BLOCKCHAIN® IN HEALTHCARE TODAY
PLATFORM APPROACHES IN HEALTHCARE
NOTAS TÉCNICAS E INFORMES BREVES

Model Complexity Reduction for ZKML Healthcare Applications (Reducción de la complejidad del modelo para aplicaciones sanitarias ZKML): Protección de la privacidad y optimización de la inferencia para aplicaciones ZKML: una implementación de referencia con el conjunto de datos sintéticos ICHOM

Sathya Krishnasamy, MS;[1] e Ilangovan Govindarajan, MD[(2)]

[1]Presidente y Director, ChainAim, Newington, Connecticut, EE.UU., [2]Director Médico, GuardianMedx, Las Vegas, Nevada Autor

correspondiente: Sathya Krishnasamy, Correo electrónico: sathya.krishnasamy@chainaim.com

DOI: https://doi.org/10.30953/bhty.v7.340

Palabras clave: blockchain, diabetes, libro mayor distribuido, modelo de reducción de la complejidad, privacidad, aprendizaje automático, ICHOM, Consorcio Internacional para la Medición de los Resultados Sanitarios, ZKML, aprendizaje automático de conocimiento cero.

Resumen

La Web 3.0 representa la próxima evolución significativa de Internet que incorpora las arquitecturas de red descentralizadas subyacentes, los libros de contabilidad distribuidos y las capacidades avanzadas de IA. Aunque las tecnologías están madurando rápidamente, existen barreras considerables para su adopción a gran escala. El autor analiza las barreras y las mitigaciones a través de tecnologías específicas que están madurando para resolver esos problemas en un documento anterior titulado Moving Beyond POCs and Pilots, publicado en 2023 en Blockchain in Healthcare Today. Entre ellas se incluyen las tecnologías de preservación de la privacidad, las optimizaciones de diseño fuera y dentro de la cadena, y el enfoque multidimensional necesario para planificar y adoptar estas tecnologías. A modo de ampliación, este artículo analiza uno de estos facilitadores, el aprendizaje automático de conocimiento cero (ZKML), que fusiona dos corrientes tecnológicas de forma única para abordar los problemas de la privacidad y el coste de la inferencia. Las pruebas de conocimiento cero (ZKP) permiten a una parte demostrar la validez de una afirmación a otra parte sin revelar ninguna información adicional sobre la propia afirmación. El ZKML combina el principio criptográfico del ZKP con técnicas de aprendizaje automático (ML). Todavía es una tecnología en fase de maduración y necesita líneas de base para su aplicación en la sanidad mundial. En este trabajo, los autores conceptualizan la viabilidad técnica y operativa del uso de ZKML e implementan una aplicación sanitaria de referencia utilizando el Consorcio Internacional para la Medición de Resultados Sanitarios (ICHOM) sintético en la fase de evaluación en un entorno sanitario global para la recopilación de datos de gran volumen, incluidos los resultados notificados por los pacientes. Se investiga e informa sobre la reducción de la complejidad del modelo para el conjunto de datos de diabetes del ICHOM con el fin de avanzar en el uso de modelos ML en estándares globales de recopilación de datos sanitarios en arquitecturas descentralizadas de red para una mayor protección y eficiencia de los datos.

Presentado: 1 de agosto de 2024; Aceptado: 25 de agosto de 2024; Publicado: 31 de agosto de 2024

Ta gran cantidad de datos recopilados en los últimos años ha generado una capacidad analítica sin precedentes. Sin embargo, con el rápido aumento de
a ingestión de datos y el aprendizaje automático (machine learning, ML), especialmente en sistemas centralizados, las preocupaciones sobre la privacidad y la seguridad de los datos han aumentado significativamente con los excesivos movimientos de datos, a veces no realmente necesarios. A medida que aumentan los repositorios centrales de datos, se convierten en objetivos atractivos. Las filtraciones de datos sanitarios son cada vez más frecuentes.

Apéndice

comunes. Se han producido muchos incidentes de gran repercusión, y la tendencia ha sido continua.

Una estrategia clave para mitigar los riesgos asociados a los datos sanitarios es restringir el acceso a los datos en origen y reducir los movimientos innecesarios de datos. Esto incluye el acceso a los datos mediante un acceso basado en funciones, el cifrado de datos, la transferencia mínima de datos y auditorías periódicas. Aunque estas medidas son primordiales, también es fundamental idear sistemas de colaboración que puedan comunicarse a través de redes descentralizadas. Las pruebas de conocimiento cero (ZKP) pueden diseñarse para pruebas y verificación entre sistemas, incluso antes que las cadenas de bloques. Estos

Las ZKP permiten a una parte demostrar la validez de una afirmación a otra parte sin revelar ninguna información adicional sobre la propia afirmación y han ido ganando popularidad a medida que la tecnología de libro mayor distribuido madura y ha ayudado a ampliar las implementaciones de blockchain.

El aprendizaje automático de conocimiento cero (ZKML) representa una fusión revolucionaria de las tecnologías criptográfica y de LD. Combina el principio criptográfico de ZKP con técnicas de ML. Al integrar la ZKP con el ML, el ZKML garantiza la confidencialidad de los datos sensibles, al tiempo que permite el desarrollo y la utilización de modelos predictivos. Esta integración es cada vez más relevante cuando la privacidad y la seguridad de los datos son primordiales, como en los sistemas sanitarios.

En la práctica, ZKML permite que varias entidades colaboren sin comprometer la confidencialidad de su información privada. Esto significa que las organizaciones pueden entrenar y utilizar modelos ML de forma colaborativa en conjuntos de datos privados sin exponer los datos. Por ejemplo, una organización de investigación médica podría agregar datos de varios hospitales para desarrollar un sólido modelo predictivo de enfermedades sin que ningún hospital comparta sus datos de pacientes. El uso de pruebas criptográficas garantiza que los datos permanezcan seguros y privados durante todo el proceso.

