

**Comparativo de funciones kernel en la predicción de enfermedades cardiovasculares en
Redes Neuronales Artificiales (ANN) y Máquinas de Soporte Vectorial (SVM)**

Michael Rafael Rodríguez Rodríguez, ✉ mrrodriguez.6789@unicesmag.edu.co

Claudia Alejandra Delgado Calpa, ✉ cadelgado.9770@unicesmag.edu.co

Universidad CESMAG

Facultad de Ingeniería

Ingeniería de Sistemas

Pasto - Colombia

2024

**Comparativo de funciones kernel en la predicción de enfermedades cardiovasculares en
Redes Neuronales Artificiales (ANN) y Máquinas de Soporte Vectorial (SVM)**

Michael Rafael Rodríguez Rodríguez, ✉ mrrodriguez.6789@unicesmag.edu.co

Claudia Alejandra Delgado Calpa, ✉ cadelgado.9770@unicesmag.edu.co

Trabajo de grado como requisito para optar el título de Ingeniero de Sistemas

Asesor: Mg. Héctor Andrés Mora Paz

Universidad CESMAG

Facultad de Ingeniería

Ingeniería de Sistemas

Pasto - Colombia

2024

Nota de Aceptación

Página de exclusión de responsabilidad intelectual

“El pensamiento que se expresa en esta obra es exclusiva responsabilidad de sus autores y no compromete la ideología de la Institución Universitaria CESMAG”

Dedicatoria

A Dios que ha sido el motor en vida en cada proyecto y acto que he realizado, a mis padres y hermanos, por su apoyo incondicional desde el momento inicial de mi carrera, por estar a mi lado en cada momento, y por siempre animarme a continuar y nunca rendirme y recordar esta frase siempre *“El éxito no es el final, el fracaso no es fatal: lo que cuenta es el coraje para continuar”*.

DH

Claudia Alejandra Delgado Calpa

Agradecimientos

Expresamos nuestra más profunda gratitud a nuestro asesor, Héctor Andrés Mora Paz; agradecemos por su acompañamiento y apoyo a lo largo de todo el proceso de elaboración del presente trabajo de grado. Su orientación y constancia fueron fundamentales para lograr nuestros objetivos académicos.

El asesor Héctor Andrés Mora Paz no solo nos brindó orientación experta en cada paso del camino, sino que también nos inspiró con su pasión por la investigación y su firme creencia en nuestro potencial y en cada actividad que se realizaba. Su paciencia y su disposición para escucharnos y ayudarnos a superar los retos que se presentaron fueron un pilar fundamental en momentos de duda, este logro académico hubiera sido imposible sin su guía y motivación.

Claudia Alejandra Delgado Calpa
Michael Rafael Rodríguez Rodríguez

TABLA DE CONTENIDO

INTRODUCCIÓN	17
1. PROBLEMA DE INVESTIGACIÓN	19
1.1 OBJETO O TEMA DE INVESTIGACIÓN	19
1.2 LÍNEA DE INVESTIGACIÓN	19
1.3 SUB-LÍNEA DE INVESTIGACIÓN	19
1.4 PLANTEAMIENTO DEL PROBLEMA.....	19
1.5 FORMULACIÓN DEL PROBLEMA	22
1.6 OBJETIVOS.....	22
1.6.1 Objetivo General.....	22
1.6.2 Objetivos Específicos.....	22
1.7 JUSTIFICACIÓN.....	23
1.8 DELIMITACIÓN	24
2. TÓPICOS DEL MARCO TEÓRICO	25
2.1 ANTECEDENTES.....	25
2.2 ENUNCIADO DE SUPUESTOS TEÓRICOS DE LA INVESTIGACIÓN	30
2.2.1 Enfermedades Cardiovasculares	30
2.2.2 El riesgo cardiovascular	31
2.2.3 Factores de riesgo cardiovasculares.....	31
2.2.4 Método para calcular el riesgo cardiovascular.....	33
2.2.5 Inteligencia artificial	34
2.2.6 Máquinas de soporte vectorial, Algoritmo de soporte vectorial (SVM).....	34
2.2.7 Funciones kernel	34
2.2.8 Identificación de patrones mediante modelos lineales.....	36
2.2.9 Base de datos.....	38

2.2.10 Métricas de medición.....	39
2.2.11 Métricas de clasificación.....	44
2.2.12 Metodología en Espiral.....	49
2.3 VARIABLES DE ESTUDIO.....	50
2.3.1 DEFINICIÓN NOMINAL DE LAS VARIABLES.....	50
2.3.2 DEFINICIÓN OPERATIVA DE LAS VARIABLES.....	52
2.3 FORMULACIÓN DE HIPÓTESIS.....	53
2.3.1 Hipótesis de investigación.....	53
2.3.2 Hipótesis nula.....	54
2.3.3 Hipótesis alterna.....	54
3. METODOLOGÍA.....	54
3.1 PARADIGMA.....	54
3.2 ENFOQUE.....	54
3.3 MÉTODO.....	55
3.4 TIPO DE INVESTIGACIÓN.....	56
3.5 DISEÑO DE INVESTIGACIÓN.....	57
3.6 POBLACIÓN.....	58
3.8 TÉCNICAS DE RECOLECCIÓN DE LA INFORMACIÓN.....	58
3.8.1 Validez de la técnica.....	59
3.8.2 Confiabilidad de la técnica.....	60
4. RESULTADOS.....	61
4.1 Definición de funciones kernel recomendadas por expertos.....	61
4.2 Funciones kernel acopladas a SVM y ANN.....	68
4.3.1 Descripción de los datos.....	73
5. ANÁLISIS DE RESULTADOS.....	96

CONCLUSIONES.....101

RECOMENDACIONES.....102

Referencias103

ANEXOS.....110

LISTA DE TABLAS

TABLA I DEFINICIÓN FORMAL FUNCIONES KERNEL.....	36
TABLA II MÉTRICAS DE REGRESIÓN EN APRENDIZAJE AUTOMÁTICO	39
TABLA III MÉTRICAS DE EVALUACIÓN.....	44
TABLA IV VARIABLES DEPENDIENTES DE ESTUDIO	50
TABLA V VARIABLES INDEPENDIENTES DE ESTUDIO	51
TABLA VI FORMULACIÓN MATEMÁTICA DE LAS FUNCIONES KERNEL UTILIZADAS POR EXPERTOS	62
TABLA VII FORMULACIÓN MATEMÁTICA DE LAS FUNCIONES KERNEL UTILIZADAS EN EL PROYECTO.....	63
TABLA VIII CODIFICACIÓN DE FUNCIONES KERNEL EN PHYTON	65
TABLA IX DESCRIPCIÓN DE LOS DATOS.....	73
TABLA X RESULTADOS GENERALES PARA SVM	75
TABLA XI MEJORES PARÁMETROS CONFIGURADOS PARA LAS FUNCIONES KERNEL EN SVM	77
TABLA XII RESULTADOS FINALES PARA SVM	78
TABLA XIII RESULTADOS GENERALES PARA ANN	79
TABLA XIV MEJORES PARÁMETROS CONFIGURADOS PARA LAS FUNCIONES KERNEL EN ANN	81
TABLA XV RESULTADOS FINALES PARA ANN	82
Tabla XVI REQUERIMIENTO FUNCIONAL RF-01	86
TABLA XVII REQUERIMIENTO FUNCIONAL RF-02.....	88
TABLA XVIII REQUERIMIENTO FUNCIONAL RF-03.....	89
TABLA XIX REQUERIMIENTO FUNCIONAL RF-04	92
TABLA XX REQUERIMIENTO FUNCIONAL RF-05.....	93
TABLA XXI REQUERIMIENTO NO FUNCIONAL RNF-01.....	95
TABLA XXII REQUERIMIENTO NO FUNCIONAL RNF-02	95

LISTA DE FIGURAS

Fig. 1. Truco kernel de dos dimensiones a tres	35
Fig. 2. Ejemplo de clasificador lineal.....	37
Fig. 3. Clasificadores Lineales	37
Fig. 4. Matriz de Confusión	46
Fig. 5. Área Bajo la Curva.....	48
Fig. 6. Esquema del ciclo CRISP-DM estándar	55
Fig. 7. Algoritmo implementado para la comparación de kernel.....	66
Fig. 8. Algoritmo exhaustivo para búsqueda de modelos	67
Fig. 9. Modelo de acoplamiento funciones kernel en scikit-learn.....	68
Fig. 10. Código fuente de acoplamiento a biblioteca scikit-learn.....	70
Fig. 11. Código fuente de acoplamiento a biblioteca scikit-learn.....	71
Fig. 12. Algoritmo para la obtención de funciones kernel	72
Fig. 13. Ejemplo de implementación de un kernel L laplacian(izq.), S sigmoid (der)	72
Fig. 14. Diagrama arquitectónico	84
Fig. 15. Instrucciones	87
Fig. 16. Formulario de datos	89
Fig. 17. Porcentaje de RCV.....	91
Fig. 18. Alertas tempranas.....	93
Fig. 19. Recomendaciones	94
Fig. 20. Tiempo en ANN.....	97
Fig. 21. Tiempo de SVM.....	97

LISTA DE ECUACIONES

Ecuación 1. Accuracy.....	52
Ecuación 2. Precision	53
Ecuación 3. Recall.....	53
Ecuación 4.F1-Score	53
Ecuación 5. Modelo matemático de la evaluación de la medida-F.....	53

RESUMEN ANALÍTICO DE ESTUDIO R.A.E

Facultad	Ingeniería
Programa	Ingeniería de Sistemas
Fecha de elaboración	2023
Autores de la investigación	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
Director de la investigación	Mg. Héctor Andrés Mora Paz
Título de la investigación	COMPARATIVO DE FUNCIONES KERNEL EN LA PREDICCIÓN DE ENFERMEDADES CARDIOVASCULARES EN REDES NEURONALES ARTIFICIALES (ANN) Y MÁQUINAS DE SOPORTE VECTORIAL (SVM)

PALABRAS CLAVE

Funciones Kernel, Predicción, Enfermedades Cardiovasculares, Redes Neuronales Artificiales, Máquinas de Soporte Vectorial, Machine Learning, Inteligencia Artificial.

DESCRIPCIÓN

En la actualidad las enfermedades cardiovasculares, constituyen la principal causa de muerte a nivel mundial. Existen desafíos, como la inoportunidad en la atención en salud, falta de acceso a las tecnologías y diagnósticos oportunos, por ende, este proyecto se enfoca en el uso de herramientas innovadoras, dándole paso a la necesidad de utilizar inteligencia artificial en el ámbito de Machine Learning, para mejorar la predicción de las enfermedades cardiovasculares, es así que la investigación se centró en determinar la función kernel más eficaz en algoritmos de Redes Neuronales Artificiales (ANN) y Maquinas de Soporte Vectorial (SVM), haciendo un comparativos ecuánime, evaluando la exactitud y el tiempo de predicción de cada función kernel propuesta.

Con base en los resultados, se integran esas nuevas funciones kernel optimas a la biblioteca scikit-learn, logrando una validación en la configuración apropiada para la predicción del riesgo de padecer alguna ECV, este enfoque innovador permite reducir el tiempo de detección, minimizando así las posibilidades de complicaciones futuras de enfermedades que se pueden prevenir, y aportar de manera oportuna en el diagnóstico y factores de riesgo con alertas tempranas que pueden ser de gran utilidad para el personal de salud.

CONTENIDO

La investigación está realizada de la siguiente manera

Capítulo 1: En este capítulo se describe la problemática de la inoportunidad en ciertos diagnósticos médicos como las ECV y la oportunidad de identificar aquellos factores de riesgo que contribuyen a estas patologías, por falta de herramientas automáticas que permitan análisis de grandes cantidades de datos y que permita la aplicación de funciones kernel en algoritmos como SVM y ANN para mejorar las predicciones, se realiza la definición de un objetivo general, y cuatro objetivos específicos, se justifica la necesidad de realizar proyectos de este tipo porque como se menciona a lo

largo del documento existen kernel 00que mejoran los resultados para SVM y empeoran para ANN, además de comprobar la hipótesis que se menciona en algunas referencias de expertos de que los kernels funcionan mejor para SVM que en otros algoritmos de clasificación y predicción lineal, porque este algoritmo utiliza el criterio del margen máximo. Finalmente se define el proyecto en un periodo de 24 meses a partir del año 2023 hasta el periodo del año 2024.

Capítulo 2: En este capítulo se proyecta antecedentes internacionales y nacionales, tomando estudios que abordan temáticas similares.

Capítulo 3: En este capítulo se define el tipo de metodología de investigación como tipo paradigma positivista puesto que su característica de medición es cuantitativo, lo que permite identificar con base en la recolección de datos medibles, como los transformadores de escalación, permitiendo la aplicación de métodos que permita analizar esos datos identificando patrones y comportamientos que sirva en este caso para comparar las mejores parámetros e identificar el kernel más apropiado para la predicción del riesgo de padecer ECV.

Capítulo 4: En este capítulo se plasma los resultados de la investigación en donde se muestra los nuevos kernel a partir de la recomendación de expertos, y cuales se implementaron para construcción del modelo, el levantamiento de requisitos funcionales y no funcionales.

Capítulo 5: En este capítulo se evidencia el análisis y discusión de los resultados en donde se plasma los experimentos que se realizaron antes de la construcción del modelo y posterior a la implementación de este, analizando los resultados obtenidos y haciendo uso de las mejores métricas con base de datos obtenidas de repositorios libre.

METODOLOGÍA

Esta investigación se enmarca dentro del paradigma positivista, un paradigma que es racional, objetivo y se basa en comprobar hechos y particularidades propias del conocimiento científico, a su vez se caracteriza por el alto interés, por la verificación del conocimiento a través de predicciones lo que permite el estudio de los algoritmos SVM y ANN a través del uso de las funciones kernel,

realizando varias configuraciones de hiperparámetros a través de una sintonización sistemática, aplicando tales métodos con datos que permitan predecir el riesgo cardiovascular presente en una persona.

LÍNEA DE INVESTIGACIÓN

Inteligencia Artificial

SUB LINEA DE INVESTIGACIÓN:

Multiple Kernel Learning (MKL)

CONCLUSIONES

La herramienta desarrollada cumple de manera satisfactoria los objetivos planteados, la visualización de los resultados permitió una evaluación objetiva, destacando la eficacia de los kernels truncated y el escalador estándar en la clasificación de datos de riesgo cardiovascular mediante máquinas de soporte vectorial. Finalmente, el desarrollo de la herramienta proporcionó una interfaz gráfica dinámica en Python, junto con el aprendizaje de las bibliotecas de análisis de datos disponibles en este lenguaje.

RECOMENDACIONES

Las recomendaciones incluyen la fácil escalabilidad de la herramienta desarrollada para la implementación de nuevos modelos y funciones, así como la necesidad de nuevas aplicaciones de visualización para la predicción de enfermedades cardiovasculares (ECV). Además, se sugiere promover entre los estudiantes la comprensión de los fundamentos de la investigación, dada la naturaleza comparativa de las funciones kernel en diferentes algoritmos de machine learning. Por último, se recomienda llevar a cabo un estudio comparativo de funciones kernel acopladas a otros algoritmos de machine learning para clasificación o regresión lineal en la predicción de enfermedades cardiovasculares, con el fin de determinar qué algoritmo funciona mejor con kernels en estos conjuntos de datos.

INTRODUCCIÓN

Las enfermedades cardiovasculares son desórdenes del corazón y de los vasos sanguíneos, y se han convertido en la principal causa de muerte en todo el mundo [1]. En Colombia, durante 2021, se registraron un total de 51.988 muertes a raíz de condiciones relacionadas con ECV, un 12% más que en el 2020, con un aumento significativo en la incidencia de casos en mujeres, de acuerdo con cifras del Departamento Nacional de Estadísticas (DANE) 4. A pesar de que el país ha avanzado en los esfuerzos para reducir la carga de estos escenarios, aún se presentan acciones muy fragmentadas dentro del sistema de salud y los sacrificios individuales de cada actor no son suficientes [2]. Existen diferentes elementos que pueden estar relacionados con la ineficiencia en la prevención oportuna de estas patologías, como puede ser la inadecuada atención en los centros de salud, difícil acceso a tecnologías, inoportunidad en los diagnósticos, escasa información sobre factores modificables que logran afectar directamente su salud y llegar a complicaciones severas a futuro.

Por tales motivos, los profesionales sanitarios se ven obligados a cambiar la forma de abordar los padecimientos de los pacientes desde los principios de la promoción de la salud y la prevención de la enfermedad como una nueva política de salud pública [2], [3].

En otras palabras, como se menciona por Arias en [4], aun si se logra asegurar el acceso a un determinado servicio de salud, especialmente en los sectores más vulnerables, el beneficio es nulo si no se tiene la capacidad de salvar vidas y mejorarlas con los recursos que se destinan para estas enfermedades, en las que una monitorización y análisis fuera del entorno clínico continúa siendo fundamental para descubrir oportunamente las complicaciones asociadas en gran parte de la población. Una rápida actuación en cualquier lugar resulta vital, es esencial reducir el tiempo de detección y atención, de forma que se minimicen las posibilidades de sufrir daños cardiovasculares, caer en estado de coma, o, en el peor de los escenarios, la muerte.

Existen tecnologías que prometen revolucionar el ámbito de la prevención de las enfermedades cardiovasculares. Así mismo, gracias a sus grandes capacidades, son la herramienta perfecta para mejorar su predicción.

En consecuencia, se han puesto en marcha diferentes investigaciones cuyos resultados determinan que el Machine Learning podría ayudar muchísimo en la determinación de sufrir episodios o enfermedades cardiovasculares en individuos sanos [5].

Lo que se pretende con la siguiente investigación es determinar de las funciones algebraicas y trascendentes, la función kernel que ofrece el mejor compromiso en la predicción de enfermedades cardiovasculares en los algoritmos de SVM y ANN, aportando a los investigadores con las funciones kernel, a partir de un estudio comparativo ecuánime, donde se mida la exactitud y el tiempo de predicción de cada función; posteriormente se acoplan estas nuevas funciones kernel a la biblioteca Scikit-learn, se valida la configuración tomando la más apropiada para la prevención de ECV teniendo en cuenta la relación error tiempo, y logrando así construir la aplicación móvil para alertas tempranas implementando el modelo predictivo.

1. PROBLEMA DE INVESTIGACIÓN

1.1 OBJETO O TEMA DE INVESTIGACIÓN

Desempeño de las funciones kernel en la predicción de enfermedades cardiovasculares

1.2 LÍNEA DE INVESTIGACIÓN

Inteligencia Artificial

1.3 SUB-LÍNEA DE INVESTIGACIÓN

Multiple Kernel Learning (MKL)

1.4 PLANTEAMIENTO DEL PROBLEMA

Las enfermedades cardiovasculares, según la OMS, son uno de los mayores problemas de salud pública a nivel mundial, siendo la enfermedad cerebro vascular la primera causa de mortalidad y morbilidad, conforme estimación se cobra 17.9 millones de vidas cada año [6].

Además, los problemas principales que se presentan en las personas son infartos, ataque cerebrovascular, síndrome metabólico, cardiopatías, e hipertensión arterial, son causados por sobrepeso, obesidad, dislipidemia, tabaquismo, inactividad física, alimentación no saludable, estos son factores de riesgo existentes que se pueden modificar, y existen aquellos que no se pueden modificar, pero que son un factor causal para el desarrollo de una enfermedad cardiovascular como son la edad, género, antecedentes personales de enfermedad cardiovascular y antecedentes de familiares con ECV prematuros solo cuando hayan ocurrido en primer grado [7].

Muchos son los retos que afrontan las personas con este tipo de patologías, que van desde de la calidad de atención en salud hasta las tecnologías concernientes para su monitorización y cuidado. Gran cantidad de usuarios con enfermedades cardiovasculares necesitan estancias hospitalarias frecuentes, puesto que la restringida tecnología concerniente a la salud está concentrada en los centros médicos, principalmente porque adquirirlos tiene un alto costo, al igual que su

mantenimiento, sin dejar a un lado el hecho de que se necesita de un conocimiento especializado para interpretar y analizar los datos capturados por estos equipos de alta complejidad [4]. Igualmente, la ausencia de tecnologías guiadas a la detección y prevención de eventos cardiovasculares peligrosos para la salud nutren la elevada tasa de muertes y hospitalizaciones.

Un precedente que se debe tener en cuenta es el aprendizaje de máquina y su aplicación en el campo de la salud, donde este se ha ido incrementando aproximadamente en un 30% a 40% año tras año con ayuda del análisis estadístico y probabilístico de otras investigaciones. Aun así, el reto de recopilación y manipulación de datos es complejo, tal cual lo es la predicción de muerte al padecer alguna enfermedad, y el desafío es aún mayor cuando se trata del ser humano y órganos tan complejos como el corazón o el sistema que lo compone. Pero el ser diagnosticado no es un final decisivo para las personas al existir varios métodos para mantener un corazón saludable [8].

Por todo lo anteriormente dicho, se busca abordar en primera instancia la problemática desde una etapa previa para contribuir a que no se origine una patología mayor, por lo tanto, se ve la necesidad de utilizar modelos de IA que resulten útiles para el apoyo diagnóstico oportuno de enfermedades cardiovasculares. Aun así, los modelos de aprendizaje automático conocidos hasta ahora que se usan para predecir el diagnóstico de insuficiencia cardíaca logran resultados con una precisión variable, entre el 80% y 90% en sus mejores casos [8].

Así pues, existen diferentes técnicas de aprendizaje automático que son idóneas para identificar las características más importantes tras procesar grandes cantidades de datos, logrando así una predicción y mejorando algunos sistemas [9].

Por tanto, el conjunto de datos a utilizar ha sido empleado para entrenar modelos con algunas técnicas de aprendizaje de máquina, pero hay técnicas que aún no se han utilizado y que podrían arrojar resultados importantes.

Por otro lado, es conveniente que las personas generen conciencia de detectar oportunamente alguna enfermedad crónica, antes de que se manifiesten algunos síntomas y llegando al caso de no obtener un diagnóstico oportuno en cualquier momento de la vida [10].

Por lo tanto, con este estudio se pretende validar la hipótesis de que es posible encontrar mejoras en los modelos de predicciones de enfermedades cardiovasculares mediante la inspección del desempeño de los algoritmos SVM y ANN, introduciendo funciones kernel que no se hayan desarrollado y utilizado para este caso, determinando que este resultado final sea útil para estudios relacionados con el objeto de estudio [11].

En efecto, el aprendizaje de máquinas ha demostrado su valor de aplicación en contextos médicos, siendo una herramienta novedosa y alternativa que apoya tareas complejas como el diagnóstico de enfermedades. Estas tecnologías pueden garantizar la seguridad de mejor manera más efectiva, mediante una detección rápida y confiable de la enfermedad, e incluso puedan aliviar las preocupaciones de toda una población.

Por consiguiente, si no se implementa el aplicativo ni se realiza el estudio de comparación de las funciones kernel en el marco de apoyo al diagnóstico médico, se perdería la oportunidad de que las personas tomen conciencia de la significancia de los efectos que sus hábitos y estilos de vida tienen hasta el momento.

Es conveniente destacar que, probablemente, se pierda la sensibilización de la población en la prevención del riesgo cardiovascular, lo cual podría llevar a un aumento en la tasa de mortalidad y de mortalidad prematura por estas enfermedades.

En definitiva, mantener el contacto entre médicos y pacientes desde un enfoque de bienestar y estado de salud permite que haya una mayor y mejor cercanía, como se menciona en algunos artículos, el uso de estas tecnologías basadas en Inteligencia Artificial, complementan el conocimiento de los médicos y permite a estos pasar más tiempo con sus pacientes y mejorar el proceso de toma de decisiones compartido [12]. De no implementar el modelo predictivo, se perdería la posibilidad también de aplicar estas nuevas técnicas a tiempo, aumentarían los riesgos y se perdería la posibilidad de realizar un diagnóstico oportuno, por otro lado, no habría probabilidad de analizar grandes cantidades de datos rápidamente con coherencia y precisión, no se podría crear ese modelo de estudio tan importante para el conocimiento y además no se podría contribuir en ese estado de salud de cada persona, y que de carecer de salud cualquier individuo le sería muy difícil lograr algún éxito.

1.5 FORMULACIÓN DEL PROBLEMA

De un conjunto de funciones trascendentes ¿Cuál función kernel tiene el mejor compromiso entre la exactitud y el tiempo de predicción, para el apoyo al diagnóstico de enfermedades cardiovasculares?

1.6 OBJETIVOS

1.6.1 Objetivo General

- Determinar de las funciones algebraicas y trascendentes, la función kernel que ofrece el mejor compromiso en la predicción de enfermedades cardiovasculares en los algoritmos de SVM y ANN a partir de un estudio comparativo ecuánime, donde se mida la exactitud y el tiempo de predicción de cada función kernel

1.6.2 Objetivos Específicos

- Definir un conjunto de funciones kernel trascendentes alternativas para la predicción de enfermedades cardiovasculares mediante la recomendación de expertos.
- Acoplar nuevas funciones kernel alternativas en los algoritmos SVM, ANN a la biblioteca Scikit-learn.
- Validar que configuración es la más apropiada para la prevención de enfermedades cardiovasculares mediante la comparación de la relación error tiempo.
- Construir una aplicación móvil para alertas tempranas de enfermedades cardiovasculares, a través de un modelo predictivo.

1.7 JUSTIFICACIÓN

Las enfermedades cardiovasculares son enfermedades de interés en salud pública y de las cuales mayormente produce muertes y deterioro progresivo de la salud, de las cuales no son diagnosticadas oportunamente, sino cuando estas se presentan en un estado avanzado, resulta de especial interés obtener un diagnóstico oportuno de estas enfermedades, y qué porcentaje de rangos de riesgo cardiovascular se presentan en mayor frecuencia, que van ligadas a factores que se pueden a futuro modificar, permitiendo adoptar medidas oportunas además que contribuye a promover la salud, y prevenir la enfermedad.

El diagnóstico temprano de ECV es decisivo para disminuir las cifras de mortalidad y morbilidad, así como los riesgos que puedan alertar de manera más oportuna, por esta razón se han propuesto varios estudios para detectar enfermedades y riesgos cardiovasculares algunos a partir del procesamiento de la señal de electrocardiografía (ECG) y otros desde mediciones generales de salud pública, en ambos casos apoyados en técnicas de aprendizaje automático (*Machine Learning*).

Dentro de los distintos tipos de aprendizaje automático, se deriva el aprendizaje basado en kernel (*Kernel Learning*) o núcleos, el cual es un método para el análisis de patrones, cuyo algoritmo de aplicación más conocido es Máquinas de Soporte Vectorial (SVM). La tarea general del análisis de patrones es encontrar y estudiar tipos generales de relaciones en conjuntos de datos [13].

Por ejemplo, por medio de diferentes modelos de aprendizaje profundo, como RNN (Red Neuronal Recurrente), LSTM (Red de memoria a corto plazo) e incluso CNN (Red neuronal convolucional) lograron resultados de hasta 96% de precisión para detectar diferentes tipos de enfermedades cardiacas, entre las que se encuentran taquicardias ventriculares, fibrilación auricular y taquicardia sinusal [14] [4], otro ejemplo en el cual proponen un método novedoso a partir de un modelo de Random Forest híbrido con una componente lineal que logró una precisión de 88,7% en la predicción de enfermedades cardiovasculares [15], [4].

Este estudio nace de la necesidad de aplicar un modelo predictivo basado en un análisis exhaustivo de las configuraciones en los algoritmos SVM y ANN, considerando que más características del kernel se aproximan al modelo óptimo de aprendizaje automático, haciendo uso de dataset previamente analizados, y recolección de información, para entrenamiento de algoritmos de predicción, para conocer el porcentaje en el rango de riesgo de enfermedades cardiovasculares, y que factores se puedan intervenir oportunamente por el médico.

La investigación será útil a nivel social porque brinda información a medida del estado de salud de la población y apoyo al profesional médico y cuando este puede de manera oportuna identificar si hasta el momento algún factor externo o interno influye en su calidad de vida, y evitar un deterioro que puede ser prevenible.

El resultado será una contribución a la inteligencia artificial en el campo del aprendizaje automático, especialmente en el área de estudios comparativos de algoritmos. Esto incluirá una ruta de comparación, la implementación de las funciones kernel, modelos entrenados para hacer predicciones de enfermedades cardiovasculares y la visualización de dichos datos.

Finalmente, este proyecto se sintoniza con la dinámica de promoción de la salud y prevención de la enfermedad que viene trabajando Colombia en sus diferentes marcos contextuales, además que la enfermedad cardiovascular no cataloga y la puede padecer cualquier persona en su mayoría de edad, por eso es importante el estudio, análisis de los posibles factores variantes según sea, el índice de masa corporal, perímetro abdominal, el nivel socioeconómico en que se encuentra, hábitos, antecedentes de enfermedades personales y familiares, entre otros, además se obtiene un mejoramiento en funciones kernel, para propiciar la integridad en los modelos predictivos.

1.8 DELIMITACIÓN

Este proyecto se desarrolló mediante experimentos sobre base de datos obtenidos de Herramientas de repositorio como datasets Kaggle, y datasets UCI, los modelos serán entrenados usando Redes Neuronales (ANN) y Máquinas de Soporte Vectorial (SVM), Para evaluar los modelos se utilizarán métricas de clasificación asociadas a la exactitud y al tiempo.

Los Kernel a evaluar serán seleccionados dentro de un conjunto de funciones trascendentes, que cumplan las condiciones Karush-Kuhn-Tucker (KKT).

La configuración de los hiperparámetros se realizará utilizando las técnicas de búsqueda aleatoria.

El proyecto se desarrollará en un tiempo estimado de 18 meses, comenzando en el periodo A del año 2023 y finalizando en el periodo B de 2024.

