y herramientas de desarrollo

En esta sección se describen las tecnologías y herramientas para la implementación de los requisitos en el sistema PersonApp y el servicio de predicción de intereses.

JavaScript: es un lenguaje de programación multiparadigma interpretado que se utiliza mayormente en navegadores web, cuenta con un solo hilo dinámico y soporta los paradigmas orientados a objetos, imperativo y funcional (Developers Mozilla (C), 2022).

HTML: es un lenguaje de marcado de hipertexto utilizado para estructurar el contenido de páginas web (Developers Mozilla (B), 2022).

CSS: Las hojas de estilo en cascada por su traducción al español es un lenguaje utilizado para dar presentación a la estructura de los documentos HTML y es una de las tecnologías más usadas en la web (Developers Mozilla (A), 2022).

React: Es una biblioteca de JavaScript diseñada para la creación de interfaces de usuario, basado en componentes con un paradigma declarativo, destaca por la facilidad de crear componentes reutilizables en diferentes sistemas (Meta Open Source, 2022).

Firestore: Es una base de datos no relacional hospedada en la nube, que permite almacenar y sincronizar información con los usuarios en tiempo real (Firestore, 2022).

Visual Studio Code: Es un IDE que cuenta con la posibilidad de agregar extensiones que añaden nuevas capacidades tanto visuales como de funcionalidad, esta herramienta es multiplataforma y cuenta con soporte para la mayoría de los lenguajes de programación más usados hoy en día (Microsoft, 2022).

Python: Es un lenguaje de programación interpretado multipropósito que se enfoca en la legibilidad del código, actualmente es uno de los más utilizados en el mundo para todo tipo de aplicaciones (Python, 2022).

Flask: Es un framework de python que permite el rápido desarrollo de aplicaciones web (Flask, 2022).

Tensorflow: Es una biblioteca de código abierto creada por Google, contiene una serie de funciones y herramientas para el desarrollo de aprendizaje máquina (Tensorflow, 2022).

Capítulo 5

5. Resultados

En este capítulo se presentan los resultados obtenidos en la generación del dataset de personalidad e intereses, de las pruebas realizadas del reconocedor de personalidad usando el dataset de videos generado y de los modelos de predicción de intereses profesionales que utilizan los cuestionarios recolectados en PersonApp.

5.1. Dataset de personalidad e intereses profesionales

En esta sección se presenta la estadística descriptiva y un análisis de la información recolectada con el sistema PersonApp, tanto de cuestionarios, experimentos en video y algunos otros datos personales y profesionales de los participantes.

Después de aplicar el filtrado de datos descrito en el desarrollo, se obtuvo un total de 128 registros que completaron las pruebas de personalidad e intereses y pasaron las pruebas de control para validar que hayan sido contestadas correctamente, también se obtuvieron 68 experimentos en videos de los cuales se revisó que cumplieran con la actividad requerida para cada experimento, así como también que la imagen y audio sean lo suficientemente claras.

En la Figura 33 se muestran la distribución de registros según el área de conocimiento de las personas que participaron, esto después de filtrar los datos según los parámetros presentados anteriormente. En el área de las ciencias exactas e ingenierías es donde se concentran la mayor cantidad de datos, representando el 40% del total de datos, también hay un número importante de personas que no proporcionaron su información profesional, por último, son en las áreas de ciencias biológicas y agropecuarias, arte, arquitectura, diseño y ciencias sociales y humanidades donde hay menos información, representando todas estas áreas juntas solo un 11%.

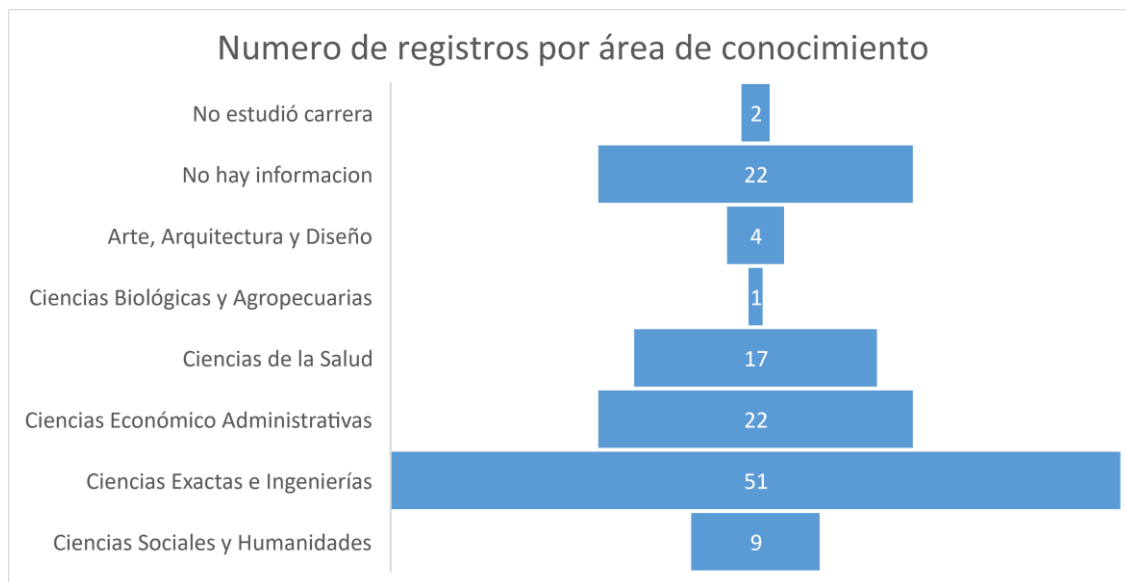


Figura 33. Número de registros por área de conocimiento.

En la Figura 34 se muestra el porcentaje de registros que se obtuvieron según el sexo de cada persona, incluyendo la opción “Prefiero no decir”, se puede observar que el porcentaje de hombres es casi el doble que el de las mujeres, obteniendo un 54 y 28 por ciento respectivamente.