En la actualidad, el ZKML se encuentra en una fase incipiente de investigación y desarrollo. Aunque los ZKP han sido objeto de investigación criptográfica desde los años 80, su aplicación al ML es nueva y compleja. Esta tecnología se enfrenta a retos relacionados con la eficiencia computacional y la demanda de recursos. La aplicación de las ZKP puede requerir muchos recursos, lo que aumenta el tiempo de procesamiento y los costes.

La reducción de la complejidad de los modelos es una técnica reciente para reducir la complejidad de los modelos y, por tanto, los tiempos de cálculo. Se ha intentado utilizando modelos más sencillos de Kaggle. El Consorcio Internacional para la Medición de los Resultados en Salud (ICHOM) se dedica a desarrollar conjuntos estándar de medidas de resultados que puedan utilizarse a escala mundial para evaluar la calidad de la atención en diversas afecciones médicas. El objetivo de esta investigación es evaluar el uso de un conjunto de datos sanitario estándar global (ICHOM) y aplicar la reducción de complejidad de modelos para un ejemplo no trivial del mundo real a un conjunto de datos sintético en el esquema ICHOM. De este modo, esta investigación servirá de referencia para desarrollar nuevos modelos y optimizaciones que permitan optimizar la complejidad y el uso de ZKML para muchos casos de uso en la atención sanitaria que requieren privacidad y colaboración en la toma de decisiones.

Un cambio de paradigma: Sistemas descentralizados, modelos de IA en origen y toma de decisiones colaborativa

A medida que nos acercamos a las arquitecturas emergentes construidas sobre sistemas descentralizados, resulta esencial restringir los datos en origen y comprender las formas de trabajar con los datos en origen a través de un ML protegido de la privacidad. Los datos son

Por ello, se necesitan mecanismos eficaces que permitan el aprendizaje federado, aborden los problemas de privacidad y seguridad y reduzcan la carga computacional. Al compartir actualizaciones de modelos en lugar de datos, el aprendizaje federado mejora la privacidad. Del mismo modo, con los libros de contabilidad distribuidos, un concepto de diseño cada vez más popular es proporcionar pruebas a los sistemas de verificación a través de ZKP. Aunque no es un concepto nuevo, este método ha escalado con éxito los sistemas blockchain y se está haciendo cada vez más popular para los ledgers distribuidos de próxima generación. Estos mecanismos también reducen el riesgo de transmisión de datos y se ajustan a la normativa sobre protección de datos y toma de decisiones colaborativa. El ZKML es una prometedora tecnología emergente que integra tecnologías de IA/ML (inteligencia artificial/aprendizaje automático) en los libros de contabilidad distribuidos.

Impacto del ZKML en la privacidad de la sanidad

Los datos sanitarios son sensibles, por lo que la privacidad es el primer principio de diseño de los sistemas sanitarios. Con ZKML, los proveedores de asistencia sanitaria pueden compartir información derivada de modelos ML sin exponer los datos subyacentes de los pacientes. Por ejemplo, un hospital podría utilizar un modelo ZKML para predecir los resultados de los pacientes. El modelo procesa los datos y genera predicciones, mientras que el ZKP garantiza que estas predicciones sean precisas sin revelar información específica del paciente. ZKML permite a distintas entidades calcular resultados en colaboración sin revelar sus datos individuales. Esto resulta especialmente útil en la investigación sanitaria, donde varias instituciones pueden desear combinar sus datos para mejorar las predicciones de enfermedades sin comprometer la confidencialidad de los pacientes. A medida que se intensifica la preocupación por la privacidad de los datos y crece la necesidad de modelos de ML seguros, resulta crucial integrar tecnologías avanzadas de preservación de la privacidad.

Bibliografía

Las ZKP son métodos criptográficos que permiten a una parte (el prover) convencer a otra (el verifier) de que una afirmación es verdadera sin revelar ninguna información adicional más allá de la validez de la propia afirmación. Introducidos por Goldwasser, Micali y Rackoff (1985)[1], establecieron el marco teórico de las pruebas interactivas y las ZKP, demostrando su viabilidad e importancia fundacional.

Trabajos posteriores de Fiat y Shamir[2]lo extendieron a las ZKP no interactivas, que no requieren múltiples rondas de comunicación entre el probador y el verificador. Avances recientes en ZKP, como los Argumentos de Conocimiento Cero Sucintos No Interactivos (zk-SNARKs) de Ben-Sasson y colegas[3,4]han mejorado su eficiencia y escalabilidad y han abierto posibilidades para aplicaciones en el mundo real.

Recientemente, el ML ha demostrado su éxito en el modelado de [tareas de predicción sanitaria, que van desde el diagnóstico de enfermedades y la

Cita: Blockchain en la sanidad hoy 2024, 7: 340 - https://doi.org/10.30953/bhty.v7.340

pronóstico al tratamiento del paciente. Guerra y sus colegas[5]revisaron la bibliografía sobre ML que preserva la privacidad para el entrenamiento y la inferencia, y concluyeron que los conjuntos de datos sanitarios son diversos y que una fracción de ellos se validan con conjuntos de datos estándar independientes. Señalaron los riesgos de la formación centralizada para el aprendizaje federado y también la necesidad de colaboración entre diferentes entidades a través de múltiples roles de científicos de ML, profesionales de la salud y expertos en privacidad y seguridad, que necesitan mecanismos de preservación de la privacidad para trabajar juntos en libros de contabilidad distribuidos.