2. TÓPICOS DEL MARCO TEÓRICO

2.1 ANTECEDENTES

Se llevaron a cabo la consulta de fuentes bibliográficas de los últimos 5 años, relacionadas con predicción del riesgo cardiovascular o en relación con enfermedades cardíacas implementando la inteligencia artificial, donde se encontraron artículos a nivel internacional y nacional, estos a su vez tienen correlación con aspectos conceptuales y técnicas asociadas para predecir el riesgo y el tipo de datos. Otra revisión realizada fue consultar fuentes que cuente con los algoritmos de aprendizaje automático utilizando funciones kernel y máquinas de soporte vectorial, en las que se obtuvieron:

Antecedentes Internacionales

- Según L. Yang y otros, en el proyecto titulado “Estudio del modelo de predicción de enfermedades cardiovasculares basado en un bosque aleatorio en el este de China”, publicado en el año 2020, (seleccionaron, 29930 sujetos con alto riesgo de enfermedades cardiovasculares (ECV)) de 101056 personas en 2014, se realizó un seguimiento regular utilizando un sistema de registro de salud electrónico. El análisis de regresión logística mostró que casi 30 indicadores estaban relacionados con la ECV, incluidos el género, la vejez, los ingresos familiares, el tabaquismo, el consumo de alcohol, la obesidad, la circunferencia de cintura excesiva, el colesterol anormal, la lipoproteína de baja densidad anormal, la glucemia en ayunas anormal y otros. Se utilizaron varios métodos para construir el modelo de predicción, incluido el modelo de regresión multivariante, el árbol de clasificación y regresión (CART), Naïve Bayes, Bagged trees, Ada Boost y Random Forest. Utilizaron el modelo de regresión multivariante como punto de referencia para la evaluación del desempeño (Área bajo la curva, $AUC = 0,7143$). Los resultados mostraron que Random Forest fue superior a otros métodos con un AUC de 0,787 y logró una mejora significativa con respecto al punto de referencia. Proporcionaron un modelo de predicción de ECV para la evaluación del riesgo de ECV a 3 años. Se basó en una gran población con alto riesgo de ECV en el este de China utilizando el algoritmo Random Forest, que proporcionaría una referencia para el trabajo de predicción y tratamiento de ECV en China. De acuerdo con esto, se necesitan más estudios poblacionales del modelo de predicción de ECV propuesto en esta

investigación, con más población, mayor tiempo de seguimiento, que cubran más lugares en China con validación externa [16], [4]. Este aporte y desarrollo del proyecto permite comparar las técnicas clásicas y como estas actúan en diversos parámetros para obtener resultados similares o diferentes a los ya estudiados, además de determinar las variables que más aportan o afectan a la variable respuesta, que para esta ocasión se obtienen los factores de riesgo significativos en predicción de enfermedades cardiovasculares, así mismo se logra la comparación de métodos usados para la evaluación de desempeño teniendo así ese resultado como referencia y determinar cuán significativa es la escogencia de las variables a trabajar para un determinado resultado.

- Por otro lado, según Chávez Olivera. O., Galindo, Honores.L., Barrientos, padilla.A., & Cuadros, Gálvez.M., en su estudio titulado “Aplicación Móvil para Predecir la Probabilidad de Pertenecer al Grupo de Riesgo Cardiovascular Utilizando Machine Learning” publicado en el año 2022., Lima, Perú [17], el presente estudio se centra en la creación de una aplicación móvil que tiene por principal funcionalidad la predicción de pertenecer al grupo de riesgo cardiovascular en personas mayores de 50 años. Para lograr esto se ha investigado sobre distintas variables y algoritmos de machine learning que permiten lograr esta tarea. Es así como se decidió que el motor de inferencia sería un modelo ensamblado, donde el metaclasificador final es un modelo de Naive Bayes y los modelos base son Random Forest y Logistic Regression. El proceso de validación de la aplicación lo realizó un especialista en cardiología, el cual comprobó el nivel de precisión del modelo. Se observó que los modelos ensamblados de Support Vector Machine, Random Forest, Naive Bayes y Logistic Regression obtienen precisiones de 87.00%, 88.00%, 87.00% y 86.00% con una estabilidad de 8.00%, 6.00%, 2.00% y 8.00% respectivamente, pese a que el modelo ensamblado Random Forest tiene una mejor precisión, este es más inestable, por lo que se escogió el modelo ensamblado Naive Bayes, ya que tiene una precisión parecida y es más estable que los demás. El aporte que se provee de esta investigación es que muestra los resultados de los entrenamientos de distintos modelos de machine learning, y la combinación de estos, el ensamblaje permite la construcción de algoritmos más precisos y estables, además que al usar modelos Random Forest y Naive Bayes se puede determinar la mejor precisión, y tener un punto de comparación con la investigación.

- Según Scavino.M., Castrillejo.A., Estragó, Mérola.V.S., Luraghi López.L.E., Muñoz.M., Álvarez Vas.R., en su informe titulado “Informe final publicable de proyecto Creación de algoritmos utilizando técnicas de clasificación supervisada y no supervisada para el diagnóstico de enfermedades cardiovasculares en una población de adultos mayores de bajos recursos en Uruguay” publicado en el año 2022, Uruguay [18]. La presente investigación tiene como propósito la generación de algoritmos de aprendizaje automático para la identificación de la patología cardíaca, fibrilación auricular, a partir de datos de la señal electrocardiográfica de una sola derivación con un dispositivo móvil de tecnología electrónica. Los algoritmos de aprendizaje profundo con las arquitecturas consideradas no mostraron un buen desempeño. Una mayor cantidad de datos podría resultar en una mejora de la capacidad de clasificación de estos algoritmos. Por otra parte, las técnicas de aprendizaje estadístico aplicadas a un conjunto de características extraídas de la señal ECG sin procesar mostraron un mejor desempeño. Es conveniente destacar que la construcción de estos algoritmos permite entender el funcionamiento de estos y cómo se llega al diagnóstico final, la capacidad de poder interpretar los mecanismos internos de los métodos para ofrecer un resultado genera una fuente de conocimiento amplia con posibilidad de desarrollo a futuro, además de encontrar posibles causas por las que el uso de diferentes algoritmos llega a la variación en los diagnósticos.
- En un cuarto estudio, según Polero.L., Garmendia. C., Echegoyen.R.E., Alves de Lima.A., Bertón.F., Lambardi.F., Ariznavarreta.P., Campos.R., Costabel.J.P., en su proyecto titulado “Predicción de riesgo de sufrir un síndrome coronario agudo mediante un algoritmo de Machine Learning (ANGINA)” publicado en el año 2020, Buenos Aires, Argentina [19]. La presente investigación pretende demostrar la capacidad de los clasificadores de machine learning para diagnosticar y predecir un SCA en pacientes que consultan de forma espontánea al SEM con dolor torácico de etiología no identificada, durante un período de seguimiento de 30 días. Se analizaron 161 pacientes que consultaron al SEM con dolor torácico. Se registró mediante un clasificador de machine learning las variables objetivas y subjetivas de caracterización del dolor. De esta manera se obtuvo que la edad promedio fue de 57 más/menos 12, 72,7% masculinos eran de sexo masculino y 17,4 % presentaban evento coronario previo. El 57,8% presentaba un síndrome coronario agudo con una incidencia de IAM de 29,8%, de los cuales requirieron revascularización por ATC el 35%, y CRM el 9,9% en el período de seguimiento a 30 días. Como

modelo de clasificación se utilizó un Random Forest Classifier que presentó un área bajo la curva ROC de 0,8991, sensibilidad de 0,8552, especificidad de 0,8588 y una precisión de 0,8441. Las variables predictoras más influyentes fueron peso ($p = 0,002$), edad ($p = 5,011e-07$), intensidad del dolor ($p = 3,0679e-05$), tensión arterial sistólica ($p = 0,6068$) y características subjetivas del dolor ($p = 1,590e-04$). Este proyecto permite conocer algunas métricas de evaluación de modelos que proporcionan medidas cuantitativas mostrando su rendimiento y las variables que pueden ser consideradas como importantes para el desarrollo del mismo manifestando a su vez el comportamiento de las variables para la problemática a resolver.

Antecedentes Nacionales

- Según Peres Tatis.J.D., en su estudio titulado “Optimización De Un Modelo De Clasificación De Enfermedades Cardiovasculares Utilizando Técnicas De Aprendizaje Profundo Supervisado Y Despliegue De Dashboard Web” publicado en el año 2021, Cartagena [20]. El desarrollo de este proyecto se basa en la optimización de un modelo previo de predicción de hipertensión, del cual se mejora el proceso de análisis de los datos, ya que el modelo inicial no generaba la predicción exacta de clasificación de enfermedades cardiovasculares debido a la mala interpretación y la falta de limpieza de los datos. Así mismo, la optimización consistió en hacer ajustes en el modelo, realizando mejoras en la construcción de la red neuronal y en el proceso de entrenamiento, identificando técnicas de activación y épocas adecuadas para lograr obtener resultados óptimos. Finalmente, la preparación de los datos permitió formular, entrenar y probar un modelo de clasificación de enfermedades cardiovasculares construido en el entorno de TensorFlow desarrollado por Google el cual se obtuvo una red neuronal recurrente compuesta con una capa de entrada con 18 nodos, dos capas ocultas con 128 nodos cada capa y una capa de salida de 3 nodos en los cuales fueron los más óptimos para la predicción la cual obtuvo una precisión de 97% en la validación del modelo, y posteriormente se desplegó haciendo uso de una aplicación web para la consulta del personal médico el modelo elaborado por los mencionados anteriormente arrojó como resultados, una precisión superior al 86%, viendo los porcentajes en las métricas de evaluación que se utilizaron como: Precisión, Recall, F1 Score y accuracy, las cuales arrojaron resultados que superan el 80% en la clasificación de los riesgos. Esta investigación es valiosa para el desarrollo del proyecto porque al comparar las métricas de

evaluación, presentan resultados considerables, que se pueden abordar con la calidad de los datos que se obtengan, así mismo se hace un recorrido sobre el entorno TensorFlow que permite la visualización de datos en distintas áreas del conocimiento.

- Como segunda investigación, según Martínez.E.J, en su investigación titulada “Predicción De Enfermedades Cardiovasculares Mediante Algoritmos De Inteligencia Artificial” publicado en el año 2020, Málaga [21]. Este proyecto se centra en implementar 5 algoritmos distintos de clasificación, analizar cómo se ajustan a los datos disponibles y, posteriormente, crear un algoritmo genético que detecte la combinación de parámetros de cada algoritmo, con la que se obtienen mejores resultados, medidos a través de la exactitud. Los algoritmos fueron implementados con la librería sci-kit learn de Python; para el algoritmo genético no se utilizó ninguna librería adicional. Los resultados mostraron que algunos algoritmos se adaptan mejor a la evolución, es decir, la exactitud aumenta con el paso de las generaciones. Otros algoritmos mostraron una disminución de este valor, sugiriendo que se necesita estudiar para cada tipo de algoritmo el impacto de cada parámetro, además de los valores que en este proyecto se consideraron constantes: número de generaciones, número de individuos por generación, probabilidad de mutación y cruce y el tamaño del conjunto de datos y de los subconjuntos de entrenamiento, validación y pruebas. Esta investigación es considerable para el proyecto debido a la comparación que hace de algoritmos de clasificación y como estos obtienen mejores resultados a través de la exactitud, aportando a uno de los objetivos específicos con el fin de tener un marco teórico amplio sobre la investigación.
- Como tercera investigación, según Florez Prias.L.A, en su investigación titulado “Modelo de inteligencia artificial como apoyo diagnóstico para la estimación de riesgo cardiovascular en pacientes atendidos bajo la modalidad de telemedicina en una IPS del departamento de Sucre 2021” publicado en el año 2021, Sucre [22]. El presente proyecto busca mediante un estudio experimental la incorporación de Inteligencia Artificial como apoyo diagnóstico en el proceso de atención a pacientes con riesgo cardiovascular. Para lo cual se aprovecharán los datos obtenidos de pacientes atendidos durante el 2018 y 2019 por una IPS del departamento de Sucre. Se presenta la problemática del aumento de mortalidades en personas con enfermedades cardiovasculares que se pueden prevenir si los diagnósticos son oportunos, para ello el proyecto busca mediante un

estudio experimental la incorporación de Inteligencia Artificial como apoyo diagnóstico en el proceso de atención de pacientes con riesgo cardiovascular. Entre las variables a monitorear se encuentran edad, sexo, peso, talla, Índice de masa corporal, presión arterial, presencia de diabetes, dislipidemias, entre otras. A su vez, bajo el concepto de Machine Learning, se desarrollará un algoritmo inteligente con la habilidad de aprender sin ser explícitamente programado que permita evaluar los riesgos potenciales del individuo y así, asistir virtualmente al personal médico en las acciones de promoción, prevención y diagnóstico de manera minuciosa y precisa para la modalidad de atención en telemedicina. Esperando que, a partir de la caracterización de las variables, definir los factores de riesgo de mayor impacto en la incidencia del riesgo cardiovascular en la población de estudio. Con la selección del modelo predictivo que más se ajusta a las características de la población y los datos de calidad suministrados, se podrá definir una clasificación de riesgo cardiovascular que podrá ser adaptado al servicio de tele monitoreo en prevención primaria y secundaria. Con el uso de estos algoritmos de manera particular, permite ver la predicción y estos algoritmos desde otra perspectiva, ya que su intención es que el algoritmo aprende sin ser explícitamente programado. El estudio de este proyecto permite confirmar y corroborar las funciones kernel utilizadas y las que mejor se acondicionen a los datos, dependiente el uso que se le brinde. Además de permitir realizar mejoras en algunos aspectos del desarrollo, para diferentes enfoques [11].

2.2 ENUNCIADO DE SUPUESTOS TEÓRICOS DE LA INVESTIGACIÓN

En este capítulo se toma en cuenta las bases teóricas del presente trabajo, de tipo conceptual.

2.2.1 Enfermedades Cardiovasculares

Son enfermedades del sistema circulatorio, de etiología y localización diversas. Se clasifican en cuatro tipos generales: enfermedades isquémicas del corazón, enfermedades cerebrovasculares, en enfermedades vasculares periféricas y otras enfermedades [23].

Importantes avances en el tratamiento de las ECV han sido facilitados por la identificación de los factores de riesgo tradicionales, pero a pesar de la evidencia clínica acumulada, la implementación

de estrategias para prevenir las enfermedades cardiovasculares aún permanecen lejos de ser óptimas [24].

2.2.2 El riesgo cardiovascular

El riesgo cardiovascular se define como la probabilidad de padecer un evento cardiovascular en un determinado período, que habitualmente se establece en 5 o 10 años, especialmente en los pacientes que no padecen enfermedad cardiovascular, es decir, en prevención primaria, es fundamental para establecer la intensidad de la intervención, la necesidad de instaurar tratamiento farmacológico y la periodicidad de las visitas de seguimiento [24].

2.2.3 Factores de riesgo cardiovasculares

Los factores de riesgo son aquellos signos biológicos o hábitos adquiridos que se presentan con mayor frecuencia en los pacientes con una enfermedad concreta. La enfermedad cardiovascular tiene un origen multifactorial, y un factor de riesgo debe ser considerado en el contexto de los otros. Los factores de riesgo cardiovascular, clásicos o tradicionales, se dividen en 2 grandes grupos: no modificables (edad, sexo y antecedentes familiares), y modificables (dislipidemia, tabaquismo, diabetes, hipertensión arterial, obesidad y sedentarismo).

Aunque el impacto de factores de riesgo individuales como la hipertensión arterial, la dislipidemia, el hábito de fumar y la diabetes, entre otros, está bien establecido y mejora la predicción del riesgo cardiovascular [24].

Cuanto mayor sea el nivel de cada factor de riesgo, mayor es el riesgo de tener una enfermedad cardiovascular aterosclerosis como la cardiopatía coronaria.

- **Obesidad:** La relación entre el peso y las enfermedades del corazón no viene tan sorpresa, ya que una persona obesa tendrá más grasa y, por lo tanto, más posibilidades de sufrir ECV. “Independientemente de la salud metabólica, las personas con sobrepeso y obesas tienen mayor riesgo de enfermedad coronaria que las personas delgadas” [25].
- **Actividad física:** El ejercicio constante y un estilo de vida saludable pueden, en general, impactar muy positivamente al tratamiento de enfermedades: prevenir o retrasar la aparición

el tipo 2 diabetes, reducir la presión arterial y ayudar a reducir el riesgo de ataque cardíaco y accidente cerebrovascular [25]. Cuanto más vigorosa la actividad, mayor el beneficio. Sin embargo, aun las actividades de intensidad moderada ayudan si se realizan de forma habitual y a largo plazo. El ejercicio puede ayudar a controlar el colesterol, la diabetes y la obesidad, así como a reducir la presión arterial en algunas personas. La actividad física debería ser una actividad diaria. Caminar entre 30 a 40 minutos, la mayor cantidad de días por semana posibles, pero no menos de 3 días es un buen ejercicio y tiene pocas contraindicaciones [26].

- **Niveles de colesterol:** la acumulación de colesterol es una de las principales causas de la aterosclerosis. Se ha demostrado consistentemente que niveles más altos de colesterol LDL a largo plazo y las concentraciones de colesterol de lipoproteínas que no son de alta densidad están asociadas con un mayor riesgo de ECV [25].
- **Glucosa / Diabetes:** la diabetes no es solo una alteración de los niveles de azúcar en sangre, pero afecta al sistema en general. Los estudios reportan una asociación positiva entre hipertensión y resistencia a la insulina [25].
- **Tabaquismo:** hay evidencia de que fumar causa aproximadamente 1 de cada 10 muertes por enfermedades cardiovasculares. El humo del tabaco contribuye a las enfermedades cardiovasculares, ya que aumenta la placa aterosclerótica y la posibilidad de trombosis [25]. El humo del tabaco es el principal factor de riesgo para la muerte súbita de origen cardíaco y los fumadores tienen de dos a cuatro veces más riesgo que los no fumadores. El riesgo cardiovascular disminuye rápidamente al dejar de fumar [26].
- **Antecedentes Familiares:** los hijos/as de padres con cardiopatía isquémica, especialmente si esta ha sido prematura (padres antes de los 65 años, madres antes de los 55 años) o con hipertensión arterial, tienen mayor probabilidad de desarrollarla. Existen formas minoritarias de colesterol muy elevado (por encima de los 350 mg/dl) llamadas hipercolesterolemia familiar, que son debidas a trastornos hereditarios y que conllevan un riesgo muy elevado, incluso antes de la menopausia. En estos casos son precisos tratamientos médicos agresivos con hipolipemiantes [26].

2.2.4 Método para calcular el riesgo cardiovascular

Existen 2 métodos de cálculo del riesgo cardiovascular: cualitativos y cuantitativos. Los cualitativos se basan en la suma de factores de riesgo o la medición de su nivel y clasifican al individuo en: riesgo leve, moderado, alto y muy alto; los cuantitativos, por su parte, están basados en ecuaciones de predicción de riesgo que nos dan un número que es la probabilidad de presentar un evento cardiovascular en un determinado tiempo, y la forma de cálculo es a través de programas informáticos o de las llamadas tablas de riesgo cardiovascular, que son unas herramientas de enorme utilidad para la toma de decisiones en la práctica clínica habitual [24].

- Algoritmo de estudio Framingham: la probabilidad de que ocurra una enfermedad cardiovascular para una variable determinada, así encontramos:
- Hombres y mujeres tienen una probabilidad distinta de llegar a padecer una enfermedad cardiovascular.
- La edad es otro determinante; a mayor edad mayor el riesgo cardiovascular.
- El hábito en el consumo de tabaco es una variable que aumenta el riesgo cardiovascular independientemente de las otras variables.
- Los niveles de colesterol, HDL y LDL son todas variables que aumentan o disminuyen (HDL) el riesgo cardiovascular, independientemente de las demás.
- Los niveles de presión arterial altos y si se tiene o no tratamiento farmacológico para la hipertensión.

Las anteriores son las variables clásicas que analiza el estudio, pero además otros investigadores buscan sumar algunas nuevas:

- Ascendencia, estado civil y educación.
- Tipo de trabajo, ritmo y horario del trabajo, apoyo de los compañeros o del supervisor.
- Perímetro abdominal en personas con diabetes o síndrome metabólico [27].

2.2.5 Inteligencia artificial

La Inteligencia Artificial (IA) podría definirse como la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano, bien sea: pensar, sentir, resolver problemas, tomar decisiones e inclusive aprender, IA comprende el área del Machine Learning, Deep Learning, Big Data y ciencia de datos.

- **Scikit Learn:** Scikit-learn es una biblioteca de aprendizaje automático gratuita para Python. Cuenta con varios algoritmos como máquina de vectores de soporte, bosques aleatorios y vecinos k, y también admite bibliotecas numéricas y científicas de Python como NumPy y SciPy [28].

2.2.6 Máquinas de soporte vectorial, Algoritmo de soporte vectorial (SVM)

Es un algoritmo de aprendizaje supervisado que se utiliza en muchos problemas de clasificación y regresión, incluidas aplicaciones médicas de procesamiento de señales, procesamiento del lenguaje natural y reconocimiento de imágenes y voz.

El objetivo del algoritmo SVM es encontrar un hiperplano que separe de la mejor forma posible dos clases diferentes de puntos de datos. “De la mejor forma posible” implica el hiperplano con el margen más amplio entre las dos clases, representado por los signos más y menos en la siguiente figura. El margen se define como la anchura máxima de la región paralela al hiperplano que no tiene puntos de datos interiores. El algoritmo solo puede encontrar este hiperplano en problemas que permiten separación lineal; en la mayoría de los problemas prácticos, el algoritmo maximiza el margen flexible, permitiendo un pequeño número de clasificaciones erróneas [29].

2.2.7 Funciones kernel

Las Funciones Kernel. Son funciones matemáticas que se emplean en las Máquinas de Soporte Vectorial. Estas funciones son las que le permiten convertir lo que sería un problema de clasificación no lineal en el espacio dimensional original, a un sencillo problema de clasificación lineal en un espacio dimensional mayor a este espacio M-dimensional se le conoce como espacio de Hilbert.

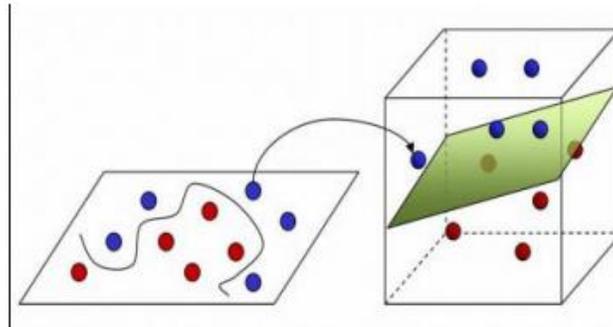


Fig. 1. Truco kernel de dos dimensiones a tres

Nota: fuente https://www.researchgate.net/figure/260283043_fig13_Figure-A15-The-non-linear-SVM-classifier-with-the-kernel-trick

En la figura 1, se puede observar cómo trabajan las funciones kernel, llevando una distribución de datos de dos dimensiones a tres dimensiones, a este funcionamiento generalmente se lo llama truco kernel. Este truco permite reducir la complejidad de una función que separe las clases de una distribución de datos como lo demuestra Peluffo-Ordóñez, la figura 1 por ejemplo en dos dimensiones se podría separar mediante funciones no lineales o segmentadas (Elipse, Círculo, Rectángulo), mientras que en tres dimensiones se podría separar mediante una función lineal (Hiperplano) como lo demuestra Baudat & Anouar en el documento [11].

Para que las funciones kernel puedan ser consideradas candidatas a kernels, deben cumplir tres condiciones iniciales fundamentales; deben ser:

- Continuas
- Simétricas
- Positivas.

Estos son los requerimientos básicos para poder ser expresadas como un producto escalar en un espacio dimensional alto [30].

Existen diversos kernel empleados comúnmente en bibliotecas de aprendizaje automático como lineal, RBF, polinomial y tangente hiperbólico cuyas definiciones se pueden expresar en la Tabla 1.

[11]

TABLA I
DEFINICIÓN FORMAL FUNCIONES KERNEL

Fuente: Comparativo de funciones kernel sobre predicción de oferta de fuentes alternativas de energía

Función kernel	Ecuación	Condición
Lineal	$k(x, x') = \langle x, x' \rangle$	$x, x' \in \mathbb{R}$
RBF	$k(x, x') = \exp \left(- \sum_{i=1}^d \lambda_i (x_i x'_i) \right)$	$\lambda_i > 0, \beta \in (0, 2]$
Polinomial	$k(x, x') = (\alpha \langle x, x' \rangle + 1)^m$	$m \in \mathbb{N}, \alpha > 0$
Tangente hiperbólica	$k(x, x') = \tanh(\alpha \langle x, x' \rangle + b)$	$a > 0, b < 0$

Aunque las funciones de la Tabla I son funciones de uso común, existen muchas más funciones kernel. Estas funciones son utilizadas en algoritmos supervisados para regresión y clasificación; y no supervisadas para detección de anomalías, análisis, clúster y extracción de características. Dentro de los algoritmos más destacados se encuentran procesos gaussianos, Spectral clustering, Kernel Linear Discriminant Analysis, Kernel Principal Components Analysis, Kernel Canonical Correlation Analysis, Kernel Independent Component Analysis, SVM, ANN y muchos más [11].

2.2.8 Identificación de patrones mediante modelos lineales

La detección de patrones es el proceso de encontrar en un conjunto de datos caóticos modelos capaces de generalizar el comportamiento de los datos para la obtención de clasificaciones, predicciones o detección de anomalías. Para la obtención de estos patrones se recurre a la sinergia del conocimiento provista por varias ciencias como las matemáticas, estadística, probabilidad, computación, entre otras. Actualmente, estas técnicas están englobadas en una rama denominada aprendizaje automático, divididas en técnicas de aprendizaje supervisado, no supervisado y por refuerzo [11].

- **Clasificadores lineales:** Un clasificador lineal es aquel capaz de encontrar la clase (y) discreta a la que pertenece un conjunto de datos basada en una combinación lineal de sus atributos (x) como se muestra en la Figura 1 [11].

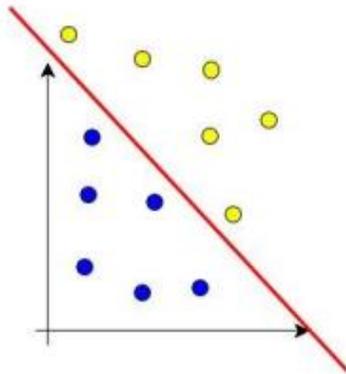


Fig. 2. Ejemplo de clasificador lineal

Nota: fuente <https://docplayer.es/192405564-Comparativo-de-kernels-sobre-prediccion-de-oferta-de-fuentes-alternativas-de-energia.html>

Como se observa en la Figura 2 el clasificador lineal ha trazado un separador, en este caso una línea que permite deducir a que clase pertenece (puntos amarillos o azules) así un nuevo registro, si se posiciona del lado superior de la línea sería clasificado como punto amarillo, de lo contrario como azul [11].

Si la entrada del clasificador es un vector de características reales \vec{x} , entonces el resultado de salida es

$$y = f(\vec{w} \cdot \vec{x}) = \left(f \sum_j w_j x_j \right),$$

Fig. 3. Clasificadores Lineales

Nota: fuente <https://docplayer.es/192405564-Comparativo-de-kernels-sobre-prediccion-de-oferta-de-fuentes-alternativas-de-energia.html>

Donde \vec{w} es un vector real de pesos y f es una función que convierte el producto punto a punto de los dos vectores en la salida deseada. El vector de pesos \vec{w} aprende de un conjunto de muestras de entrenamiento. A menudo f es una función simple que mapea todos los valores por encima de un cierto umbral a la primera clase y el resto a la segunda clase [31].

Algunos de los algoritmos de clasificación lineal más utilizados para encontrar estos patrones de clasificación se encuentran: análisis de discriminante lineal, clasificados de Bayes lineal, regresión

- **Aprendizaje supervisado:** es una rama de Machine Learning, un método de análisis de datos que utiliza algoritmos que aprenden iterativamente de los datos para permitir que los ordenadores encuentren información escondida sin tener que programar de manera explícita dónde buscar. El aprendizaje supervisado es uno de los tres métodos de la forma en que las máquinas "aprenden": supervisado, no supervisado y optimización. El aprendizaje supervisado resuelve problemas conocidos y utiliza un conjunto de datos etiquetados para entrenar un algoritmo para realizar tareas específicas [32].
- **Aprendizaje no supervisado:** es un tipo de Machine Learning que se utiliza para identificar nuevos patrones y detectar anomalías. Los datos que se introducen en los algoritmos de aprendizaje no supervisados no están etiquetados. El algoritmo (o modelos) intentan dar sentido a los datos por sí mismos mediante la búsqueda de características y patrones [32]. Es importante destacar que para este proyecto se centrara específicamente en las técnicas de aprendizaje supervisado para clasificación y predicción mediante modelos lineales.
- **Clasificadores lineales:** Un clasificador lineal es aquel capaz de encontrar la clase (y) discreta a la que pertenece un conjunto de datos basados en una combinación lineal de sus atributos (x) [11].

2.2.9 Base de datos

Se denomina base de datos a un conjunto de información perteneciente a un mismo contexto, ordenada de modo sistemático para su posterior recuperación, análisis y transmisión. Hoy en día, las bases de datos se presentan de diferentes formas y tamaños, esto de acuerdo al lugar donde son empleados, por ejemplo, en una biblioteca o en cuentas de una empresa. Las bases de datos se originaron para cubrir la necesidad de almacenar grandes cantidades de información, es decir, de preservarla contra el tiempo y el deterioro, para acudir a ella posteriormente. Es ese sentido, la aparición de la electrónica y la computación brindaron el elemento digital indispensable para almacenar enormes cantidades de datos en espacios físicos limitados, gracias a su conversión en señales eléctricas o magnéticas [33].

2.2.10 Métricas de medición

Las métricas de regresión en aprendizaje automático y cada modelo de Machine Learning que la utiliza intenta resolver un problema con otro objetivo utilizando un conjunto de datos diferente [34].

TABLA II
MÉTRICAS DE REGRESIÓN EN APRENDIZAJE AUTOMÁTICO
MÉTRICAS DE REGRESIÓN EN APRENDIZAJE AUTOMÁTICO

NOMBRE	DEFINICIÓN	FÓRMULA
<p>(MSE) Error cuadrático medio</p>	<p>Es una métrica utilizada en estadística y aprendizaje automático para evaluar el rendimiento de un modelo de predicción. Se utiliza para medir la diferencia entre los valores predichos por el modelo y los valores reales observados.</p> <p>El MSE se calcula tomando la diferencia entre cada valor predicho y el valor real correspondiente, elevando al cuadrado esta diferencia y luego calculando el promedio de todos estos errores cuadráticos. En otras palabras, se obtiene la media de los errores cuadrados.</p> <p>Cuanto menor sea el valor del MSE, mejor será el modelo, ya que indica que las predicciones se acercan más a los valores reales. Un MSE de cero indicaría un modelo perfecto en el que las predicciones coinciden exactamente con los valores observados [34].</p>	$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$ <p style="text-align: center;">Error Cuadrático Medio</p> <p>Donde n es el número de observaciones en el conjunto de datos, y_i es el valor real observado y \hat{y}_i es el valor predicho por el modelo.</p>

<p>(RMSE) Raíz del error cuadrático medio</p>	<p>Es una medida estadística comúnmente utilizada para evaluar la precisión de un modelo de regresión. Es una métrica que se deriva del Error Cuadrático Medio (MSE) y se calcula tomando la raíz cuadrada del MSE [34].</p>	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} = \sqrt{MSE}$ <p style="text-align: center;">RMSE</p> <p>Donde MSE es el Error Cuadrático Medio.</p> <p>Al igual que el MSE, el RMSE mide la diferencia promedio entre los valores predichos por el modelo y los valores reales observados. La principal diferencia es que el RMSE tiene la misma unidad de medida que la variable objetivo, lo que lo hace más interpretable y fácil de comparar con los valores reales.</p> <p>El RMSE se utiliza ampliamente en problemas de regresión para evaluar qué tan cerca están las predicciones del modelo de los valores reales. Cuanto menor sea el valor del RMSE, mejor será la capacidad predictiva del modelo. Por ejemplo, si se tiene dos conjuntos de predicciones, A y B, y se dice que el MSE de A es mayor que el MSE de B, entonces se puede estar seguro de que RMSE de A es mayor que RMSE de B. Y también funciona en la dirección opuesta.</p>
<p>(MAE) -Error absoluto medio</p>	<p>Se calcula como la media de las diferencias absolutas entre los valores predichos por el modelo y los valores reales observados. El MAE proporciona una medida de la magnitud promedio de los errores de predicción sin tener en cuenta su dirección. Es una métrica comúnmente utilizada debido a su simplicidad y facilidad de interpretación.</p> <p>El MAE se expresa en las mismas unidades que la variable objetivo y</p>	$MAE = \frac{1}{N} \sum_{i=1}^N y_i - \hat{y}_i $ <p style="text-align: center;">Error Absoluto Medio</p> <p>Donde:</p> <ul style="list-style-type: none"> • MAE: Mean Absolute Error • N: Número de observaciones en el conjunto de datos. • \sum: Sumatoria. • y_i: Valor real observado.

