



UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
ESCUELA DE POSGRADO

MAESTRÍA EN ESTADÍSTICA

TESIS

MÁQUINA DE SOPORTE VECTORIAL EN EL ESTUDIO DE FACTORES DE
EMBARAZO ADOLESCENTE DEL DISTRITO DE SANTA ANA, 2022

PARA OPTAR AL GRADO ACADÉMICO DE MAESTRO EN ESTADÍSTICA

AUTOR:

Br. PASCUAL BARCENA CARRASCO

ASESOR:

Dr. CLETO DE LA TORRE DUEÑAS

CÓDIGO ORCID:

[0000-0003-0921-7217](https://orcid.org/0000-0003-0921-7217)

CUSCO – PERÚ

2024

INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro. CU-303-2020-UNSAAC)

El que suscribe, asesor del trabajo de investigación/tesis titulada: **MÁQUINA DE SOPORTE VECTORIAL EN EL ESTUDIO DE FACTORES DE EMBARAZO ADOLESCENTE DEL DISTRITO DE SANTA ANA, 2022** presentado por: **Br. BARCENA CARRASCO PASCUAL** con N° de DNI: 42477510 Para optar el título profesional/grado académico de: **MAESTRO EN ESTADISTICA.**

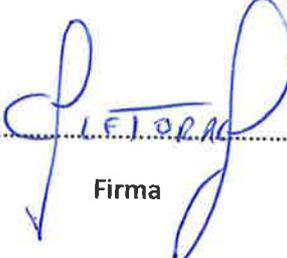
Informo que el trabajo de investigación ha sido sometido a revisión por 3 Veces, mediante Software Antiplagio, conforme al Art. 6° del Reglamento para Uso de Sistema Antiplagio de la UNSAAC y de la evaluación de originalidad se tiene un porcentaje de...9.....

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1% al 10%	No se considera plagio.	X
Del 11% al 30%	Devolver al usuario para las correcciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que corresponden de acuerdo a Ley.	

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y **adjunto** la primera hoja del reporte del Sistema Antiplagio.

Cusco, 08 de...Enero.....de 2025


.....
Firma

Post firma...Dr CLETO DE LA TORRE DUEÑAS

N° de DNI.....23988416

ORCID del asesor.....<https://orcid.org/0000-0003-0921-7217>

Se adjunta:

1. Reporte generado por el sistema antiplagio.
2. Enlace del reporte generado por el sistema antiplagio... oid: 27259:419799507

PASCUAL BÁRCENA CARRASCO

MÁQUINA DE SOPORTE VECTORIAL EN EL ESTUDIO DE FACTORES DE EMBARAZO ADOLESCENTE DEL DISTRITO DE S...

 Universidad Nacional San Antonio Abad del Cusco

Detalles del documento

Identificador de la entrega

trn:oid:::27259:419799507

131 Páginas

Fecha de entrega

8 ene 2025, 11:22 a.m. GMT-5

18,883 Palabras

Fecha de descarga

8 ene 2025, 12:09 p.m. GMT-5

105,483 Caracteres

Nombre de archivo

En Limpio tesis pascual 19.docx

Tamaño de archivo

26.7 MB

9% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- ▶ Bibliografía
- ▶ Texto citado
- ▶ Texto mencionado
- ▶ Coincidencias menores (menos de 8 palabras)
- ▶ Trabajos entregados

Exclusiones

- ▶ N.º de coincidencias excluidas

Fuentes principales

- 9%  Fuentes de Internet
- 3%  Publicaciones
- 0%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alerta de integridad para revisión

-  **Texto oculto**
1 caracteres sospechosos en N.º de página
El texto es alterado para mezclarse con el fondo blanco del documento.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.



UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
ESCUELA DE POSGRADO

INFORME DE LEVANTAMIENTO DE OBSERVACIONES A TESIS

Dra. NELLY AYDE CAVERO TORRE, Directora (e) General de la Escuela de Posgrado, nos dirigimos a usted en condición de integrantes del jurado evaluador de la tesis intitulada **MÁQUINA DE SOPORTE VECTORIAL EN EL ESTUDIO DE FACTORES DE EMBARAZO ADOLESCENTE DEL DISTRITO DE SANTA ANA, 2022** de la Br. Br. PASCUAL BÁRCENA CARRASCO. Hacemos de su conocimiento que el (la) sustentante ha cumplido con el levantamiento de las observaciones realizadas por el Jurado el día **ONCE DE DICIEMBRE DE 2024**.