ZKML, una tecnología de reciente aparición, aplica ZKP a ML para la privacidad, garantizando que los datos sensibles sigan siendo confidenciales al tiempo que permite desarrollar y utilizar modelos predictivos para la privacidad y la colaboración. Las ZKP, como tales, son costosas computacionalmente y se vuelven aún más costosas computacionalmente para las pruebas de inferencia de ML. Un trabajo reciente de Alejandro Martínez Gator[6]ha creado una biblioteca de reductores de complejidad de modelos (MCR) y ha ilustrado sus implementaciones de referencia en conjuntos de datos de muestra. Sin embargo, desde el punto de vista de la asistencia sanitaria global, es necesario validar y comparar este esfuerzo con un esquema sanitario global estandarizado a gran escala.

Métodos

Objetivos de la investigación

Guardian Medx es un plan de asistencia integral que ofrece atención médica personalizada con seguimiento y asistencia continuos. El objetivo es mejorar el bienestar de los mayores y reducir las hospitalizaciones en la región del sur de la India. Los directores se propusieron seguir un proceso de evaluación para, en primer lugar, aprender de trabajos anteriores realizados en la atención a diabéticos en el sur de la India[7], e identificar los elementos culturales que intervienen en la dinámica de la enfermedad, desde el diagnóstico al tratamiento, pasando por la adherencia y el mantenimiento continuo, y encontrar el formato estandarizado adecuado para recoger los resultados clínicos y los comunicados por los pacientes para obtener una base holística. La intención es encontrar el esquema de recopilación de datos que conduzca a la colaboración para la preservación de la privacidad en las tecnologías emergentes de la Web 3, incluido el ML para el análisis de datos y el consentimiento y la cooperación para la preservación de la privacidad.

El punto de partida de esta investigación es encontrar un esquema sanitario global estandarizado a gran escala que aborde los retos indicados anteriormente, que permita una recogida de datos adecuada y que identifique los constructos de privacidad preservada necesarios para fines de aprendizaje colaborativo a nivel global.

Los objetivos específicos de la investigación son los siguientes

- Aprender de investigaciones anteriores y crear un enfoque de recopilación de datos que capte los elementos culturales de la atención diabética.
- Identificar un esquema de datos específico que esté estandarizado y pueda utilizarse para la recopilación de datos y el aprendizaje preservando la privacidad a escala global.
- Explorar ZKML como modelo para ese esquema de datos global.
- Identificar los parámetros, limitaciones y mitigaciones utilizando MCR.
- Comparar el esquema de datos con datos sintéticos.

Tras revisar la bibliografía sobre la prestación de atención diabética en la India[7], la recopilación de big data y el esfuerzo de ML[8]y los formatos de estandarización e informes de adaptabilidad[9,10]se decidió que el esquema de datos fuera el esquema ICHOM, con la adaptación necesaria.

Los conjuntos de datos del ICHOM y su impacto en la atención sanitaria global de la diabetes

El ICHOM se dedica a desarrollar conjuntos estándar de medidas de resultados que puedan utilizarse en todo el mundo para evaluar la calidad de la atención de diversas afecciones médicas. En el caso de la diabetes, una enfermedad crónica que presenta variaciones significativas en su presentación y tratamiento en diferentes culturas y sistemas sanitarios, esta estandarización es crucial.

La atención diabética puede variar significativamente debido a las diferencias culturales, los factores socioeconómicos y la infraestructura sanitaria. Por ejemplo, las estrategias de tratamiento y los resultados de los pacientes en un país de renta alta pueden no aplicarse directamente a un entorno de renta baja con diferentes recursos y actitudes culturales hacia la salud. El conjunto de datos de ICHOM aborda este reto permitiendo a los sistemas sanitarios adaptar las medidas estandarizadas a los contextos locales. Esta adaptación cultural garantiza que las medidas de resultados sean pertinentes y prácticas en diversos entornos, mejorando así la utilidad del conjunto de datos y su impacto en los resultados sanitarios mundiales.

Al proporcionar un conjunto estándar de métricas, ICHOM ayuda a identificar lagunas en la atención y los resultados en diferentes regiones y poblaciones. Al adoptar un enfoque uniforme para medir resultados como los niveles de glucosa en sangre, la calidad de vida y las tasas de complicaciones, el conjunto de datos sobre diabetes de ICHOM permite a los proveedores de atención sanitaria comparar su rendimiento con los estándares mundiales, identificar las mejores prácticas y mejorar la atención al paciente y las cohortes de salud de la población. Dada la naturaleza de los datos y el esfuerzo de estandarización, los directores iniciaron la evaluación con conjuntos de datos sintéticos. Utilizaron los conjuntos de datos de población mayor y diabéticos de ICHOM para comprender la compleja naturaleza de la atención diabética y su impacto en los resultados de los pacientes.

Esta investigación ofrece un enfoque integral de la gestión de la diabetes orientado a los datos mediante la recopilación y el análisis de datos sobre demografía, diagnóstico, estilo de vida y factores sociales, métodos de tratamiento, control de la diabetes, acontecimientos agudos, complicaciones crónicas y resultados comunicados por los pacientes.

informados por los pacientes. Pretende arrojar luz sobre la intrincada interacción de los factores que afectan a la gestión de la diabetes y optimizar la atención a los dia-betes de forma rentable mediante el diagnóstico precoz y la monitorización remota y continua a escala.

Como parte del ejercicio de evaluación de datos para la investigación, se elaboran conjuntos de datos sintéticos con valores hipotéticos a partir de bibliografía anterior y de las experiencias de los investigadores. Se establecen la adecuación de los datos, la línea de base y los mapeos de datos significativos utilizando los conjuntos de datos de diabetes ICHOM V5.0 para despliegues prácticos en el entorno cultural indio. Estos conjuntos de datos se seleccionaron del conjunto de datos de diabetes ICHOM V5. Se establecieron datos sintéticos para 100 pacientes, con varias iteraciones de datos y validaciones clínicas. Se desarrollaron y analizaron análisis exploratorios de datos con análisis univariantes y bivariantes y correlaciones.

