Corbot

Bernardo Arturo Garcia Espinosa. Francisco José Velasco Casillas, Lía Dayana Zúñiga Medina, Ramiro Lupercio Coronel.

*CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍAS, (CUCEI, UDG)*

bernardo.garcia@alumnos.udg.mx

francisco.vcasillas@alumnos.udg.mx lia.zuniga4454@alumnos.udg.mx

ramiro.lupercio3413@academicos.udg.mx

**En los años recientes, el desarrollo del campo de las ciencias computacionales ha tenido un incremento significativo, impactando positivamente otros sectores, tales como la salud, proporcionando herramientas tanto para la manufactura de medicamentos, como para el diagnóstico y tratamiento de enfermedades. CorBot se presenta como una aplicación de consulta mediante chat, ofreciendo un diagnóstico preliminar a padecimientos cardiovasculares de manera accesible, rápida y gratuita apoyándose de tecnologías de IA para el análisis y predicción de riesgos mediante el uso de una extensa base de datos, con el propósito de generar un impacto positivo en la salud mexicana.**

**Enfermedades cardiovasculares, diagnósticos preliminares, Chat/Mensajería, redes neuronales profundas, NoSQL.**

**Repositorio de código: ttps://github.com/ArtureishonSt/CorBot\_Modular\_2024A.git**

**Versión actual del código: 1.8.5**

**Licencia legal código: Propietario**

I. INTRODUCCIÓN

La salud es considerada uno de los problemas sociales de mayor relevancia y campo de acción, a su vez es una de las áreas que más crecimiento ha tenido en los últimos años, apoyándose del desarrollo de otras áreas de la ciencia como química [Ref], electrónica [Ref], computación [Ref], robótica [Ref], entre otras. Estos avances han permitido mejorar en la prevención, detección y tratamiento de diversas enfermedades. De acuerdo con la Secretaría de Salud de México, las enfermedades relacionadas con afecciones cardíacas representan alrededor 220 mil fallecimientos anuales en el país [1]. Por otra parte, y pese a estos números, se ha demostrado que una consulta temprana puede reducir los riesgos de un padecimiento cardiovascular un X% [Ref], dependiendo el caso. Sin embargo, factores como la distancia, disponibilidad, y costos de las consultas genera que algunos sectores de la población desistan de realizarse un diagnóstico temprano ante la incertidumbre de si costo-beneficio es redituable.

En este contexto que surge el proyecto, "CorBot", una herramienta de chat que implementa diferentes tecnologías de software que mejoren la accesibilidad en el ámbito de la atención médica, solucionando los problemas previamente mencionados.

II. TRABAJOS RELACIONADOS

Uno de los sectores que más se ha beneficiado de los avances tecnológicos, es la medicina [Ref]. En los los últimos años el desarrollo en ciencias de computación ha permitido

La idea de apoyarse herramientas computacionales para

Existen varias aplicaciones de chat que utilizan inteligencia artificial para el diagnóstico preventivo en México. Algunas de las aplicaciones más populares son IMPAI [4], GitMind Chat [5] y Google Bard [6]. IMPAI es un Chatbot impulsado por un internista de la Sociedad Española de Medicina Interna que facilita el diagnóstico y orienta sobre la fase de infección por SARS-CoV-2 a través de una serie de preguntas clave a los potenciales pacientes. GitMind Chat es un chatbot equipado con la última tecnología LLM e impulsado por inteligencia artificial, que ofrece a individuos y empresas una variedad y gama integral de servicios de inteligencia artificial. Google Bard es una herramienta de IA generativa que se puede utilizar de forma gratuita. Sin embargo, tienen varias desventajas desde que son desconocidas, de difícil acceso, no confiables o de paga.

III. DESCRIPCIÓN DEL DESARROLLO DEL PROYECTO MODULAR

El proyecto CorBot tiene como objetivo principal el de proporcionar una herramienta confiable y de fácil acceso para la consulta y diagnóstico de enfermedades cardiovasculares. Con este propósito en mente, se plantearon objetivos específicos a menor escala que permitieran llevar a cabo el objetivo final del proyecto:

* Generar un sistema de chat que el paciente pueda consultar de manera fácil en cualquier momento.
* Obtener una fuente de información confiable para realizar los análisis y predicciones correspondientes.
* Implementar herramientas de Machine Learning que sean capaces de interpretar de manera eficaz los datos obtenidos.
* Almacenar la información obtenida de los usuarios de manera confiable, tal que análisis posteriores para la mejora de la plataforma sean posibles.