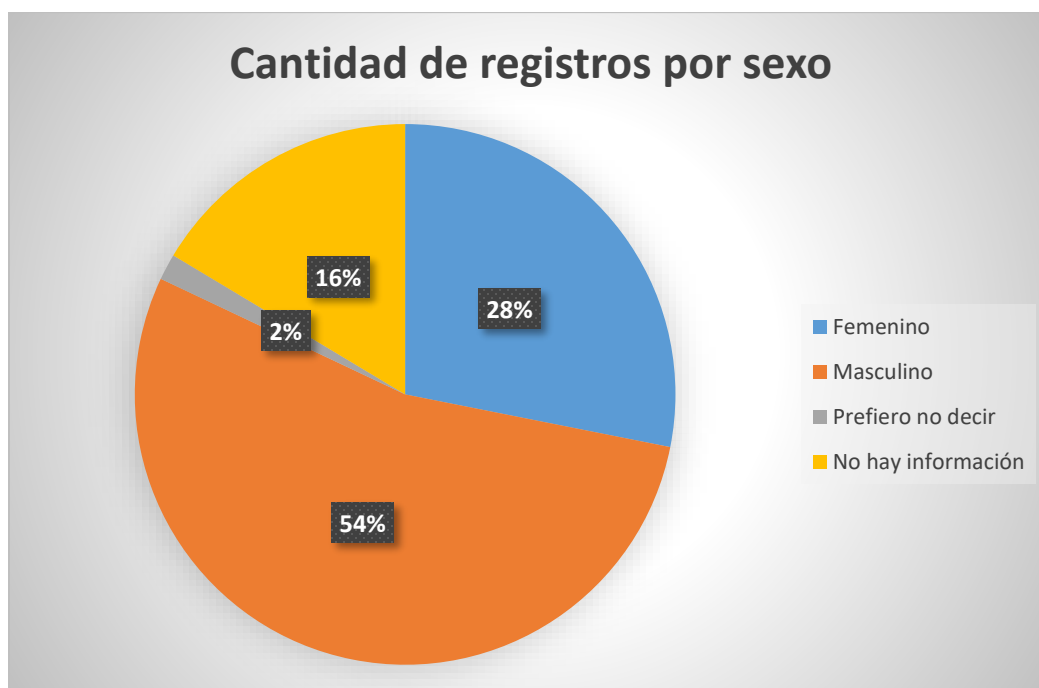


Figura 34. Cantidad de registros por sexo.

En la Tabla 10 se presentan algunas estadísticas sobre el dataset, en la primera columna se listan todos los atributos tanto del modelo HEXACO y de RIASEC, esta tabla se realizó a partir del dataset filtrado por lo que el número de participantes tomados en cuenta fueron 128 para cada atributo, la media de cada atributo se encuentra alrededor de 0.5, la desviación estándar varía según el rasgo aunque presentan un valor muy bajo, los mínimos y máximos de cada rasgo se encuentran mayormente un poco por debajo o por encima de la media y solo algunos llegan al máximo y mínimo reales, los cuales son 1 y 0 respectivamente.

Tabla 10. Estadística descriptiva por atributo.

Atributo	Participantes	Media	Desv. Est.	Mínimo	Máximo
Apertura	128	0.68	0.08	0.48	0.84
Escrupulosidad	128	0.65	0.07	0.44	0.76
Neuroticismo	128	0.44	0.11	0.16	0.70
Emocionalidad	128	0.67	0.09	0.46	0.90
Amabilidad	128	0.68	0.08	0.52	0.88
Extraversión	128	0.68	0.09	0.42	0.90
Honestidad	128	0.63	0.07	0.48	0.76
Realista	128	0.51	0.18	0.00	1.00
Investigativo	128	0.61	0.20	0.20	1.00
Artístico	128	0.58	0.19	0.15	1.00
Social	128	0.60	0.18	0.10	0.97
Empresarial	128	0.58	0.20	0.00	1.00
Convencional	128	0.54	0.20	0.00	1.00

En la Tabla 11 se muestra el porcentaje de datos de cada atributo que se encuentra dentro de los intervalos de una típica distribución normal, los cuales corresponden a la columna 2 y 3 y representan que tanto se ajustan los datos a una curva normal, esto se ve comparando el porcentaje de cada atributo con el valor encima de cada columna, este valor es la distribución de probabilidad alrededor de la media en una distribución normal.

En el anexo A se presenta una tabla de correlación donde aparecen todos los rasgos RIASEC y HEXACO, aquí se puede encontrar que el rasgo de investigativo tiene una correlación moderada con la apertura, amabilidad y honestidad, también artístico tiene una correlación

moderada con la apertura, se presenta también una correlación negativa entre la emocionalidad y neuroticismo.

Tabla 11. Porcentaje de datos por desviación estándar.

	68,26	95,44
Rasgo	1 desviación estándar	2 desviación estándar
Realista	0.70	0.94
Investigativo	0.62	0.98
Artístico	0.65	0.93
Social	0.67	0.98
Emprendedor	0.72	0.95
Convencional	0.69	0.95

En la Figura 35 se muestra el histograma de los valores obtenidos por los participantes en el atributo de escrupulosidad, donde se muestra una tendencia hacia el intervalo de 0.65 y 0.75; también se puede ver que el rango de valores está incompleto ya que solo abarca de 0.45 a 0.75.

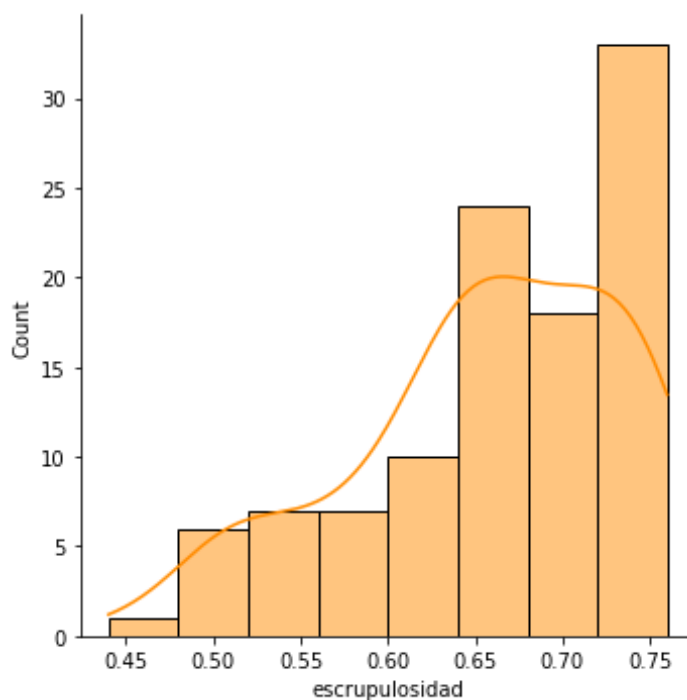


Figura 35. Histograma del atributo escrupulosidad.

En la Figura 36 se muestra el histograma de los valores obtenidos por los participantes en el atributo de apertura, el rango de los valores se presenta de 0.5 a 0.85; mostrando una tendencia alrededor de 0.70.