El modelo se construyó utilizando un regresor Light Gradient Boost-ing Machine (LightGBM), un marco de trabajo de gradiente de refuerzo de código abierto y alto rendimiento diseñado para tareas de ML eficientes y escalables. Está especialmente diseñado para ofrecer velocidad y precisión, lo que lo convierte en una opción popular para datos estructurados y no estructurados en diversos dominios. Entre las principales características de LightGBM destacan su capacidad para manejar grandes conjuntos de datos con millones de filas y columnas, su compatibilidad con la computación paralela y distribuida, y sus algoritmos optimizados de gradiente-boosting mediante técnicas basadas en histogramas y crecimiento en árbol por hojas.

Un aspecto crucial de ZKML es la reducción de la complejidad del modelo, que es fundamental con los costes de inferencia actuales y la escalabilidad de la tecnología de libro mayor distribuido para hacer que estos modelos sean más eficientes y prácticos para las aplicaciones del mundo real. Este artículo explora el concepto de reducción de la complejidad de los modelos en el contexto de ZKML, con ejemplos específicos y centrándose en su papel fundamental en la asistencia sanitaria. La reducción de la complejidad del modelo utiliza los conceptos de poda (eliminación de partes innecesarias del modelo que contribuyen mínimamente a la decisión final), cuantización (reducción de la precisión de los pesos y las activaciones para un cálculo eficiente) y destilación del conocimiento (transferencia del conocimiento a un modelo más simple que conserva las capacidades predictivas). Este modelo reducido se puede utilizar en un marco ZKML para realizar cálculos de forma eficiente, al tiempo que se garantiza la privacidad mediante ZKP. Además, un objetivo adicional era determinar la viabilidad técnica de utilizar los datos sintéticos generados para las aplicaciones ZKML de forma eficaz en sistemas de prueba y verificación, dado el potencial de aprendizaje cruzado. sin perder privacidad.

Reducción de la complejidad del modelo para el conjunto de datos sintéticos de diabetes ICHOM

La reducción de la complejidad del modelo en ZKML es esencial para mitigar el sobreajuste, mejorar la interpretabilidad y aumentar la eficiencia computacional, reduciendo los recursos computacionales para el entrenamiento y la inferencia. El software ZKML y su biblioteca de reducción de modelos utilizada para esta investigación fue GIZA ZK Cook.

El algoritmo de reducción de complejidad ejecuta los siguientes pasos.

1. Análisis de correlación e importancia de características para la selección y reducción de características: Las características con alta correlación son candidatas que pueden contribuir a la redundancia y la reducción. Utilizando técnicas como la eliminación recursiva de características, se eliminan las características de menor importancia.
2. La regularización L1 (Lasso) reduce a cero los coeficientes de características menos importantes, y la regularización L2 (Ridge) penaliza los coeficientes grandes para reducir la complejidad sin eliminar características. La regularización L1 y L2 se combinan para equilibrar las ventajas de ambas. En los modelos basados en árboles, la poda elimina las ramas que contribuyen mínimamente a las predicciones y consolida y divide los nodos.
3. Para reducir la dimensionalidad, se aplican el análisis de componentes principales (ACP) y la incrustación estocástica de vecinos distribuida en t (t-SNE).
4. Validación cruzada multipase y ajuste de hiperparámetros para equilibrar la complejidad y la precisión del modelo.

Se utilizó el MCR de la biblioteca Giza. El modelo de análisis se ejecuta en un entorno Python.

Resultados y discusión

La calibración de datos de la fase de evaluación muestra la viabilidad del uso práctico y sustancial de los conjuntos de datos de ICHOM v5 en la monitorización periódica y continua para prevenir la progresión de las condiciones, reduciendo la calidad de vida (Figura 1), cómo se relaciona con las puntuaciones de calidad de vida (Figura 2), y cómo la adherencia puede ser mapeada granularmente (Figura 3).

A partir de la evaluación clínica de Guardian Medx de los datos sintéticos producidos sobre la base de la extracción anónima de resultados agregados anteriores para la población del sur de la India, se considera que el modelo ayuda significativamente a avanzar en la telemedicina basada en resultados y rentable.

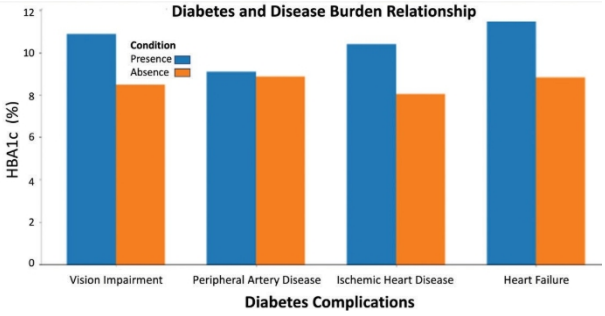
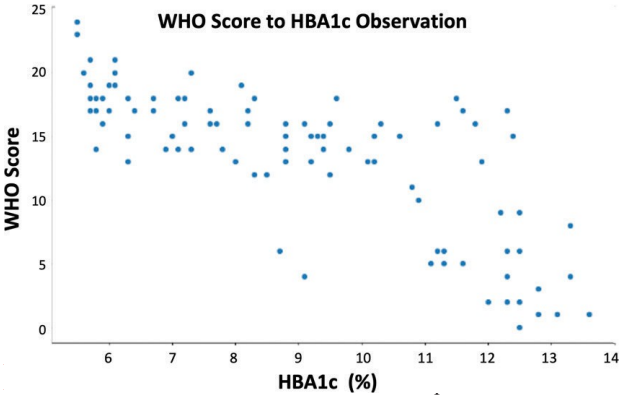