Es todo cuanto informamos a usted fin de que se prosiga con los trámites para el otorgamiento del grado académico de MAESTRO EN ESTADÍSTICA.

Cusco, 27 de Diciembre 2024

DRA. YHENI FARFAN MACHACA
Primer Replicante

MTRO. ARTURO ZUÑIGA BLANCO
Segundo Replicante

DRA. KARLA ZELMIRA APARICIO ARENAS
Primer Dictaminante

MTRO. YARIANI BARRETO SERRANO
Segundo Dictaminante

Dedicatoria

Este trabajo de investigación se lo dedico en primer lugar a Dios, por haber permitido llegar hasta este punto y haberme dado salud para lograr mis objetivos, además de su infinita bondad y amor.

A mis padres Mario y María, por inculcarme buenos valores y disciplina. A mis hermanas Laura y Gina por demostrarme que si se puede seguir estudiando. A mi compañera de vida Yurema por su comprensión incondicional y a mi amado hijo Jefferson.

A la UNSAAC por permitirme lograr dar un paso más hacia el éxito. Por convertirme en un profesional competitivo en la sociedad. De igual forma a todos los profesores que me enseñaron, por su tiempo y dedicación.

Pascual Bárcena Carrasco.

Agradecimiento

En la presente investigación agradezco a Dios, por guiarnos en cada paso que hemos dado a lo largo de nuestra vida y por alcanzar cada una de nuestras metas trazadas, siendo nuestra inspiración espiritual para elegir esta opción profesional y formarnos para el servicio de la sociedad de nuestra región Cusco.

Agradezco a mi asesor Dr. Cleto De La Torre Dueñas, le doy mi eterno agradecimiento por su constante apoyo profesional y moral en el desarrollo de la maestría en la especialidad de Estadística. A mis maestros que, con sus sabias enseñanzas y formación profesional, nos encaminaron en el logro de nuestro perfil profesional para alcanzar las metas de la carrera profesional de Matemáticas mención Estadística.

Al Director y a la Comunidad Red de Servicios de Salud La Convención y la Microred de Salud Santa Ana, por el apoyo y las facilidades dadas en el desarrollo de esta investigación, al personal de obstetricia y al área de estadística que en todo momento me apoyaron en las actividades que se realizaron, mi agradecimiento a la Señora Gisela Janett Cruz Silva Funcionaria Responsable de entregar la información Pública del Ministerio de Educación y Doña Claudia Paola Lisboa Vásquez jefa de la Unidad de Estadística del Ministerio de Educación, por la autorización de datos MINEDU SIAGIE.

A toda mi familia, por su apoyo moral, estando siempre a mi lado. A mis amigos que incondicionalmente me dieron su apoyo y valoraron el esfuerzo que puse en la carrera profesional.

Índice General

Portada	i
Dedicatoria.....	ii
Agradecimiento	iii
Índice General.....	iv
Lista de Tablas.....	vii
Lista de Figuras	ix
Presentación.....	x
Resumen	xi
Abstract.....	xii
CAPITULO I.....	1
PLANTEAMIENTO DEL PROBLEMA.....	1
1.1 Situación Problemática.....	1
1.2 Formulación del Problema	3
1.2.1 Problema General	4
1.2.2 Problema Específicos.....	4
1.3 Justificación de la Investigación.....	5
1.3.1 Justificación Teórica.....	5
1.3.2 Justificación Metodológica.....	5
1.4 Objetivos de la Investigación	5
1.4.1 Objetivo General.....	5
1.4.2 Objetivos Específicos	5
CAPITULO II.....	6
MARCO TEÓRICO CONCEPTUAL.....	6
2.1 Bases Teóricas del Método de Máquinas de Soporte Vectorial	6
2.1.1 Método de Máquinas de Soporte Vectorial.	6
2.1.2 Ventajas de Máquinas de Soporte Vectorial.....	7
2.1.3 Formulación del Problema de Máquinas de Soporte Vectorial.	8
2.1.4 Ancho de Margen	8