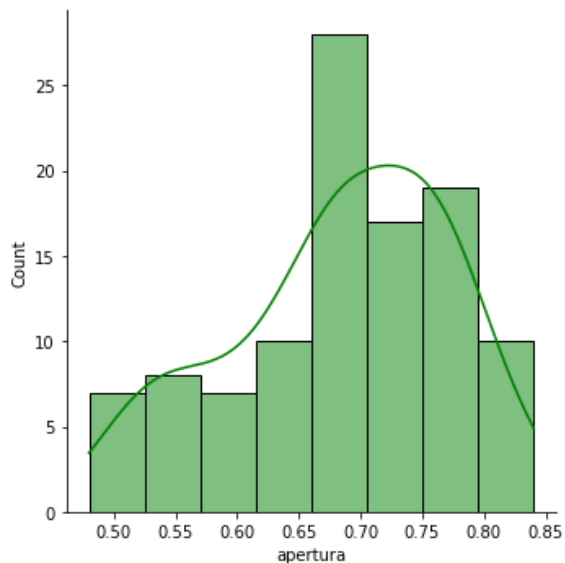


Figura 36. Histograma del atributo apertura.

En la Figura 37 se muestra el histograma de los valores obtenidos por los participantes en el atributo de emocionalidad, con un rango de valores entre el 0.5 y 0.9; mostrando un mayor volumen de datos en torno a 0.65.

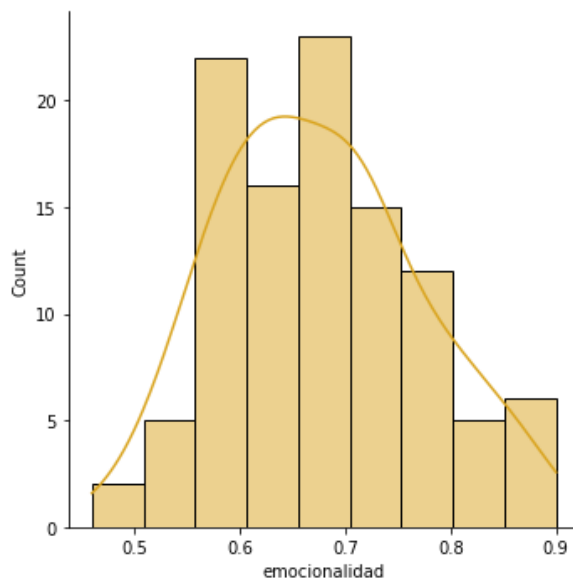


Figura 37. Histograma del atributo emocionalidad.

En la Figura 38 se muestra el histograma de los valores obtenidos por los participantes en el atributo de neuroticismo, con un rango de valores que va de 0.2 a 0.7; concentrándose mayormente en 0.55.

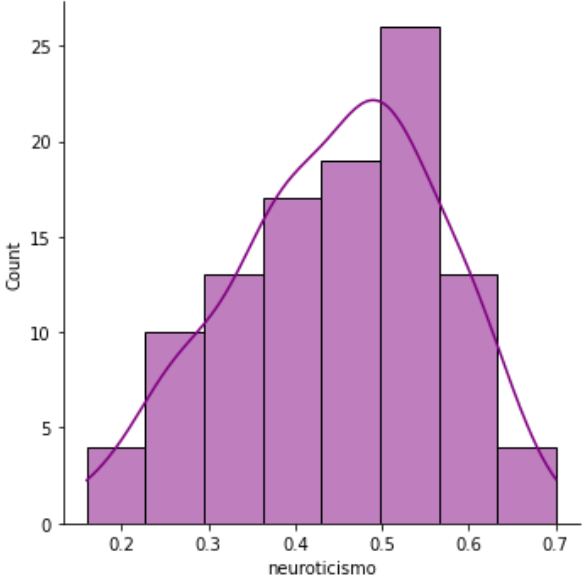


Figura 38. Histograma del atributo neuroticismo.

En la Figura 39 se muestra el histograma de los valores obtenidos por los participantes en el atributo de amabilidad, con un intervalo de valores de 0.55 a 0.85; con una tendencia a 0.70.

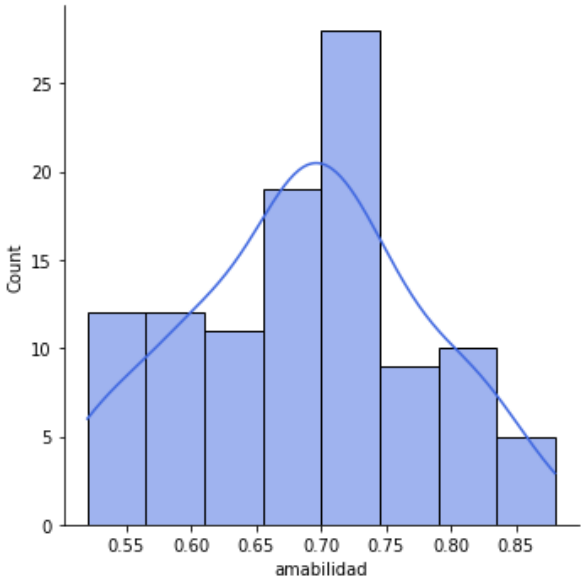


Figura 39. Histograma del atributo amabilidad.

En la Figura 40 se muestra el histograma de los valores obtenidos por los participantes en el atributo de honestidad, con un rango de valores bastante corto que va de 0.5 a 0.75; concentrándose mayormente en 0.65.

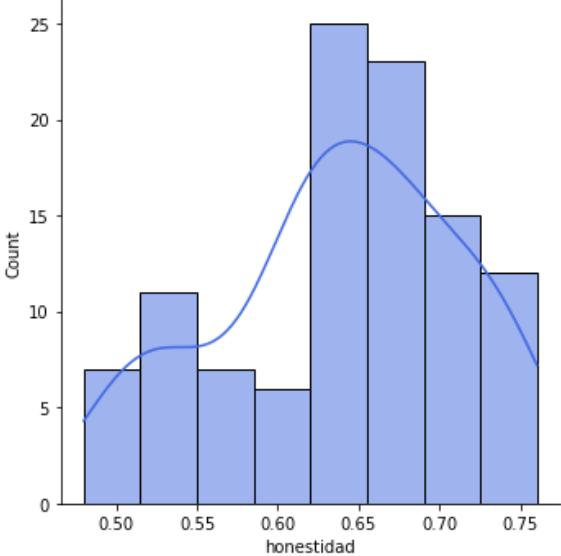


Figura 40. Histograma del atributo honestidad.

En la Figura 41 se muestra el histograma de los valores obtenidos por los participantes en el atributo de extraversión, con un rango de valores que va de 0.45 a 0.9; concentrándose mayormente en 0.7.