	<p>cuanto menor sea su valor, mayor será la precisión del modelo en términos de predicción.</p> <p>Es importante tener en cuenta que el MAE no considera la magnitud relativa de los errores, por lo que todos los errores se tratan por igual en la métrica. Esto puede ser adecuado en ciertos escenarios donde se desea penalizar de manera uniforme todos los errores de predicción [34].</p>	<ul style="list-style-type: none"> • \hat{y}_i: Valor predicho por el modelo.
<p>(R²) – R al cuadrado</p>	<p>El coeficiente de determinación, o R² (a veces leído como R-dos), es una medida estadística utilizada en el análisis de regresión para evaluar qué tan bien se ajusta un modelo a los datos observados. El R-cuadrado es un valor que varía entre 0 y 1, y se interpreta como el porcentaje de la variabilidad de la variable dependiente que es explicada por el modelo.</p> <p>un valor de R² cercano a 1 indica que el modelo explica una gran proporción de la variabilidad de la variable dependiente y se ajusta bien a los datos. Por otro lado, un valor de R² cercano a 0 indica que el modelo no explica de manera adecuada la variabilidad de la variable dependiente y no se ajusta bien a los datos [34].</p>	$R^2 = 1 - \frac{\text{MSE}(\text{modelo})}{\text{MSE}(\text{baseline})}$ <p style="text-align: center;">R Cuadrado</p>
<p>R cuadrado ajustado (R²)</p>	<p>Es una medida estadística relacionada con el coeficiente de determinación (R-cuadrado) que tiene en cuenta la cantidad de variables predictoras en un modelo de regresión y ajusta el R-cuadrado en función del número de</p>	$R_{\text{adj}}^2 = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right]$ <p style="text-align: center;">R Cuadrado Ajustado</p> <p>Donde:</p> <ul style="list-style-type: none"> • R² ajustado: Coeficiente de determinación

	<p>predictores utilizados. El R-cuadrado ajustado se utiliza para evaluar y comparar modelos de regresión que tienen diferentes números de variables predictoras.</p> <p>El R-cuadrado ajustado penaliza el uso de variables predictoras adicionales que no aportan información significativa al modelo. A medida que se agregan más variables predictoras al modelo, el R-cuadrado ajustado disminuirá si esas variables no mejoran de manera sustancial la capacidad de explicación del modelo. Por lo tanto, el R-cuadrado ajustado tiende a ser más conservador que el R-cuadrado estándar y proporciona una medida más realista del ajuste del modelo [34].</p>	<p>ajustado.</p> <ul style="list-style-type: none"> • R²: Coeficiente de determinación (R-cuadrado). • n: es el número total de observaciones • k: es el número de regresores independientes, es decir, el número de variables en su modelo, excluyendo la constante.
<p>(MSPCE) – Error porcentual cuadrático medio</p>	<p>Es una medida de error utilizada para evaluar la precisión de un modelo de predicción en relación con los valores reales. Es una métrica que combina la magnitud de los errores y la proporción relativa de los errores en relación con los valores reales.</p> <p>El MSPE se expresa como un porcentaje, lo que facilita la interpretación de la magnitud del error en relación con los valores reales. Un valor más bajo de MSPE indica una mejor precisión del modelo, mientras que un valor más alto indica una mayor discrepancia entre las predicciones y los valores reales [34].</p>	$MSPE = \frac{100\%}{N} \sum_{i=1}^N \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2$ <p>Error de Porcentaje Cuadrático Medio (MSPCE)</p> <p>Donde:</p> <ul style="list-style-type: none"> • n es el número total de muestras o datos. • y_i: son los valores reales o verdaderos. • \hat{y}_i: son los valores predichos por el modelo.
<p>(MAPE) – Error porcentual</p>	<p>Es una medida de error comúnmente utilizada para evaluar la precisión de un</p>	

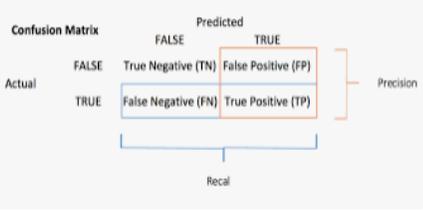
<p>absoluto medio</p>	<p>modelo de predicción en relación con los valores reales. El MAPE calcula el promedio del porcentaje absoluto de error entre las predicciones y los valores reales.</p> <p>El MAPE se expresa como un porcentaje, lo que facilita la interpretación de la magnitud del error en relación con los valores reales. Un valor más bajo de MAPE indica una mejor precisión del modelo, mientras que un valor más alto indica una mayor discrepancia entre las predicciones y los valores reales [34].</p>	$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left \frac{y_i - \hat{y}_i}{y_i} \right $ <p>Error Porcentual Absoluto Medio (MAPE)</p> <p>Donde:</p> <ul style="list-style-type: none"> • n es el número total de muestras o datos. • y_i: son los valores reales o verdaderos. • \hat{y}_i: son los valores predichos por el modelo.
<p>(RMSLE) – Error logarítmico cuadrático medio</p>	<p>Es solo un RMSE calculado en escala logarítmica. Es una métrica útil cuando los valores tienen una amplia gama y hay una gran variabilidad en los datos. El RMSLE toma el logaritmo de los valores reales y los valores predichos antes de calcular el error cuadrático medio. Esto es útil cuando los valores objetivo abarcan un rango amplio y se desea penalizar de manera más equitativa los errores en diferentes magnitudes.</p> <p>El RMSLE se calcula como la raíz cuadrada del promedio de los errores cuadráticos de los logaritmos. Al tomar la raíz cuadrada, se obtiene una medida de error en la misma escala que los valores originales [34].</p>	<p>RMSLE</p> $= \sqrt{\frac{1}{N} \sum_{i=1}^N (\log(y_i + 1) - \log(\hat{y}_i + 1))^2} =$ $= \text{RMSE}(\log(y_i + 1), \log(\hat{y}_i + 1)) =$ $= \sqrt{\text{MSE}(\log(y_i + 1), \log(\hat{y}_i + 1))} =$ <p>Error Logarítmico Cuadrático Medio (RMSLE)</p> <p>Por lo tanto, esta métrica se usa generalmente en la misma situación que MSPCE y MAPE, ya que también conlleva errores relativos más que errores absolutos.</p> <p>RMSLE penaliza una estimación poco predicha mayor que una estimación sobre pronosticada. RMSLE se puede calcular sin la operación raíz, pero la versión rooteada se usa más ampliamente.</p>

2.2.11 Métricas de clasificación

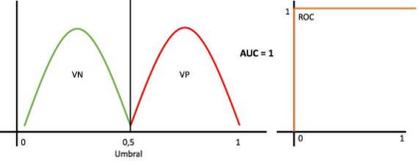
TABLA III
MÉTRICAS DE EVALUACIÓN

MÉTRICAS DE EVALUACIÓN		
NOMBRE	DEFINICIÓN	FÓRMULA
Precision (Precisión)	<p>Es una medida de evaluación utilizada en problemas de clasificación para medir la exactitud de un modelo al predecir correctamente las instancias positivas. Representa la proporción de predicciones positivas que son verdaderamente positivas en comparación con todas las predicciones positivas realizadas por el modelo.</p> <p>La precisión se expresa como un valor entre 0 y 1, donde 1 representa una precisión perfecta y 0 indica una precisión nula.</p> <p>La precisión es una métrica útil cuando el objetivo principal es minimizar los falsos positivos, es decir, evitar clasificar incorrectamente instancias negativas como positivas. Es especialmente importante en casos donde los falsos positivos tienen un impacto significativo o costoso.</p> <p>Es importante tener en cuenta que la precisión no tiene en cuenta los casos negativos que fueron correctamente clasificados como negativos (verdaderos negativos) [35].</p>	$\text{precision} = \frac{TP}{TP + FP}$ <p>Fórmula de Precisión (Precisión)</p> <p>Donde:</p> <ul style="list-style-type: none"> • Verdaderos positivos (TP) son los casos positivos que fueron correctamente identificados por el modelo. • Falsos positivos (FP) son los casos negativos que fueron incorrectamente clasificados como positivos por el modelo.

<p>RECALL (Exhaustividad)</p>	<p>El recall, también conocido como sensibilidad o tasa de verdaderos positivos, es una medida de evaluación utilizada en problemas de clasificación para medir la capacidad de un modelo de identificar correctamente las instancias positivas.</p> <p>El recall se expresa como un valor entre 0 y 1, donde 1 representa un recall perfecto y 0 indica un recall nulo.</p> <p>El recall es una métrica importante cuando el objetivo principal es minimizar los falsos negativos, es decir, evitar clasificar incorrectamente instancias positivas como negativas. Es especialmente relevante en casos donde los falsos negativos tienen un impacto significativo o costoso [35].</p>	$\text{recall} = \frac{TP}{TP + FN}$ <p>Exhaustividad (recall)</p> <p>Donde:</p> <ul style="list-style-type: none"> • Verdaderos positivos (TP): son los casos positivos que fueron correctamente identificados por el modelo. • Falsos negativos (FN): son los casos positivos que fueron incorrectamente clasificados como negativos por el modelo.
<p>F1-SCORE (Valor-F)</p>	<p>El F1-score también conocido como puntuación F1, es una medida de evaluación utilizada en problemas de clasificación que combina la precisión y el recall en una sola métrica. Representa el equilibrio entre la precisión y el recall de un modelo.</p> <p>El F1-score es una medida que varía entre 0 y 1, donde 1 representa un F1-score perfecto y 0 indica un rendimiento nulo. Un F1-score más alto indica un mejor equilibrio entre la precisión y el</p>	$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ <p>F1-Score</p> <ul style="list-style-type: none"> • Precision es la proporción de verdaderos positivos (TP) sobre el total de instancias clasificadas como positivas por el modelo, es decir, la capacidad del modelo para evitar falsos positivos. • Recall es la proporción de verdaderos positivos (TP) sobre el total de instancias positivas reales, es decir, la capacidad del modelo para evitar falsos negativos.

	<p>recall.</p> <p>El F1-score es especialmente útil cuando hay un desequilibrio entre las clases en los datos de entrenamiento [35].</p>	
<p>Accuracy (Exactitud)</p>	<p>La exactitud (accuracy) representa la proporción de instancias clasificadas correctamente sobre el total de instancias en el conjunto de datos [35].</p> <p>La exactitud se calcula dividiendo el número de predicciones correctas (verdaderos positivos y verdaderos negativos) entre el número total de instancias en el conjunto de datos.</p> <p>La exactitud es la cantidad de aciertos del modelo sobre el total de instancias. Una exactitud de 1 indica que el modelo clasifica todas las instancias correctamente, mientras que una exactitud de 0 indica que el modelo no acierta ninguna instancia [35].</p>	$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$ <p>Fórmula Accuracy</p>
<p>CONFUSION MATRIX (Matriz de confusión)</p>	<p>Confusión o error Matrix es una tabla que describe el rendimiento de un modelo supervisado de Machine Learning en los datos de prueba, donde se desconocen los verdaderos valores. Se llama “matriz de confusión” porque hace que sea fácil detectar dónde el sistema está confundiendo dos clases.</p> <ul style="list-style-type: none"> • True Positives (TP): cuando la clase real del punto de datos 	 <p>Fig. 4. Matriz de Confusión</p>

	<p>era 1 (Verdadero) y la predicha es también 1 (Verdadero).</p> <ul style="list-style-type: none"> • Verdaderos Negativos (TN): cuando la clase real del punto de datos fue 0 (Falso) y el pronosticado también es 0 (Falso). • False Positives (FP): cuando la clase real del punto de datos era 0 (False) y el pronosticado es 1 (True). • False Negatives (FN): Cuando la clase real del punto de datos era 1 (Verdadero) y el valor predicho es 0 (Falso) [34]. 	
<p>ESPECIFICIDAD O TNR (Tasa negativa real)</p>	<p>Es el número de ítems correctamente identificados como negativos fuera del total de negativos.</p> <p>La especificidad se calcula dividiendo el número de verdaderos negativos (TN) entre la suma de los verdaderos negativos (TN) y los falsos positivos (FP).</p> <p>La especificidad proporciona información sobre la capacidad del modelo para evitar clasificar incorrectamente los casos negativos. Un valor alto de especificidad indica que el modelo tiene una alta tasa de aciertos en la identificación de los casos negativos [34].</p>	$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$ <p>Fórmula Especificidad Área</p>
<p>ÁREA BAJO LA CURVA DE FUNCIONAMIENTO</p>	<p>Es una métrica utilizada en problemas de clasificación binaria</p>	

<p>DEL RECEPTOR (ROC) (AUC)</p>	<p>para evaluar la capacidad de un modelo para discriminar entre clases positivas y negativas.</p> <p>La curva ROC representa la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) para diferentes umbrales de clasificación. El área bajo esta curva (AUC) proporciona una medida de la capacidad de discriminación del modelo: cuanto mayor sea el valor del AUC, mejor será la capacidad del modelo para distinguir entre las clases.</p> <p>El AUC se calcula integrando el área bajo la curva ROC, que varía de 0 a 1. Un valor de AUC de 0.5 indica un rendimiento aleatorio o equivalente al azar, mientras que un valor de AUC cercano a 1 indica un rendimiento excelente del modelo en la clasificación.</p> <p>El AUC es una métrica popular en la evaluación de modelos de clasificación, especialmente cuando los conjuntos de datos están desequilibrados. Proporciona una medida agregada de la precisión del modelo en todos los umbrales de clasificación posibles y es independiente del umbral de decisión específico utilizado [34].</p>	 <p>Fig. 5. Área Bajo la Curva</p>
<p>PÉRDIDA LOGARÍTMICA</p>	<p>Es una métrica utilizada en problemas de clasificación para evaluar la calidad de las predicciones de un modelo de</p>	

	<p>aprendizaje automático.</p> <p>La pérdida logarítmica mide la discrepancia entre las probabilidades predichas por el modelo y las etiquetas verdaderas de los ejemplos de entrenamiento. Cuanto más baja sea la pérdida logarítmica, mejor será el modelo en términos de su capacidad para asignar probabilidades adecuadas a las clases correctas.</p> <p>La pérdida logarítmica penaliza de manera más pronunciada las predicciones incorrectas con probabilidades cercanas a 0 o 1. En problemas de clasificación binaria, se busca minimizar la pérdida logarítmica para obtener un modelo con mejores predicciones [36].</p>	$-(y \log(p) + (1 - y) \log(1 - p))$ <p style="text-align: center;">Pérdida Logarítmica</p> <p>Donde:</p> <ul style="list-style-type: none"> • y es la etiqueta verdadera del ejemplo (0 o 1). • p es la probabilidad predicha por el modelo de que el ejemplo pertenezca a la clase positiva.
--	--	---

2.2.12 Metodología en Espiral

El desarrollo en espiral es un modelo de ciclo de vida del software definido por primera vez por Barry Boehm en 1986,1 utilizados generalmente en la ingeniería de software.

Las actividades de este modelo consisten en la confirmación en un espiral, en la que cada bucle o iteración representa un conjunto de actividades. Las actividades no están fijadas a ninguna prioridad, sino que las siguientes se eligen en función del análisis de riesgo, comenzando por el bucle interior. Un modelo de ciclo de vida en espiral tiene en cuenta fuertemente el riesgo que aparece a la hora de desarrollar software. Para ello, se comienza mirando las posibles alternativas de desarrollo, se opta por la de riesgo más asumible y se hace un ciclo de la espiral. Si el cliente quiere seguir haciendo mejoras en el software, se vuelve a evaluar las distintas nuevas alternativas y riesgos y se realiza otra vuelta de la espiral, así hasta que llegue un momento en el que el producto software desarrollado sea aceptado y no necesite seguir mejorándose con otro nuevo ciclo [37].

2.3 VARIABLES DE ESTUDIO

2.3.1 DEFINICIÓN NOMINAL DE LAS VARIABLES

En las siguientes tablas se muestran las variables dependientes e independientes presentes en el estudio de la predicción.

TABLA IV
VARIABLES DEPENDIENTES DE ESTUDIO

Nombre	Descripción	Tipo de Variable
Riesgo Cardiovascular	El riesgo cardiovascular es la probabilidad que tiene cada persona de sufrir una enfermedad cardiovascular, es decir, un infarto de miocardio, hemorragias cerebrales, embolias, etcétera.	Dependiente
Tiempo de entrenamiento	Es el tiempo que se toma un algoritmo en detectar los patrones de un conjunto de datos.	Dependiente
Tiempo de predicción	Conjunto de actuaciones que, siguiendo una metodología determinada y a través de los resultados de los modelos numéricos de predicción, van dirigidas a definir el valor más probable de los parámetros de tiempo.	Dependiente
Exactitud	La exactitud (accuracy) mide el porcentaje de casos que el modelo ha acertado [35].	Dependiente

TABLA V
VARIABLES INDEPENDIENTES DE ESTUDIO

Nombre	Descripción	Unidad de Medida	Tipo de Variable	Valores Normales
Edad	Edad de la persona	Años	Independiente	0-120 años
Peso	El peso es una medida de la fuerza gravitatoria que actúa sobre un objeto.	kilogramos	Independiente	IMC es entre 18.5 y 24.9, está dentro de los valores "normales" o de peso saludable. Si su IMC es entre 25.0 y 29.9, está dentro de los valores correspondientes a "sobrepeso". Si su IMC es 30.0 o superior, está dentro de los valores de "obesidad".
Talla	Medida de la persona en estatura	Centímetros	Independiente	1-200 centímetros
Género	El género se refiere a los roles, las características y oportunidades definidos por la sociedad que se consideran apropiados para los hombres, las mujeres, los niños, las niñas y las personas con identidades no binarias.	No aplica	Independiente	No Aplica
Tensión Arterial	Medida de la fuerza o presión de la sangre sobre las arterias cuando el corazón bombea.	Milímetro de mercurio	Independiente	Presión sistólica de menos de 120 y una presión diastólica de menos de 80.

Nivel de colesterol en sangre	Mide la cantidad de colesterol y de ciertos lípidos en la sangre.	Miligramos de colesterol por decilitro de sangre	Independiente	Menos de 200 mg/dL
Nivel de glucosa en sangre	Mide los niveles normales de azúcar en sangre	Miligramos de glucosa por decilitro de sangre	Independiente	Menor que 100 mg/dl

2.3.2 DEFINICIÓN OPERATIVA DE LAS VARIABLES

Las variables independientes corresponden a variables que ya han sido medidas en grupo de datos proporcionados por investigadores expertos de importancia académica y de investigación, en la cual se confirma en el documento SEA 2022 para el control global del riesgo cardiovascular. Para obtener los resultados de este estudio, se ha analizado un dataset (conjunto de datos) obtenido a través de la plataforma Kaggle, plataforma de Data Science que permite a los usuarios adquirir y publicar conjuntos de datos.

Las variables dependientes se pueden medir después de que se entrene cada modelo, partiendo de ello la exactitud (accuracy) representa el porcentaje de predicciones correctas frente al total. Por tanto, es el cociente entre los casos bien clasificados por el modelo (verdaderos positivos y verdaderos negativos), y la suma de todos los casos [20].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Ecuación 1. Accuracy

La precisión (precision) se refiere a lo cerca que está el resultado de una predicción del valor verdadero. Por tanto, es el cociente entre los casos positivos bien clasificados por el modelo y el total de predicciones positivas. Esta es priorizada en los casos donde tener una gran cantidad de falsos positivos tiene un mayor coste [20].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Ecuación 2. Precision

La sensibilidad (recall) representa la tasa de verdaderos positivos (TP). Es la proporción entre los casos positivos bien clasificados por el modelo, respecto al total de positivos [20].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Ecuación 3. Recall

El valor-F (F1-score) se utiliza para combinar las medidas de precision y recall en un solo valor, donde 0 es la peor puntuación y 1 la mejor. Se utiliza cuando es mejor reducir tanto falsos positivos como falsos negativos [20].

$$\text{F1score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Ecuación 4.F1-Score

$$\text{F1score} = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{\text{tp}}{\text{tp} + \frac{1}{2}(\text{fp} + \text{fn})}$$

Ecuación 5. Modelo matemático de la evaluación de la medida-F

2.3 FORMULACIÓN DE HIPÓTESIS

2.3.1 Hipótesis de investigación

Las funciones kernel alternativas mejoran la relación error tiempo en la predicción de enfermedades cardiovasculares.

2.3.2 Hipótesis nula

Las funciones kernel alternativas no mejoran la relación error tiempo en la predicción de enfermedades cardiovasculares.

2.3.3 Hipótesis alterna

Las funciones kernel disminuyen el error, pero tardan más en dar una predicción.

3. METODOLOGÍA

3.1 PARADIGMA

Esta investigación se enmarca dentro del paradigma positivista, un paradigma que es racional, objetivo y se basa en comprobar hechos y particularidades propias del conocimiento científico, característico de la esencia disciplinar ingenieril, a su vez se caracteriza por el alto interés, por la verificación del conocimiento a través de predicciones. Algunos lo llaman el “paradigma prediccionalista”, ya que lo importante es plantearse una serie de hipótesis como predecir que algo va a suceder y luego verificarlo o comprobarlo. En las ciencias exactas y naturales es en donde tiene mayor aplicación [38]. El presente proyecto permite estudiar el desempeño de los algoritmos SVM y ANN a través del uso de las funciones kernel, realizando varias configuraciones de hiperparámetros a través de una sintonización sistemática, aplicando tales métodos con datos. Además, que la presencia de las variables en los datasets influye de manera significativa para una posterior predicción y posterior análisis.

3.2 ENFOQUE

El enfoque a usar en esta investigación es empírico-analítico, es un método de observación utilizado para profundizar en el estudio de los fenómenos, pudiendo establecer leyes generales a partir de la conexión que existe entre la causa y el efecto en un contexto determinado, además es uno de los modelos para describir el método científico, que se basa en la experimentación y la lógica empírica [39]. Aborda la realidad de los hechos que son observables, cuantificables y medibles. Es un método

que contrasta sus hipótesis de una forma rigurosa a través de la demostración científica que determina si dicha hipótesis es verdadera o falsa [38]. El actual proyecto permite observar y analizar las causas y efectos de la problemática estudiada, plantear procesos de experimentación y pruebas, que servirán de base para proponer una solución. De esta manera se obtendrá modelos apropiados en la predicción de enfermedades cardiovasculares.

3.3 MÉTODO

El método será cuasiexperimental, porque su objetivo es poner a prueba una hipótesis causal manipulando una variable independiente para posterior a ello analizar las consecuencias que la manipulación tiene sobre una o más variables dependientes.

Esto se verá reflejado en el apoyo al diagnóstico clínico en la predicción de enfermedades cardiovasculares, fomentando a su vez de hiperparámetros para la obtención de predicciones en los algoritmos SVM, ANN frente al conjunto de datos de factores de enfermedades cardiovasculares.

La metodología que se utilizara en el proyecto es **CRISP-DM**, que son las siglas de Cross-Industry Standard Process for Data Mining, es un método probado para orientar trabajos de minería de datos. Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas [40].



Fig. 6. Esquema del ciclo CRISP-DM estándar

Nota: fuente <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>

La metodología CRISP-DM establece un proyecto de minería de datos como una secuencia de fases [41]:

1. Comprensión del negocio: El objetivo de esta fase es alinear los objetivos del proyecto de data mining con los objetivos del negocio.
2. Compresión de los datos: Dos puntos clave en esta fase: conocer los datos, estructura y distribución, y la calidad de los mismos.
3. Preparación de los datos: El objetivo final de esta fase es obtener los datos finales sobre los que aplicarán los modelos: El objetivo último de esta fase es construir un modelo que nos permita alcanzar los objetivos del proyecto.
4. Modelado: El objetivo último de esta fase es construir un modelo que nos permita alcanzar los objetivos del proyecto.
5. Evaluación: En esta fase nos centramos en evaluar el grado de acercamiento del modelo a los objetivos del proyecto.
6. Despliegue: El objetivo último de esta fase es realizar el despliegue de los resultados obtenidos de forma que sea propagado a los usuarios finales, así como el mantenimiento del mismo una vez el despliegue haya finalizado.

De momento en la actualidad los proyectos que tienen que ver con desarrollos de modelos inteligentes se concentran en dos metodologías específicas una es KDD y otra es CRISP-DM donde esta última es una metodología más dinámica, que permite la interacción del ciclo del modelo y también del software, logrando un desarrollo interactivo, permitiendo así que el modelo se pueda escalar, tiene ese dinamismo de retroalimentación que tiene de algunas fases que componen la metodología.

3.4 TIPO DE INVESTIGACIÓN

El tipo de investigación que se desarrolla a lo largo del proyecto es determinado por el tipo de investigación aplicada. Se persigue la creación de conocimiento aplicándolo directamente a los problemas sociales. Se basa en los resultados tecnológicos de la investigación básica, que se ocupa de la relación entre concepto y producto.

3.5 DISEÑO DE INVESTIGACIÓN

El diseño de investigación es de ensayos debido a que se realizarán distintos tipos de pruebas, como el análisis en cuanto a rendimiento, la confianza por el entrenamiento de algoritmos otorgando una predicción, el proyecto se basa en dos herramientas que tenga en cuenta las etapas de proceso de máquinas de soporte vectorial y redes neuronales artificiales, de esta manera se plantea las siguientes fases.

- **Conexión a los datos:** principalmente es necesario tener un repositorio que nos provee de datos de diferentes autores, permitiendo así una estimación de los datos que se obtienen sean suficientes y relevantes para construcción del modelo.
- **Definición de un criterio de evaluación:** se trata de una medida de error, comúnmente el uso del error cuadrático medio que permita su uso en medidas como la precisión y exhaustividad.
- **Preparación de los datos:** datos completos, combinación de datos de varias fuentes, formato adecuado a los datos, tener presente las características relevantes, y posteriormente la construcción del modelo.
- **Análisis de errores:** usar modelos de SVM Y ANN diferentes, revisar los datos si se requieren en su totalidad o solo se toma algunos.
- **Fase de entrenamiento:** al seleccionar el tipo de neurona artificial a utilizar en una red neuronal y a su vez determinada la topología, inicia el proceso de entrenamiento de la red para que pueda ser utilizada. A partir de un conjunto de pesos sinápticos aleatorios, el proceso de aprendizaje busca un conjunto de pesos que permitan a la red realizar correctamente una función específica.
- **Fase de operación:** Una vez finalizada la fase de aprendizaje, el modelo puede que se ajuste demasiado a las particularidades presentes en los patrones de entrenamiento, perdiendo su habilidad de generalizar su aprendizaje a casos nuevos [42].
- **Análisis de resultados y descubrimiento de conocimiento:** Se analizarán los resultados de los métodos mediante visualizaciones, de tal forma que pueda determinar qué métodos tuvieron un mejor rendimiento en los diferentes conjuntos de datos, esto permite que los investigadores prueben sus métodos en SVM, ANN y permita así una predicción segura de las enfermedades cardiovasculares y que esta herramienta permita ser usada como ayudas diagnósticas.

3.6 POBLACIÓN

Esta investigación tiene como fin el estudio de la parte epistemológica y matemática del objeto de estudio, que en este caso son las funciones kernel aplicadas a SVM y ANN, por lo tanto, no se requiere población ni muestras, como se menciona en la investigación de distinción máxima Evaluación De Métodos De Reducción De Dimensión Para La Preservación Topológica De Los Datos Mediante Métricas Rnx desarrollada por (Correa Lozano.C.D., Lozano Thomé.J.A., Urrea Burgos.D.F.), ya que la naturaleza del proyecto exige un tipo de investigación de Data Driven Approach, en el cual las bases de datos son los principales en el desarrollo del proyecto. Los Datasets que se trabajaran serán probados en diferentes conjuntos de datos, en este proyecto no se utilizara datos focales, sino que se hará uso de diferentes bases de datos ya preestablecidos por otros investigadores.

3.8 TÉCNICAS DE RECOLECCIÓN DE LA INFORMACIÓN

Este proyecto no requiere la recolección de información, ya que existen bases de datos con características estándar que se usan por diferentes investigadores para probarlos con diferentes métodos de predicción de riesgo cardiovascular. Pero se tendrá en cuenta las bases de datos utilizadas relacionadas a continuación:

Herramientas de repositorio de datasets Kaggle

Un instituto nacional encargado de la evaluación de la educación, con el objetivo de facilitar el uso y consulta de las bases de datos de los resultados de las pruebas que aplica, tiene a disposición de la comunidad académica el FTP digital para investigadores. Esta herramienta es un repositorio para que grupos de investigación y estudiantes de niveles avanzados, interesados en participar en convocatorias de investigación, tengan la información de primera mano para desarrollar sus propuestas. A partir del cruce de bases disponibles, se logra recopilar información que permite clasificar al estudiante en diferentes niveles socioeconómicos. Es de interés en este proyecto predecir el nivel socioeconómico del estudiante a partir de la información de la base de datos.

Para lo anterior, use todas las técnicas de predicción vistas (algunas por ver) en clase:

- Knn - Vecinos más cercanos
- Árboles de clasificación
- Redes Neuronales
- Random Forest
- Boosting, Gradient Boosting
- LDA, QDA, FDA [43].