2.1.5	Optimización.....	10
2.1.6	Formulación del Problema Primal y Dual	10
2.1.7	Observaciones de la Ecuación	12
2.1.8	Problemas de la Ecuación.....	12
2.1.9	Margen Suave y Variable Holgura.	13
2.1.10	Margen Suave Formulación.....	15
2.1.11	Método del Kernel	18
2.2	Bases Teóricas de los Factores de Embarazo en Adolescencia.....	24
2.2.1	La Adolescencia.....	24
2.2.2	Embarazo en la Adolescencia.....	25
2.2.3	Factores para el Embarazo Adolescente.	28
2.3	Antecedentes de la Investigación	30
2.3.1	Antecedentes Nacionales	30
2.3.2	Antecedentes Internacionales	33
CAPITULO III		38
HIPOTESIS Y VARIABLES		38
3.1	Hipótesis.....	38
3.1.1	Hipótesis General.....	38
3.1.2	Hipótesis Específicas	38
3.2	Identificación de Variables e Indicadores	38
3.3	Operacionalización de Variables.....	39
CAPITULO IV		41
METODOLOGÍA.....		41
4.1	Ámbito de Estudio: Localización Política y Geográfica	41
4.2	Tipo y Nivel de Investigación.	45
4.3	Unidad de Análisis	46
4.4	Población de Estudio	46
4.5	Tamaño de Muestra	46
4.6	Técnicas de Selección de Muestra.....	47
4.6.1	Técnica.....	47
4.6.2	Instrumento	47
4.7	Validez y Confiabilidad de Instrumentos	47

4.8 Técnicas de Análisis e Interpretación de la Información.....	48
4.9 Técnicas para Demostrar la Verdad o Falsedad de la Hipótesis Planteada	50
CAPITULO V	51
RESULTADOS Y DISCUSIÓN	51
5.1 Procesamiento, Análisis, Interpretación y Discusión de Resultados.....	51
5.1.1 Separación de los Datos en Entrenamiento y Prueba.	59
5.1.2 Análisis de Máquinas de Soporte Vectorial Lineal	60
5.1.3 Análisis de Máquinas de Soporte Vectorial Radial	63
5.1.4 Análisis de Máquinas de Soporte Vectorial Polinómica	67
5.1.5 Comparando los Indicadores de Precisión.....	72
5.2 Discusión de Resultados.....	74
5.3 Contrastación de Resultados	78
CONCLUSIONES.....	80
RECOMENDACIONES	81
BIBLIOGRÁFIAS.....	82
ANEXOS	87
Anexo 1: Matriz de Consistencia.....	88
Anexo 2: Instrumento de Recolección de Información.....	91
Anexo 3: Medios de Verificación.....	99
Anexo 4: Otros	109

Lista de Tablas

Tabla 1	<i>Operacionalización de las variables.</i>	39
Tabla 2	<i>Distritos de la provincia de La Convención 2022.</i>	42
Tabla 3	<i>Validez y confiabilidad de Instrumento</i>	48
Tabla 4	<i>Confiabilidad.</i>	48
Tabla 5	<i>Plan de análisis de datos.</i>	50
Tabla 6	<i>Tipo de Gestión.</i>	51
Tabla 7	<i>Nivel Educativo</i>	52
Tabla 8	<i>Grado de Instrucción.</i>	52
Tabla 9	<i>Quintil de Pobreza.</i>	53
Tabla 10	<i>Pobreza Monetaria.</i>	53
Tabla 11	<i>Planificación Familiar.</i>	54
Tabla 12	<i>Distrito.</i>	54
Tabla 13	<i>Migración.</i>	54
Tabla 14	<i>Microrred de Salud.</i>	55
Tabla 15	<i>Categoría de salud por lugar.</i>	55
Tabla 16	<i>Embarazo.</i>	56
Tabla 17	<i>Valores faltantes según variable</i>	56
Tabla 18	<i>Valores faltantes en las variables.</i>	56
Tabla 19	<i>Combinación de valores Faltantes.</i>	57
Tabla 20	<i>Variabes ordenadas por valores faltantes.</i>	59
Tabla 21	<i>Separación de los datos en Entrenamiento y Prueba.</i>	60
Tabla 22	<i>Máquinas de Soporte Vectorial con Kernel Lineal</i>	60
Tabla 23	<i>Matriz de confusión y estadísticas</i>	61
Tabla 24	<i>ROC Curva de importancia de Variable.</i>	62