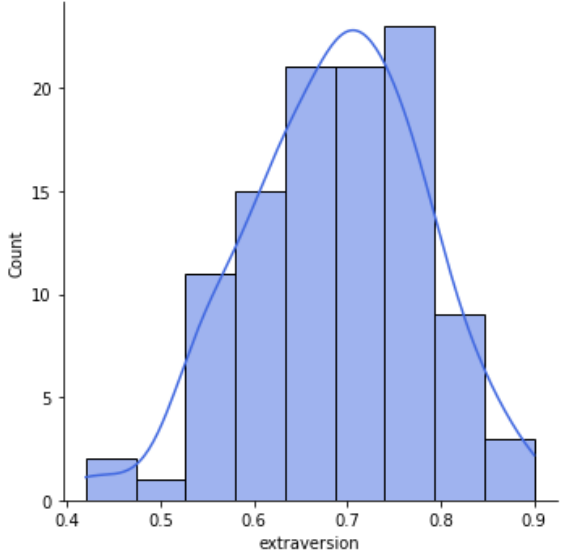


Figura 41. Histograma del atributo extraversión.

En la Figura 42 se muestra el histograma de los valores obtenidos por los participantes en el atributo de investigativo. Con un rango de valore de 0.2 a 1, lo que abarca prácticamente todo; la distribución tiene a los valores entre el 0.5 y 0.8.

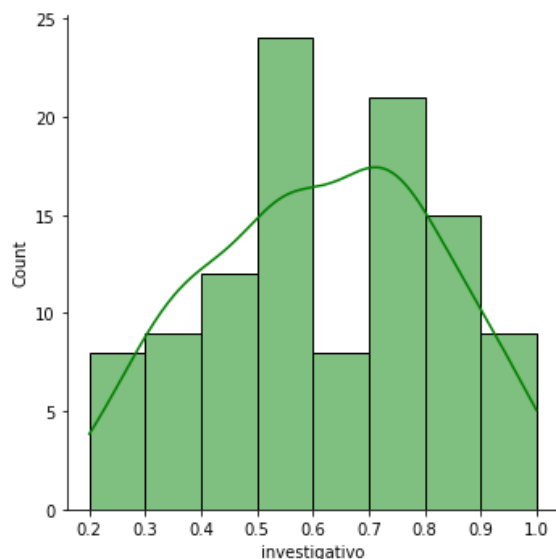


Figura 42. Histograma del atributo investigativo.

En la Figura 43 se muestra el histograma de los valores obtenidos por los participantes en el atributo de social, con valores de 0.2 a 1, mostrando un gran volumen de datos en 0.5; aunque los demás valores están distribuidos en todo el rango.

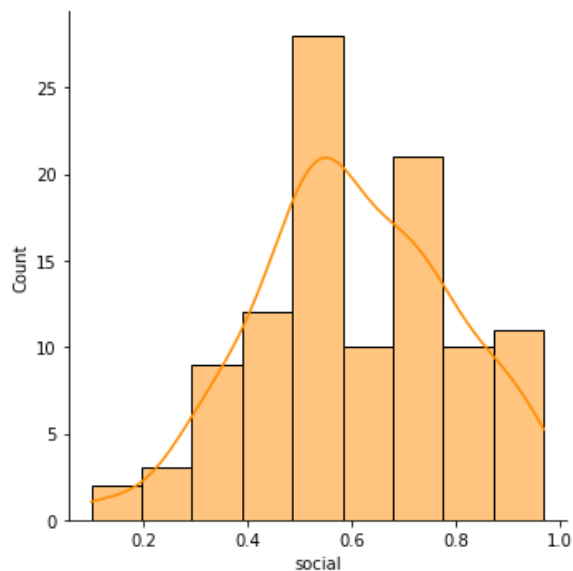


Figura 43. Histograma del atributo social.

En la Figura 44 se muestra el histograma de los valores obtenidos por los participantes en el atributo de artístico, con un rango de valores de 0.2 a 1 y mostrando una clara tendencia hacia el 0.6 que es donde se ubican el mayor volumen de datos.

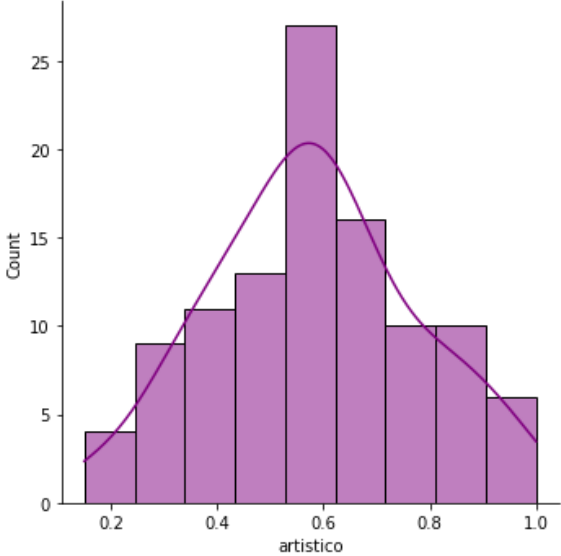


Figura 44. Histograma del atributo artístico.

En la Figura 45 se muestra el histograma de los valores obtenidos por los participantes en el atributo de empresarial, con un rango de valores completo de 0 a 1, con un mayor volumen de datos en torno a 0.4 y 0.8, con un pico en 0.5.

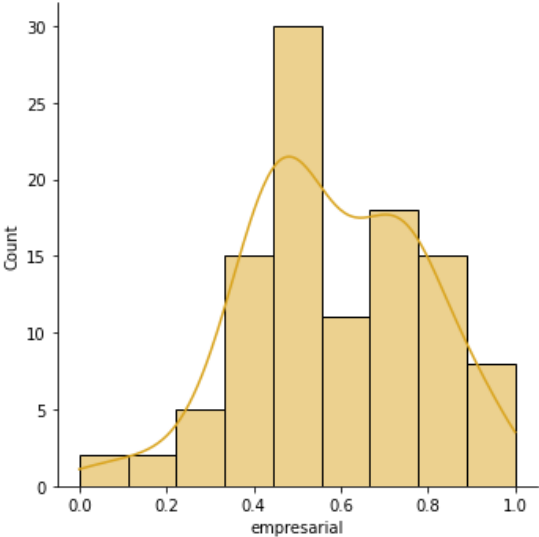


Figura 45. Histograma del atributo empresarial.

En la Figura 46 se muestra el histograma de los valores obtenidos por los participantes en el atributo de convencional, con un rango completo de valores de 0 a 1, y una clara tendencia hacia el 0.5.