Figura 1. Complicaciones Complicaciones de la calidad de vida relacionadas con la diabetes. HBA1c: hemoglobina glucosilada. Fuente: Copyright de los autores

Cita: Blockchain en la sanidad hoy 2024, **7**: 340 - https://doi.org/10.30953/bhty.v7.340

atención de seguimiento, según el análisis de datos y los modelos de regresores para los datos analizados en la fase 1. Esto es particularmente en el control de la HBA1c (hemoglobina glucosilada), la frecuencia de seguimiento y la relación con las puntuaciones de cumplimiento y calidad de vida. Los datos ponen de manifiesto la coexistencia de otras enfermedades crónicas. Los datos del ICHOM recogen los resultados comunicados por los pacientes con las puntuaciones de la OMS, que mostraron una tendencia negativa en las puntuaciones de calidad de vida con el aumento de los valores de HBA1C comunicados. Todas las métricas de adherencia también mostraron la relación esperada con los resultados del HBA1c.



los Resultados Sanitarios: Puntuaciones de la OMS (Organización Mundial de la Salud) frente a HBA1c (hemoglobina glucosilada). Fuente: Copyright de los autores

El estudio abre la posibilidad de alertas para la acción basadas en predicciones de escalada a medida que las alimentaciones de monitorización remota llegan para la gestión de pacientes a escala para intervenciones tempranas.

Complejidad del modelo antes de la reducción del modelo

El aspecto clave es evaluar las líneas de base de reducción de la complejidad de los modelos para este conjunto de datos sintéticos y comparar la

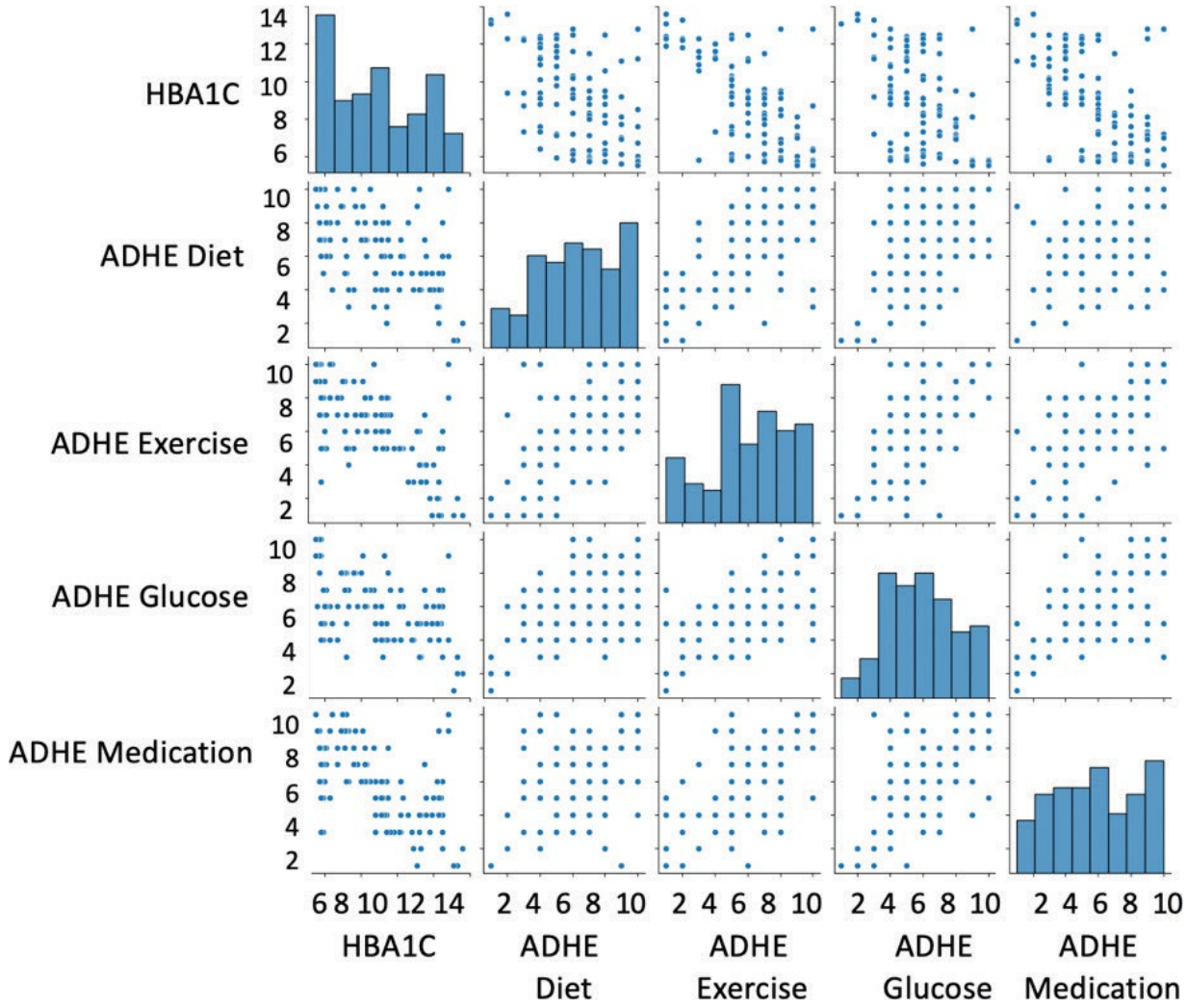


Figura 3. Efecto de la adherencia Efecto de la adherencia (ADHE) en el control de la HBA1c (hemoglobina glucosilada). Fuente: Copyright de los autores