Con base en lo anteriormente mencionado, CorBot fue construido sobre las bases de tres conceptos fundamentales: un sistema de consulta accesible, una gestión segura de la información y su almacenamiento, y un nivel de predicción confiable.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Fig. 1 Proyecto CorBot

1. *Sistema de consulta*

Para el sistema de consulta se emplearon aplicaciones de mensajería como WhatsApp, Telegram

Las ventajas que se obtienen mediante estas aplicaciones de mensajería es:

* Simplicidad
* Accesibilidad
* Difusión
* Customización

Considerando estas facilidades, se diseña un formulario que sirva para llenar los campos de la base de datos.

De manera interna estos datos son procesados para que entren al modelo de IA en el formato adecuado.

Finalmente, la predicción de salida es dada con valor probabilístico, dado que la finalidad no es reemplazar la asistencia médica sino concientizar al usuario de su salud cardiovascular, sin mencionar que los datos permiten evaluar el riesgo, no determinar una enfermedad en específico, al igual que haría un doctor.

1. *Información y almacenamiento*

Este punto puede ser dividido en dos subsecciones, que consisten en la calidad y estructura de los datos obtenidos, y como segundo punto el método de almacenamiento.

La información empleada, para el análisis y posterior entrenamiento de la IA fue obtenida del Centro para el Control y Prevención de Enfermedades (CDC) [Ref], y está compuesta por 319,795 registros, y 18 parámetros que corresponden a la información de cada paciente. Los parámetros que se toman en consideración son:

* Sexo: Variable binaria que indica a qué genero pertenece el paciente (masculino o femenino).
* Edad: Variable Categórica que indica a que grupo de edad pertenece la persona. Está dividida en grupos de 5 años desde 18-24 años hasta 85+ años.
* BMI: Variable numérica que indica el índice de masa corporal (BMI) del paciente.
* Etnia: Variable Categórica que indica a qué etnia pertenece el paciente (caucásico, afroamericano, latino, asiático, indio americano, otro).
* Fumador: Variable binaria que indica si el paciente es fumador (si, no).
* Bebedor: Variable binaria que indica si el paciente es bebedor (si, no).
* Ejercicio: Variable binaria que indica si el paciente se ejercita (si, no).
* Horas de sueño: Variable numérica que indica cuantas horas de sueño tiene el paciente.
* Malestares de salud física: Variable numérica que indica cuantos días del último mes el paciente ha presentado molestias en su salud física.
* Malestares de salud mental: Variable numérica que indica cuantos días del último mes el paciente ha presentado molestias en su salud mental.
* Dificultad para caminar: Variable binaria que indica si al paciente se le dificulta caminar (si, no).
* Salud general: Variable Categórica que indica como el paciente considera su propia salud (excelente, muy buena, buena, regular, mala)
* Infarto: Variable binaria que indica si el paciente ha sufrido de un paro cardíaco (si, no).
* Asmático: Variable binaria que indica si el paciente es asmático (si, no).
* Diabetes: Variable binaria que indica si el paciente padece de diabetes (si, no).
* Enfermedad del riñón: Variable binaria que indica ha padecido de alguna enfermedad del riñón (si, no).
* Cáncer de piel: Variable binaria que indica si el paciente ha padecido de cáncer de piel (si, no).
* Enfermedad Cardíaca: Variable binaria que indica padece o ha padecido de alguna enfermedad del corazón. Esta es la variable objetivo, y es comparada con la salida del modelo de IA para determinar el error.

Una ventaja importante que tiene esta base de datos sobre otras bases de datos es que todos los campos, a excepción del BMI, son de conocimiento generalmente para el usuario. Por su parte, la información del BMI se obtiene de manera indirecta mediante el peso, la estatura y las semanas de embarazo (en caso de que aplique) del usuario.

Adicionalmente, con el objetivo de mantener esta información segura, al igual que la nueva información recopilada durante las consultas, ambos conjuntos de datos son almacenados en una base de datos confiable y segura. El gestor de base de datos implementado para esto es MongoDB, el cual es una base de datos NoSQL que permite ciertas bondades en relación con una base de datos SQL tradicional:

* Mayor adaptabilidad a las reestructuras y cambios en el diseño y arquitectura de la base de datos.
* Los datos que se almacenan no están restringidos a un formato específico.
* Escala horizontalmente, lo cual implica que se puede aumentar la capacidad y potencia de almacenamiento al añadir más nodos al sistema.