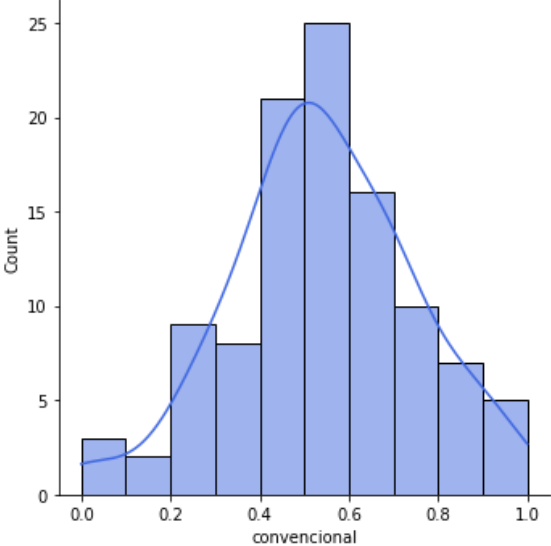


Figura 46. Histograma del atributo convencional.

En la Figura 47 se muestra el histograma de los valores obtenidos por los participantes en el atributo de realista, con un rango completo de valores de 0 a 1, y una clara tendencia hacia el 0.5.

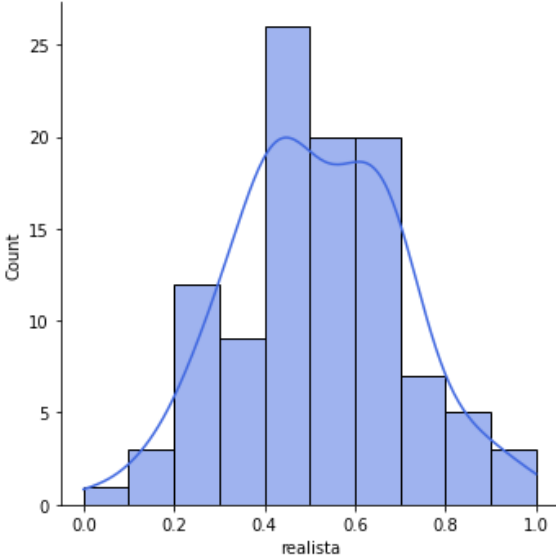


Figura 47. Histograma del atributo realista.

5.2. Reconocedor de intereses profesionales

El reconocedor de intereses profesionales se entrenó utilizando el modelo presentado en la sección 4.6.2 y a partir del dataset filtrado, se utilizaron los valores de HEXACO para que la red aprendiera a predecir RIASEC, durante el entrenamiento se realizaron 100 épocas teniendo como métrica el error absoluto medio por atributo, se tomaron 100 datos para el entrenamiento y 28 para la validación.

En la Tabla 12 se muestra en la segunda columna el error absoluto medio por cada atributo, donde todos obtuvieron un error entre 0.14 y 0.16, en las posteriores columnas se muestra el porcentaje de datos cuyo error estuvo en cierto rango para la tercera columna es el error menor a 0.1 y para la cuarta y quinta columna 0.2 y 0.3 respectivamente.

Tabla 12. Métricas de error por atributo RIASEC.

Rasgo	Mae	Error <= 0.1	Error <= 0.2	Error <= 0.3
Realista	0.1457	46.15	73.07	88.46
Investigativo	0.1690	30.70	57.69	88.46
Artístico	0.1586	38.46	73.07	80.76
Social	0.1657	34.61	65.38	88.46
Emprendedor	0.1393	46.15	69.23	88.46
Convencional	0.1643	34.61	76.92	92.30

5.3. Reconocedor de personalidad

El reconocedor de personalidad fue entrenado con el modelo presentado en la sección 4.6.1 usando el dataset de experimentos en video después de ser filtrado y también otros videos recolectados con la versión anterior de PersonApp, en total se usaron 131 videos de los cuales se extrajeron 6180 imágenes para el entrenamiento, los resultados se muestran a continuación.

En la Tabla 13 se muestra el error absoluto medio por cada atributo, en cada uno el error está alrededor de 0.17.

Tabla 13. Error absoluto medio por atributo OCEAN.

Rasgo	Mae
Extraversión	0.1748
Amabilidad	0.1886
Apertura	0.1602
Neuroticismo	0.1893
Escrupulosidad	0.1886

Capítulo 6

6. Conclusiones y trabajo futuro

En este capítulo se presentan las conclusiones de este trabajo de investigación, así como las aportaciones que se realizaron en el área y el trabajo futuro que se planea realizar para mejorar los resultados obtenidos tanto del dataset generado como de los modelos de predicción.

6.1. Conclusiones

En este proyecto se logró la generación de una base de datos con información de personalidad e intereses profesionales, la implementación de sistemas para la aplicación de cuestionarios y predicción de personalidad e intereses profesionales, las conclusiones individuales de cada parte se detallan a continuación.

Con la implementación de los nuevos requerimientos sobre el sistema PersonApp se logró una plataforma donde se pueden aplicar pruebas estandarizadas de personalidad e intereses usando el modelo HEXACO y RIASEC respectivamente, también es capaz de almacenar información académica y profesional de los estudiantes y experimentos de personalidad en video.

A partir del uso de esta plataforma se logró recolectar 128 registros que completaron las pruebas de personalidad e intereses y pasaron las pruebas de control para validar que hayan sido contestadas correctamente, también se obtuvieron 68 experimentos en video, de los cuales se revisó que cumplieran con la actividad requerida para cada experimento. El análisis de estos datos mostró que hay una participación principalmente masculina, que representa el 52 por ciento de los datos contra el 28 por ciento correspondiente a las mujeres. También se presenta un desbalance en las áreas del conocimiento de las cuales participaron en las pruebas, siendo las ciencias exactas e ingenierías donde participaron el mayor porcentaje de personas. Este desbalance junto con la naturaleza misma de los fenómenos psicológicos, causan una distribución normal en los valores obtenidos, a pesar de que esta distribución es

natural por el tipo de muestra, provoca problemas para los propósitos con los que se utiliza este *dataset*.

Los resultados obtenidos de los modelos de predicción de personalidad e intereses muestran, primero, que los modelos están aprendiendo a minimizar el error, a esto se le conoce como regresión a la media. Esto pasa ya que se presenta una distribución normal en los datos, causando que los modelos aprendan que probabilísticamente obtendrán un error bajo con valores cercanos a la media y un error alto cuanto más se alejen de ella, en lugar de aprender a reconocer los atributos a partir de rasgos en las imágenes en el caso del modelo de predicción de personalidad, y relaciones entre la personalidad e intereses en el caso del modelo de predicción de intereses.