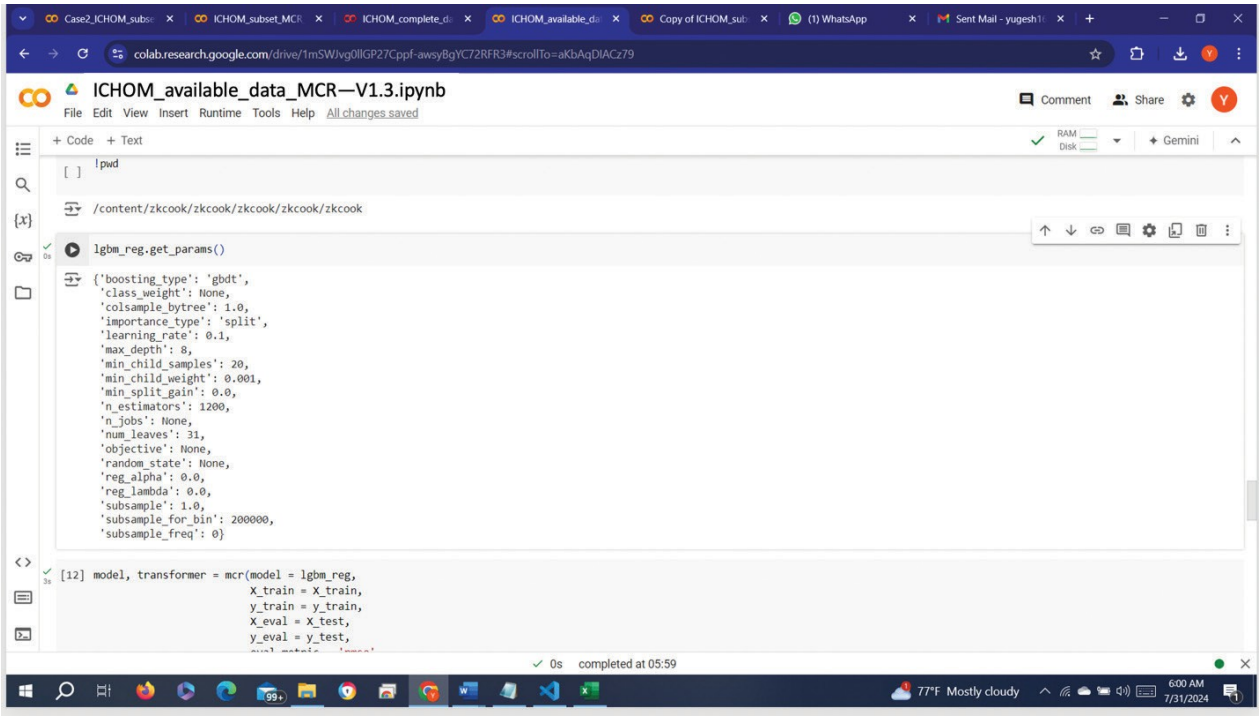


Figura 4 Complejidad del modelo antes de la reducción de la complejidad del modelo. ICHOM: Consorcio Internacional para la Medición de los Resultados en Salud; MCR: reductor de la complejidad del modelo.
Fuente: Copyright de los autores

complejidad de los parámetros del modelo antes y después. Esto es crucial para optimizar el coste en tiempo de ejecución y ver cómo se comporta este modelo para su uso en ML distribuido y sistemas colaborativos sobre libros de contabilidad distribuidos.

El regresor LGBM se configuró con los parámetros n-estimators: 1.200 y max-depth: 8 (Figura 4). En cambio, la complejidad del modelo se reduce a n-estimators: 150 y max-depth: 4, tras la reducción de la complejidad del modelo después de pasarlo por la biblioteca ZKCook (Figura 5).

Representando esto en términos de nodos:

- Número de nodos= Número de árboles* ($2^{\text{profundidad}}$-1)
- El número de nodos antes de la reducción de complejidad se evalúa en 1.200 * ($2^{(8)-1}$) = 306.000
- El número de nodos antes de la reducción de complejidad se evalúa en 150 * ($2^{(4)-1}$) = 2.250

Complejidad del modelo Después de la reducción de complejidad del modelo

La diferencia supone una reducción del 99,26%, y también está en línea con otros ejemplos de referencia. La elección del regresor para el problema y la posterior reducción mediante MCR basada en los datos de evaluación muestra que el conjunto de datos de diabetes ICHOM V5 puede utilizarse para capturar datos de forma que se pueda mejorar la interoperabilidad para estudiar e informar de los resultados en entornos colaborativos para su uso en aplicaciones ZKML.

Estos resultados implican que la adaptabilidad necesaria para el estudio de la diabetes podría capturarse en la plataforma global de ICHOM y resultan prometedores según los datos de la evaluación. Además, estos resultados son prometedores en la generación de pruebas protegidas por la privacidad para sistemas de verificación y el intercambio colaborativo de pruebas de diagnóstico basadas en el consentimiento del paciente de forma protegida por la privacidad con otras partes colaboradoras. Dados estos resultados en la fase de evaluación, las pruebas pueden generarse optimizadas para la eficiencia computacional una vez que el estudio avance.