Adicionalmente, MongoDB permite un almacenamiento en la nube, mediante plataformas como AWS y Azure, lo que permite mantener la información segura y respaldada a un bajo costo.

1. *Predicción de diagnóstico*

Finalmente, el proceso con el que CorBot realiza una predicción del diagnóstico es mediante un modelo de redes neuronales profundas (RNP) con tres capas intermedias. La Figura 2 muestra una versión simplificada de la RNP implementada.

Diagrama, Gráfico de burbujas

Descripción generada automáticamente

Fig. 2. Esquema simplificado de la red neuronal profunda implementada.

Las funciones de activación entre cada capa, tanto las capas ocultas, como las capas de entrada y salida, están construidas de la misma manera, y consisten en las siguientes funciones.

**Sigmoidal.** La función sigmoidal (Figura 3), también conocida como función probabilística [Ref] está determinada mediante la ecuación

Donde el exponente corresponde a la forma base de la función lineal. Si se explora la función a detalle se puede apreciar que para el rango de valores en la parte exponencial, el dominio de la función oscila entre .

Imagen que contiene Diagrama

Descripción generada automáticamente

Fig. 3. Función Sigmoide cuyo exponente es igual a .

De esta manera, los valores de la función se mantienen en un rango probabilístico, mientras que la función lineal perteneciente a la parte exponencial de la función sigmoidal se encarga de ajustar los pesos para que la salida del modelo coincida con el valor deseado.

**Batch normalization**. Una vez obtenidos los valores de la sigmoide, estos son normalizados con la finalidad de asegurar que los datos queden completamente distribuidos en el dominio de la función. Esta normalización se realiza mediante batch [Ref], que tiene la ventaja

**Dropout.** El “Dropout” es una técnica ampliamente utilizada para evitar el sobre entrenamiento de la RNP [Ref]. Este proceso consiste en “apagar” un porcentaje de las neuronas en cada capa con la finalidad de balancear la carga e importancia que cada neurona ejerce sobre el valor de salidas, evitando tanto la sobrecarga de trabajo como la inactividad de este. Este proceso de “apagar” neuronas no es estático, y cambia con cada iteración.

Una vez establecida la arquitectura de la RNP, es necesario establecer los pesos de la red neuronal que proporcionen una salida satisfactoria. Para ello, el conjunto de datos es dividido en dos partes, el conjunto de entrenamiento, y el conjunto de validación en una proporción de 70-30 respectivamente. El primer conjunto se implementa para el entrenamiento de la red neuronal empleando la técnica de retro propagación [Ref] por medio de mini-batch [Ref] durante 300 épocas.

Diagrama

Descripción generada automáticamente

Fig. 4. RNN sin dropout (izquierda) versus RNN con dropout (derecha).

La Figura 5 muestra la disminución del error de raíz cuadrada media (RMSE) tras cada época llegando a un valor final de 0.2706.

Por su parte, el conjunto de validación fue empleado para verificar que la RNP funcionase con datos no observados previamente. Empleando la misma métrica, el conjunto de validación obtuvo un RSME de 0.2683, que dado su similitud con el RSME del conjunto de entrenamiento se acepta como valida siendo apta para generalizar información desconocida.

Gráfico, Histograma

Descripción generada automáticamente

Fig. 5. Disminución del RSME durante el entrenamiento.

*Modulo I Justificación de Arquitectura Programación de Sistemas*

El proyecto CorBot emplea Python 3 como lenguaje de programación principal debido a su versatilidad y amplio soporte en el desarrollo de aplicaciones.

* 1. CorBot utiliza MongoDB como base de datos, aprovechando su capacidad para manejar grandes volúmenes de datos no estructurados y su integración eficiente con aplicaciones Python.
  2. El equipo ha optado por la metodología de desarrollo conocida como "SCRUM". Esta metodología se caracteriza por su enfoque incremental, donde cada fase del desarrollo añade funcionalidades y mejoras de manera secuencial, permitiendo una adaptación flexible a cambios y una entrega temprana de versiones operativas.
  3. La elección de Python se justifica por su legibilidad y mantenibilidad, elementos cruciales en el desarrollo a largo plazo. La utilización de MongoDB se respalda por su capacidad de escalabilidad y su estructura de documentos flexible. La metodología en escalera se selecciona para asegurar una evolución gradual y controlada del proyecto.
  4. CorBot puede involucrar la creación de diagramas de flujo que ilustren las interacciones entre el usuario y el sistema, así como los procesos internos del CorBot para un entendimiento claro de su funcionamiento.Principio del formulario