Tabla 25 <i>Máquinas de Soporte Vectorial con Kernel Radial</i>	64
Tabla 26 <i>Matriz de confusión y estadísticas</i>	65
Tabla 27 <i>ROC Curva de importancia de Variable</i>	66
Tabla 28 <i>Máquinas de Soporte Vectorial con Kernel Polinomial</i>	68
Tabla 29 <i>Matriz de confusión y estadísticas</i>	69
Tabla 30 <i>ROC Curva de importancia de Variable</i>	70
Tabla 31 <i>Comparación de los indicadores de precisión</i>	72
Tabla 32 <i>Comparación de los indicadores de precisión</i>	73

Lista de Figuras

Figura 1	<i>Varios clasificadores</i>	6
Figura 2	<i>Máximo margen</i>	7
Figura 3	<i>Clasificación del máximo margen</i>	9
Figura 4	<i>Sensibilidad del máximo margen</i>	13
Figura 5	<i>Valores de la variable holgura</i>	14
Figura 6	<i>Hiperplano con el costo C aumentado</i>	16
Figura 7	<i>Hiperplano con el costo C disminuido</i>	16
Figura 8	<i>Clasificación no lineal</i>	17
Figura 9	<i>Método de Kernel</i>	18
Figura 10	<i>Separación en un espacio de dimensión mayor</i>	19
Figura 11	<i>Clasificación con el método de las Máquinas de Soporte Vectorial con núcleo polinomial</i>	23
Figura 12	<i>Clasificación con Máquinas de Soporte Vectorial con núcleo radial</i>	24
Figura 13	<i>Forma esquemática de los factores institucionales y de género relacionados con el embarazo en adolescentes</i>	28
Figura 14	<i>Mapa de la provincia La Convención con 14 distritos documento de gestión 2017</i>	43
Figura 15	<i>Mapa de la provincia de La Convención</i>	44
Figura 16	<i>Mapa de la provincia de La Convención por distritos y Microredes de Salud</i>	45
Figura 17	<i>Proporción de valores faltantes y combinación</i>	58
Figura 18	<i>Variables más importantes</i>	63
Figura 19	<i>Variables más importantes</i>	67
Figura 20	<i>Variables más importantes</i>	71

Presentación

Señora Directora General de la Escuela de Posgrado de la institución Dra. Nelly Ayde Cavero Torre.

Distinguidos Docentes Miembros del Jurado Examinador. Dando cumplimiento a las disposiciones de las normas actuales del Reglamento de Grados y Títulos de la Escuela de Posgrado concerniente a la obtención del grado académico de Maestro en Estadística, pongo a vuestra disposición la investigación con el título: Máquina de soporte vectorial en el estudio de factores de embarazo adolescente, del distrito de Santa Ana, 2022, con el propósito de optar el grado académico de Maestro en Estadística.

La presente investigación tiene como propósito determinar el método de minería de datos: Determinar el método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022. que se encuentran afiliados en el SIS de la Microred de Salud.

La tesis está estructurada en 5 capítulos de la siguiente manera: planteamiento del problema, marco teórico conceptual, hipótesis y variables, metodología de la investigación, resultados y discusión.

Por lo indicado señores miembros del jurado, recibo con aceptación vuestro aportes y sugerencias, para poder mejorar esta investigación, a la vez deseamos valga de aporte a quien desea continuar un estudio en este tema. La presente investigación se desenvolverá en enfoque cuantitativo, cualitativo y según el propósito de la investigación es básica y no experimental.