Esta adaptación específica incluyó un subconjunto del diccionario completo del conjunto de datos ICHOM, ya que se adaptó a este entorno clínico, lo que supone una limitación. Por lo tanto, vemos esto como un punto de partida para otros trabajos para utilizar conjuntos de datos estandarizados como los conjuntos de datos ICHOM en diferentes cohortes y otras enfermedades de los conjuntos de datos ICHOM. Algunas de esas situaciones tendrán un mayor número de columnas de datos y requisitos de análisis de datos, lo que nos dará puntos de referencia adicionales para la complejidad antes y después de las líneas de base y su impacto en los sistemas de prueba y verificación. También es importante señalar que se trata de una línea de base preliminar, ya que la tecnología madura muy rápidamente desde todos los ángulos: desarrollo de normas, técnicas de reducción de modelos, reducción de pruebas y aceleración de los sistemas de verificación tanto a nivel de software como de hardware. Por lo tanto, se convertirá en

Cita: Blockchain en la sanidad hoy 2024, 7: 340 - https://doi.org/10.30953/bhty.v7.340

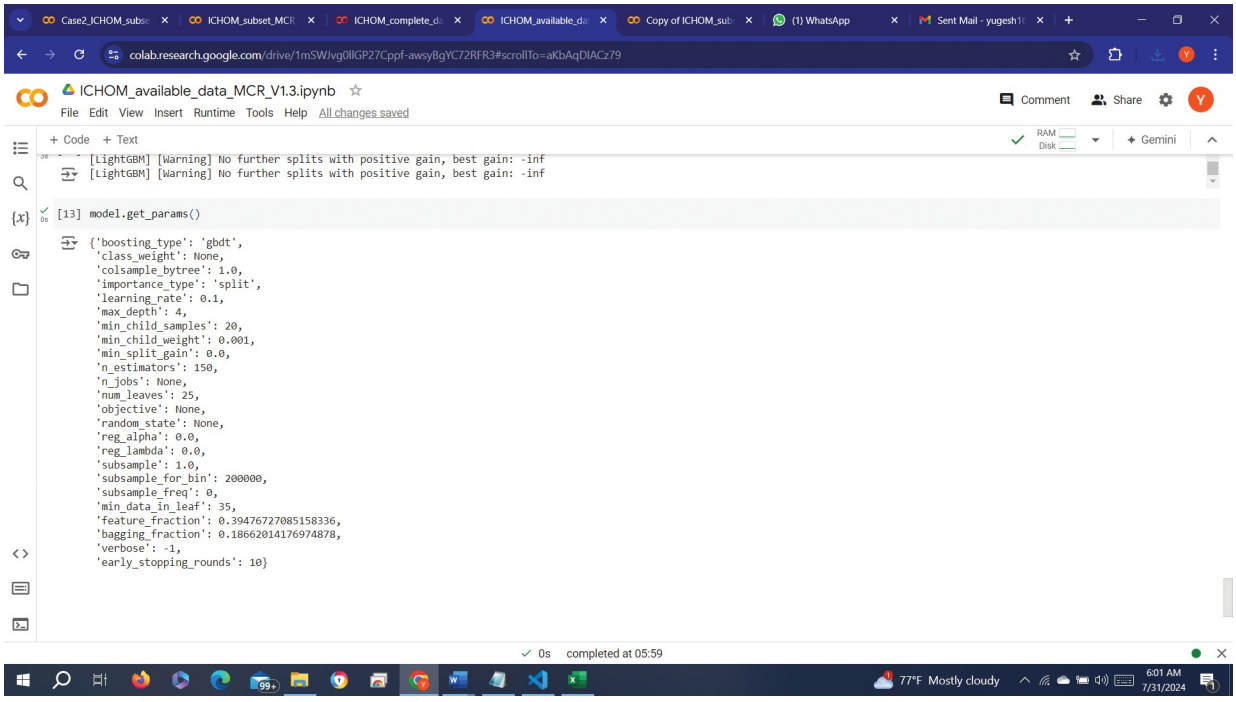


Figura 5. Complejidad del modelo antes de la reducción de la complejidad del modelo. Complejidad del modelo antes de la reducción de la complejidad del modelo. ICHOM: Consorcio Internacional para la Medición de los Resultados Sanitarios; MCR: reductor de la complejidad de los modelos.
Fuente: Copyright de los autores

importante disponer de un registro de los desarrollos de ZKML a través de estos parámetros.

Conclusiones y trabajo futuro

Basándose en el análisis de datos y modelos regresores para los datos analizados en la fase de evaluación, la evaluación clínica de GuardianMedx considera que el modelo ayuda a avanzar significativamente en la atención de monitorización remota basada en resultados y rentable. La aplicación del algoritmo de reducción de la complejidad del modelo Giza ZKcook a los datos de diabetes de ICHOM dio como resultado modelos más interpretables y eficientes desde el punto de vista computacional. Los tiempos de computación se redujeron significativamente en un conjunto de datos ICHOM estandarizado para utilizar ML y configuraciones protegidas de la privacidad para retener los datos en la fuente. Esto aumenta la seguridad y proporciona pruebas verificables para cualquier modelo de predicción que manejan los agentes y para utilizar modelos ML junto con led-gers distribuidos descentralizados para abrir posibilidades de colaboración sin dar a conocer los detalles internos de los datos. Dados estos resultados en la fase de evaluación, una vez que el estudio avance, las pruebas pueden generarse optimizadas para la eficiencia computacional. El trabajo posterior puede ampliarse a otras adaptaciones del marco ICHOM para la diabetes en otra cohorte en otro entorno para comparar los resultados, así como utilizarlos en otros conjuntos de datos de enfermedades del ICHOM. En breve, el equipo tiene la intención de ampliar las funcionalidades de ZKML para alimentar a los agentes con el fin de seguir procesando y avanzando en los conocimientos multipartitos protegidos por la privacidad.

Financiación

Ninguna.

Conflictos de intereses

Ninguno.