El servicio de predicción de intereses funciona correctamente para procesar las peticiones de los usuarios, descargando los videos de los experimentos, procesándolos para obtener un conjunto de imágenes, estas las pasa por los modelos de predicción y contesta las peticiones de los usuarios con los valores guardados en formato JSON. En el caso de la plataforma se permite al usuario grabar un video de personalidad y al finalizar utiliza el servicio de predicción de intereses para mostrarle un reporte de sus resultados, los cuales incluyen una lista de carreras recomendadas por el sistema.

Con la realización de este trabajo se llega a la conclusión que para construir modelos que realicen predicciones de personalidad e intereses con un error bajo es necesario un dataset más grande, compuesto por muestras de personas de diversas áreas del conocimiento para mantener un balance en los valores, también sería importante realizar un preprocesado a los datos para asegurarse que la distribución de estos sea la óptima. El uso de una plataforma que ayuda a los estudiantes a elegir una carrera puede ser de gran ayuda en la sociedad, pero es necesario que la precisión de esta sea similar al de las pruebas estandarizadas; para que pueda funcionar correctamente y aún se requiere de más investigación para llegar a ese punto.

6.2. Aportaciones y limitaciones

Las principales aportaciones de este trabajo fueron, primero, desarrollar una nueva versión de la plataforma PersonApp, ampliando las pruebas estandarizadas que se pueden aplicar,

mejorando la interfaz al hacerla más intuitiva y clara para los usuarios y agregando un nuevo módulo para la predicción de intereses profesionales.

Segundo, La generación de un *dataset* de personalidad, intereses y experimentos en video, compuesto de 128 registros de participantes de varias áreas del conocimiento.

Tercero, se implementaron modelos de predicción de intereses profesionales y personalidad, entrenados con el *dataset* generado con PersonApp 2.0.

Por último, el desarrollo del sistema CareerHelp, que utiliza los modelos de predicción para predecir carreras profesionales afines a cada persona.

Como limitaciones se tiene la cantidad de datos recabada en el *dataset*, lo que dificulta el entrenamiento de los modelos de predicción. ya que se considera que aumentar considerablemente la muestra permitirá obtener mejores resultados.

6.3. Trabajo futuro

Como trabajo futuro, se propone aumentar el tamaño del *dataset* tanto de cuestionarios como de videos, esto con la finalidad de conseguir una muestra mejor balanceada y con suficientes registros para que los modelos puedan aprender. También se pueden mejorar los modelos de aprendizaje máquina, aplicando diversas técnicas para mitigar el problema regresión a la media.

Finalmente, se sugiere delimitar más el problema para enfocarse únicamente a la parte de predicción de intereses, utilizando instrumentos más cortos para la aplicación de los cuestionarios, esto facilitará conseguir más personas que participen en los experimentos, resultando en un *dataset* más grande con el cual poder trabajar.

Referencias

- Acosta-Amaya, M. M. La orientación profesional: Revisión del modelo RIASEC y la teoría social cognitiva del desarrollo de carrera. In Arias Cantor, Magda Yolima (Ed.), Colección Cuadernos de Ciencias Sociales - Investigación en Psicología (pp. 47-68). Fondo Editorial Universidad Católica de Oriente, 2018.
- Alfaro-Barquero, A., & Chinchilla-Brenes, S. (2019). Diseño de un instrumento de preferencias vocacionales en Administración, Materiales y Biotecnología. *Revista Costarricense de Psicología*, 38(2), 99-124.
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, 11(2), 150–166.
- Aslan, S., & Güdükbay, U. (2019). Multimodal video-based apparent personality recognition using long short-term memory and convolutional neural networks. arXiv preprint arXiv:1911.00381.
- Bátiz-Beltrán, V. M. (2021). Sistema de evaluación de la personalidad y de las emociones en el proceso cognitivo. Tesis de Maestría en Ciencias de la Computación. Instituto Tecnológico de Culiacán.
- Blanco, M., & Frutos, J. (2001). Orientación Vocacional: propuesta de un instrumento de autoorientación. *Educación y Futuro: Revista de Investigación Aplicada y Experiencias Educativas*, 4, 4.
- Celdrán, A. G. (1990). Técnicas para la orientación psicopedagógica. *Ciencias de la Educación Preescolar y Especial (CEPE)*.
- Cruz, R. Z. (2011). Revisión sistemática del cuestionario de personalidad de Eysenck (Eysenck Personality Questionnaire-EPQ). *Liberabit*, 17(2), 147–155.
- Developers Mozilla (A). (11 de julio de 2022). CSS. <https://developer.mozilla.org/es/docs/Web/CSS>.
- Developers Mozilla (B). (11 de julio de 2022). HTML. https://developer.mozilla.org/es/docs/Learn/Getting_started_with_the_web/HTML_basics.
- Developers Mozilla (C). (11 de julio de 2022). JavaScript. <https://developer.mozilla.org/es/docs/Web/JavaScript>.
- Duane, P., & Sydney, S. (2010). *Teorías de la personalidad*. São Paulo: Thomsom Learning.
- Durupinar, F., Pelechano, N., Allbeck, J., Güdükbay, U., & Badler, N. I. (2011). How the ocean personality model affects the perception of crowds. *IEEE Computer Graphics and Applications*, 31(3), 22–31.
- Ehrhart, K. H., & Makransky, G. (2007). Testing vocational interests and personality as predictors of person-vocation and person-job fit. *Journal of Career Assessment*, 15(2),

206–226.