El investigador

Resumen

El presente trabajo tiene por finalidad determinar el método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022, El estudio se realizó bajo un paradigma cuantitativo y cualitativo, el tipo de investigación según el propósito de la investigación es básica y es una investigación no experimental. La población es de 11200 adolescentes mujeres afiliados al SIS de la Red de Servicios de Salud La Convención, de estas 8537 cumplen con los criterios de integridad de información, y la muestra se consideró 2993 mayor a lo establecido en el tamaño de la muestra ($n \geq 367.7$). Se empleó la técnica de investigación documental referida a las historias clínicas y una ficha de información sociodemográficas de cada adolescente y el instrumento más adecuado para el registro de información que se ha diseñado es la historia clínica. Se estableció que la máquina de soporte lineal presenta una exactitud de 0.9665, sensibilidad de 0.86 y especificidad de 0.92, las máquinas de soporte radial presentan una exactitud de 0.9602, sensibilidad de 0.7401 y especificidad de 0.9364 y máquina de soporte polinómico presenta una exactitud de 0.9788, sensibilidad de 0.7742 y especificidad de 0.9409. Se concluye que el modelo de máquinas de soporte vectorial con kernel polinómico de grado 2, presenta mayor exactitud en la predicción de embarazo adolescente con una precisión de 97.88%.

Palabras claves: Embarazo adolescente, máquinas de soporte vectorial, clasificación.

Abstract

The purpose of this work is to determine the support vector machine method that presents the greatest accuracy in predicting adolescent pregnancy factors, in the Santa Ana district in 2022. The study was carried out under a quantitative and qualitative paradigm, the type of research according to the purpose of the research is basic and it is a non-experimental research. The population is 11,200 adolescent women affiliated with the SIS of the La Convención Health Services Network, of these 8,537 meet the information integrity criteria, and the sample was considered 2,993 larger than established in the sample size ($n \geq 367.7$). The documentary research technique was used referring to the clinical histories and a sociodemographic information sheet of each adolescent and the most appropriate instrument for recording information that has been designed is the clinical history. It was established that the linear support machine presents an accuracy of 0.9665, sensitivity of 0.86 and specificity of 0.92, the radial support machines present an accuracy of 0.9602, sensitivity of 0.7401 and specificity of 0.9364 and polynomial support machine presents an accuracy of 0.9788, sensitivity of 0.7742 and specificity of 0.9409. It is concluded that the support vector machine model with polynomial kernel of degree 2, presents greater accuracy in the prediction of teenage pregnancy with an accuracy of 97.88%.

Keywords: Teenage pregnancy, support vector machines, classification.

CAPITULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 Situación Problemática

El embarazo en la adolescente cambia la forma de vida que tiene planificado a futuro, estos cambios son de forma individual y de su entorno familiar, son varios los impactos que se asocian al embarazo en la adolescente cuando no es planificado, estos impactos tienen consecuencias significativas para las jóvenes madres y para sus bebés. Los riesgos para las madres son complicaciones obstétricas, infecciones sistémicas, salud mental y para los bebés son el bajo peso al nacer, prematuridad y afecciones neonatales graves, también existe un impacto social y económico; el embarazo adolescente afecta en sus estudios, las oportunidades laborales y la calidad de vida de las jóvenes madres.

Según la Organización Mundial de la Salud (2024), en las regiones en desarrollo, alrededor de 21 millones de mujeres de entre 15 y 19 años quedan embarazadas anualmente, y aproximadamente 12 millones de ellas dan a luz. Según el informe, en el año 2019, el 55% de los embarazos no planificados en adolescentes resultaron en interrupción voluntaria del embarazo.

Según la Organización Mundial de la Salud (2023), el riesgo de mortalidad materna disminuye en las mujeres alrededor de los 20 años de edad y alcanza su punto máximo en aquellas que dan a luz antes de los 15 años.

Es esencial garantizar la prevención del embarazo durante la adolescencia y ofrecer cuidados maternos de excelencia para alcanzar beneficios saludables a lo largo de toda la vida. Es crucial comprender a fondo los elementos que influyen en el embarazo en la juventud, con el fin de brindar un respaldo continuo a los jóvenes, permitiéndoles tomar decisiones bien fundamentadas acerca de su bienestar sexual y reproductivo.