Herramientas de repositorio de datasets UCI

El repositorio de aprendizaje automático de UCI es una colección de bases de datos, teorías de dominio y generadores de datos que utiliza la comunidad de aprendizaje automático para el análisis empírico de los algoritmos de aprendizaje automático. El archivo fue creado como un archivo ftp en 1987 por David Aha y sus compañeros estudiantes de posgrado en UC Irvine. Desde entonces, ha sido ampliamente utilizado por estudiantes, educadores e investigadores de todo el mundo como fuente principal de conjuntos de datos de aprendizaje automático. Como indicación del impacto del archivo, ha sido citado más de 1000 veces, lo que lo convierte en uno de los 100 "artículos" más citados en toda la informática. La versión actual del sitio web fue diseñada en 2007 por Arthur Asunción y David Newman, y este proyecto es en colaboración con Rexa.info en la Universidad de Massachusetts Amherst. Se agradece el apoyo financiero de la National Science Foundation [44].

3.8.1 Validez de la técnica

Las variables que van a tener el conjunto de datos de los repositorios van a ser preprocesados con algoritmos de limpieza, transformadores de escalamiento y algoritmos de reducción de dimensión. Para almacenar los resultados de cada experimentación que se va a realizar, en especial para las funciones kernel, se almacenara en un dataframe cuyos datos van a ser aterrizados en un archivo CSV.

3.8.2 Confiabilidad de la técnica

Este proyecto como se mencionó en secciones anteriores adopta técnicas de aprendizaje automático para generar datos que puedan reducir significativamente el potencial margen de error. Esto se logrará ajustando la configuración del algoritmo de manera apropiada. Además, se utilizarán los resultados obtenidos de estas pruebas para validar o refutar la hipótesis planteada.

4. RESULTADOS

A continuación, se presenta en este capítulo los resultados de la investigación que corresponden al desarrollo de cada uno de los objetivos planteados.

4.1 Definición de funciones Kernel recomendadas por expertos

4.1.1 Resultados para la definición de funciones kernel con recomendación de expertos

Para ejecutar el resultado del primer objetivo, que es definir funciones Kernel con base en la recomendación de expertos, se procedió a la búsqueda de dos documentos importantes teniendo una temática diferente entre sí, y a la vez una estrecha relación con el uso de funciones Kernel su comportamiento con variedad de datos y cuáles al final son las más adecuadas para acoplarse al modelo creado. Se comienza así con una revisión de la literatura científica relacionada con el uso de funciones Kernel, citando a estos expertos, se demuestra que existe una base teórica para explorar funciones Kernel alternativas.

Uno de los artículos de los cuales se toma en cuenta para el desarrollo de esta investigación es el artículo de Mora.H., en su proyecto titulado “Comparación de funciones Kernel para la predicción de la oferta energética fotovoltaica” publicado en el año 2020, España [45], y el otro experto Belanche.L., en su proyecto titulado “Kernel Functions for Categorical Variables with Application to Problems in the Life Sciences” publicado en el año 2013, Barcelona, Spain [46], se puede evidenciar que hacen uso de funciones Kernel alternativas, generalmente en las implementaciones de las bibliotecas que hay en el aprendizaje automático y que a su vez se implementa 4 algoritmos de funciones Kernel Lineal, Polinómica, RBF, Sigmoide, tomando este último Belanche.L que con su proyecto presenta que el enfoque propuesto puede superar notablemente a los Kernel Estándar, por lo que puede ser utilizado como una buena alternativa a otras funciones comunes del Kernel (al menos para la clasificación SVM) para obtener una mayor exactitud, con funciones alternativas se puede adquirir mejores resultados que la RBF, ya que se destaca que el Kernel Categórico aprovecha información que no es utilizada por los otros Kernels, como K_0 o el Kernel RBF. Esto sugiere que el Kernel Categórico es capaz de capturar relaciones y patrones en los datos que los otros Kernels no pueden capturar.

Es importante destacar también que en el documento se menciona que en casos en los que los resultados de los Kernels Estándar son buenos, el Kernel Categórico K_1 , mejora ligeramente en comparación con K_0 , y ambos superan al Kernel RBF. Esto indica que el enfoque propuesto es beneficioso incluso cuando los resultados ya son aceptables. Se menciona que el Kernel RBF muestra un bajo rendimiento en algunos casos, posiblemente debido a su susceptibilidad al sobre ajuste en datos de muestra pequeña en representaciones de alta dimensión.

Y el primer artículo mencionado de Mora, H., utiliza un par de funciones Kernel; la Gaussiana y Polinómica, de esta investigación se destaca que los Kernel alternativos solo han mejorado los resultados para SVM y han empeorado para ANN. Se tiene la hipótesis de que los Kernels funcionan mejor para SVM que en otros algoritmos de clasificación y predicción lineal, porque este algoritmo utiliza el criterio del margen máximo, abriendo la discusión para realizar nuevos experimentos teniendo en cuenta más algoritmos de aprendizaje automático que usan funciones Kernel, realizando dicha evaluación con distintas bases de datos para clasificación y predicción. Los resultados del análisis reflejan que SVM con los normalizadores Standard y Min-Max y los Kernel Triangular y Rational Quadratic son los más indicados para realizar regresión sobre los datos Landsat y MODIS. A su vez se tiene en cuenta las funciones Kernel que se proponen tanto en el de Belanche como en el de Mora son las que están definidas en la siguiente tabla.

TABLA VI
FORMULACIÓN MATEMÁTICA DE LAS FUNCIONES KERNEL UTILIZADAS POR EXPERTOS

Fuente: Comparativo de funciones kernel sobre predicción de oferta de fuentes alternativas de energía

Kernel	
Gaussiano (RBF)	Triangular (tr)
$e^{-\sum_{i=1}^d \gamma (x_i - x_i')^\beta}$ <p>$\gamma > 0, \beta \in (0, 2]$</p>	$\begin{cases} \ x - x'\ \leq a \rightarrow 1 - \frac{\ x - x'\ }{a} \\ \ x - x'\ > a \rightarrow 0 \end{cases}$ <p>$a > 0$</p>
Radial Basic (Rb)	Rational Quadratic (Rq)
$\left(\sum_{i=1}^d e^{-\gamma (x_i - x_i')^2} \right)^m$ <p>$\gamma > 0, m \in \mathbf{N}$</p>	$1 - \frac{\ x - x'\ ^2}{\ x - x'\ ^2 + a}$ <p>$a > 0$</p>

Canberra (Can)	Truncated (Tru)
$1 - \frac{1}{d} \sum_i \gamma \frac{ x_i - x_i' }{ x_i + x_i' }$ $\gamma \in (0, 1]$	$\frac{1}{d} \sum_i \max\left(0, \frac{ x_i - x_i' }{\gamma}\right)$ $\gamma > 0$

Y de la cual adicionalmente basada en la definición de belanche:

$$k(x, x') = \langle \phi(x), \phi(x') \rangle$$

De la cual se propone como exploración los siguientes kernels:

- **Un kernel Logarítmico Radial Basic:** Es una función de similitud que se utiliza para medir la distancia entre puntos de datos en un espacio de características. El kernel RBF puede mapear los datos a un espacio de características de mayor dimensionalidad donde los datos pueden ser más fácilmente separables mediante un hiperplano. Esto permite a los algoritmos como las SVM realizar clasificación y regresión en conjuntos de datos que no son linealmente separables en su espacio original. Dado que el kernel RBF puede capturar patrones no lineales, es útil en situaciones donde los datos tienen relaciones complejas y no lineales [47].

$$k_{ln rbf} = \ln \sum \gamma \langle x, x' \rangle^b$$

$$k_{ln rb} = \sum \ln \gamma \langle x, x' \rangle^b$$

Todo este análisis anteriormente dicho nos llevó a definir los Kernel inspirados en la concepción de funciones de activación alternativas en redes neuronales y son:

TABLA VII
FORMULACIÓN MATEMÁTICA DE LAS FUNCIONES KERNEL UTILIZADAS EN EL PROYECTO

Kernel	
Polynomial (pk)	Triangular (tk)
$K(x, x') = (x, x' + c)^d$	$\begin{cases} x - x' \leq a \rightarrow 1 - \frac{ x - x'}{a} \\ x - x' > a \rightarrow 0 \end{cases}$ $a > 0$
Hyperbolic (hk)	Rational Quadratic (rqk)

$K(x, y) = \tanh(\alpha x, y + c)$	$1 - \frac{\ x - x'\ ^2}{\ x - x'\ ^2 + a}$ $a > 0$
Radial Basic (rbk)	Truncated (trk)
$\left(\sum_{i=1}^d e^{-\gamma(x_i - x_i')^2} \right)^m$ $\gamma > 0, m \in \mathbb{N}$	$\frac{1}{d} \sum_i \max\left(0, \frac{ x_i - x_i' }{\gamma}\right)$ $\gamma > 0$
Canberra (ck)	Laplacian (lpl)
$1 - \frac{1}{d} \sum_i \gamma \frac{ x_i - x_i' }{ x_i + x_i' }$ $\gamma \in (0, 1]$	$k(x, y) = e\left(\frac{-\ x - y\ }{\gamma}\right)$
Sigmoid (sigm)	Cosine (cos)
$k(x, y) = \tanh(\gamma x, y)$	$k(x, y) = \frac{\ x-y\ }{\ x\ \cdot \ y\ }$

Se representa, a su vez, en la TABLA VII la función matemática del Kernel y en la siguiente tabla, la TABLA VIII, su respectiva codificación en el lenguaje Python.

4.1.2 Codificación de funciones Kernel

TABLA VIII
CODIFICACIÓN DE FUNCIONES KERNEL EN PHYTON

Kernel	Kernel
Polynomial (pk)	Triangular (tk)
<pre>#POLYNOMIAL KERNEL m E N, a>0 def polynomial(x,y, degree=3, gamma=0.01,coef0=0.0): m=degree a=gamma return (a*np.dot(x,y)+1)**m</pre>	<pre>#Triangle a>0 def triangle(x,y, degree=3, gamma=0.01,coef0=0.0): a=gamma norm=la.norm(np.subtract(x, y)) if norm<=a: return 1-norm/a return 0</pre>
Hyperbolic (hk)	Rational Quadratic (rqk)
<pre>#Hyperbolic tangent kernel a0>0, b<0 def hyperbolic(x,y, degree=3, gamma=0.01,coef0=0.0): b=coef0 a=gamma return np.tanh(a*np.dot(x,y)+b)</pre>	<pre>#Rational quadratic a>0 def rquadratic(x,y, degree=3, gamma=0.01,coef0=0.1): a=coef0 norm=la.norm(np.subtract(x, y)) return 1-(norm**2)/(norm**2+a)</pre>
Radial Basic (rbk)	Truncated (trk)
<pre>def radial_basic(x, y, degree=3, gamma=0.01,coef0=0.0): m=degree sm=0 sm=np.sum(np.exp(-gamma*((x-y)**2))) return sm**m</pre>	<pre>#Truncated gamma>0 def truncated(x,y, degree=3, gamma=0.01,coef0=0.0): sm=0 d=x.shape[0] val=1-np.abs(x-y)/gamma sm=np.sum(val[val>0]) return sm/d</pre>
Canberra (ck)	Laplacian (lpl)
<pre>#CANBERRA def c_canberra_kernel(gamma): def ck(X,Y): return canberra(X,Y,gamma=gamma) return ck</pre>	<pre>#LAPLACIAN def c_laplacian_kernel(gamma): def lpl(X,Y): return laplacian(X,Y,gamma=gamma) return lpl</pre>
Sigmod (sigm)	Cosine (cos)

```

# sigmoid
def sigmoid(x,y, degree=3, gamma=0.01,coef0=0.0):
    #print('x',x,'y',y)
    sm=np.tanh(gamma*x@y+coef0)

```

```

def cosine(x,y, degree=3, gamma=0.01,coef0=0.0):
    #print('x',x,'y',y)
    sm=(x@y)/(np.linalg.norm(x)*np.linalg.norm(y))
    if np.isnan(sm) or np.isinf(sm):
        sm=0
    return sm

```

Por lo tanto, las funciones de activación de las redes neuronales están diseñadas para activar y representar características particulares en los datos. En el proyecto de predicción del riesgo cardiovascular, se diseñó funciones kernel alternativas que capturen patrones específicos en los datos clínicos o de salud relacionados con factores de riesgo cardiovascular, como presión arterial, peso, talla, índice de masa corporal entre otros, una vez diseñados, estos Kernels alternativos requerirían optimización y ajuste en función de conjuntos de datos específicos de esta manera se podría evaluar si estos Kernels mejoran la capacidad de predicción y generalización en comparación con los Kernels tradicionales en modelos de aprendizaje automático.

Para llevar a cabo la codificación de esas funciones kernel se tuvo en cuenta la Fig. 7 que hace relación al algoritmo de comparación utilizando conjunto de datos.

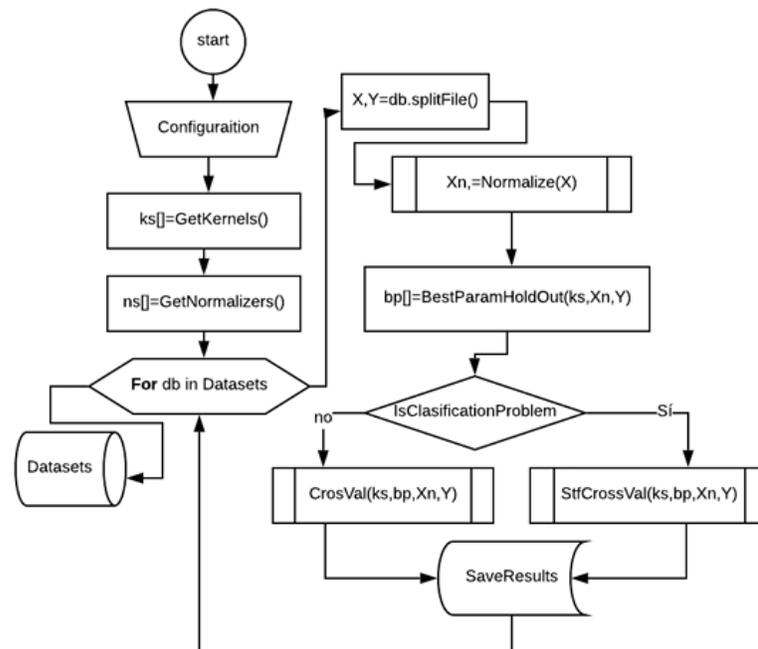


Fig. 7. Algoritmo implementado para la comparación de kernel

Nota: fuente <https://docplayer.es/192405564-Comparativo-de-kernels-sobre-prediccion-de-oferta-de-fuentes-alternativas-de-energia.html>

Se realiza una configuración manual primero se define: la carpeta de entrada, aquí están los datos a inspeccionar, la carpeta de salida es donde se colocaron los resultados de la comparación y el tipo de problema a procesar para este caso es de clasificación. Luego, se obtiene un vector de funciones del kernel y normalizadores, se procede a la lectura de los datos ubicados en el directorio de entrada, seguido de la ejecución de las siguientes actividades:

- Se realiza una partición de los datos respectivamente.
- Se realiza la normalización de los datos y se identifican los parámetros donde el Kernel respectivo obtuvo un puntaje superior o un menor error.
- Se realiza una validación con estos parámetros, como el problema es de clasificación, se realiza una validación cruzada dividida; cuando finalizan las pruebas, los resultados se registran en la carpeta de resultados.

Este método inicial proporciona una estimación aproximada de qué función del Kernel tiene el mejor rendimiento, pero para obtener una aproximación más precisa, se requiere una sintonización de hiperparámetros utilizando técnicas avanzadas, por lo que se establece el siguiente algoritmo.

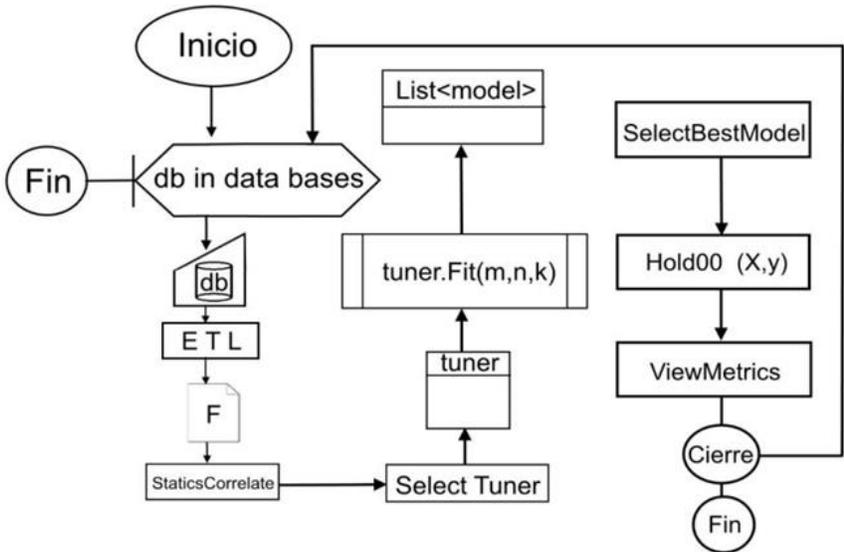


Fig. 8. Algoritmo exhaustivo para búsqueda de modelos

Nota: fuente <https://docplayer.es/192405564-Comparativo-de-kernels-sobre-prediccion-de-oferta-de-fuentes-alternativas-de-energia.html>

En la Fig. 8 se muestra que primero se ejecuta un ciclo para recorrer la base de datos, extrayendo, transformando y cargando datos, los cuales finalmente se sintetizan en un archivo en la que se

encuentras las estadísticas y aspectos de correlación, después de elegir un sintonizador predeterminado, configurar un motor de búsqueda evolutivo (algoritmo genético), este sintonizador selecciona la mejor configuración de parámetros, los mejores son seleccionados y reentrenados.

4.2 Funciones Kernel acopladas a SVM y ANN

4.2.1 Acoplamiento con biblioteca Scikit-learn

Para llevar a cabo y cumplir con este segundo objetivo de acoplar nuevas funciones Kernel alternativas en los algoritmos SVM, ANN a la biblioteca Scikit-learn. Se tomó como base lo siguiente:

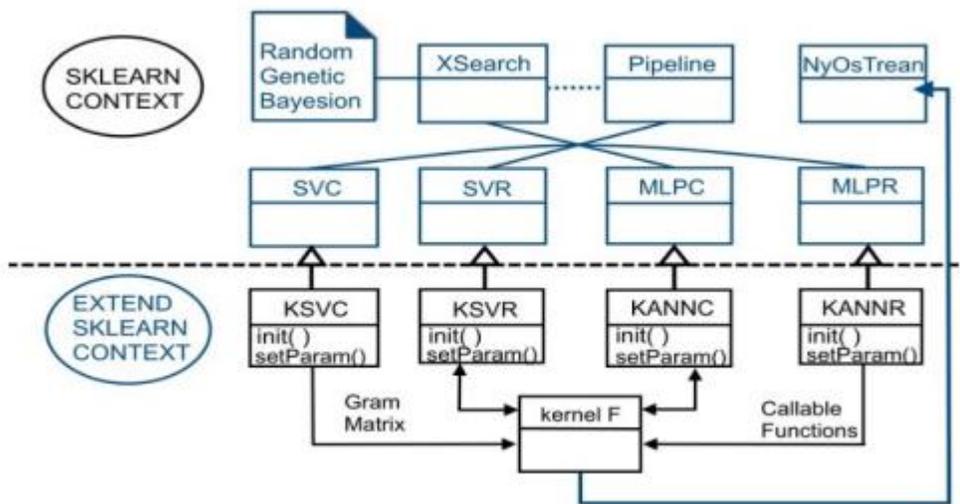


Fig. 9. Modelo de acoplamiento funciones Kernel en Scikit-learn

Nota: fuente <https://docplayer.es/192405564-Comparativo-de-kernels-sobre-prediccion-de-oferta-de-fuentes-alternativas-de-energia.html>

El truco Kernel se introduce en SVM para clasificación en la clase SVC y se nutre de más funciones kernel con la clase KSVC mediante la matriz de Gram como se muestra en las ecuaciones (1) y (2):

$$\max: L(a) = \sum_1^n a_i - \frac{1}{2} \sum_{i=1}^n a_i a_j y_i y_j \mathbf{k}(x_i x_j) \quad (1)$$

$$\text{s. a: } \sum_1^n a_i y_i = 0, 0 \leq a_i \leq C, i = 1, 2, \dots, n \quad (2)$$

Un proceso similar se sigue para el algoritmo de regresión provista con la clase KSVR cuya expresión formal se muestra en las ecuaciones (3) y (4):

$$\begin{aligned} \max: L(a) &= \sum_1^n (a_i^- - a_i^+) y_i - \epsilon \sum_1^n (a_i^- + a_i^+) - \frac{1}{2} \sum_{ij=1}^n (a_i^- - a_i^+) (a_j^- - a_j^+) k(x_i, x_j) \quad (3) \\ \text{s. a: } \sum_1^n (a_i^- - a_i^+) &= 0, 0 \leq (a_i^-, a_i^+) \leq C, i = 1, 2, \dots, n \quad (4) \end{aligned}$$

Para ANN se utilizó la aproximación de Nystroem (SI et al., 2017) mediante una transformación de base sobre el vector de características. Formalmente:

$$Z_L = W_L \cdot a_{L-1} + b_L \quad (5)$$

$$N_t = \text{NyosTream}(kf) \quad (6)$$

$$X = N_t * X \quad (7)$$

$$a_0 = X \quad (8)$$

$$a_L = f(Z_L) \quad (9)$$

$$\hat{y} = \text{predict}(X_{batch} * Nt^T) \quad (10)$$

Donde la ecuación (5) representa la suma ponderada, mientras que (6) la transformación de Nystroem preparada con una función Kernel (Buitinck et al., 2013). La ecuación (7) es la generación de la transformación, conjuntamente (8) es la activación inicial o capa de entrada y (9) la función de activación que se ha definido lineal por defecto, buscando que las funciones kernel sean las que cumplan el oficio de la transformación del espacio de características, en este caso un espacio de Hilbert de dimensión infinita. Cabe resaltar que es necesario sobrescribir el método predict del clasificador o regresor en KANNC y KANNR para realizar la transformación sobre cada batch del descenso del gradiente estocástico SGD (10).

En la siguiente Fig. 10 se muestra cómo se acopla la clase KSVC a la biblioteca Scikit-learn la cual sigue un procedimiento análogo para KSVR, KANNC, KANNR.

```

class KSVC(SVC):

    def __init__(self, C=1.0, kernel='rbf', degree=2, gamma='auto',
                 coef0=0.0, shrinking=True, probability=False,
                 tol=1e-3, cache_size=200, class_weight=None,
                 verbose=False, max_iter=-1, decision_function_shape='ovr',
                 random_state=None, a=2):

        super().__init__(
            kernel=kernel, degree=degree, gamma=gamma,
            coef0=coef0, tol=tol, C=C, shrinking=shrinking,
            probability=probability, cache_size=cache_size,
            class_weight=class_weight, verbose=verbose, max_iter=max_iter,
            decision_function_shape=decision_function_shape,
            random_state=random_state)

        self.a=a

    def fit(self, X, y, sample_weight=None):

        if(self.kernel=="linear" or self.kernel== "poly" or self.kernel== "rbf"):
            super().fit(X,y)
            return self
        else:
            if(self.kernel=="mrbf"):
                self.kernel=mrbf_kernel(degree=self.degree, gamma=self.gamma)
                super().fit(X,y)
                return self
            if(self.kernel=="hyperbolic"):
                self.kernel=hyperbolic_kernel(self.gamma, self.coef0)
                super().fit(X,y)
                return self
            if(self.kernel=="triangle"):
                self.kernel=triangle_kernel(self.gamma)
                super().fit(X,y)
                return self
            if(self.kernel=="radial_basic"):
                self.kernel=radial_basic_kernel(self.degree, self.gamma)
                super().fit(X,y)
                return self

```

Fig. 10. Código fuente de acoplamiento a biblioteca Scikit-learn

```

if(self.kernel=="rquadratic"):
    self.kernel=rquadratic_kernel(self.degree, self.gamma, self.coef0)
    super().fit(X,y)
    return self
if(self.kernel=="can"):
    self.kernel=canberra_kernel(self.gamma)
    super().fit(X,y)
    return self
if(self.kernel=="tru"):
    self.kernel=truncated_kernel(self.gamma)
    super().fit(X,y)
    return self
if(self.kernel=="chisq"):
    self.kernel=additive_chi2_kernel
    super().fit(X,y)
    return self
if(self.kernel=="chi2"):
    self.kernel=chi2_kernel
    super().fit(X,y)
    return self
if(self.kernel=="laplacian"):
    self.kernel=laplacian_kernel
    super().fit(X,y)
    return self
if(self.kernel=="sigmoid"):
    self.kernel=sigmoid_kernel
    super().fit(X,y)
    return self
if(self.kernel=="cosine"):
    self.kernel=cosine_similarity
    super().fit(X,y)
    return self

def set_params(self, **parameters):
    for parameter, value in parameters.items():
        setattr(self, parameter, value)
    return self

```

Fig. 11. Código fuente de acoplamiento a biblioteca Scikit-learn

4.2.2 Obtención de funciones Kernel (procedimiento GetKernels)

Para obtener este procedimiento se construyó la matriz de Gram correspondiente a la función Kernel K (ver Fig. 12) tomando como Kernel la formulación de la TABLA VII enviando los parámetros correspondientes a cada Kernel.

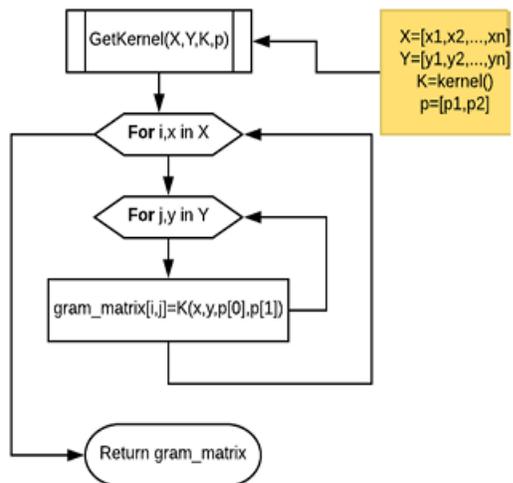


Fig. 12. Algoritmo para la obtención de funciones kernel

Nota: fuente <https://docplayer.es/192405564-Comparativo-de-kernels-sobre-prediccion-de-oferta-de-fuentes-alternativas-de-energia.html>

Para los recursos realizados en la Fig. 13 bastaría con enviar como parámetro al algoritmo de aprendizaje (SVM, ANN).

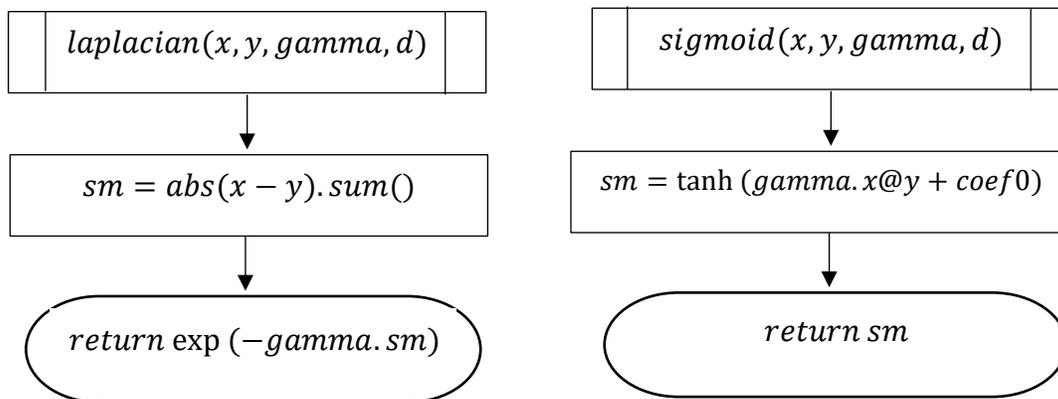


Fig. 13. Ejemplo de implementación de un kernel L laplacian(izq.), S sigmoid (der)

Sin embargo, en la mayoría de los casos, estos algoritmos están diseñados para recibir una función en su inicializador (constructor) y enviar los datos de entrenamiento X, Y (funciones de retorno).

4.3 Validando la configuración apropiada de los Kernel

Con el fin de alcanzar el tercer objetivo, que consiste en validar la configuración más apropiada para la prevención de enfermedades cardiovasculares mediante la comparación de la relación entre el error y el tiempo se registraron los resultados de cada experimento en un dataframe, para su posterior análisis.

4.3.1 Descripción de los datos

La base de datos que se usó para el desarrollo del proyecto fue un dataset que consta de 70.000 registros de datos de pacientes, contiene 11 características más objetivo, y a su vez consta de 3 tipos de entradas:

- Objetivo: información fáctica
- Examen: resultados de examen médico
- Subjetivo: información dada por el paciente

En la TABLA IX se puede observar las características de los datos

TABLA IX
DESCRIPCIÓN DE LOS DATOS

Edad	Característica de objetivo	años
Altura	Característica de objetivo	Cm
Peso	Característica de objetivo	Kg
Género	Característica de objetivo	Código categórico 1: mujer 2: hombre
Presión arterial sistólica	Función del examen	Ap_hi
Presión arterial diastólica	Función del examen	Ap_lo
Colesterol	Función del examen	1: normal 2: por encima de lo normal 3: muy por encima de lo normal
Glucosa	Función del examen	1: normal

		2: por encima de lo normal 3: muy por encima de lo normal
Fumar	Característica subjetiva	Si el paciente fuma 1: Si 2: No
Consumo de alcohol	Característica subjetiva	Alco Binario 1: Si 2: No
Actividad física	Característica subjetiva	Activo Binario 1: Si 2: No
Presencia o ausencia de enfermedad cardiovascular	Variable objetivo	Cardio Binario Variable Objetivo 1: Presencia de ECV 2: Ausencia de ECV

4.3.2 Obtención del mejor parámetro de entrenamiento para SVM y ANN

En esta fase, se emplea un método de búsqueda exponencial para explorar los diferentes valores de los parámetros asociados a un Kernel específico dentro de un rango definido. Posteriormente, se lleva a cabo una división de los datos en conjuntos de entrenamiento y prueba, reservando una muestra del 14.29%. Una vez completada esta etapa, se calcula la puntuación del modelo utilizando los datos de prueba, identificando así los mejores parámetros y su correspondiente puntuación óptima.

4.3.3 Resultados inspección de funciones Kernel en clasificación para SVM

En la siguiente TABLA X se presentan los resultados de la evaluación de las funciones en problemas de clasificación para determinar la adecuación de las funciones Kernel. Este análisis busca validar la afirmación de Mora.H., y Belanche.L., la cual sugiere que las funciones consideradas en este estudio pueden conducir a resultados mejorados.