- Firestore. (11 de julio de 2022). Firestore. <https://firebase.google.com/>.
- Flask Pallets Projects. (11 de julio de 2022). Flask. <https://flask.palletsprojects.com/en/2.1.x/>.
- Flores, L. Y., Spanierman, L. B., Armstrong, P. I., & Velez, A. D. (2006). Validity of the strong interest inventory and skills confidence inventory with Mexican American high school students. *Journal of Career Assessment*, 14(2), 183–202.
- Galilea, V. Orientación Vocacional. In Anónimo (Ed.), *Rehabilitación laboral de personas con enfermedad mental crónica: programas básicos de intervención* (pp. 73-82). Consejería de servicios sociales comunidad de Madrid, 2001.
- Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In CHI'11 extended abstracts on human factors in computing systems (pp. 253-262).
- Goldberg, X. (2009). Introduction to semi-supervised learning. In *Synthesis Lectures on Artificial Intelligence and Machine Learning* (Vol. 6).
- Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's "IPIP" Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39(2), 317–329.
- Hansen, J. I., & Neuman, J. L. (1999). Evidence of Concurrent Prediction of the Campbell Interest and Skill Survey (CISS) for College Major Selection. *Journal of Career Assessment*, 7(3), 239–247.
- Hansen, J.-I. C. (2019). Interest inventories. In G. Goldstein, D. N. Allen, & J. DeLuca (Eds.), *Handbook of psychological assessment* (pp. 169–190). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-802203-0.00006-7>.
- IBM Cloud Education (A). (2020). Convolutional Neural Networks. Recuperado el 19 de Julio de 2022 de <https://www.ibm.com/cloud/learn/convolutional-neural-networks>.
- IBM Cloud Education (B). (2020). Supervised Learning. Recuperado el 19 de Julio de 2022 de <https://www.ibm.com/cloud/learn/supervised-learning>.
- IBM Cloud Education (C). (2020). Unsupervised Learning. Recuperado el 19 de Julio de 2022 de [https://www.ibm.com/cloud/learn/unsupervised-learning?mhsrc=ibmsearch_a&mhq=unsupervised learning](https://www.ibm.com/cloud/learn/unsupervised-learning?mhsrc=ibmsearch_a&mhq=unsupervised%20learning).
- Jacques, J. C., Gucluturk, Y., Perez, M., Guclu, U., Andujar, C., Baro, X., Escalante, H. J., Guyon, I., van Gerven, M. A., van Lier, R., & Escalera, S. (2022). First Impressions: A Survey on Vision-Based Apparent Personality Trait Analysis. *IEEE Transactions on Affective Computing*, 13(1), 75–95.
- Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78–89.
- Lee, K., & Ashton, M. C. (2018). Psychometric Properties of the HEXACO-100. *Assessment*, 25(5), 543–556.

- Linares, R. N., & Hernández, M. M. (2017). Las estrategias para abordar el abandono escolar en una institución de educación superior tecnológica en México. In Congresos CLABES.
- Martínez, G. S., García, C. F., Sáinz, M., Prieto, M. F., Sánchez, M. D., García, M. R., & Hernández, D. (2011). Características psicométricas del cuestionario de personalidad BFQ-NA (Big Five Questionnaire-Niños y Adolescentes). *Aula Abierta*, 39(1), 13-24.
- Meta Open Source. (11 de julio de 2022). ReactJs. <https://es.reactjs.org/>.
- Microsoft. (11 de julio de 2022). Visualstudio. <https://code.visualstudio.com/>.
- Morales, P. P., & Monzón, A. F. (2014). La orientación vocacional, y su influencia en la elección de carrera universitaria de estudiantes que cursan el último grado de nivel diversificado, en una institución educativa privada.
- Naylor, F. D., & Kidd, G. J. (1991). The predictive validity of the Investigative scale of the Career Assessment Inventory. *Educational and psychological measurement*, 51(1), 217-226.
- Negnevitsky, M. (2005). *Artificial Intelligence: A Guide to Intelligent Systems*. In Pearson Education Limited (Vol. 2).
- Oracle México. (11 de julio de 2022). ¿Qué es la inteligencia artificial—IA?. <https://www.oracle.com/mx/artificial-intelligence/what-is-ai/>.
- Pakinee, A., & Puritat, K. (2021). Designing a gamified e-learning environment for teaching undergraduate ERP course based on big five personality traits. *Education and Information Technologies*, 26(4), 4049–4067.
- Patterson, J., & Gibson, A. (2017). *Deep learning: A practitioner's approach*. " O'Reilly Media, Inc."
- Pavithran, A., & Ashraf, Z. A. (2020). Identification of Career Interest using Text Mining Techniques. *Shabdbooks.Com*, IX(V), 27–33.
- Python Software Foundation. (11 de julio de 2022). Python. <https://www.python.org/>.
- Ramadhan, I., Akbar, H., Firmansyah, G., & Mulyo, A. (2021). Predictions of Six Personality Characters (HEXACO) from Social Media using Random Forest Classifier and Particle Swarm Optimization.
- Ramos, M. C., & González, B. A. (2020). Orientación Vocacional, Aprendizaje Socio-Emocional y Sentido de Vida en la Educación Superior. *Dilemas contemporáneos: educación, política y valores*, 8(SPE5).
- Real Academia Española. (s.f.). Cultura. En *Diccionario de la lengua española*. Recuperado en 8 de julio de 2022, de <https://dle.rae.es/inteligencia>.
- Reyes, J. (2013). Modelo de decisión multicriterio para la selección de carrera universitaria. *Investigación & Desarrollo*, 6(1), 25-32.
- Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2010). O* NET Interest Profiler Short Form

psychometric characteristics: summary and supporting evidence. Department of Labor O* NET Resource Center.

Secretaría de Educación Pública. (2020). Principales cifras, Sistema Educativo de los Estados Unidos Mexicanos. Secretaría de Educación Pública, 288.

Sinisterra, M. M., Cruz, J. P., & Gantiva, C. (2009). Teorías de la personalidad. Un análisis histórico del concepto y su medición. *Psychologia. Avances de la disciplina*, 3(2), 81-107.

Song, Q. C., Shin, H. J., Tang, C., Hanna, A., & Behrend, T. (2022). Investigating machine learning's capacity to enhance the prediction of career choices. *Personnel Psychology*.

Swamynathan, M. (2017). *Mastering Machine Learning with Python in Six Steps*. doi:10.1007/978-1-4842-2866-1

Tensorflow. (11 de julio de 2022). Tensorflow. <https://www.tensorflow.org/federated>.

Vidal Ledo, L. M., & Fernández Oliva, B. (2009). Orientación vocacional. *Educación Médica Superior*, 23(2), 0-0.

William, P., & Badholia, A. (2020). Evaluating Efficacy of Classification Algorithms on Personality Prediction Dataset. *Ilkogretim Online*, 19(4), 3362-3375.