Según la Encuesta Demográfica y de Salud familiar en Perú, realizada por el Instituto Nacional de Estadística e Informática (2022), se ha observado que el 9.2% de las jóvenes de 15 a 19 años han experimentado un embarazo, lo que implica que ya son madres o están esperando su primer hijo. En las zonas rurales, esta cifra asciende al 18.4%, mientras que en las áreas urbanas se sitúa en un 6.8%.

En la provincia de La Convención las adolescentes embarazadas que se registra en la Red de Servicios de Salud La Convención son 103 en el año 2022,

Por las consideraciones señaladas, es muy importante identificar los factores asociados al embarazo adolescente empleando las técnicas estadísticas con los algoritmos de la máquina de soporte vectorial (SVM). Con este método se clasificará los datos de los factores mediante un hiperplano.

La máquina de soporte vectorial fue desarrollada por Vladimir Vapnik y su equipo en los laboratorios de AT&T Bell, son algoritmos de aprendizaje supervisado para clasificación, el método busca un hiperplano que separe de forma óptima los datos de una clase (factores) de los de otra clase, este hiperplano tiene la máxima distancia (margen) con los datos más cercanos a él.

El objetivo es determinar el método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo adolescente, en el distrito de Santa Ana 2022. Estableciendo los factores asociados al embarazo adolescente obtenidos de la aplicación del método de máquinas de soporte vectorial. Determinar la sensibilidad y la especificidad de las máquinas de soporte vectorial lineal polinomial y radial, en la predicción de factores de embarazo adolescente.

1.2 Formulación del Problema

El embarazo en adolescentes es un problema significativo en Perú, afectando la salud, la educación y el bienestar social de las jóvenes.

En el año 2022, se registraron más de 50,000 partos en adolescentes de 10 a 19 años, con un notable incremento en las tasas de maternidad entre las más jóvenes (10 a 14 años) en comparación con años anteriores (Mesa de Concentracion para la Lucha contra la Pobreza, 2023) (Renato, 2024). La Encuesta Demográfica y de Salud Familiar (ENDES) indica que alrededor del 13% de las adolescentes de 15 a 19 años han experimentado un embarazo, cifra que se ha mantenido relativamente estable en los últimos años, aunque con variaciones significativas entre las áreas urbanas y rurales.

Según investigaciones, este fenómeno no solo se asocia con problemas de salud pública, sino también con factores socioeconómicos y culturales que influyen en la vida de las adolescentes (Sanchez Vargas, 2020) (Prado Juscamaita & Reves Huapaya, 2018) La falta de apoyo social y la violencia sexual son factores críticos que contribuyen a la alta incidencia de embarazos no deseados en este grupo (Fondo de las Naciones Unidas para la Infancia, 2014).

La maternidad adolescente está fuertemente asociada a factores como la violencia sexual, la falta de educación sexual integral, y la desigualdad de género. En muchas comunidades, el embarazo adolescente es resultado de una serie de violencias y abusos, lo que resalta la necesidad de un enfoque multisectorial que aborde tanto la prevención como la atención adecuada a estas jóvenes (Rodríguez Bernate & Tobar, 2022) (UNFPA, 2023)

En la provincia de La Convención se presenta altos índices de embarazos en adolescentes, principalmente estas están asociadas factor económico, familiar por otra parte

en las comunidades nativas el principal factor es la cultura asociada a la formación de familia en periodo de adolescencia.

La importancia del estudio de estos factores radica en la prevención del embarazo en adolescentes para garantizar la calidad de resultados se deben emplear metodologías estadísticas actuales más robustas como el caso de las máquinas de soporte vectorial.

Las máquinas de soporte vectorial son técnicas de aprendizaje automático utilizadas principalmente para predecir fenómenos de carácter cuantitativo o cualitativo; en el presente estudio se utiliza máquinas de soporte vectorial para fines de clasificación es decir predicción de fenómeno cualitativo como el caso de embarazo y la predicción de factores asociados a este.

Este estudio busca contribuir a la formulación de políticas públicas efectivas que no solo aborden la prevención del embarazo en adolescentes, sino que también promuevan el empoderamiento de las jóvenes y su acceso a educación y servicios de salud adecuados.