*Modulo II Justificación de Sistemas Inteligentes*

* 1. El proyecto incorpora redes neuronales para potenciar la capacidad de aprendizaje y adaptabilidad del CorBot, permitiéndole mejorar sus respuestas y comportamientos a lo largo del tiempo.
  2. Se implementa aprendizaje automático para que el CorBot pueda analizar patrones en las interacciones pasadas, optimizando así sus respuestas futuras y adaptándose a las preferencias de los usuarios.
  3. CorBot opera como un asistente virtual, proporcionando información, realizando tareas específicas y brindando asistencia personalizada mediante la integración de inteligencia artificial.
  4. En el caso de las redes neuronales y el aprendizaje automático, se emplean modelos matemáticos complejos, como las funciones de activación en redes neuronales o las ecuaciones que gobiernan los algoritmos de aprendizaje automático. El asistente virtual también se basa en modelos matemáticos para comprender y responder de manera efectiva a las consultas de los usuarios.
  5. La elección de utilizar redes neuronales se justifica por su capacidad para procesar datos complejos y aprender patrones, lo que mejora la calidad de las respuestas del CorBot con el tiempo. El aprendizaje automático se selecciona para permitir al CorBot adaptarse a nuevas situaciones y optimizar su desempeño continuamente. En cuanto al asistente virtual, se eligen algoritmos que permiten una comprensión semántica avanzada para interpretar y responder de manera coherente a las consultas de los usuarios.

Principio del formulario

*Modulo III Justificación de Sistemas Distribuidos*

* 1. El sistema CorBot implementa un enfoque descentralizado para distribuir el procesamiento de cálculos, permitiendo un rendimiento eficiente y escalable al procesar las consultas de los usuarios.
  2. En el contexto de CorBot, se implementa un modelo cliente-servidor para gestionar las solicitudes de los usuarios y proporcionar respuestas a través de una interfaz amigable en de chat/mensajero.
  3. Se establece una comunicación efectiva entre el dispositivo del usuario (cliente) y el servidor que aloja el CorBot, posibilitando la interacción bidireccional y el intercambio de información.
  4. La elección de implementar un sistema distribuido radica en la necesidad de ofrecer un entorno de interacción descentralizado entre el usuario y CorBot. Esta arquitectura distribuida permite la compartición eficiente de recursos y la gestión de las interacciones de manera escalonada. La selección de protocolos como el de chat/mensajero se fundamenta en su capacidad para habilitar una comunicación ágil y accesible en tiempo real, aspectos esenciales para potenciar la eficacia y la adaptabilidad de nuestro sistema distribuido.

IV. RESULTADOS OBTENIDOS DEL PROYECTO

***Implementación Exitosa de Tecnologías de Inteligencia Artificial:***

La integración de redes neuronales Machine Learning y la funcionalidad de asistente virtual en el proyecto CorBot ha resultado exitosa, ofreciendo respuestas más precisas y adaptativas a medida que se interactúa con los usuarios.

1. **Sistema Distribuido Eficiente:** La adopción de un modelo descentralizado ha demostrado ser eficaz en la distribución del procesamiento de cálculos, permitiendo un rendimiento escalable y mejorando la capacidad de respuesta del CorBot.
2. **Interacción Usuario-Servidor Fluida:** La implementación del modelo cliente-servidor facilita una interacción fluida entre los usuarios y el CorBot a través de chat/mensajero, proporcionando una experiencia de usuario intuitiva y accesible.
3. **Optimización Continua con Aprendizaje Automático:** El uso de algoritmos de aprendizaje automático ha permitido al CorBot analizar patrones de interacción y optimizar sus respuestas con el tiempo, mejorando la calidad del servicio ofrecido.
4. **Seguridad Garantizada mediante Protocolos Establecidos:** La elección de protocolos, tales como chat/mensajero ha garantizado la seguridad en la transmisión de datos sensibles, resultando en una experiencia segura y protegida para los usuarios.

Los objetivos previamente planteados permiten que Corbot se diseñe como una herramienta eficiente, adaptable y segura que satisface las necesidades de los usuarios mientras demuestra la viabilidad de las tecnologías de inteligencia artificial en aplicaciones prácticas.