En la TABLA X se detallan los resultados de la validación mediante la división en conjuntos de entrenamiento y prueba realizada con el algoritmo SVM para clasificación, donde se evaluaron

diversas funciones Kernel. Este proceso se llevó a cabo utilizando el conjunto de datos previamente mencionado y aplicando técnicas de normalización y escalación.

TABLA X
RESULTADOS GENERALES PARA SVM

SCALER	KERNEL	MEAN_TEST_SCORE	BEST_PARAM
SScaler	rquadratic	0.710	γ : auto coef0: 545.5594 C: 46.415
SScaler	rbf	0.698	γ : 0.0006 C: 2154.4346
SScaler	tru	0.718	γ : 0.0001 C: 1.0
SScaler	can	0.717	γ : 0.0006 C: 2154.4346
SScaler	radial_basic	0.708	γ : 0.0006 C: 2154.4346
SScaler	triangle	0.712	γ : 545.5594 C: 46.4158
SScaler	hyperbolic	0.703	γ : 0.0006 C: 2154.4346
SScaler	laplacian	0.718	γ : 0.0001 C: 1.0
SScaler	sigmoid	0.601	γ : 0.0001 C: 1.0
SScaler	cosine	0.708	C: 46.4158
MMScaler	rquadratic	0.706	γ : auto coef0: 11.2883 C: 2154.4346
MMScaler	rbf	0.696	γ : 0.0006 C: 2154.4346
MMScaler	tru	0.718	γ : 0.0001 C: 1.0
MMScaler	can	0.717	γ : 0.0006 C: 2154.4346
MMScaler	radial_basic	0.703	γ : 0.0006 C: 2154.4346

MMScaler	triangle	0.704	γ : 4.2813 C: 12.9154
MMScaler	hyperbolic	0.692	γ : 0.0006 C: 2154.4346
MMScaler	laplacian	0.716	γ : 0.0001 C: 1.0
MMScaler	sigmoid	0.658	γ : 4.2813 C: 12.9154
MMScaler	cosine	0.715	C: 2154.4346
NMScaler	rquadratic	0.716	γ : auto coef0: 0.2335 C: 7742.6368
NMScaler	rbf	0.717	γ : 11.2883 C: 2154.4346
NMScaler	tru	0.702	γ : 0.0048 C: 3.5938
NMScaler	can	0.701	γ : 4.2813 C: 12.9154
NMScaler	radial_basic	0.722	γ : 11.2883 C: 2154.4346
NMScaler	triangle	0.71	γ : 11.2883 C: 2154.4346
NMScaler	hyperbolic	0.54	γ : 0.2335 C: 7742.6368
NMScaler	laplacian	0.703	γ : 0.0006 C: 2154.4346
NMScaler	sigmoid	0.536	γ : 0.2335 C: 7742.6368
NMScaler	cosine	0.718	C: 100000.0

En la TABLA X se han resaltado los Kernels que mejor score obtuvieron y a su vez el comparador de escalación y el Kernel correspondiente.

Este comparador de escalación toma el mejor score de la TABLA X y lo reevalúa por validación cruzada, para este caso la función Kernel Truncated. En general para este conjunto se produjeron muy buenos resultados a excepción del Kernel Hyperbolic, por lo que se visualizara las salidas del

normalizador Normalizer Scaler, Standard Scaler ya que producen mejores resultados en cada Kernel.

Por consiguiente, se puede notar que el valor del parámetro alcanza un score máximo, con este parámetro se realiza una validación cruzada con diferentes métricas. En la TABLA XI se puede observar los parámetros que se usan para la próxima etapa del algoritmo.

TABLA XI
MEJORES PARÁMETROS CONFIGURADOS PARA LAS FUNCIONES KERNEL EN SVM

KERNEL	SCALER	BEST_PARAM	VALOR	SCORE
rbf	SScaler	C γ	2154.4346 0.0006	0.698
Can	SScaler	C γ	2154.4346 0.0006	0.717
Tru	SScaler	C γ	1.0 0.0001	0.718
hyperbolic	SScaler	C γ	2154.4346 0.0006	0.703
triangle	SScaler	C γ	46.4158 545.5594	0.712
Radial basic	SScaler	C γ	2154.4346 0.0006	0.708
Rquadratic	SScaler	C γ Coef0	46.4158 auto 545.5594	0.710
Laplacian	SScaler	C γ	1.0 0.0001	0.718
Sigmoid	SScaler	C γ	11.0 0.0001	0.601
Cosine	SScaler	C	46.4158	0.708
rquadratic	NMScaler	C γ Coef0	7742.6368 auto 0.2335	0.716
rbf	NMScaler	C γ	2154.4346 11.2883	0.717
tru	NMScaler	C	3.5938	0.702

		γ	0.0048	
can	NMScaler	c	12.9154	0.701
		γ	4.2813	
radial_basic	NMScaler	c	2154.4346	0.722
		γ	11.2883	
triangle	NMScaler	c	2154.4346	0.71
		γ	11.2883	
hyperbolic	NMScaler	c	7742.6368	0.54
		γ	0.2335	
laplacian	NMScaler	c	2154.4346	0.703
		γ	0.0006	
sigmoid	NMScaler	c	7742.6368	0.536
		γ	0.2335	
Cosine	SScaler	c	100000.0	0.708

4.3.4 Clasificación de los resultados

Los resultados finales de los Kernel previa configuración de los mejores parámetros descritos anteriormente se sintetiza en la TABLA XII, donde se muestra el Kernel y la métrica de exactitud (Accuracy). Esta métrica indica la proporción de predicciones correctas realizadas por el modelo en comparación con el total de predicciones, de esta manera se valida lo expuesto por los expertos Mora.H., y Belanche.L., en la que se afirma que las funciones a considerar en el presente estudio pueden obtener mejores resultados para las Máquinas de Soporte Vectorial, culminando así la primera fase experimental.

TABLA XII
RESULTADOS FINALES PARA SVM

Kernel	Exactitud (Accuracy)
rbf	0.58
can	0.725
tru	0.71
hyperbolic	0.445
triangle	0.61

radial basic	0.685
Rquadratic	0.605
Laplacian	0.575
Sigmoid	0.445
Cosine	0.705

Se observa en la TABLA XII que la mayor exactitud la obtuvo el Kernel Canberra para este conjunto de datos. Con una exactitud promedio del 72 %.

4.3.5 Resultados inspección de funciones Kernel en clasificación para ANN

En la siguiente TABLA XIII, se presentan los resultados de la evaluación de las funciones en problemas de clasificación para ANN para determinar la adecuación de las funciones Kernel. Este análisis busca validar la afirmación por los expertos Mora.H., y Belanche.L., donde se afirma que las funciones a considerar empeoran los resultados para ANN.

En la TABLA XIV se detallan los resultados de la validación mediante la división en conjuntos de entrenamiento y prueba realizada con el algoritmo ANN para clasificación, donde se evaluaron diversas funciones Kernel. Este proceso se llevó a cabo utilizando el conjunto de datos previamente mencionado y aplicando técnicas de normalización y escalación.

TABLA XIII
RESULTADOS GENERALES PARA ANN

SCALER	KERNEL	MEAN_TEST_SCORE	BEST_PARAM
SScaler	rquadratic	0.684	γ : auto coef0: 4.2813 C: 12.9154
SScaler	rbf	0.703	γ : 0.2335 C: 7742.6368
SScaler	tru	0.702	γ : 0.2335 C: 7742.6368
SScaler	can	0.712	γ : 0.2335 C: 7742.6368

<i>S</i> Scaler	<i>radial_basic</i>	0.716	γ : 4.2813 C: 12.9154
SScaler	triangle	0.704	γ : 4.2813 C: 12.9154
SScaler	hyperbolic	0.688	γ : 4.2813 C: 12.9154
SScaler	laplacian	0.703	γ : 0.2335 C: 7742.6368
SScaler	sigmoid	0.703	γ : 545.5594 C: 46.4158
SScaler	cosine	0.704	C: 1.0
MMScaler	rquadratic	0.661	γ : auto coef0: 0.0018 C: 3.5938
MMScaler	rbf	0.658	γ : 11.2883 C: 2154.4346
MMScaler	tru	0.712	γ : 0.2335 C: 7742.6368
MMScaler	can	0.73	γ : 206.9138 C: 3.5938
MMScaler	radial_basic	0.711	γ : 4.2813 C: 12.9154
MMScaler	triangle	0.636	γ : 4.2813 C: 12.9154
MMScaler	hyperbolic	0.622	γ : 4.2813 C: 12.9154
MMScaler	laplacian	0.688	γ : 0.2335 C: 7742.6368
MMScaler	sigmoid	0.618	γ : 0.0048 C: 3.5938
MMScaler	cosine	0.618	C: 100000.0
NMScaler	rquadratic	0.528	γ : auto coef0: 545.5594 C: 46.4158
NMScaler	rbf	0.7	γ : 3792.6901 C: 27825.5940

NMScaler	tru	0.681	γ : 0.0048 C: 3.5938
NMScaler	can	0.687	γ : 4.2813 C: 12.9154
NMScaler	radial_basic	0.643	γ : 3792.6901 C: 27825.5940
NMScaler	triangle	0.711	γ : 0.0048 C: 3.5938
NMScaler	hyperbolic	0.522	γ : 0.0006 C: 2154.4346
NMScaler	laplacian	0.703	γ : 545.5594 C: 46.4158
NMScaler	sigmoid	0.522	γ : 0.0006 C: 2154.4346
NMScaler	cosine	0.528	C: 27825.5940

En la TABLA XIII se han resaltado los Kernel que mejor score obtuvieron y a su vez el comparador de escalación y el Kernel correspondiente.

Este comparador toma el mejor score de la TABLA XIII y lo reevalúa por validación cruzada para este caso la función Kernel Radial Basic. En general, para este conjunto no se produjeron muy buenos resultados a excepción del Kernel mencionado, por lo que se visualizara solo las salidas del Standard Scaler ya que produce mejores resultados en cada Kernel.

Por consiguiente, se puede notar qué el valor del parámetro alcanza un score máximo, con este parámetro se realiza una validación cruzada con diferentes métricas. En la TABLA XI se puede observar los parámetros que se usan para la entrada para la próxima etapa del algoritmo.

TABLA XIV
MEJORES PARÁMETROS CONFIGURADOS PARA LAS FUNCIONES KERNEL EN ANN

KERNEL	BEST_PARAM	VALOR	SCORE
rquadratic	c	12.9154	0.684
	γ	auto	
	Coef0	4.2813	
rbf	c	7742.6368	0.703
	γ	0.2335	

tru	c	7742.6368	0.702
	γ	0.2335	
can	c	7742.6368	0.712
	γ	0.2335	
<i>radial_basic</i>	c	12.9154	0.716
	γ	4.2813	
triangle	c	12.9154	0.704
	γ	4.2813	
hyperbolic	c	12.9154	0.688
	γ	4.2813	
laplacian	c	7742.6368	0.703
	γ	0.2335	
sigmoid	c	46.4158	0.703
	γ	545.5594	
cosine	c	1.0	0.704

4.3.6 Clasificación de los resultados

Los resultados finales de los Kernel previa configuración de los mejores parámetros descritos anteriormente se sintetiza en la TABLA XII, donde se muestra el Kernel y la métrica de exactitud (Accuracy). Esta métrica indica la proporción de predicciones correctas realizadas por el modelo en comparación con el total de predicciones, de esta manera se valida lo expuesto por los expertos Mora.H., y Belanche.L., en la que se afirma que las funciones a considerar en el presente estudio pueden obtener mejores resultados para las Máquinas de Soporte Vectorial, pero no para las redes neuronales artificiales culminando así la primera fase experimental.

TABLA XV
RESULTADOS FINALES PARA ANN

Kernel	Exactitud (Accuracy)
rbf	0.445
<i>can</i>	0.695
tru	0.67
hyperbolic	0.445
triangle	0.445

radial basic	0.645
Rquadratic	0.445
Laplacian	0.445
Sigmoid	0.445
Cosine	0.445

Se observa en la tabla anterior que la mayor exactitud la obtuvo el Kernel Canberra para este conjunto de datos. Con una exactitud promedio del 69 %.

Una vez obtenidos los resultados de experimentación tanto para SVM y ANN se creó el modelo con el Kernel que mejor Score tuvo, en este caso y para estos datos es más favorable utilizar algoritmos de clasificación basados en SVM dado su rendimiento.

4.4 Construir una aplicación web para alertas tempranas de enfermedades cardiovasculares, a través de un modelo predictivo.

4.4.1 Construcción de la herramienta de software

Dado que puede haber usuarios que no tengan habilidades de programación y, por lo tanto, no puedan utilizar el modelo para evaluar los Kernel, se ha desarrollado una herramienta que proporciona un entorno interactivo y fácil de usar para realizar la evaluación de los nuevos Kernels inspirados en la metodología y mediante, esta se pueda hacer una evaluación del riesgo cardiovascular y obtener una alerta temprana.

La metodología que se implementó en todo el desarrollo de la investigación y de los objetivos es la metodología CRISP-DM, no fue necesario una metodología adicional para desarrollar el software porque este perfectamente hace parte de la etapa del despliegue, puesto que el aplicativo lo único que va a realizar es consumir el modelo y mostrar la interfaz sencilla al usuario final, para ello se desarrolló varias actividades presentadas a continuación.

4.4.2 Diagrama arquitectónico

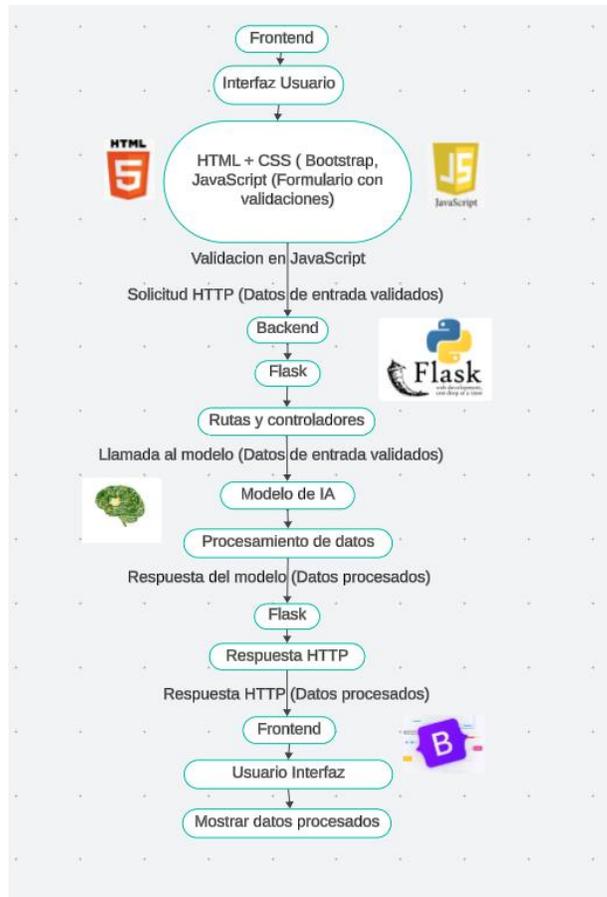


Fig. 14. Diagrama arquitectónico

Arquitectura

Interacción del usuario: El usuario interactúa con la interfaz, construida con HTML y Bootstrap, introduciendo datos en un formulario.

Validación en JavaScript: Antes de enviar el formulario, JavaScript valida los datos introducidos.

Envío del formulario: Una vez validados los datos, el formulario se envía al backend usando el atributo action del formulario.

Procesamiento en Flask:

- Flask recibe la solicitud y extrae los datos.
- Flask envía estos datos al modelo de IA para su procesamiento.

Modelo de IA: El modelo de IA procesa los datos y devuelve los resultados a Flask.

Respuesta al frontend:

- Flask toma los resultados del modelo de IA y crea una respuesta HTTP.
- Esta respuesta se envía de vuelta al frontend.

Mostrar resultados: El navegador recibe la plantilla HTML con los resultados procesados y los muestra al usuario.

4.4.3 Creación del modelo

En esta etapa, se construye el modelo predictivo utilizando técnicas de aprendizaje supervisado, específicamente utilizando un algoritmo de Máquina de Soporte Vectorial (SVC) con el Kernel Truncated. Se seleccionan las técnicas de modelado apropiadas para este caso basándose en las características de los datos y los requisitos del problema. Como se mencionó anteriormente, se ha optado por utilizar un modelo de Máquina de Soporte Vectorial (SVC) debido a su capacidad para manejar conjuntos de datos lineales como no lineales, y su eficacia en problemas de clasificación. Además, al habilitar la probabilidad (`probability=True`), el modelo proporciona estimaciones de probabilidad para las predicciones, lo cual es útil en la evaluación de riesgos como lo es en nuestro caso.

4.4.4 Abstracción del escenario

Dentro del marco de este proyecto centrado en el riesgo cardiovascular, el escenario de enfoque en la salud se establece en la identificación, prevención y gestión de las enfermedades cardiovasculares. Este entorno de investigación aplica a toda la población como individuos con factores de riesgo, personas con antecedentes cardiovasculares, y aquellos que podrían beneficiarse de intervenciones preventivas y que no padezcan de alguna enfermedad cardiovascular. Además, se considera la calidad de la atención médica, la disponibilidad de tecnologías de monitoreo y las barreras que enfrentan las personas en la detección temprana y prevención de eventos cardiovasculares. El objetivo es mejorar la salud cardiovascular a través de la aplicación de modelos de inteligencia artificial y el análisis de datos para la predicción temprana y la implementación de estrategias preventivas.

4.4.5 Visualización de instrucciones

En esta sección se establecen descripciones útiles que permitan una mayor comprensión del aplicativo web y sus funciones.

- **Análisis: Requerimientos funcionales para la visualización de instrucciones**

Tabla XVI
REQUERIMIENTO FUNCIONAL RF-01

<i>RF-01</i>	Visualizar instrucciones
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
<i>Descripción:</i>	La aplicación web debe mostrar al usuario las instrucciones de uso del formulario y su entorno virtual.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

- Desarrollo

La herramienta cuenta con instrucciones claras y concisas para guiar al usuario a través del proceso de ingreso de datos y visualización de resultados. Estas instrucciones son fáciles de entender por qué genera ejemplos de medición y está disponible en un formato accesible para todos los usuarios.

CRK
☰

Instrucciones para el uso del aplicativo

1. Sobre CRK CardioRiskMeter:

Esta aplicación contiene un modelo entrenado con el mejor kernel que permite la identificación del riesgo cardiovascular identificando a su vez aquellos factores de riesgo predominantes logrando brindar alertas tempranas.

2. Instrucciones para el uso de la aplicación:

En el primer módulo se encuentra un formulario con varios ítems de los cuales están disponibles para recolectar la información que ud proporcione, es importante recordar que debe contar con resultados previos de glucosa en ayuno y

Ruta de Atención Clínica:

La Ruta de Atención Clínica es fundamental para la implementación efectiva de la Iniciativa CRK(CardioRiskMeter), mejorando la gestión clínica y autoconocimiento. Además, capacita al usuario y a su familia al educarlos sobre el manejo completo de los factores de riesgo cardiovascular. Esta aplicación abarca la ruta de atención clínica para el manejo integral del riesgo cardiovascular.

3. Acerca de CRK(CardioRiskMeter):

CRK es una iniciativa que surge de un proyecto de grado dirigida a fortalecer los sistemas de atención primaria de salud. Su enfoque se centra en mejorar la prevención y el control de las enfermedades cardiovasculares y sus factores de riesgo. Esta iniciativa tiene como meta fomentar un proceso de mejora continua en la calidad de atención. Esto se logra mediante un mejor control de la hipertensión y la implementación de un enfoque integral en el manejo del riesgo cardiovascular.

Evaluación del riesgo cardiovascular:

La evaluación del riesgo cardiovascular comienza por la identificación de esos datos proporcionados. Estas indagaciones sobre: Ingresar el género, la edad, tabaquismo, actividad física, presencia o ausencia de alguna enfermedad cardiovascular, peso, altura, los valores de colesterol total y presión arterial sistólica y diastólica, glucosa de ayuno. El dato de riesgo cardiovascular será un porcentaje visible mediante un diagrama de barras que indicará la presencia del riesgo de padecer alguna enfermedad cardiovascular.

Modificación del riesgo corrigiendo factores:

Una vez obtenida la estimación de riesgo, se desplegarán unas alertas tempranas que nos permiten identificar el factor de riesgo que se ve alterado, para que se pueda acudir algún especialista en la salud para que aborde ese factor de riesgo. Se considera óptimo evitar fumar,

4. Uso:

En ninguna circunstancia esta aplicación pretende sustituir la consulta con un profesional de la salud ni su criterio clínico. Su objetivo es proporcionar un apoyo al diagnóstico médico, una herramienta para evaluar rápidamente el riesgo cardiovascular y facilitar la comunicación con los usuarios sobre las posibles modificaciones. Además, busca asistir a las personas preocupadas por su salud, ayudándoles a determinar cuándo es necesario buscar atención médica si su riesgo no es bajo. Las recomendaciones en hábitos de vida saludables están descritas a nivel de información y no deben ser utilizadas como guía sin la supervisión de un personal de salud, ya que esto podría resultar peligroso.

Contacto

Correo electrónico:
mrrodriguez.6789@unicesmag.edu.co,
cadelgado.9770@unicesmag.edu.co

Teléfono: 3219671739, 3157155716

Fig. 15. Instrucciones

4.4.6 Selección de datos

El aplicativo web implementa un formulario que permite recolectar los datos relevantes para la predicción de enfermedades cardiovasculares.

- **Análisis: Requerimientos funcionales para la evaluación de riesgo cardiovascular**

TABLA XVII
REQUERIMIENTO FUNCIONAL RF-02

<i>RF-02</i>	Ingresar los datos
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
<i>Descripción:</i>	La aplicación web debe permitir recolectar los datos de aquellos factores de riesgo incluidas en las medidas antropométricas
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

- Desarrollo

El aplicativo web cuenta con un formulario interactivo que le permite al usuario ingresar los datos antropométricos y factores de riesgo cardiovascular. Estos datos son necesarios para realizar una evaluación precisa del riesgo de enfermedades cardiovasculares.

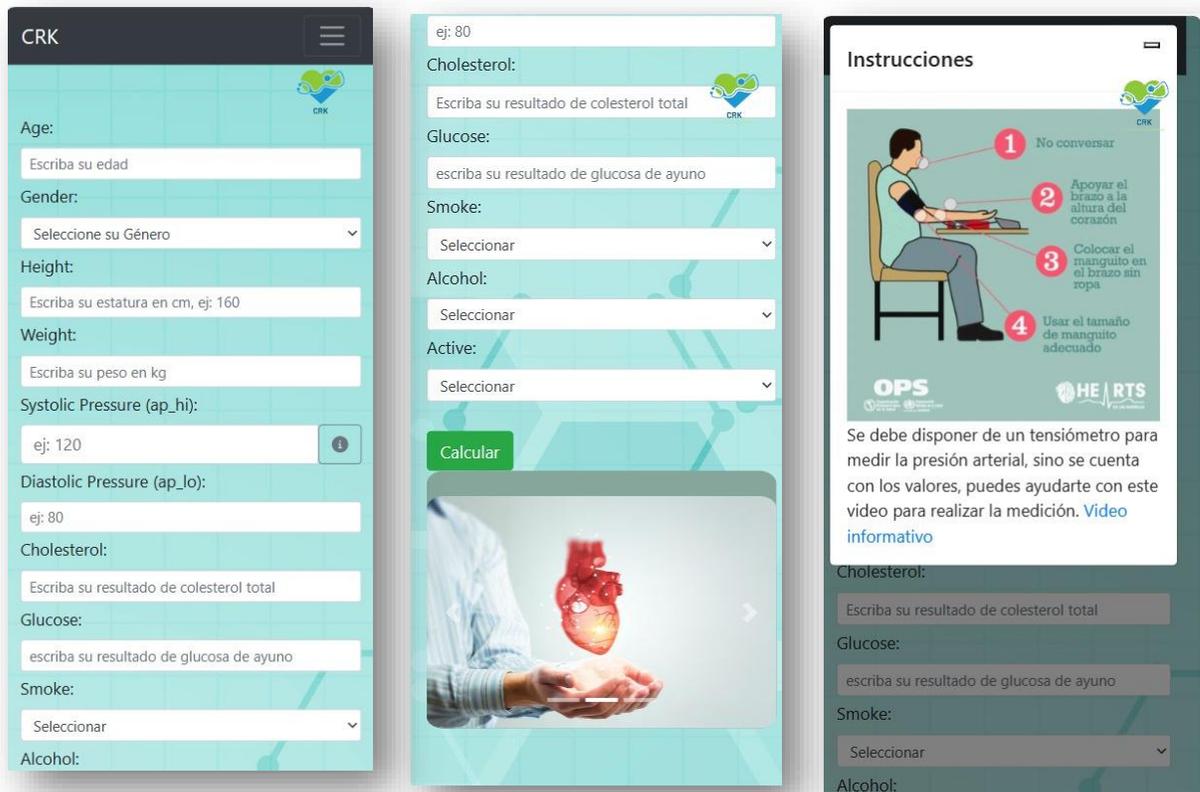


Fig. 16. Formulario de datos

4.4.7 Visualización del Riesgo Cardiovascular (RCV)

En esta etapa el usuario podrá visualizar los resultados que se obtuvieron al complementar el formulario y el porcentaje de padecer alguna enfermedad cardiovascular (ECV).

Análisis: Requerimientos funcionales para la visualización de ECV

TABLA XVIII
REQUERIMIENTO FUNCIONAL RF-03

<i>RF-03</i>	Visualizar el porcentaje de RCV
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
<i>Descripción:</i>	La aplicación web debe indicar al usuario el riesgo de presentar alguna ECV con su porcentaje.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata

<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

- Desarrollo

En este caso se evalúan los datos recopilados del formulario, se visualiza el porcentaje de riesgo de padecer alguna ECV y además se podrá visualizar por medio de una gráfica el resultado.

Para ello se desarrolló un modelo avanzado para evaluar el riesgo de enfermedades cardiovasculares, utilizando una base de datos de 10,000 registros de usuarios. Este modelo fue entrenado con un enfoque binario, donde un resultado de 1 indica la presencia de riesgo y un resultado de 0 indica ausencia de riesgo. Durante el proceso de entrenamiento, el modelo aprendió a identificar patrones y factores de riesgo asociados con enfermedades cardiovasculares basándose en los datos proporcionados.

El funcionamiento del modelo se basa en la generación de un porcentaje de riesgo. Si este porcentaje supera el umbral del 50%, el modelo clasifica al individuo como en riesgo de padecer una enfermedad cardiovascular. Por el contrario, si el porcentaje es inferior al 50%, el modelo determina que el individuo no está en riesgo. Este método permite una evaluación clara y directa del riesgo cardiovascular, facilitando la toma de decisiones informadas sobre la salud del usuario.

Este aplicativo ha sido diseñado para ser intuitivo y accesible, permitiendo a los usuarios comprender fácilmente su nivel de riesgo sin necesidad de conocimientos técnicos avanzados. La información proporcionada por el modelo puede ser fundamental para la prevención y el tratamiento temprano de enfermedades cardiovasculares, mejorando así la salud y el bienestar de los usuarios.

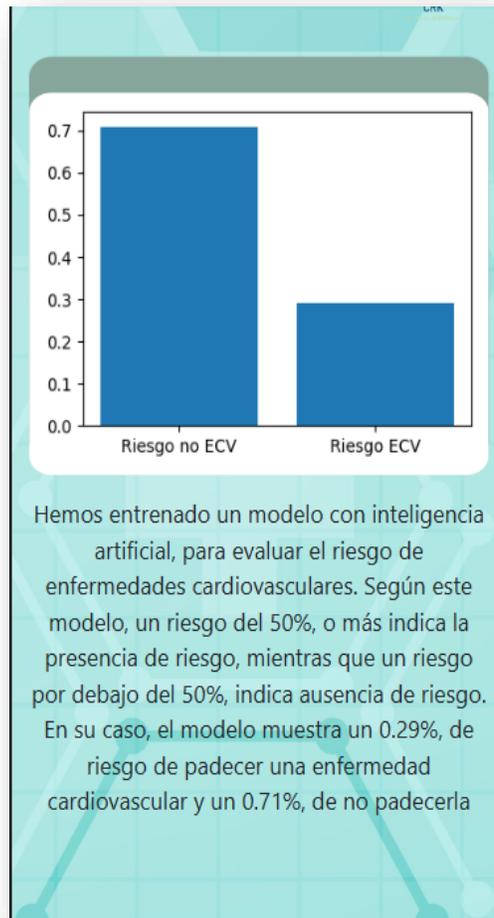


Fig. 17. Porcentaje de RCV

4.4.8 Interpretación de los factores de riesgo generando alertas tempranas

Esta sección va orientada a evaluar cada factor de riesgo, permitiendo generar una alerta temprana con información útil de evaluación.

- **Análisis: Requerimientos funcionales para la visualización de alertas tempranas**

TABLA XIX
REQUERIMIENTO FUNCIONAL RF-04

<i>RF-04</i>	Visualizar de alertas tempranas
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
<i>Descripción:</i>	La aplicación web debe permitir la visualización de las alertas presentadas en cada factor de riesgo, con recomendaciones útiles para su implementación.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

- Desarrollo

Las alertas tempranas se generan cuando existan variables del formulario alteradas que permitan de cierta manera generar estas alertas con recomendaciones útiles para los usuarios.



Fig. 18. Alertas tempranas

4.4.9 Indicaciones prácticas para llevar un estilo de vida saludable

Este módulo contiene consejos prácticos adaptados por recomendaciones de expertos para llevar un estilo de vida saludable.

- **Análisis: Requerimientos funcionales de indicaciones prácticas**

TABLA XX
REQUERIMIENTO FUNCIONAL RF-05

<i>RF-05</i>	Módulo de sugerencias prácticas
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez

	Claudia Alejandra Delgado Calpa
<i>Descripción:</i>	La aplicación web debe proporcionar indicaciones prácticas para la prevención de ECV
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

- Desarrollo

El módulo de Indicaciones prácticas tiene como objetivo proporcionar a los usuarios información detallada y consejos prácticos sobre cómo adoptar y mantener un estilo de vida saludable. Esto incluye recomendaciones sobre nutrición, actividad física, descanso, manejo del estrés, abandono de hábitos nocivos como el tabaquismo y el consumo excesivo de alcohol, así como pautas generales para la prevención de enfermedades y la promoción del bienestar general.