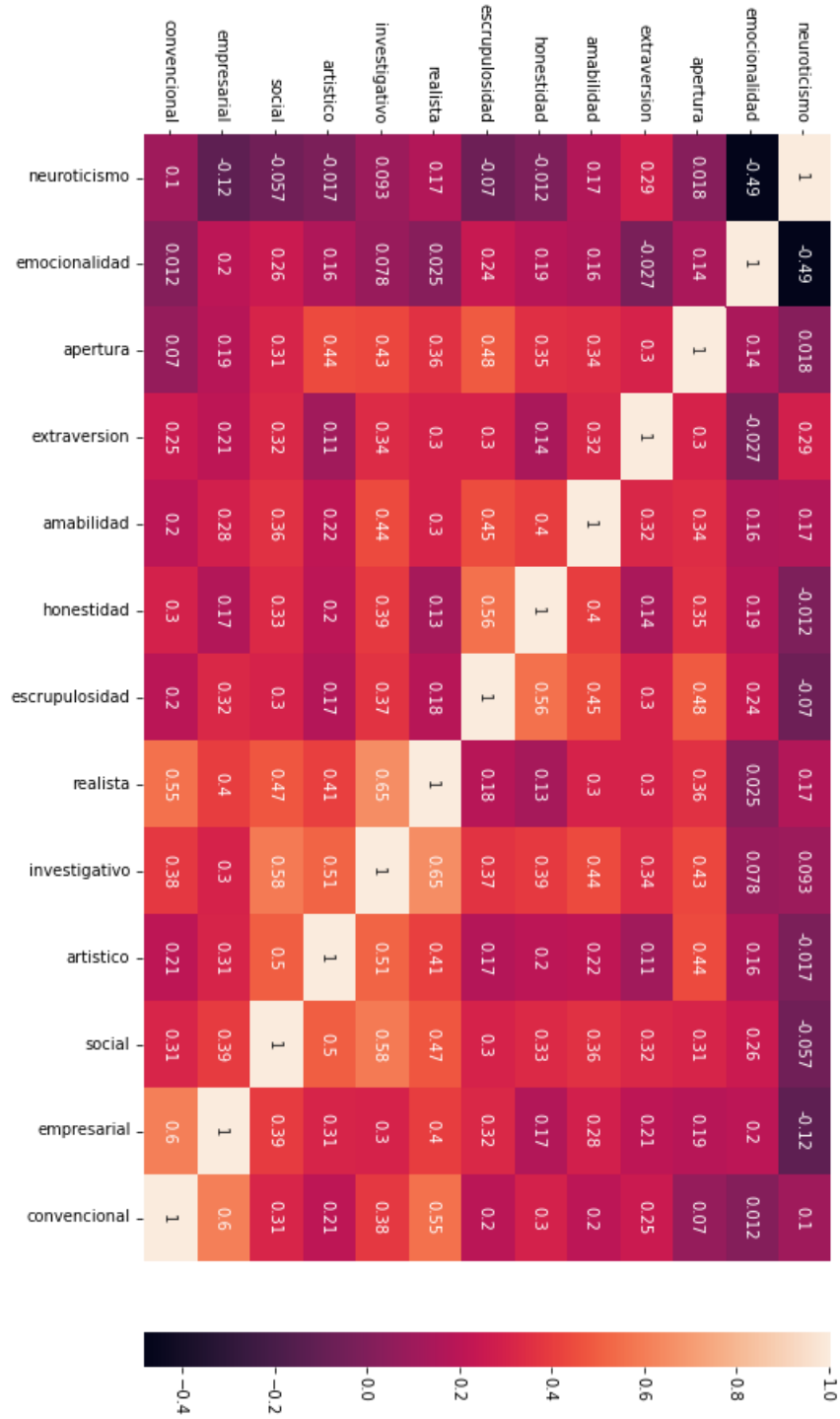
Xu, J., Tian, W., Lv, G., Liu, S., & Fan, Y. (2021). Prediction of the Big Five Personality Traits Using Static Facial Images of College Students with Different Academic Backgrounds. *IEEE Access*, 9, 76822–76832.

Yeye, W., Deyuan, C., Baobin, L., Xiaoyang, W., Xiaoqian, L., & Tingshao, Z. (2020). Predicting Personality based on Self-Introduction Video. *IFAC-PapersOnLine*, 53(5), 452–457.

Zatarain-Cabada, R., Barron-Estrada, M. L., Jair-Escalante, H., Cardenas-Lopez, H. Batiz-Beltran, V. (2022). Self-reported vs apparent personality recognition using deep learning in the wild.

Anexos

A. Tabla de correlaciones entre atributos



B. Casos de uso para el sistema PersonApp 2.0

CU-01. Presentar prueba estandarizada

CU-01. Presentar prueba estandarizada.	
Descripción	Proceso en el que un usuario contesta una prueba estandarizada.
Actores	Participante.
Precondiciones	<ul style="list-style-type: none">• Contar con conexión a internet.• Haber hecho log in en el sistema.
Escenario principal	<ol style="list-style-type: none">1. El usuario selecciona una prueba en el menú principal.2. El sistema abre una página con todas las preguntas del cuestionario.3. El usuario responde todas las preguntas requeridas y pulsa el botón enviar.4. El sistema muestra un mensaje de agradecimiento.
Postcondiciones	<ul style="list-style-type: none">• El sistema registra las respuestas del usuario en la base de datos.• El sistema redirige al usuario al menú principal.

CU-02. Registrar información académica y profesional

CU-02. Registrar información académica y profesional.	
Descripción	Proceso en el que un usuario registra su información académica o profesional en el sistema.
Actores	Participante.
Precondiciones	<ul style="list-style-type: none">• Contar con conexión a internet.• Haber hecho log in en el sistema.
Escenario principal	<ol style="list-style-type: none">1. El usuario escoge la prueba 1 (información académica/laboral).2. El sistema muestra una ventana flotante donde pregunta si el usuario es un estudiante actualmente.3. El usuario selecciona su respuesta de la caja de opciones.4. El sistema muestra las preguntas correspondientes a la opción elegida.5. El usuario contesta las preguntas requeridas y pulsa enviar.
Postcondiciones	<ul style="list-style-type: none">• El sistema registra las respuestas del usuario en la base de datos.• El sistema redirige al usuario al menú principal.

CU-03. Visualizar pruebas completadas.

CU-03. Visualizar pruebas completadas.	
Descripción	Proceso en el que un usuario visualiza las pruebas que ha contestado.
Actores	Participante.
Precondiciones	<ul style="list-style-type: none">• Contar con conexión a internet.• Haber hecho log in en el sistema.
Escenario principal	<ol style="list-style-type: none">1. El usuario accede al menú principal.2. El sistema muestra las pruebas, coloreando de color verde las completadas.
Postcondiciones	

CU-04. Grabar video.

CU-04. Grabar video.	
Descripción	Proceso en el que un usuario graba experimentos en video.
Actores	Participante.
Precondiciones	<ul style="list-style-type: none">• Contar con conexión a internet.• Haber hecho log in en el sistema.
Escenario principal	<ol style="list-style-type: none">1. El usuario selecciona la prueba 4.2. El sistema muestra una pantalla con la explicación de los experimentos.3. El usuario selecciona un experimento.4. El sistema muestra una pantalla con un cuadro de video, las instrucciones y un botón de empezar.5. El usuario pulsa el botón iniciar.6. El sistema comienza la grabación de video.7. El usuario realiza el experimento.8. El sistema muestra un mensaje de agradecimiento.
Postcondiciones	<ul style="list-style-type: none">• El sistema sube el video a la base de datos.• El sistema registra la prueba como completada.• El sistema redirige al usuario al menú principal.