Colaboradores

Sathya Krishnasamy es presidente y director de ChainAim Technologies. Sus 25 años de trayectoria abarcan una amplia experiencia en entornos de pagadores de atención gestionada en empresas sanitarias líderes de EE. UU., incluidas Aetna y Anthem. Se centra en las tecnologías emergentes, incluidos los sistemas AI/ML y las tecnologías de libro mayor distribuido. También es asesor en muchos esfuerzos de la industria en la colaboración entre pagadores y proveedores, organizaciones de normalización y esfuerzos como Account Aggregators en los sectores de Fintech, Healthcare y Skills en India. Actualmente es Presidente y Director de ChainAim, que ofrece servicios de consultoría de estrategia técnica y desarrollo de aplicaciones.

Sathya Krishnasamy ayudó a conceptualizar el uso de ICHOM para datos, evaluar el conjunto de datos sintéticos, establecer una línea de base de complejidad del modelo y evaluar el caso de uso sanitario de ZKML.

El Dr. Govindarajan es el Director Médico de GuardianMedX. Es un ejecutivo sanitario con una sólida formación médica y conocimientos tecnológicos de vanguardia. Cuenta con 35 años de amplia experiencia en medicina interna.

En los últimos años ha trabajado en medicina geriátrica y geriatría en la India y es asesor de cuidados geriátricos y paliativos para muchas entidades gubernamentales del país. Ha gestionado y administrado una atención de calidad centrada en el paciente siguiendo un con-tinuum único de atención en clínicas, hospitales, residencias de ancianos, hospicios y hogares.

El Dr. Govindarajan puso en marcha la iniciativa y llevó a cabo la investigación sobre las necesidades de recopilación de datos, el diseño y la evaluación del ICHOM para datos sobre diabetes, y la evaluación clínica de los conjuntos de datos sintéticos.

Declaración de disponibilidad de datos (DAS), intercambio de datos, reproducibilidad y repositorios de datos

El diccionario de datos de diabetes ICHOM V5 está disponible en https://www.ichom.org/ patient-centered-outcome-measure/diabetes/

Aplicación de texto generado por IA o tecnología relacionada

Ninguna.

Agradecimientos

Yugesh Panta, estudiante de Máster en Ciencias del Departamento de Ingeniería Eléctrica e Informática de la Escuela de Ingeniería Tandon de la Universidad de Nueva York, ayudó a los directores en la recopilación de datos de investigación, la validación, el análisis y el ajuste del modelo.

Referencias

1. Goldwasser S, Micali S, Rackoff C. The knowledge complexity of interactive proof systems. Actas del decimoséptimo simposio anual de la ACM sobre teoría de la computación-STOC '85. 1985.
2. Fiat A, Shamir A. How to prove yourself: practical solutions to identification and signature problems. Avances en criptología-CRYPTO' 86 [Internet]. 2019;186-94. Disponible en: https://link.springer.com/chapter/10.1007%2F3-540-47721-7_12 [citado 2024 31 de julio].
3. Ben-Sasson E, Chiesa A, Tromer E, Virza M. Conocimiento cero sucinto no interactivo para una arquitectura von neumann [Internet]. 2019. Disponible en: https://eprint.iacr. org/2013/879.pdf [citado 2024 31 de julio].
4. Ben-Sasson E, Bentov I, Horesh Y, Riabzev M. Scalable, transpar-ent, and post-quantum secure computational integrity [Internet]. ePrint IACR. 2018. Disponible en: https://eprint.iacr.org/2018/046 [citado 2024 31 julio]
5. Guerra-Manzanares A, Lechuga J, Maniatakos M, Shamout FE. Privacy-preserving machine learning for healthcare: open challenges and future perspectives. ICLR 2023 Workshop on Trustworthy Machine Learning for Healthcare. arXiv. 2023; 1-13. https://arxiv.org/abs/2303.15563
6. Gotor AM. Maximizing model efficiency with model-com-plexity-reducer (MCR). zkcook/docs/mcr.pdf at main giza-techxyz/zkcook [Internet]. GitHub. [citado 2024 Ago 1]. Disponible en: https://github.com/gizatechxyz/zkcook/blob/ main/docs/mcr.pdf
7. Das AK, Saboo B, Maheshwari A, Nair VM, Banerjee S, Jay-akumar C, et al. Health care delivery model in India with rele-vance to diabetes care. Heliyon. 2022 Oct;8(10):e10904. https:// doi.org/10.1016/j.heliyon.2022.e10904
8. Musacchio N, Giancaterini A, Guaita G, Ozzello A, Pellegrini MA, Ponzani P, et al. Artificial intelligence and big data in diabetes care: a position statement of the Italian Associa-tion of Medical Diabetologists. J Med Internet Res. 2020 Jun 22;22(6):e16922. https://doi.org/10.2196/16922
9. Diabetes [Internet]. ICHOM. [citado 2024 Ago 1]. Disponible en: https://www.ichom.org/patient-centered-outcome-measure/ diabetes/
10. Benning L, Das-Gupta Z, Fialho LS, Wissig S, Tapela N, Gaunt S. Balancing adaptability and standardisation: insights from 27 routinely implemented ICHOM standard sets. BMC Health Serv Res. 2022 Nov 28;22(1):1424. https://doi.org/10.1186/ s12913-022-08694-9

APÉNDICE

Definición de acrónimos

AI/ML: Inteligencia Artificial / Aprendizaje Automático

HBA1c: Hemoglobina Glicoxilada

ICHOM: Consorcio Internacional para la Medición de los Resultados en Salud (International Consortium for Health Outcomes Measurement)

LightGBM: Light Gradient Boosting Machine MCR: Model Complexity Reducer (Reductor de la complejidad del modelo)

ML: Aprendizaje automático

zk-SNARKs: Zero-Knowledge Succinct Non-Interactive Arguments of Knowledge (argumentos sucintos no interactivos de conocimiento cero)

ZKML: Aprendizaje automático de conocimiento cero ZKP: Pruebas de conocimiento cero