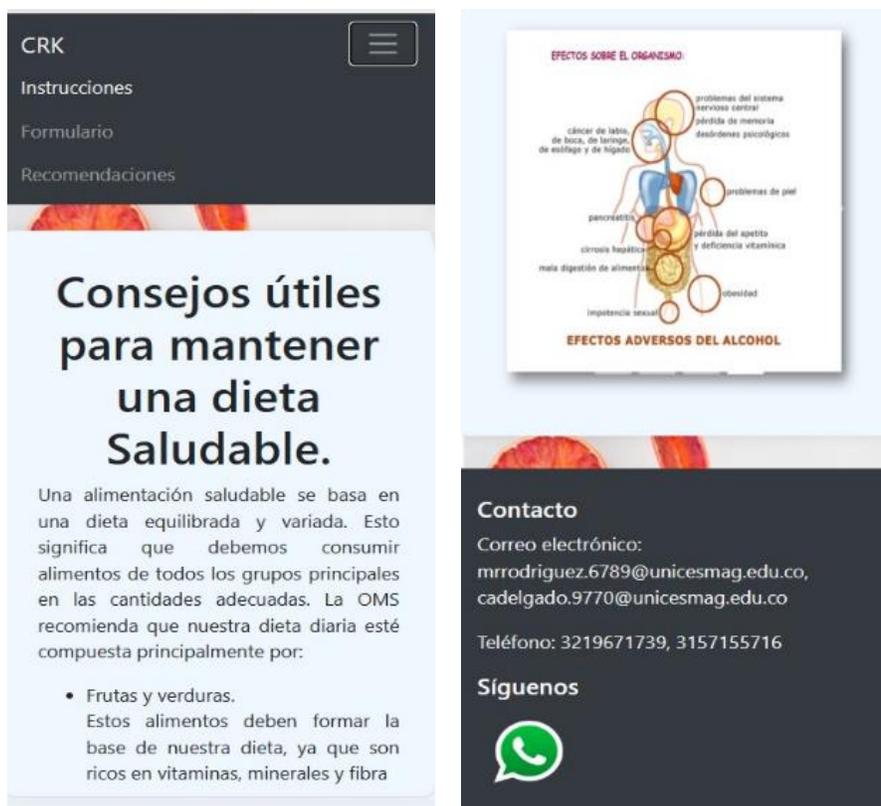


Fig. 19. Recomendaciones

- **Requerimientos no funcionales**

Debido a que la aplicación es útil para ejecutarla en diferentes escenarios, es importante tener un buen proceso de ejecución con su interfaz gráfica visible, impidiendo que esta no se adapte al dispositivo generando al usuario una mala experiencia.

TABLA XXI
REQUERIMIENTO NO FUNCIONAL RNF-01

<i>RNF-01</i>	Usabilidad
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
<i>Descripción:</i>	La aplicación debe permitir ser ejecutada en distintos dispositivos.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

TABLA XXII
REQUERIMIENTO NO FUNCIONAL RNF-02

<i>RNF-02</i>	Eficiencia
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
<i>Descripción:</i>	La aplicación debe tener tiempos de ejecución en el cálculo de RCV no mayor a 5 segundos en promedio.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

Además, en el Anexo A están los diagramas UML que hacen referencia al control lógico y visual de la herramienta, donde se explica la relación de los componentes gráficos, además en el Anexo B, se incluye el manual de usuario.

5. ANÁLISIS DE RESULTADOS

En esta sección se llevaron a cabo diferentes entrenamientos con diferente cantidad de datos, los entrenamientos realizados fueron una parte fundamental de la investigación, puesto que se sabe que no hay un kernel superior a los demás, sino que cada uno tiene sus virtudes y comportamientos dependiendo del conjunto de datos al que se esté aplicando y su uso con cualquier algoritmo de aprendizaje automático, es importante destacar que para este caso se usó un algoritmo de clasificación. Seguido a esto, se creó el modelo con el Kernel que mejor exactitud obtuvo y que fue usado con un conjunto de datos mencionados en la TABLA IX. Finalmente, la evaluación fue realizada con el modelo y con los mejores parámetros como se describe en TABLA XI.

En el análisis de los datos que se obtuvieron durante la experimentación, se utilizaron métricas que permitieron medir el tiempo necesario para proporcionar resultados, para ello se utilizó tiempo de entrenamiento (Training Time), esta métrica mide el tiempo total requerido para entrenar el modelo desde el inicio hasta que se completa el entrenamiento. Incluye el tiempo necesario para procesar los datos, ajustar los parámetros del modelo y realizar las iteraciones necesarias, se observó que las redes neuronales artificiales (ANN) tomaron menos tiempo que las máquinas de soporte vectorial (SVM), para todos los Kernel, debido a la aproximación de Nystroem, proporcionando los resultados como se muestra en la Fig. 20 y en la Fig. 21,

```
start_time = time.time()

annc=KANNC(kernel='can',gamma=206.913808111479,random_state=0)
modelo_predic2 = annc.fit(Xtrain, ytrain)

end_time = time.time()

training_time = end_time - start_time

print('clfk',annc.score(Xtest,ytest))
print(f'Tiempo de entrenamiento: {training_time:.2f} segundos')

clfk 0.69
Tiempo de entrenamiento: 3.67 segundos
```

Fig. 20. Tiempo en ANN

```
start_time = time.time()

svc=KSVC(kernel='can', gamma=0.0006951927961775605,C=2154.4346,probability=True)
modelo_predic2 = svc.fit(Xtrain, ytrain)

end_time = time.time()

training_time = end_time - start_time

print('cmfvc',svc.score(Xtest,ytest))
print(f'Tiempo de entrenamiento: {training_time:.2f} segundos')

cmfvc 0.725
Tiempo de entrenamiento: 11.65 segundos
```

Fig. 21. Tiempo de SVM

Por consiguiente, es importante destacar el acoplamiento de los modelos mencionados en la Fig. 9 a la biblioteca Scikit-learn. Esto permitió ahorrar tiempo en la construcción de los algoritmos de búsqueda de hiperparámetros, logrando así un aporte a la comunidad de software basada en la construcción de modelos basados en funciones Kernel.

De acuerdo con lo enunciado por los expertos Mora.H., y Belanche.L., las funciones Kernel como la Canberra, Truncated, y adicionalmente Coseno proporcionan mejores resultados que la RBF para cierto tipo de datos, confirmando esta hipótesis en lo que respecta a esta investigación y para

comprobar lo expuesto por los expertos se puede observar los resultados en la TABLA XII, donde la creación de esos nuevos kernel se destacan por la obtención de mejores resultados para SVM, que, para ANN, e independiente de la configuración de los mejores parámetros, comprobando que el kernel Canberra obtuvo un 72% de rendimiento demostrando ser la opción más eficiente, frente a otros Kernels, que se puede visualizar en la TABLA XII.

De esta experimentación se puede destacar que según la configuración de los parámetros se puede aproximar, a lograr mejores resultados, logrando varias comparaciones, incluyendo la de Kernels más escalables, la de algoritmos y la de transformadores de escalación. En el caso de las ANN, la comparación de transformadores de escalación en el Standar Scaler con el Kernel Canberra mostró un rendimiento del 71% con el Kernel Radial Basic, 72%, en cuanto al Migmag Scaler con el Kernel Canberra, se obtuvo un porcentaje del 73% y con el Normalizer Scaler se obtuvo un porcentaje del 71% con el Kernel Triangular, como se puede evidenciar en la TABLA XIII.

Para las SVM, el rendimiento con el Standar Scaler el Kernel Truncated fue del 71%, mientras que con Migmag Scaler el Kernel Truncated fue del 71%. con el transformador Normalizer Scaler el Kernel Radial Basic, el rendimiento fue del 72%. Se observó una similitud en los porcentajes en estas configuraciones como se muestra en la TABLA X. Sin embargo, en las SVM, el Kernel Canberra demostró ser la opción más eficiente evidenciada en la TABLA XII.

Concluyendo así que las ANN mostraron un rendimiento inferior al 50%, mientras que las SVM superaron el 70%. Esto indica que es más favorable utilizar algoritmos de clasificación basados en SVM.

Una de las características de los Kernels que permitió esta conclusión fue el estudio de la exactitud en Máquinas de Soporte Vectorial donde se muestra en la TABLA XII, cada Kernel con su exactitud observando que un modelo con valores de exactitud mayores al 50% se clasifica en aceptable lo que permite el desarrollo de un buen modelo, y empeoran para ANN porque como se muestra en la TABLA XV, cada Kernel con su exactitud muestra valores que están por debajo del 50% en su mayoría, por lo tanto un modelo con menos del 50% de exactitud se considera muy pobre esto significa que el modelo está prediciendo incorrectamente más de la mitad del tiempo.

Dentro del marco de este proyecto centrado en la comparación de funciones kernel en SVM y ANN en la predicción del riesgo cardiovascular, se ve la necesidad de que ese modelo sea desplegado mediante el aplicativo. Para ello este entorno de investigación abarca poblaciones como individuos con factores de riesgo, personas con antecedentes cardiovasculares y aquellos que podrían

beneficiarse de intervenciones preventivas sin padecer alguna enfermedad cardiovascular. Además, se considera la calidad de la atención médica, la disponibilidad de tecnologías de monitoreo y las barreras que enfrentan las personas en la detección temprana y prevención de eventos cardiovasculares. El objetivo es mejorar la salud cardiovascular a través de la aplicación de modelos de inteligencia artificial y el análisis de datos para la predicción temprana y la implementación de estrategias preventivas.

La herramienta gráfica desarrollada en este proyecto se basa en la visualización interactiva de datos relacionados con el riesgo cardiovascular. Esta herramienta fue implementada utilizando tecnologías avanzadas de visualización de datos como Python, integrándose con modelos de inteligencia artificial que procesan y analizan datos clínicos de diversas fuentes. La implementación incluyó las siguientes etapas:

1. Se recopilaron datos de diferentes fuentes. Estos datos fueron preprocesados para garantizar su calidad y coherencia.
2. Utilizando técnicas de Machine Learning, se desarrolló un modelo predictivo para evaluar el riesgo cardiovascular de los usuarios. Estos modelos fueron entrenados y validados utilizando conjuntos de datos relevantes.
3. La herramienta gráfica se integró con el modelo predictivo para permitir una visualización intuitiva de los resultados. Se desarrollaron módulos interactivos que permiten a los usuarios explorar los datos y entender mejor los factores de riesgo y las predicciones del modelo.

La visualización gráfica de los datos permite a los profesionales de la salud tomar decisiones informadas basadas en la comprensión clara de los riesgos individuales y colectivos, siendo esta herramienta un apoyo al diagnóstico médico. Esto es particularmente útil para identificar patrones y tendencias que no son fácilmente discernibles a través de datos tabulares, a su vez la herramienta ayuda a identificar a individuos en alto riesgo de desarrollar enfermedades cardiovasculares antes de que se presenten síntomas clínicos evidentes. Esto facilita la implementación de estrategias preventivas y de intervención temprana, mejorando los resultados de salud a largo plazo.

La interfaz gráfica está diseñada para ser intuitiva y accesible, incluso para usuarios con poca experiencia en análisis de datos. Esto permite a los usuarios y a los profesionales de la salud entender mejor la información presentada y tomar decisiones basadas en datos. Además, sirve como una

herramienta educativa que puede ilustrar claramente los efectos de diferentes factores de riesgo y las ventajas de las intervenciones preventivas.

CONCLUSIONES

El escenario experimental propuesto permitió confirmar lo postulado por los expertos Mora.H., y Belanche.L., y validar el proyecto de investigación de que las principales características de los kernel investigados se ajustan y se comportan mejor ante el tipo de datos suministrados para Maquinas de Soporte Vectorial (SVM) y empeoran para Redes Neuronales Artificiales (ANN).

El acoplamiento de las funciones kernel a la biblioteca Scikit-Learn facilita una relación cercana con algoritmos desarrollados, contribuyendo así a la comunidad de software libre e investigación académica. Esto sirve como punto de referencia para construir modelos que se ajusten a esta biblioteca que es utilizada globalmente.

La visualización de los resultados de los experimentos permitió evaluar objetivamente el rendimiento de cada Kernel. Esto nos permitió identificar los mejores parámetros para las máquinas de soporte vectorial, utilizando normalizadores de escalación como Standar Scaler y Kernel Truncated para la clasificación de datos de riesgo cardiovascular, logrando un buen equilibrio entre tiempo de procesamiento y exactitud.

Finalmente, el desarrollo de la herramienta permitió desplegar de manera gráfica el modelo y facilito el aprendizaje en la implementación de interfaces gráficas dinámicas, de usuario en Python, además del uso de bibliotecas de análisis de datos disponibles en Python.

RECOMENDACIONES

La herramienta desarrollada es fácilmente escalable, se recomienda su utilización para implementar nuevos modelos y funciones. Además, es importante desarrollar nuevas aplicaciones de visualización que faciliten la predicción de enfermedades cardiovasculares.

Dado que la investigación incluye un comparativo de funciones kernel tanto en Máquinas de Soporte Vectorial (SVM) como en Redes Neuronales Artificiales (ANN), donde fue necesario profundizar en la teoría de métricas utilizadas y matemáticas, se considera necesario promover entre los estudiantes la importancia de comprender los fundamentos de estos principios para abordar las bases de cualquier tema de investigación.

Por último, se recomienda profundizar en los estudios comparativos de funciones kernel acopladas a otros algoritmos de Machine Learning para clasificación o regresión lineal de predicción de enfermedades cardiovasculares, para saber qué algoritmo trabaja mejor con Kernels en estos conjuntos de datos.

Referencias

- [1] A. Beade Ruelas y C. E. García Soto, «Gobierno de México,» No rompas más tu corazón. Salud cardiovascular, 15 02 2017. [En línea]. Available: <https://www.gob.mx/profeco/documentos/no-rompas-mas-tu-corazon-salud-cardiovascular?state=published#:~:text=De%20acuerdo%20con%20la%20Organizaci%C3%B3n,muerte%20en%20todo%20el%20mundo..> [Último acceso: 30 04 2023].
- [2] Cemprende, «innpulsacolombia,» Enfermedades Cardiovasculares, un gran desafío para la salud pública en Colombia, 29 09 2022. [En línea]. Available: [https://www.innpulsacolombia.com/cemprende/noticias/enfermedades-cardiovasculares-un-gran-desafio-para-la-salud-publica-en-colombia#:~:text=En%20Colombia%2C%20durante%202021%2C%20se,de%20Estad%C3%ADsticas%20\(DANE\)%204..](https://www.innpulsacolombia.com/cemprende/noticias/enfermedades-cardiovasculares-un-gran-desafio-para-la-salud-publica-en-colombia#:~:text=En%20Colombia%2C%20durante%202021%2C%20se,de%20Estad%C3%ADsticas%20(DANE)%204..) [Último acceso: 30 04 2023].
- [3] D. Molina de Salazar, «Propuesta en prevención del riesgo cardiovascular,» *Revista Colombiana de Cardiología*, vol. 5, n° 5, 20888.
- [4] C. . F. Patiño Zambrano, «DISPOSITIVO VESTIBLE INTELIGENTE PARA LA GENERACIÓN DE ALERTAS TEMPRANAS DE EVENTOS CARDIOVASCULARES DE RIESGO,» Envigado, 2022.
- [5] E. MAT, «Sanofi Campus,» Sanofi, 03 10 2022. [En línea]. Available: <https://campus.sanofi.es/es/noticias/machine-learning-prediccion-enfermedades-cardiovasculares>. [Último acceso: 25 02 2023].
- [6] L. A. Gómez, «Las enfermedades cardiovasculares: un problema de salud pública y un reto global,» *SCIELO*, p. 1, 2011.
- [7] L. E. Pérez Leal y J. A. Buitrago Cárdenas, «Predicción del diagnóstico de diabetes a partir de perfiles clínicos de pacientes utilizando aprendizaje automático,» Universidad Antonio Nariño, Bogotá, 2021.
- [8] D. A. Gallego Valcárcel y D. F. Lucas Monsalve, «MODELOS DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DEL RIESGO DE FATALIDAD POR INSUFICIENCIA CARDIACA CON DATOS CLÍNICOS,» Bogotá, 2021.

- [9] M. Álvarez Vega, L. M. Quirós Mora y M. V. Cortés Badilla, «Inteligencia artificial y aprendizaje automático en medicina,» *Revista Médica Sinergia*, vol. 5, n° 8, p. 12, 2020.
- [10] D. d. P. Ferreira Guerrero, M. Díaz Vera y C. P. Bonila Ibañez, «Factores de riesgo cardiovascular modificables en adolescentes escolarizados de Ibagué 2013,» *Repositorio Institucional Universidad de Antioquia*, p. 1, 2017.
- [11] H. A. Mora Paz, «Comparativo de Kernels sobre predicción de oferta de fuentes alternativas de energía,» unir La universidad en Internet, Pasto, 2019.
- [12] P. A. Friedman , S. Kapa, F. López Jiménez y P. A. Noseworthy, «Mayo Clinic,» Centro y departamentos médicos, 14 01 2023. [En línea]. Available: <https://www.mayoclinic.org/es-es/departments-centers/ai-cardiology/overview/ovc-20486648>. [Último acceso: 22 mayo 2023].
- [13] Aprende IA, «Aprende IA,» [En línea]. Available: <https://aprendeia.com/kernel-maquinas-vectores-de-soporte-clasificacion-regresion/>. [Último acceso: 06 04 2023].
- [14] V. Sowmya , K. Sanjana, E. Gopalakrishnan y K. P. Dr. Soman, «Inteligencia artificial explicable para la variabilidad de la frecuencia cardíaca en la señal de ECG,» *Tecnología de la salud Lett*, vol. 7, n° 6, p. 146, 2020.
- [15] S. Mohan, C. Thirumalai y G. Srivastava, «Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques,» *IEEE*, vol. 7, n° 18770357, pp. 81542 - 81554, 2019.
- [16] H. Wu, L. Yang, X. Jin y P. Zheng, «Study of cardiovascular disease prediction model based on random forest in eastern China,» *Scientific Reports*, vol. 1, n° 5245, p. 10, 2020.
- [17] O. Chavez Olivera, L. Galindo Honores, A. Barrientos Padilla y M. Cuadros Galvez, «Aplicación Móvil para Predecir la Probabilidad de Pertenecer al Grupo de Riesgo Cardiovascular Utilizando Machine Learning,» Memorias de la Décima Segunda Conferencia Iberoamericana de Complejidad, Informática y Cibernética, Lima, Perú, 2022.
- [18] M. SCAVINO, . A. CASTRILLEJO, V. S. ESTRAGÓ MÉROLA, L. E. LURAGHI LOPEZ, M. MUÑOZ y R. ÁLVAREZ VAZ, «Informe final publicable de proyecto Creación de algoritmos utilizando técnicas de clasificación supervisada y no supervisada para el diagnóstico de enfermedades cardiovasculares en una población de adultos mayores de bajos recursos en Uruguay,» Uruguay, 2022.

- [19] L. D. Polero , C. M. Garmendia, R. E. Echegoyen, A. Alves de Lima, F. Bertón, F. Lambardi, P. Ariznavarreta, R. Campos y J. Costabel, «Predicción de riesgo de sufrir un síndrome coronario agudo mediante un algoritmo de Machine Learning (ANGINA),» *Scielo - Revista argentina de cardiología*, vol. 88, n° 1, 2020.
- [20] J. D. Perez Tatis, «OPTIMIZACIÓN DE UN MODELO DE CLASIFICACIÓN DE ENFERMEDADES CARDIOVASCULARES UTILIZANDO TÉCNICAS DE APRENDIZAJE PROFUNDO SUPERVISADO Y DESPLIEGUE DE DASHBOARD WEB,» Cartagena, 2021.
- [21] E. J. Martínez, «Predicción De Enfermedades Cardiovasculares Mediante Algoritmos De Inteligencia Artificial,» Málaga, 2020.
- [22] F. L. Carrascal Porras y L. A. Florez Prias, «Modelo de inteligencia artificial como apoyo diagnóstico para la estimación de riesgo cardiovascular en pacientes atendidos bajo la modalidad de telemedicina en una IPS del departamento de Sucre 2021,» UNAD Universidad Nacional Abierta y a Distancia, Sucre, 2021.
- [23] C. Dolores y J. Ordovás, «GENES, DIETA, Y ENFERMEDADES CARDIOVASCULARES,» *GENETICA*, vol. 5, pp. 71-118, 2007.
- [24] J. Vega Abascal, M. Guimará Mosqueda y L. Vega Abascal , «Riesgo cardiovascular, una herramienta útil para la prevención de las enfermedades cardiovasculares,» *Scielo*, vol. 27, n° 1, p. 3, 2011.
- [25] E. J. Martínez, «PREDICCIÓN DE ENFERMEDADES CARDIOVASCULARES,» Universidad de Malaga, Malaga, 2020.
- [26] S. Sans Menéndez, «Enfermedades Cardiovasculares,» Institut d' Estudis de la Salut, Barcelona, Barcelona, 2011.
- [27] Diabetrics, «Diabetrics 361 en funcion de la vida,» Diabetrics, [En línea]. Available: <https://blog.diabetrics.com/escala-framingham>. [Último acceso: 28 03 2023].
- [28] J. D. Perez Tatis, «OPTIMIZACIÓN DE UN MODELO DE CLASIFICACIÓN DE ENFERMEDADES CARDIOVASCULARES UTILIZANDO TÉCNICAS DE APRENDIZAJE PROFUNDO SUPERVISADO Y DESPLIEGUE DE DASHBOARD WEB,» Universidad del Sinú, Cartagena, 2021.

- [29] Mathworks, «Mathworks,» Mathworks, [En línea]. Available: <https://es.mathworks.com/discovery/support-vector-machine.html>. [Último acceso: 28 03 2023].
- [30] Ecured, «Ecured,» Ecured, [En línea]. Available: https://www.ecured.cu/Funci%C3%B3n_Kernel. [Último acceso: 27 03 2023].
- [31] Clasificador lineal, «Wikipedia,» La enciclopedia libre, 30 08 2019. [En línea]. Available: https://es.wikipedia.org/wiki/Clasificador_lineal. [Último acceso: 30 04 2023].
- [32] Tibco Data Science, «Tibco cloud,» [En línea]. Available: <https://www.tibco.com/es/reference-center/what-is-supervised-learning>. [Último acceso: 28 03 2023].
- [33] A. Montiel de Jesús, «Desarrollo de una aplicación para dispositivos móviles para la detección temprana de enfermedades cardiovasculares,» TECNM tecnologico Nacional de Mexico, Orizaba México, 2022.
- [34] Sitio bigdata , «Sitio bigdata,» 27 05 2019. [En línea]. Available: <https://sitiobigdata.com/2019/05/27/modelos-de-machine-learning-metricas-de-regresion-mse-parte-2/#>. [Último acceso: 23 04 2023].
- [35] J. Mat3nez Heras, «IArtificial,» 09 10 2020. [En línea]. Available: <https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/>. [Último acceso: 28 04 2023].
- [36] L. Quintanilla , G. Warren, S. Kirsch, V. Youssef, N. Kershaw, S. Killeen, N. Schonning y J. Ormont, «Learn Microsoft,» 23-25 05 2023. [En línea]. Available: <https://learn.microsoft.com/es-es/dotnet/machine-learning/resources/metrics>. [Último acceso: 30 04 2023].
- [37] B. Boehm, «Wikipedia Desarrollo en espiral,» Enciclopedia libre, 23 05 2012. [En línea]. Available: https://es.wikipedia.org/wiki/Desarrollo_en_espiral. [Último acceso: 30 04 2023].
- [38] F. Ballina R3os, «PARADIGMAS Y PERSPECTIVAS TE3RICO-METODOL3GICAS EN EL ESTUDIO DE LA ADMINISTRACI3N,» uvmx, 2013.
- [39] M. Radrigan, «Wikipedia la enciclopedia libre,» Wikipedia, 30 11 2022. [En línea]. Available: https://es.wikipedia.org/wiki/M%C3%A9todo_emp%C3%A9rico-anal%C3%A9tico#:~:text=El%20m%C3%A9todo%20emp%C3%A9rico%2Danal%C3%A9

- tico%20es,sociales%20y%20en%20las%20matem%C3%A1ticas.. [Último acceso: 22 05 2023].
- [40] Conceptos básicos de ayuda de CRISP-DM, «IBM,» 17 08 2021. [En línea]. Available: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>. [Último acceso: 30 04 2023].
- [41] J. F. Vallalta Rueda, «IA Health Data Miner,» CRISP-DM: una metodología para minería de datos en salud, [En línea]. Available: <https://healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/>. [Último acceso: 30 04 2023].
- [42] J. A. Toral Barrera, «Redes Neuronales,» cucei, España, 2019.
- [43] Kaggle, *Herramientas de repositorio*, 2018.
- [44] A. Asunción y D. Newman, *Repositorios*, Massachusetts Amherst: Rexa.info, 2007.
- [45] H. Mora Paz, J. A. Riascos, J. A. Salazar Castro, G. Mora y A. Pantoja, «Comparación de funciones kernel para la predicción de la oferta energética fotovoltaica,» *RISTI Revista Ibérica de Sistemas e Tecnologías de Información*, nº E38, pp. 310-324, 2020.
- [46] L. A. Belanche y M. A. Villegas, «Kernel Functions for Categorical Variables with Application to Problems in the Life Sciences,» nº 08034, pp. 1-3, 2023.
- [47] ArcGIS Pro, «esri,» [En línea]. Available: <https://pro.arcgis.com/es/pro-app/latest/help/analysis/introduction/spatial-analysis-in-arcgis-pro.htm>. [Último acceso: 15 08 2023].
- [48] Scriptología, 01 2024. [En línea]. Available: <https://scriptologia.com/tutorial-de-flask-desarrollando-aplicaciones-web-en-python/>.
- [49] J. Hunter, D. Dale, E. Firing y M. Droettboom, 2012. [En línea]. Available: <https://es.matplotlib.net/stable/tutorials/introductory/pyplot.html>.
- [50] Python, 2001. [En línea]. Available: <https://docs.python.org/es/3/library/pickle.html>.
- [51] DataScientest, 2023. [En línea]. Available: <https://datascientest.com/es/pandas-python>.
- [52] N. Manav, 2023. [En línea]. Available: <https://www.delftstack.com/es/howto/python/write-bytes-to-file-python/>.
- [53] [En línea]. Available: <https://docs.python.org/es/dev/library/base64.html>.

- [54] S. Navarro, abril 2024. [En línea]. Available: <https://keepcoding.io/blog/para-que-sirve-el-train-test-split/>.
- [55] Scikit Learn, 2007. [En línea]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.kernel_approximation.Nystroem.html.
- [56] B. Li, P. Lu y v. chmcc1, 01 06 2023. [En línea]. Available: <https://learn.microsoft.com/es-es/azure/machine-learning/component-reference/multiclass-neural-network?view=azureml-api-2>.
- [57] A. Imbert y G. Lemaitre, 01 2024. [En línea]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>.
- [58] M. Ruiz, 12 2018. [En línea]. Available: <https://openwebinars.net/blog/que-es-redux/>.
- [59] T. Moes, Julio 2023. [En línea]. Available: <https://softwarelab.org/es/blog/que-es-windows/>.
- [60] D. A., 01 2023. [En línea]. Available: <https://www.hostinger.mx/tutoriales/que-es-bootstrap#:~:text=Bootstrap%20es%20un%20framework%20de%20desarrollo%20web%20gratuito,una%20colecci%C3%B3n%20de%20sintaxis%20para%20dise%C3%B1os%20de%20plantillas..>
- [61] R. Monk, 11 2023. [En línea]. Available: <https://www.coursera.org/mx/articles/what-is-python-used-for-a-beginners-guide-to-using-python>.
- [62] F. T. Valiente, «icc Instituto de Ingenieria del conocimiento,» UAM, [En línea]. Available: [https://www.iic.uam.es/inteligencia-artificial/aprendizaje-por-refuerzo/#:~:text=El%20aprendizaje%20por%20refuerzo%20\(o,de%20optimizaci%C3%B3n%20basada%20en%20datos..](https://www.iic.uam.es/inteligencia-artificial/aprendizaje-por-refuerzo/#:~:text=El%20aprendizaje%20por%20refuerzo%20(o,de%20optimizaci%C3%B3n%20basada%20en%20datos..) [Último acceso: 28 03 2023].
- [63] A. M. Arceo Vilas, «Estado nutricional y adherencia a la dieta mediterránea en población mayor de 40 años. Utilidad de las técnicas de inteligencia artificial versus técnicas estadísticas clásicas.,» UDC Universidad de Coruña, Coruña, 2020.
- [64] J. M. Sánchez Santos y P. L. Sánchez Fernández, «Predicción de eventos cardiovasculares y hemorrágicos en pacientes con doble antiagregación con modelos machine learning,» CREDOS, Salamanca, 2020.

- [65] D. A. Gallego Valcárcel y D. F. Lucas Monsalve, «modelos de aprendizaje automático para la predicción del riesgo de fatalidad por insuficiencia cardiaca con datos clínicos,» UAN Universidad Antonio Nariño, Bogotá, 2021.
- [66] P. Fernández García, G. Vallejo Seco, P. E. livacic-Rojas y E. Tuero Herrero, «Validez Estructurada para una investigación cuasi-experimental de calidad. Se cumplen 50 años de la presentación en sociedad de los diseños cuasi-experimentales,» *Scielo*, vol. 30, nº 2, p. 5, 2014.
- [67] J. Lozada, «Investigación Aplicada,» *Dialnet unirioja*, vol. 3, nº 1, pp. 47-50, 2014.
- [68] M. Londoño Ocampo, «Definición de un modelo de clasificación de riesgo cardiovascular para una población de adultos mayores usando técnicas de aprendizaje de máquinas,» Universidad Nacional de Colombia, Medellín, 2020.
- [69] F. d. J. Núñez Cardenas , I. D. C. Zavaleta Chi, A. M. Felipe Redondo y J. Meléndez Hernández, «Aplicación de técnicas de minería de datos para la tipificación de enfermedades cardiovasculares en alumnos universitarios,» México, 2018.
- [70] J. Martínez, «Datasmarts,» 19 11 2019. [En línea]. Available: <https://datasmarts.net/es/mas-alla-del-accuracy-precision-recall-y-f1/>. [Último acceso: 23 04 2023].
- [71] J. M. Mostaza, X. Pintó, P. Armario, L. Masana, J. T. Real, P. Valdivielso, T. Arrobas-Velilla, R. Baeza-Trinidad, P. Calmarza, J. Cebollada, A. M.Civera, J. I. Cuende Melero, J. L. Díaz Díaz, J. Fernández Pardo, C. Guijarro, C. Jericó, M. Laclaustra, C. Lahoz, J. López Miranda, S. Martínez Hervás, O. Muñiz Grijalvo, J. A. Páramo, V. Pascual, J. Pedro Botet, P. Pérez Martínez, N. Plana, J. Puzo, M. Á. Sánchez Chaparro y L. Vila, «Estándares SEA 2022 para el control global del riesgo cardiovascular,» *Clínica e Investigación en Arteriosclerosis*, vol. 34, nº 3, pp. 130-179, 2022.
- [72] J. Calvo Martin, «Linked in,» 18 08 2022. [En línea]. Available: <https://es.linkedin.com/pulse/la-importancia-de-las-funciones-activaci%C3%B3n-en-una-red-calvo-martin>. [Último acceso: 16 08 2023].