V. CONCLUSIONES Y TRABAJO A FUTURO

***Conclusiones:***

**Eficiencia en la Implementación de Tecnologías AI:** La exitosa implementación de tecnologías de inteligencia artificial en CorBot demuestra su eficacia en mejorar la calidad de las respuestas y adaptarse a las necesidades cambiantes de los usuarios.

**Distribución Descentralizada para Rendimiento Óptimo:** La adopción de un modelo descentralizado ha resultado en un procesamiento de cálculos eficiente, mejorando la capacidad de respuesta del CorBot y garantizando una experiencia de usuario mejorada.

**Interacción Segura y Accesible:** La combinación de protocolos seguros y la interfaz amigable de chat/mensajero resultan en una interacción segura y accesible entre los usuarios y el CorBot.

***Trabajo a Futuro:***

**Mejora Continua con Retroalimentación del Usuario:** Implementar un sistema de retroalimentación para que los usuarios puedan evaluar y proporcionar comentarios al CorBot, permitiendo mejoras continúas basadas en las necesidades y preferencias de los usuarios.

**Explorar Nuevas Funcionalidades AI:** Investigar y añadir nuevas funcionalidades de inteligencia artificial, como procesamiento de lenguaje natural mejorado, reconocimiento de voz y análisis de sentimientos, para enriquecer la experiencia del usuario.

**Ampliación de Plataformas de Mensajería:** Extender la compatibilidad del CorBot a otras plataformas de mensajería popular, aumentando su accesibilidad y alcance.

**Integración de Tecnologías Emergentes:** Mantenerse al tanto de las tendencias tecnológicas y evaluar la integración de tecnologías emergentes para mantener el CorBot actualizado y competitivo en el futuro.

En conclusión, el proyecto proporciona una base sólida para un CorBot efectivo y adaptable, mientras que las áreas identificadas para el trabajo futuro permitirán mantenerlo a la vanguardia de las expectativas y avances tecnológicos.

RECONOCIMIENTOS

Quisiéramos expresar nuestro profundo agradecimiento a todas las personas que han contribuido a este proyecto:

* **Equipo de Desarrollo:** A cada miembro del equipo cuya dedicación y arduo trabajo fueron y siguen siendo esenciales para la realización de este proyecto. Su colaboración y habilidades individuales fueron fundamentales.
* **Asesores y Mentores:** Agradecemos sinceramente a nuestros asesores y mentores, cuya orientación y experiencia fueron invaluables durante todo el proceso. Sus consejos y retroalimentación contribuyeron y contribuyen significativamente al desarrollo exitoso del proyecto.
* **Usuarios Beta:** Agradecemos a los usuarios beta que participaron activamente en las pruebas y proporcionaron valiosos comentarios. Su involucramiento fue esencial para perfeccionar el rendimiento y la usabilidad del CorBot.
* **Institución Académica:** Agradecemos a nuestra institución académica por proporcionar los recursos y el entorno propicio para llevar a cabo este proyecto. Su apoyo continuo ha sido fundamental para el logro de nuestros objetivos.

Este reconocimiento no solo celebra los logros tangibles, sino también el esfuerzo y la colaboración que hicieron y hacen posible este proyecto. A todos, ¡muchas gracias!

REFERENCIAS

[1] Secretaría de Salud. (2022, septiembre 28). [Comunicado de prensa]. gob.mx. <https://www.gob.mx/salud/prensa/490-cada-ano-220-mil-personas-fallecen-debido-a-enfermedades-del-corazon>

[2] TN.com.ar. (2021, 30 de septiembre). Las enfermedades cardiovasculares siguen liderando el ranking mundial de mortalidad de adultos. <https://tn.com.ar/salud/noticias/2021/09/30/las-enfermedades-cardiovasculares-siguen-liderando-el-ranking-mundial-de-mortalidad-de-adultos/>

[3] Organización Mundial de la Salud [OMS]. (2019, 4 de abril). El acceso desigual a los servicios de salud genera diferencias en la esperanza de vida [Comunicado de prensa]. Recuperado de <https://www.who.int/es/news/item/04-04-2019-uneven-access-to-health-services-drives-life-expectancy-gaps-who>

[4] IMPAI: Sociedad Española de Medicina Interna. (2022). IMPAI. https://www.fesemi.org/informacion/prensa/semi/impai-un-chatbot-que-impulsa-un-internista-de-semi-facilita-el-diagnostico-y

[5] GitMind Chat: GitMind. (2022). GitMind Chat. https://gitmind.com/faq/use-gitmind-on-mobile-phone.html

[6] Google Bard: Google. (2022). Bard. https://bard.google.com/