ANEXOS

Anexo A: Manual del sistema



CRK

CardioRiskMeter

CRK CardioRiskMeter / Manual del Sistema

**Michael Rodríguez
Claudia Delgado
Héctor Mora
Jorge Rivera**

2024

INTRODUCCIÓN

CRK-WEB, es un sistema de información que se adapta a cualquier tipo de dispositivo para su visualización y uso, puesto que maneja un diseño responsive. Permite predecir el riesgo de padecer enfermedades cardiovasculares, generar alertas tempranas en caso de que se identifiquen factores de riesgo potenciales, además ofrece recomendaciones útiles sobre como promover un estilo de vida saludable. Para llevar a cabo esta herramienta fue importante realizar un comparativo de funciones kernel en redes neuronales artificiales (ANN), y máquinas de soporte vectorial (SVM), logrando desarrollar un modelo preciso y confiable.

Se estableció un formulario con varios campos para recopilar la información introducida por el usuario, estos campos solicitan información como medidas antropométricas (peso, altura), factores de riesgo modificables (presión arterial, colesterol, glucosa, tabaquismo, consumo de alcohol, actividad física) y factores de riesgo no modificables (edad, género). La herramienta puede ser utilizada por cualquier persona que quiera conocer su riesgo de padecer alguna enfermedad cardiovascular.

La herramienta permite de manera interactiva y fácil conocer el estado de salud, factores que pueda modificar, y evitar a futuro enfermedades que la comprometan, está disponible para todos aquellos usuarios que tengan interés en utilizarla sean como apoyo al diagnóstico médico, o por autonomía propia siempre y cuando tenga conocimiento en terminología de salud.

El presente documento detalla el diseño, la metodología de desarrollo, las tecnologías utilizadas y la implementación del Sistema CRK-WEB.

Los Autores de esta producción de software, a saber: Michael Rodríguez, Claudia Delgado

Héctor Mora, Jorge Rivera

Aplicativo CRK-Web

CRK-WEB, es desarrollado como aplicativo de despliegue para consumir el modelo creado con el kernel que mejor precisión y comportamiento obtuvo con el tipo de datos implementados, en este caso el mejor accuracy se obtuvo para máquinas de soporte vectorial con el kernel Canberra. Este sistema permite optimizar procesos de salud para cualquier tipo de población sin padecer alguna enfermedad cardiovascular, en el desarrollo de este sistema, se abarcó todo el proceso que va desde la fase de consultar la recomendación de expertos, creación de nuevos kernel, arquitectura, desarrollo hasta la fase de pruebas. Además, se utilizó la misma metodología que se implementó en el desarrollo de la investigación, que fue la que más se adaptó al proyecto, utilizando así una metodología ágil y moderna.

En la implementación de esta herramienta, se desarrolló dos módulos importantes, uno de recolección de información por medio de un formulario, y el otro de recomendaciones útiles para la prevención de ECV, que de cierta manera permiten ejecutar el objetivo del proyecto.

El objetivo del desarrollo de CRK-WEB es, como se mencionó anteriormente, consumir el modelo predictivo, con el fin de optimizar procesos, y presentar un diseño visualmente agradable para el usuario. En el desarrollo de este aplicativo se utilizan herramientas que permiten la ejecución de los procesos que llevan a cumplir el objetivo final propuesto, además comprender la estructura del sistema y lo que implica la creación del modelo haciendo uso del kernel, exige el uso de esta herramienta para proporcionar el despliegue que perfectamente se adhiere a la etapa que provee la metodología.

Se hace énfasis en la metodología, ya que para este proyecto no se implementó una metodología aparte para el desarrollo del aplicativo, sino que se consolidó en una de las etapas de la metodología del desarrollo del proyecto, la etapa de despliegue dando cumplimiento con el último objetivo propuesto, generando resultados óptimos y eficaces que estén a disposición incluso para nuevas investigaciones.

El aplicativo web permite visualizar y administrar la información relevante para los usuarios que hagan uso de la plataforma.

Marco del Proyecto

CRK-WEB, es una aplicación que le permite estudiar la información recolectada, cuenta con el uso de tecnologías actualizadas que permite que se acople a las necesidades del proyecto. Las tecnologías más relevantes utilizadas en este desarrollo se describen a continuación.

FLASK: la biblioteca Flask es un marco de desarrollo web ligero y flexible para Python. Permite a los desarrolladores construir aplicaciones web rápidas y escalables con facilidad. Flask es el nombre de la clase principal de Flask. Al importar Flask, puedes crear instancias de esta clase para crear una aplicación web [48].

MATPLOTLIB: la biblioteca Matplotlib.pyplot es una parte de Matplotlib, que es una biblioteca de trazado en 2D de Python que produce figuras de calidad de publicación en una variedad de formatos impresos y entornos interactivos a través de interfaces de usuario, Matplotlib.pyplot es una colección de funciones de estilo de comando que hacen que Matplotlib funcione como MATLAB. Cada función de pyplot hace algún cambio en una figura [49].

IMPORT DILL: la biblioteca dill es una extensión de la biblioteca estándar pickle de Python. Mientras que pickle es capaz de serializar y deserializar objetos Python simples, dill puede manejar, de manera más flexible y robusta, objetos más complejos, como funciones definidas por el usuario, instancias de clases, generadores, y más.

La principal ventaja de dill es la capacidad para serializar objetos que contienen referencias a funciones definidas por el usuario y otros objetos que no son compatibles con la serialización estándar. Esto hace que dill sea útil en situaciones donde se necesita guardar y cargar objetos complejos en Python, como la persistencia de modelos de aprendizaje automático [50].

PANDAS: la biblioteca panda es una herramienta poderosa y fácil de usar para el análisis y manipulación de datos en Python. Proporciona estructuras de datos flexibles y rápidas, ideales para trabajar con datos tabulares y series temporales [51].

BYTESIO: es una clase que permite crear un objeto similar a un archivo en memoria que actúa como un búfer de bytes. Este objeto se puede utilizar para leer y escribir datos binarios en

memoria, como bytes, en lugar de interactuar con un archivo real en el sistema de archivos. Al importar BytesIO, puedes crear instancias de esta clase y utilizarlas para manipular datos binarios en memoria de manera eficiente, como leer y escribir archivos binarios, trabajar con datos de imágenes, y más [52].

BASE64: en Python proporciona funciones para codificar y decodificar datos utilizando el algoritmo de codificación Base64. La codificación Base64 se utiliza comúnmente para representar datos binarios de una manera que sea segura para su transporte a través de protocolos que pueden manejar solo datos de texto, como el correo electrónico o HTTP. Al importar base64, tienes acceso a las siguientes funciones principales: `base64.b64encode(data)`: Esta función toma datos binarios como entrada y devuelve una versión codificada en Base64 como una cadena de bytes. `base64.b64decode(encoded_data)`: Esta función toma una cadena de bytes codificada en Base64 como entrada y devuelve los datos binarios decodificados. Por ejemplo, puedes usar `b64encode` para convertir datos binarios en una cadena Base64, y `b64decode` para convertir la cadena Base64 de vuelta a los datos binarios originales [53].

TRAIN_TEST_SPLIT: permite acceder a la función `train_test_split()` de Scikit-learn (sklearn). Esta función es muy útil en el aprendizaje automático para dividir un conjunto de datos en subconjuntos de entrenamiento y prueba. La función `train_test_split()` toma como entrada una o varias matrices que representan características (X) y opcionalmente una matriz que representa las etiquetas (y). También puedes especificar el tamaño del conjunto de prueba, o proporcionar una semilla aleatoria para la división, entre otros parámetros [54].

NYSTROEM: sirve como aproximación del núcleo utilizando el método de Nystroem. Este método puede ser útil en situaciones donde se necesita aproximar la matriz de núcleo para un conjunto de datos grande, lo que puede ser más eficiente computacionalmente que calcular la matriz de núcleo completa [55].

MLPCLASSIFIER: esta clase se utiliza para construir modelos de redes neuronales artificiales para tareas de clasificación. Utiliza la técnica de retropropagación (backpropagation) para entrenar la red neuronal y puede trabajar con datos de entrada tanto de características continuas como categóricas [56].

SVC: esta clase se utiliza para construir modelos de SVM para tareas de clasificación. Los modelos de SVC buscan encontrar un hiperplano en un espacio dimensional superior que separe de la mejor manera posible las diferentes clases de datos. Puedes especificar diversos parámetros, como el tipo de kernel (lineal, polinómico, radial, etc.), la penalización por error de clasificación, el parámetro de regularización C, entre otros [57].

REDUX: biblioteca para administrar el estado actual de la aplicación. Todos los datos que se vayan a procesar sus cambios se verán reflejados en toda la aplicación. Esta herramienta sirve para la gestión de estados en apps, en este desarrollo permite incluir las respuestas de servidor y los datos verificados. Además, se puede gestionar el estado de las interfaces, las rutas activas y los controles de paginación. El fin del uso de esta herramienta es no limitar la gestión de estados cuando la aplicación comience a crecer, especialmente cuando los valores de un componente pueden afectar a los valores de otros componentes, así la gestión se hace de una manera correcta y no se corre el riesgo de perder el control del cuándo, cómo y por qué de los estados. Asimismo, el propósito de Redux es hacer predecibles los cambios de estado, permitiendo incorporar nuevas características a las aplicaciones en desarrollo [58].

WINDOWS: es el sistema operativo informático de Microsoft. Existe desde 1985 y viene preinstalado en muchos PC. Permite interactuar con el ordenador, navegar por Internet, jugar, enviar correos electrónicos, ver vídeos, procesar palabras y mucho más [59].

BOOSTRAP 5.3.3: es un framework de desarrollo web gratuito y de código abierto. Está diseñado para facilitar el proceso de desarrollo de los sitios web responsivos y orientados a los dispositivos móviles, proporcionando una colección de sintaxis para diseños de plantillas [60].

PHYTON: es un lenguaje de programación informático que se utiliza a menudo para crear sitios web y software, automatizar tareas y realizar análisis de datos. Python es un lenguaje de propósito general, lo que significa que se puede utilizar para crear una variedad de programas diferentes y no está especializado en ningún problema específico. Esta versatilidad, junto con su facilidad para los principiantes, lo ha convertido en uno de los lenguajes de programación más utilizados en la actualidad [61].

1. Diagrama Arquitectónico

Arquitectura

El diagrama arquitectónico de la aplicación consta de los siguientes módulos y capas:

- **Interfaz de Usuario (UI):** se compone de la página principal de la aplicación, que permite a los usuarios interactuar con ella, en esta capa de interfaz se hizo uso de flask para manejar las solicitudes HTTP definir las rutas, renderizar más plantillas HTML y las solicitudes entrantes, logrando gestionar la interacción con los usuarios a su vez se implementó en el desarrollo Bootstrap proporcionando estilos CSS y componentes JavaScript para el diseño de la interfaz de usuario.
- **Capa de Modelos:** contiene el modelo de predicción cardiovascular entrenado. El modelo con el kernel que mejor accuracy obtuvo, usando bibliotecas como MLPClassifier y SVC (Support Vector Classifier) de scikit-learn, estas son bibliotecas que se utilizaron para construir y entrenar el modelo de aprendizaje automático.
- **Capa de Servicios (Services):** se encarga de procesar los datos de los controladores y usar el modelo de predicción para calcular el riesgo cardiovascular de una persona. En la capa de servicios, pandas se utilizó para procesar y manipular los datos de entrada antes de alimentarlos a los modelos de predicción. A su vez, las librerías Dill, BytesIO, Base64: en esta capa son útiles para manejar la entrada y salida de datos, especialmente al transmitir datos entre diferentes partes de la aplicación.

Cabe destacar que Dill proporcionó capacidades de serialización avanzadas, permitiendo guardar y cargar objetos Python de manera eficiente. La biblioteca Dill, se la utilizó para guardar el modelo entrenado en este caso se cargó un modelo entrenado en una variable utilizando `dill.dump()` y luego se lo utilizo con `dill.load()`. Esta capacidad de guardar y cargar objetos complejos en Python facilitó el almacenamiento y la reutilización del modelo y otros datos importantes en el proyecto.

Metodología de diseño

La metodología utilizada en el desarrollo de este sistema es CRISP-DM, establece fundamentalmente un modelo de proceso que es de naturaleza jerárquica, comenzando por el nivel fase y luego bajando a tareas genéricas, tareas específicas e instancias de proceso.

CRISP-DM, es una metodología estándar que incluye un modelo y una guía, estructurados en seis fases, algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede volver a una fase anterior para poder revisarla, por lo que la sucesión de fases no tiene por qué ser ordenada desde la primera hasta la última. Presenta una buena adaptabilidad durante todo el ciclo de vida del proyecto.

Si es necesario, la flexibilidad de CRISP-DM le permite volver atrás y concentrarse en cualquier etapa del proceso. Esto no solo evita problemas, sino que también puede ahorrar tiempo al abordar proactivamente cualquier problema o área que necesite atención adicional.

La comunicación efectiva en el proyecto CRISP-DM es fundamental. Los miembros del equipo se comunican constantemente y de manera transparente, lo que garantiza que toda la información relevante se comparta y entienda completamente en todo momento.

Desarrollo del sistema de información

Las siguientes actividades se toman en cuenta durante el proceso de desarrollo del sistema de información.

- Plataforma de recolección de datos replicable y escalable con el objetivo de recopilar datos relevantes sobre la predicción de enfermedades cardiovasculares. El objetivo de esta plataforma es recopilar datos provenientes de varias fuentes, como exámenes clínicos, y factores de riesgo cardiovascular que provee el usuario.
- Se desarrolla un sistema de información integrado para registrar, almacenar y analizar datos relacionados con la salud cardiovascular. Este sistema permitirá generar alertas tempranas de los parámetros relevantes para la predicción de enfermedades cardiovasculares y la gestión eficiente de la información recopilada.

Infraestructura de software y hardware

El desarrollo de la plataforma requerirá la implementación de tecnologías que brinden escalabilidad, seguridad y rendimiento, entre otras características, con el fin de satisfacer las demandas futuras. La Tabla siguiente organiza los elementos a nivel de software y hardware.

Elemento de la infraestructura Software.

TABLA I
ELEMENTO DE LA INFRAESTRUCTURA DE SOFTWARE

ELEMENTO	DESCRIPCIÓN
Software	
Sistema operativo	Windows
Sistema de desarrollo	VSCODE
Backend	
Servidor WEB	Logico
Lenguaje de programación	Python
Frontend	
Maquetación	HTML5
Styles	CSS3
Lenguaje de Programación	Python
Framework de diseño	Bootstrap v5.3.3
Bibliotecas relevantes	Flask, Dill, Scikit-learn
Hardware	
Procesador	Intel i5
Memoria RAM	Mínimo 1GB

Configuración del Entorno de Despliegue

Instalación de IDE VSCODE

- Descargar instalador de VSCODE en:

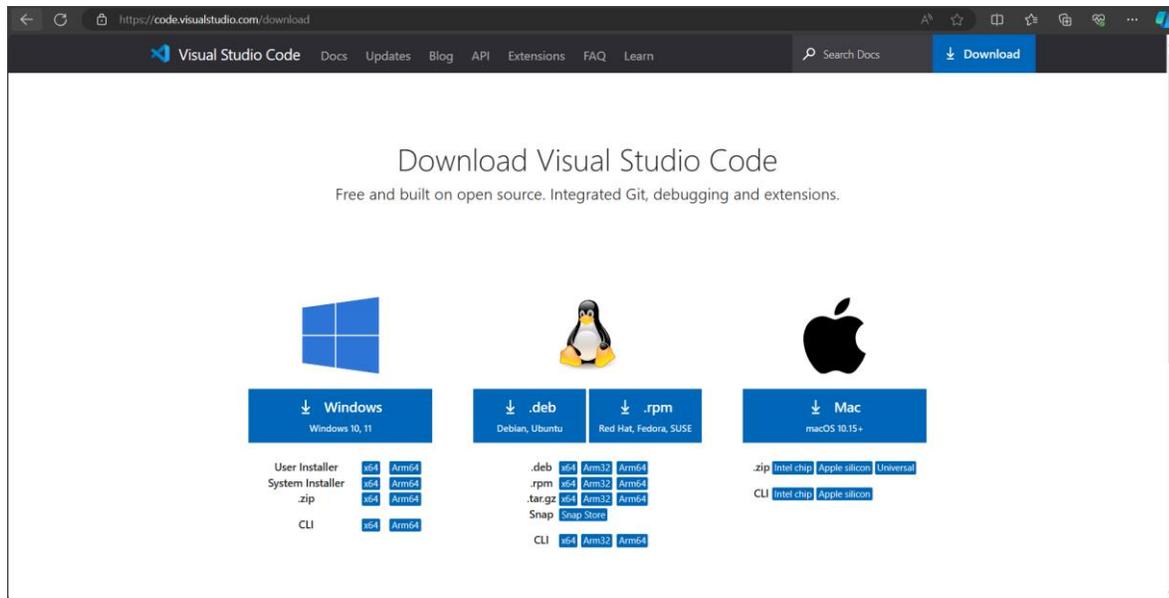


Fig. 1. Descargar VSCODE

- Seguir instrucciones de instalación del ejecutable que previamente se descargó.



Fig. 2. Instalación de VSCODE

1. Configuración del entorno de programación

- Abrir desde el editor el proyecto

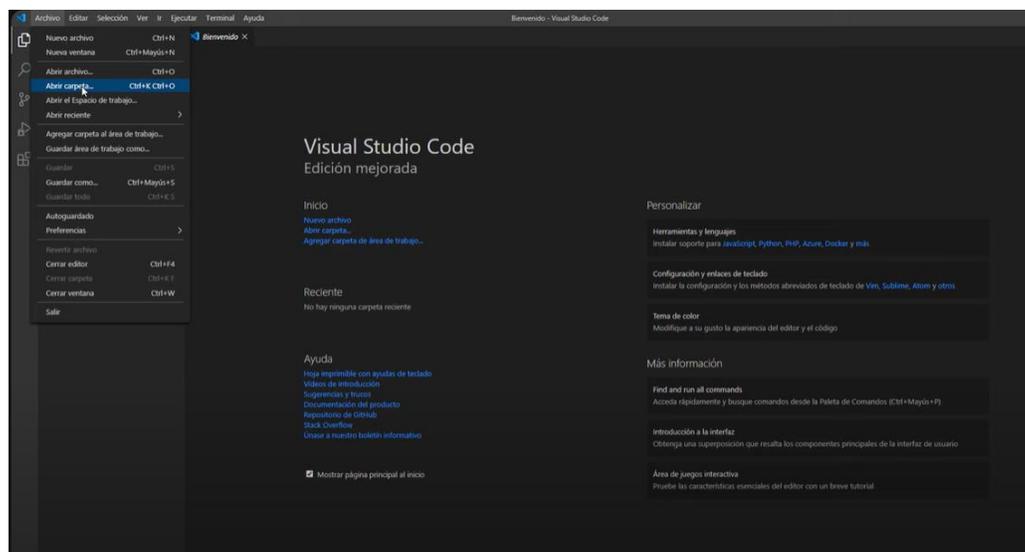


Fig.3. Entorno de programación

- Escoger el intérprete para la ejecución del proyecto

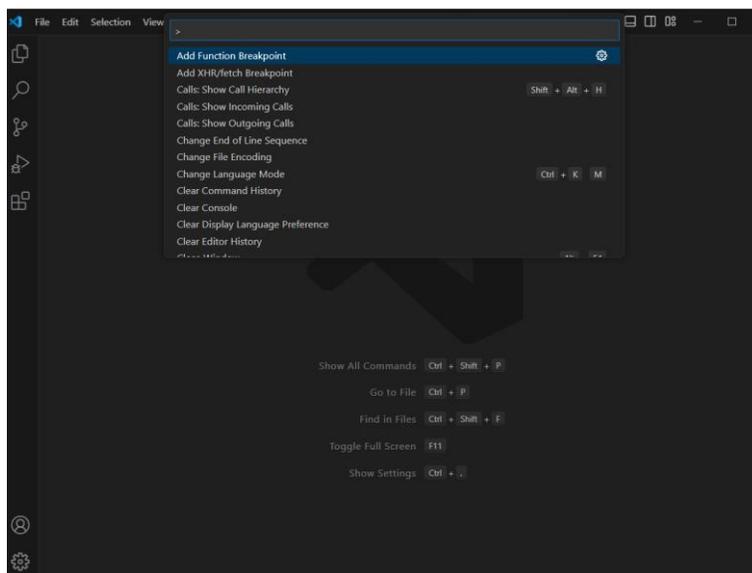


Fig. 4. Seleccionar intérprete de ejecución

2. Participantes

Se describe los participantes e interesados en el desarrollo CRK

Organizaciones Participantes

TABLA II

ORGANIZACIONES PARTICIPANTES

ORG-01	
Organización	Universidad CESMAG Ingeniería de Sistemas
Dirección	Cra. 20a # 14-54, Pasto, Nariño
Teléfono	(602) 7216535 – 7244434
Descripción	Director del programa: Carlos Fernando González Guzmán

Personas participantes

TABLA III
PERSONAS PARTICIPANTES

PART-01	
Participante	Michael Rafael Rodríguez Rodríguez
Teléfono	3219671739
E-mail	mrrodriguez.6789@unicesmag.edu.co
Rol	Desarrollador e investigador

PART-02	
Participante	Claudia Alejandra Delgado Calpa
Teléfono	3157155716
E-mail	cadelgado.9770@unicesmag.edu.co
Rol	Desarrollador e investigador

PART-03	
Participante	Héctor Andrés Mora Paz
Teléfono	3172537641
E-mail	hamora@unicesmag.edu.co
Rol	Director de la investigación

PART-04	
Participante	Jorge Albeiro Rivera Rosero
Teléfono	3148045960
E-mail	Jarivera1@unicesmag.edu.co
Rol	Analista de la investigación

3. Objetivos del sistema

TABLA IV
OBJETIVOS DEL SISTEMA

<i>OBJ-001</i>	Disposición para cargar los datos
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Fuentes:</i>	Grupo de investigación Tecnofilia
<i>Descripción:</i>	Permitir cargar los datos proporcionados por el usuario para procesamiento y análisis.
<i>Sub-objetivos:</i>	Ninguno
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estado:</i>	Implementado
<i>Estabilidad:</i>	Alta
<i>Comentarios:</i>	Ninguno

<i>OBJ-002</i>	Implementar el modelo predictivo desarrollado en la literatura científica
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Fuentes:</i>	Grupo de investigación Tecnofilia
<i>Descripción:</i>	Permitir procesar la información haciendo uso del modelo predictivo para detectar el riesgo de padecer ECV.
<i>Sub-objetivos:</i>	Ninguno

COMPARATIVO DE FUNCIONES KERNEL EN LA PREDICCIÓN DE ECV EN (ANN) Y (SVM)

<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estado:</i>	Implementado
<i>Estabilidad:</i>	Alta
<i>Comentarios:</i>	Ninguno

<i>OBJ-003</i>	Evaluar el riesgo generando alertas tempranas
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Fuentes:</i>	Grupo de investigación Tecnofilia
<i>Descripción:</i>	Evaluar el riesgo generando alertas tempranas frente a riesgos potenciales.
<i>Sub-objetivos:</i>	Ninguno
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estado:</i>	Implementado
<i>Estabilidad:</i>	Alta
<i>Comentarios:</i>	Ninguno

<i>OBJ-004</i>	Interacción rápida y sencilla con la herramienta
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Fuentes:</i>	Grupo de investigación Tecnofilia

COMPARATIVO DE FUNCIONES KERNEL EN LA PREDICCIÓN DE ECV EN (ANN) Y (SVM)

<i>Descripción:</i>	Permitir una interacción de fácil manejo para el usuario por medio de sugerencias e instrucciones.
<i>Sub-objetivos:</i>	Ninguno
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estado:</i>	Implementado
<i>Estabilidad:</i>	Alta
<i>Comentarios:</i>	Ninguno

<i>OBJ-005</i>	Visualizar los diversos resultados generados en la herramienta
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Fuentes:</i>	Grupo de investigación Tecnofilia
<i>Descripción:</i>	Disponer de visualizaciones mediante gráficos estadísticos que permitan interpretar rápida y eficientemente los resultados.
<i>Sub-objetivos:</i>	Ninguno
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estado:</i>	Implementado
<i>Estabilidad:</i>	Alta
<i>Comentarios:</i>	Ninguno

4. Catálogo de requisito del sistema

11.1 Requisitos funcionales

TABLA V
REQUISITOS FUNCIONALES

<i>RF-01</i>	Visualizar instrucciones
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Descripción:</i>	La aplicación web debe mostrar al usuario las instrucciones de uso del formulario y su entorno virtual.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

<i>RF-02</i>	Ingresar los datos
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Descripción:</i>	La aplicación web debe permitir recolectar los datos de aquellos factores de riesgo incluidas en las medidas antropométricas
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata

COMPARATIVO DE FUNCIONES KERNEL EN LA PREDICCIÓN DE ECV EN (ANN) Y (SVM)

<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

<i>RF-03</i>	Visualizar el porcentaje de RCV
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Descripción:</i>	La aplicación web debe indicar al usuario el riesgo de presentar alguna ECV con su porcentaje
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

<i>RF-04</i>	Visualizar de alertas tempranas
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Descripción:</i>	La aplicación web debe permitir la visualización de las alertas presentadas en cada factor de riesgo, con recomendaciones útiles para su implementación.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta

<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

<i>RF-05</i>	Módulo de sugerencias prácticas
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Descripción:</i>	La aplicación web debe proporcionar indicaciones prácticas para la prevención de ECV
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

11.2 Requisitos no funcionales

TABLA VI
REQUISITOS NO FUNCIONALES

<i>RNF-01</i>	Usabilidad
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Descripción:</i>	La aplicación debe permitir ser ejecutada en distintos dispositivos.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata

COMPARATIVO DE FUNCIONES KERNEL EN LA PREDICCIÓN DE ECV EN (ANN) Y (SVM)

<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

<i>RNF-02</i>	Eficiencia
<i>Versión:</i>	1.0 (01-04-2023)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero
<i>Descripción:</i>	La aplicación debe tener tiempos de ejecución en el cálculo de RCV no mayor a 5 segundos en promedio.
<i>Importancia:</i>	Esencial
<i>Urgencia:</i>	Inmediata
<i>Estabilidad:</i>	Alta
<i>Estado:</i>	Implementado
<i>Comentarios:</i>	Ninguno

5. Diagrama General de Casos de Uso

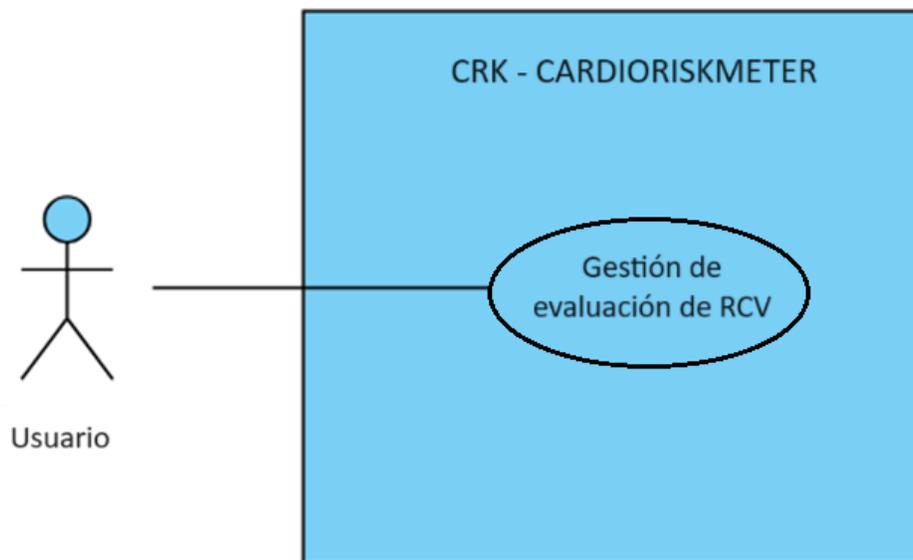


Fig. 5. Diagrama general de casos de uso

12.1 Definición de Actores

TABLA VII
DEFINICIÓN DE ACTORES

<i>AC-001</i>	Usuario
<i>Versión:</i>	1.0 (01-04-2024)
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa
<i>Fuentes:</i>	Grupo de investigación Tecnofilia
<i>Descripción:</i>	Usuario el cual pueden interactuar con el entorno virtual que provee la herramienta web.
<i>Comentarios:</i>	Ninguno

12.2 Descripción de Casos de Uso

12.2.1 Caso de Uso Gestión de evaluación de RCV

TABLA VIII
DESCRIPCIÓN DE CASO DE USO

<i>CU-001</i>	Gestión de evaluación de RCV																				
<i>Versión:</i>	1.0 (01-04-2024)																				
<i>Autores:</i>	Michael Rafael Rodríguez Rodríguez Claudia Alejandra Delgado Calpa Héctor Andrés Mora Paz Jorge Albeiro Rivera Rosero																				
<i>Fuentes:</i>	Grupo de investigación Tecnofilia																				
<i>Descripción:</i>	El sistema deberá comportarse como se describe en el siguiente caso de uso cuando un usuario decida acceder a la aplicación.																				
<i>Actores:</i>	Usuario																				
<i>Precondición:</i>	El sistema debe estar disponible y funcional																				
<i>Secuencia normal:</i>	<table border="0"> <thead> <tr> <th style="text-align: left;">Paso</th> <th style="text-align: left;">Acción</th> </tr> </thead> <tbody> <tr> <td>1.</td> <td>El sistema muestra la página principal</td> </tr> <tr> <td>2.</td> <td>El sistema muestra las instrucciones</td> </tr> <tr> <td>3.</td> <td>El usuario ingresa al módulo formulario</td> </tr> <tr> <td>4.</td> <td>El usuario completa cada uno de los campos del formulario</td> </tr> <tr> <td>5.</td> <td>El usuario da clic en el botón evaluar riesgo</td> </tr> <tr> <td>6.</td> <td>El sistema muestra una gráfica con la información del riesgo cardiovascular</td> </tr> <tr> <td>7.</td> <td>El sistema muestra a su vez alertas tempranas según los factores de riesgo alterados.</td> </tr> <tr> <td>8.</td> <td>El sistema muestra otro módulo de información saludable</td> </tr> <tr> <td>9.</td> <td>El usuario ingresa a ese módulo y puede observar recomendaciones útiles para llevar</td> </tr> </tbody> </table>	Paso	Acción	1.	El sistema muestra la página principal	2.	El sistema muestra las instrucciones	3.	El usuario ingresa al módulo formulario	4.	El usuario completa cada uno de los campos del formulario	5.	El usuario da clic en el botón evaluar riesgo	6.	El sistema muestra una gráfica con la información del riesgo cardiovascular	7.	El sistema muestra a su vez alertas tempranas según los factores de riesgo alterados.	8.	El sistema muestra otro módulo de información saludable	9.	El usuario ingresa a ese módulo y puede observar recomendaciones útiles para llevar
Paso	Acción																				
1.	El sistema muestra la página principal																				
2.	El sistema muestra las instrucciones																				
3.	El usuario ingresa al módulo formulario																				
4.	El usuario completa cada uno de los campos del formulario																				
5.	El usuario da clic en el botón evaluar riesgo																				
6.	El sistema muestra una gráfica con la información del riesgo cardiovascular																				
7.	El sistema muestra a su vez alertas tempranas según los factores de riesgo alterados.																				
8.	El sistema muestra otro módulo de información saludable																				
9.	El usuario ingresa a ese módulo y puede observar recomendaciones útiles para llevar																				

	<p>una vida saludable con recomendación de expertos.</p> <p>10. El usuario puede nuevamente regresar al escenario principal.</p>
<i>Post condición:</i>	<p>La información ingresada por el usuario se registra correctamente en el sistema, y se muestran las gráficas, y por consiguiente las alertas tempranas según los factores de riesgo alterados, logrando así acceder a los diferentes módulos si requiere más información.</p>
<i>Excepciones:</i>	<p>Paso Acción</p> <p>1. Si el sistema no puede mostrar la página principal o las instrucciones, se debería mostrar un mensaje de error.</p> <p>4. Si el usuario no completa todos los campos del formulario, el sistema muestra un mensaje indicando los campos faltantes.</p> <p>7. Si hay algún error en la evaluación del riesgo cardiovascular o en la generación de las alertas tempranas, se debería notificar al usuario.</p>
<i>Comentarios:</i>	<p>Ninguno</p>

13. Protocolo de pruebas e implementación del modelo

Para este desarrollo de este protocolo de implementación se realiza la creación del modelo con el kernel Canberra, entrenado previamente.

```
svc=K SVC(kernel='can', gamma=0.0006951927961775605,C=2154.4346,probability=True)
modelo_predic2 = svc.fit(Xtrain, ytrain)
print('cmlfc',svc.score(Xtest,ytest))

cmlfc 0.712
```

Fig. 6. Modelo con kernel Canberra

Se crea un archivo con permisos de escritura para guardar el modelo que hemos creado previamente, a su vez instalamos las librerías necesarias y se procede a guardar el modelo de esta manera modelo_predic2, en el archivo de escritura que se ha creado **¡Error! No se encuentra el origen de la referencia..**

```
✓ creación del modelo
[ ] # Creamos un archivo con permisos de escritura
    archivo_modelo = open("Modelo_predictivo.sav", "wb")

[ ] !pip install dill

collecting dill
  Downloading dill-0.3.8-py3-none-any.whl (116 kB)
    116.3/116.3 kB 2.5 MB/s eta 0:00:00
Installing collected packages: dill
Successfully installed dill-0.3.8

# Con la función dump guardamos el modelo que se encuentra en modelo_predic2, en el archivo_modelo
from pickle import dump
import dill

dill.dump(modelo_predic2, archivo_modelo)
```

Fig. 7. Creación del modelo

Se importa nuevos datos para probar este modelo que ha sido entrenado como se muestra en la Fig.8.

```
# importamos los nuevos datos a analizar con el modelo

dfx = pd.read_csv('/content/drive/MyDrive/noveno semestre tareas/Nuevos.csv', delimiter=";")
dfx.shape
dfx
```

	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active
0	18393	2	168	62.0	110	80	1	1	0	0	1
1	20228	1	156	85.0	140	90	3	1	0	0	1
2	18857	1	165	64.0	130	70	3	1	0	0	0
3	17623	2	169	82.0	150	100	1	1	0	0	1
4	17474	1	156	56.0	100	60	1	1	0	0	0

```
# X_nuevos es una nueva variable que ahora contiene estos valores en forma de un arreglo NumPy. para poder trabajar mejor con los datos
X_nuevos = dfx.values
```

Fig. 8. Datos entrenados

Se carga el modelo en una variable para poder trabajar con los nuevos datos (X_nuevos), y se obtiene los resultados que nos otorga el modelo, comprobando su funcionamiento **¡Error! No se encuentra el origen de la referencia..**

```
# Cargamos el modelo guardado utilizando la función load

from pickle import load

archivo_modelo22 = 'Modelo_predictivo.sav'

with open(archivo_modelo22, 'rb') as f:
    modelo_cargado = dill.load(f)

# Realizamos la predicción con los nuevos datos.

prediccion_nuevos = modelo_cargado.predict(X_nuevos)
prediccion_nuevos

array([0, 1, 1, 1, 0])
```

Fig. 9. Resultados nuevos datos

14. Desarrollo de la aplicación

Para este desarrollo (Back End), se utilizó un entorno de ejecución VSCODE, basado en el lenguaje de programación Python, en la **¡Error! No se encuentra el origen de la referencia.** se evidencia la estructura de carpetas de creación de la aplicación.

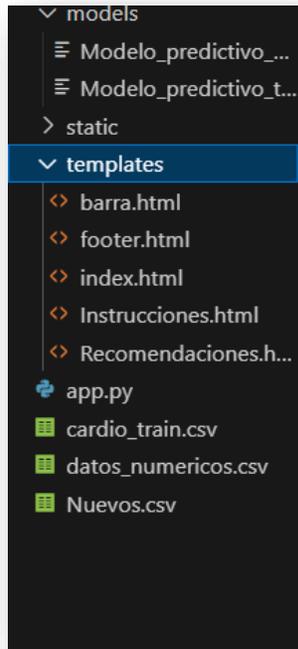


Fig. 10. Estructura de carpetas

15. Pruebas de campo

Después de terminada la aplicación se realizaron pruebas con el equipo haciendo uso de un servidor local, supervisando de manera remota la conectividad.

La plataforma por el momento esta alojada de manera local <https://0jwqv74-5000.use.devtnnls.ms/> como se muestra en la **¡Error! No se encuentra el origen de la referencia.** aquí es donde se realizó el testeo, logrando una implementación eficiente y estable lo cual permite la recepción de los campos que otorga el formulario para su posterior análisis.

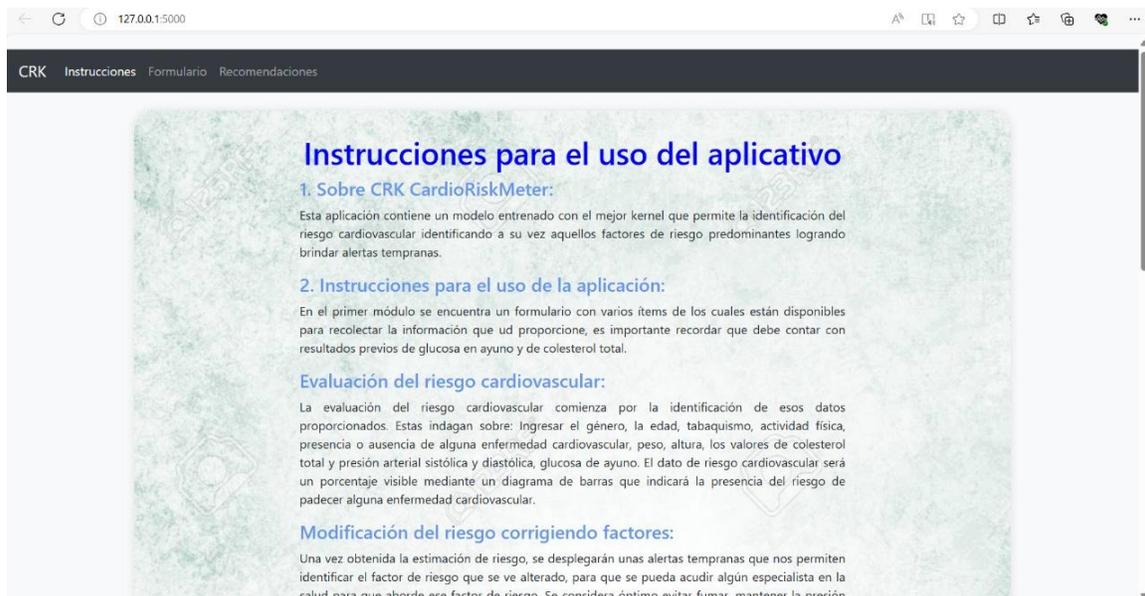


Fig. 11. Módulo de instrucciones

16. Ejecución de la aplicación

El módulo formulario permite llevar los registros necesarios para lograr obtener información que permita la predicción del riesgo cardiovascular. Para llevar a cabo este proceso se cuenta con una serie de campos para el registro de la información requerida y se puede evidenciar los resultados mediante unas gráficas y alertas tempranas, donde la idea de esto es entender el riesgo cardiovascular que presente el usuario y alertar sobre los factores de riesgo que se pueden modificar.

COMPARATIVO DE FUNCIONES KERNEL EN LA PREDICCIÓN DE ECV EN (ANN) Y (SVM)

The figure displays three screenshots of a mobile application interface for cardiovascular risk prediction. The first screenshot shows a data entry form with fields for Age, Gender, Height, Weight, Systolic Pressure, Diastolic Pressure, Cholesterol, Glucose, Smoke, and Alcohol. The second screenshot shows the same form with a 'Calcular' button and a 3D heart model. The third screenshot shows an 'Instrucciones' (Instructions) screen with a diagram of a person sitting and having their blood pressure measured, with numbered steps: 1. No conversar, 2. Apoyar el brazo a la altura del corazón, 3. Colocar el manguito en el brazo sin ropa, 4. Usar el tamaño de manguito adecuado. Logos for OPS and HEARTS are visible.

Fig. 12. Formulario

Una vez llenos estos campos en la misma interfaz de formulario se obtiene un diagrama de barra con la información en porcentaje de presentar una enfermedad cardiovascular, generando a su vez alertas tempranas en la misma interfaz de formulario sobre los factores de riesgo a modificar.

COMPARATIVO DE FUNCIONES KERNEL EN LA PREDICCIÓN DE ECV EN (ANN) Y (SVM)

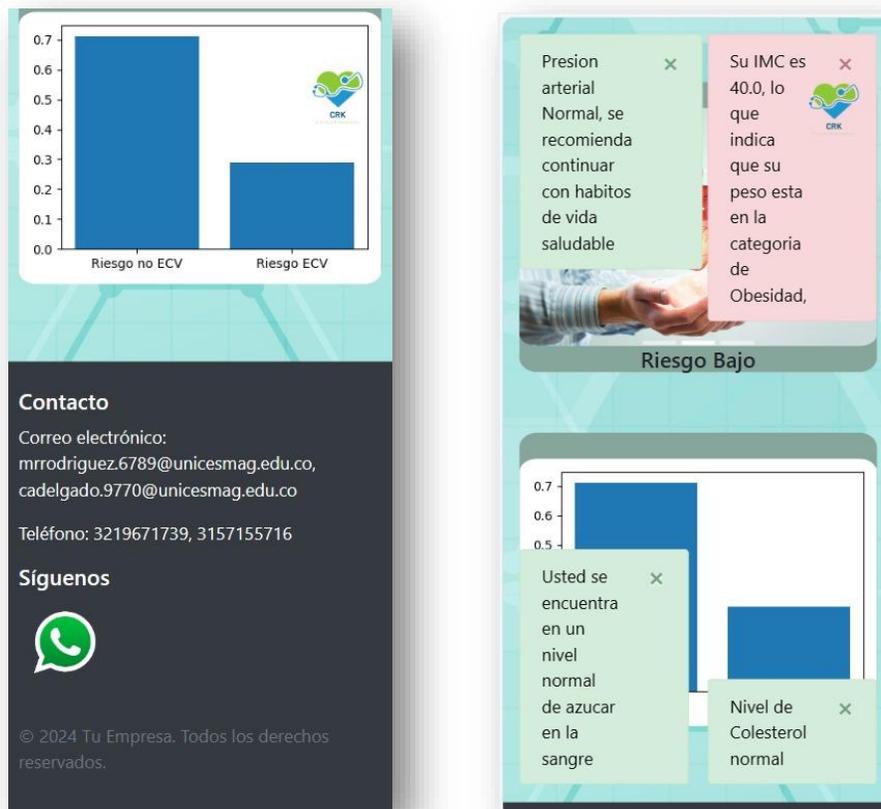


Fig. 13. Resultados

Anexo B: Manual del Usuario



**CRK CardioRiskMeter / Manual del
Usuario**

**Michael Rodríguez
Claudia Delgado
Héctor Mora
Jorge Rivera**

2024

Introducción

Este manual tiene como objetivo servir de guía para el uso del aplicativo CRK-WEB. Se ha creado con la intención de proporcionar al usuario una comprensión y manejo adecuados de la aplicación, facilitando la navegación a través de sus funciones mediante imágenes intuitivas para una mejor comprensión. Además, este aplicativo permite la evaluación del modelo del kernel óptimo propuesto.

CRK-WEB es un Sistema de Información que funciona en la web, y se adapta a diferentes dispositivos por su diseño responsive. Este aplicativo permite predecir el riesgo de padecer enfermedades cardiovasculares, generar alertas tempranas en caso de que se identifiquen factores de riesgo potenciales, además ofrece recomendaciones útiles sobre como promover un estilo de vida saludable.

La herramienta permite conocer el riesgo cardiovascular mostrando el porcentaje, siendo visible mediante un diagrama de barras que indicará la presencia del riesgo de padecer alguna enfermedad cardiovascular.

Una vez obtenida la estimación de riesgo, se despliegan unas alertas tempranas que permiten identificar el factor de riesgo que se ve alterado, para que se pueda acudir algún especialista en la salud de manera oportuna para que aborde ese factor de riesgo. Además, permite visualizar recomendaciones útiles proporcionadas por la OMS sobre hábitos y estilos de vida saludable.

La herramienta puede ser utilizada por cualquier persona que quiera conocer su riesgo de padecer alguna enfermedad cardiovascular esta herramienta permite de manera interactiva y fácil conocer el estado de salud, factores que pueda modificar, y evitar a futuro enfermedades que la comprometan, está disponible para todos aquellos usuarios que tengan interés en utilizarla sean como apoyo al diagnóstico médico, o por autonomía propia.

Aplicativo CRK-WEB

CRK-WEB, es desarrollado como aplicativo de despliegue para consumir el modelo creado con el kernel que mejor precisión y comportamiento obtuvo con el tipo de datos implementados, en este caso el mejor accuracy se obtuvo para máquinas de soporte vectorial. Este sistema permite optimizar procesos de salud para cualquier tipo de población sin padecer alguna enfermedad cardiovascular, en el desarrollo de este sistema, se abarcó todo el proceso que va desde la fase de consultar la recomendación de expertos, creación de nuevos kernel, arquitectura, desarrollo hasta la fase de pruebas. Además, se utilizó la misma metodología que se implementó en el desarrollo de la investigación, que fue la que más se adaptó al proyecto, utilizando así una metodología ágil y moderna.

En la implementación de esta herramienta, se desarrolló dos módulos importantes, uno de recolección de información por medio de un formulario, y el otro de recomendaciones útiles para la prevención de ECV, que de cierta manera permiten ejecutar el objetivo del proyecto.

El objetivo del desarrollo de CRK-WEB es, como se mencionó anteriormente, consumir el modelo predictivo, con el fin de optimizar procesos, y presentar un diseño visualmente agradable para el usuario. En el desarrollo de este aplicativo se utilizan herramientas que permiten la ejecución de los procesos que llevarán a cumplir el objetivo final propuesto, además comprender la estructura del sistema y lo que implica la creación del modelo haciendo uso del kernel, exige el uso de esta herramienta para proporcionar el despliegue que perfectamente se adhiere a la etapa que provee la metodología.

Se hace énfasis en la metodología, ya que para este proyecto no se implementó una metodología aparte para el desarrollo del aplicativo, sino que se consolidó en una de las etapas de la metodología del desarrollo del proyecto, la etapa de despliegue dando cumplimiento con el último objetivo propuesto, generando resultados óptimos y eficaces que estén a disposición incluso para

nuevas investigaciones.

El aplicativo webg permitirá visualizar y administrar la información relevante para los usuarios que hagan uso de la plataforma.

Acceso a la aplicación

CRK-WEB, Para acceder a la aplicación y lograr conocer su riesgo cardiovascular y por consiguiente consumir el modelo creado se dispondrá de un enlace web allí encontrará la página de inicio con la primera información instrucciones, esto hace parte del módulo de instrucciones.

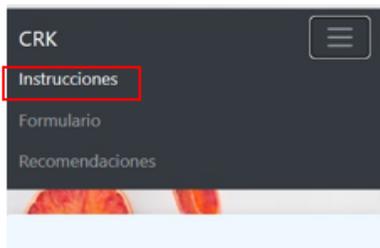


Fig. 2. Menú instrucciones

CRK

Instrucciones para el uso del aplicativo

1. Sobre CRK CardioRiskMeter:

Esta aplicación contiene un modelo entrenado con el mejor kernel que permite la identificación del riesgo cardiovascular identificando a su vez aquellos factores de riesgo predominantes logrando brindar alertas tempranas.

2. Instrucciones para el uso de la aplicación:

En el primer módulo se encuentra un formulario con varios ítems de los cuales están disponibles para recolectar la información que ud proporcione, es importante recordar que debe contar con resultados previos de glucosa en ayuno y

Evaluación del riesgo cardiovascular:

La evaluación del riesgo cardiovascular comienza por la identificación de esos datos proporcionados. Estas indagán sobre: Ingresar el género, la edad, tabaquismo, actividad física, presencia o ausencia de alguna enfermedad cardiovascular, peso, altura, los valores de colesterol total y presión arterial sistólica y diastólica, glucosa de ayuno. El dato de riesgo cardiovascular será un porcentaje visible mediante un diagrama de barras que indicará la presencia del riesgo de padecer alguna enfermedad cardiovascular.

Modificación del riesgo corrigiendo factores:

Una vez obtenida la estimación de riesgo, se desplegarán unas alertas tempranas que nos permiten identificar el factor de riesgo que se ve alterado, para que se pueda acudir algún especialista, en la salud para que aborde ese factor de riesgo. Se considera óptimo evitar fumar.

Ruta de Atención Clínica:

La Ruta de Atención Clínica es fundamental para la implementación efectiva de la Iniciativa CRK(CardioRiskMeter), mejorando la gestión clínica y autoconocimiento. Además, capacita al usuario y a su familia al educarlos sobre el manejo completo de los factores de riesgo cardiovascular. Esta aplicación abarca la ruta de atención clínica para el manejo integral del riesgo cardiovascular.

3. Acerca de CRK(CardioRiskMeter):

CRK es una iniciativa que surge de un proyecto de grado dirigida a fortalecer los sistemas de atención primaria de salud. Su enfoque se centra en mejorar la prevención y el control de las enfermedades cardiovasculares y sus factores de riesgo. Esta iniciativa tiene como meta fomentar un proceso de mejora continua en la calidad de atención. Esto se logra mediante un mejor control de la hipertensión y la implementación de un enfoque integral en el manejo del riesgo cardiovascular.

4. Uso:

En ninguna circunstancia esta aplicación pretende sustituir la consulta con un profesional de la salud ni su criterio clínico. Su objetivo es proporcionar un apoyo al diagnóstico médico, una herramienta para evaluar rápidamente el riesgo cardiovascular y facilitar la comunicación con los usuarios sobre las posibles modificaciones. Además, busca asistir a las personas preocupadas por su salud, ayudándoles a determinar cuándo es necesario buscar atención médica si su riesgo no es bajo. Las recomendaciones en hábitos de vida saludables están descritas a nivel de información y no deben ser utilizadas como guía sin la supervisión de un personal de salud, ya que esto podría resultar peligroso.

Contacto
 Correo electrónico: mirroldguzar5770@unicesmag.edu.co, cardelgado5770@unicesmag.edu.co
 Teléfono: 3219671739, 3157155716

Fig. 1. Instrucciones

Módulo de formulario

Para acceder a una de las funcionalidades de evaluación del riesgo procedemos rellenar los campos que aparecen en el módulo de formulario. Escribimos edad (age), género (gender), peso (weight), talla (height) se colocará en centímetros, presión sistólica (systolic pressure), presión diastólica (diastolic pressure) para esta medida se opta por instrucciones adicionales con un mensaje de despliegue de información que son útiles al proporcionar dicho valor, colesterol (cholesterol), glucosa (glucosa), en estos campos colocamos manualmente los datos, si es fumador (smoke), consume alcohol, si realiza actividad física (active), la opción será seleccionar mediante el menú desplegable la opción sí o no y damos clic en calcular, como se muestra en la Fig.2.

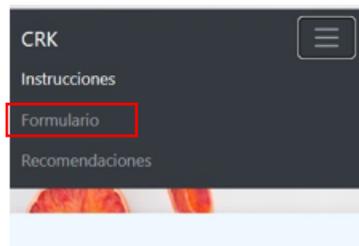


Fig. 4. Menú formulario

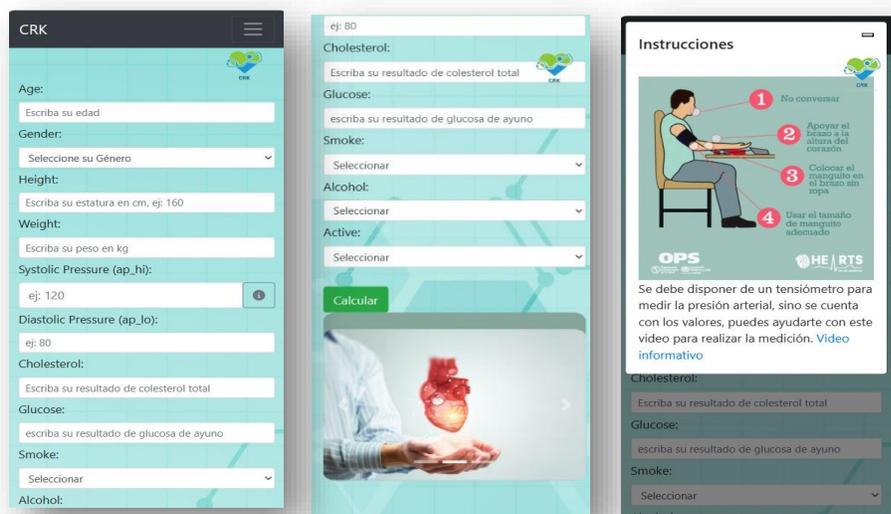


Fig. 3. Formulario

A continuación, en la misma interfaz de formulario, en la Fig.3 , se presenta el resultado obtenido a partir de los datos ingresados, se puede observar una gráfica con una barra con el porcentaje de riesgo de padecer alguna enfermedad cardiovascular y otra barra con el porcentaje de no padecer riesgo cardiovascular.

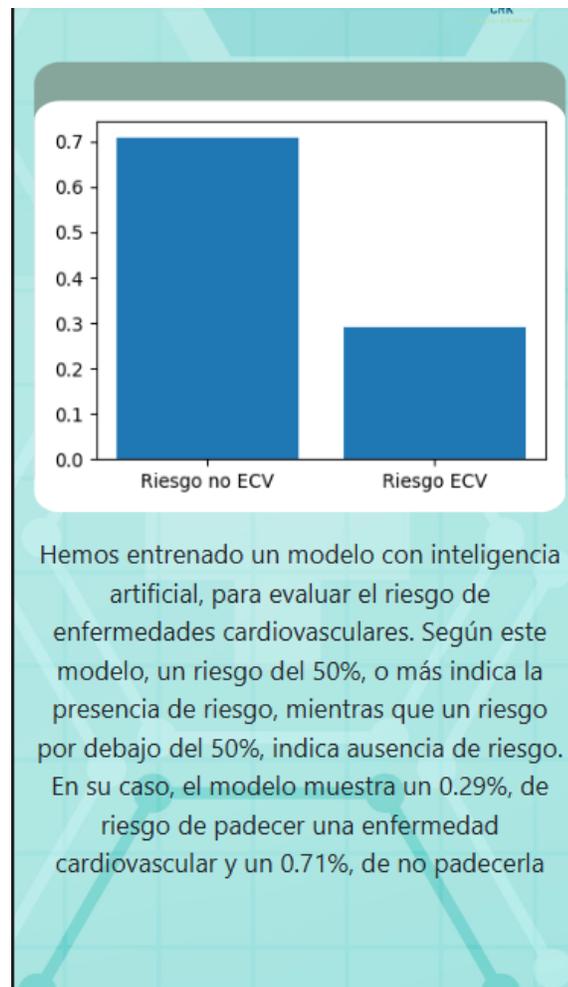


Fig. 5. Porcentaje de RCV

Alertas tempranas

Una vez obtenida la evaluación del riesgo, en la misma interfaz de formulario se genera unas alertas tempranas como se muestra en la **¡Error! No se encuentra el origen de la referencia.** de los factores que se pueden modificar con una breve interpretación de los resultados de factores como lo

son índice de masa corporal, presión arterial, colesterol, glucosa, tabaquismo, alcohol, actividad física, sugiriendo así la vía clínica para el caso de que el usuario quiera conocer más sobre su salud.

Aparte de ser un apoyo al diagnóstico médico, permite instruir al usuario sobre el manejo integral de sus factores de riesgo cardiovascular.

Se ha identificado rangos normales, según la OMS, de factores como el índice de masa corporal, este un parámetro pronóstico relevante. Se ha preferido facilitar su cálculo (requiere ingresar el peso y la altura), para luego poder compararlo con los parámetros recomendados. Se considera un IMC normal cuando sus valores se encuentran entre 18,5 y 24,9; sobrepeso, entre 25 y 29,9 y obesidad cuando es igual o mayor a 30 kg/m².

A su vez, otro factor importante es mantener los valores de presión arterial sistólica por debajo de 140 mmHg. Es deseable también que los valores de colesterol se encuentren por debajo de 200 mg/dl, pero, en pacientes de alto riesgo, por lo general estos valores deberían ser mucho más bajos y para la glucosa se tiene en cuenta los valores que nos proporciona la sociedad americana de la diabetes que nos dice que la diabetes se diagnostica cuando el nivel de glucosa (azúcar) en sangre en ayunas es superior o igual a 126 mg/dl



Fig. 6. Alertas tempranas

Recomendaciones útiles para una vida saludable

En este nuevo módulo vamos a observar la información adecuada a nivel general sobre nutrición, actividad física, tabaquismo, consumo de alcohol, cabe resaltar que la información que se destaca en este módulo fue tomada de la organización mundial de la salud, recomendaciones de la asociación americana de diabetes y la asociación americana del corazón.

CRK

Instrucciones

Formulario

Recomendaciones

Consejos útiles para mantener una dieta Saludable.

Una alimentación saludable se basa en una dieta equilibrada y variada. Esto significa que debemos consumir alimentos de todos los grupos principales en las cantidades adecuadas. La OMS

EFECTOS SOBRE EL ORGANISMO:

problemas del sistema nervioso central
pérdida de memoria
trastornos psicológicos

problemas de piel

pérdida del apetito y deficiencia vitamínica

obesidad

impotencia sexual

mala digestión de alimentos

cirrosis hepática

pancreatitis

cáncer de labio, de boca, de esófago y de hígado

EFECTOS ADVERSOS DEL ALCOHOL

Contacto

Correo electrónico:
mrrodriguez.6789@unicesmag.edu.co,
cadelgado.9770@unicesmag.edu.co

Fig. 7. Recomendaciones

 <p>UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)	CÓDIGO: AAC-BL-FR-032
		VERSIÓN: 1
		FECHA: 09/JUN/2022

San Juan de Pasto, 3 de septiembre del 2024

Biblioteca
REMIGIO FIORE FORTEZZA OFM. CAP.
Universidad CESMAG
Pasto

Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominado **Comparativo de funciones kernel en la predicción de enfermedades cardiovasculares en Redes Neuronales artificiales (ANN) y Máquinas de Soporte Vectorial (SVM)**, presentado por el (los) autor(es) **Claudia Alejandra Delgado Calpa y Michael Rafael Rodríguez Rodríguez** del Programa Académico Ingeniería de Sistemas al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesor(a), que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,



(Firma del Asesor)

HÉCTOR ANDRÉS MORA PAZ
1085251119
Ingeniería de sistemas
317 2537641
hamora@unicesmag.edu.co



UNIVERSIDAD
CESMAG

NIT: 800.109.387-7
VIGILADA MIMEDUCACIÓN

**CARTA DE ENTREGA TRABAJO DE GRADO O
TRABAJO DE APLICACIÓN – ASESOR(A)**

CÓDIGO: AAC-BL-FR-032

VERSIÓN: 1

FECHA: 09/JUN/2022

 UNIVERSIDAD CESMAG <small>NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</small>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

INFORMACIÓN DEL (LOS) AUTOR(ES)	
Nombres y apellidos del autor: Claudia Alejandra Delgado Calpa	Documento de identidad: 1085309770
Correo electrónico: cadelgado.9770@unicesmag.edu.co	Número de contacto: 3157155716
Nombres y apellidos del autor: Michael Rafael Rodríguez Rodríguez	Documento de identidad: 1086896789
Correo electrónico: mrrodriguez.6789@unicesmag.edu.co	Número de contacto: 3219671739
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del autor:	Documento de identidad:
Correo electrónico:	Número de contacto:
Nombres y apellidos del asesor: Héctor Andrés Mora Paz	Documento de identidad: 1085251119
Correo electrónico: hamora@unicesmag.edu.co	Número de contacto: 3172537641
Título del trabajo de grado: Comparativo de funciones kernel en la predicción de enfermedades cardiovasculares en Redes Neuronales artificiales (ANN) y Máquinas de Soporte Vectorial (SVM)	
Facultad y Programa Académico: Ingeniería – Ingeniería de sistemas	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

- a) La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.

 <p>UNIVERSIDAD CESMAG NIT: 800.109.387-7 VIGILADA MINEDUCACIÓN</p>	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

- b) Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.
- c) Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación, distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.
- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
- e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

NOTA: En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permiso(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG, por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 3 días del mes de septiembre del año 2024

<p><i>Michael Rodriguez</i> Firma del autor</p>	<p><i>Claudia Delgado</i> Firma del autor</p>
Nombre del autor:	Nombre del autor:
Firma del autor	Firma del autor
Nombre del autor:	Nombre del autor:
Firma del asesor	



UNIVERSIDAD
CESMAG
NIT: 800.109.387-7
VIGILADA MINEDUCACIÓN

AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL

CÓDIGO: AAC-BL-FR-031

VERSIÓN: 1

FECHA: 09/JUN/2022


Nombre del asesor.