1.2.1 Problema General

¿Cuál de los métodos de máquinas de soporte vectorial presenta mejores niveles de predicción en los factores de embarazo de adolescentes en el distrito de Santa Ana 2022?

1.2.2 Problema Específicos

- a) ¿Cuáles son los factores asociados en el embarazo adolescente en el distrito de Santa Ana 2022, obtenidos de la aplicación del método máquina de soporte vectorial?
- b) ¿Cuál es la sensibilidad y la especificidad de las máquinas de soporte vectorial en la predicción de factores de embarazo de adolescentes en el distrito de Santa Ana 2022?

1.3 Justificación de la Investigación

1.3.1 Justificación Teórica

El estudio es importante porque permitirá conocer los factores de embarazo adolescentes, esto permitirá tomar mejores decisiones para la población adolescentes.

1.3.2 Justificación Metodológica

En el presente trabajo se diseñaron fichas de registros de datos que pueden ser utilizados en otros estudios. Por otra parte, para determinar los factores se aplicó la técnica de máquinas de soporte vectorial la misma, que puede ser empleada en otros estudios de salud.

1.4 Objetivos de la Investigación

1.4.1 Objetivo General

Determinar el método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022.

1.4.2 Objetivos Específicos

- a) Establecer los factores asociados al embarazo de adolescentes en el distrito de Santa Ana 2022, obtenidos de la aplicación del método máquina de soporte vectorial.
- b) Determinar la sensibilidad y la especificidad de las máquinas de soporte vectorial en la predicción de factores de embarazo de adolescentes en el distrito de Santa Ana 2022.

CAPITULO II

MARCO TEÓRICO CONCEPTUAL

2.1 Bases Teóricas del Método de Máquinas de Soporte Vectorial

2.1.1 Método de Máquinas de Soporte Vectorial.

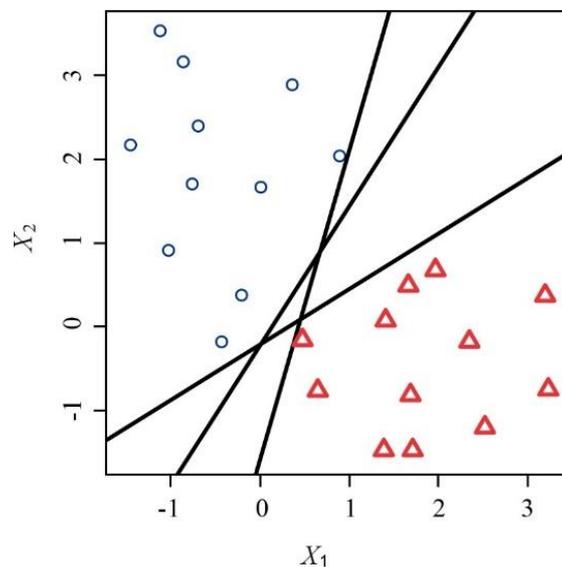
Para un problema de clasificación donde las clases son linealmente separables, por lo general existen infinitos hiperplanos separadores (Bishop, 2006).

Como se puede ver en la siguiente figura, la asignación de clase es similar en la mayoría del espacio salvo la región cercana al límite de las clases.

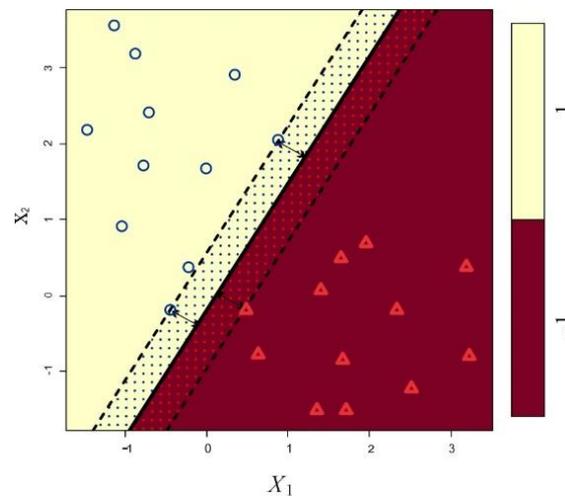
En la figura se observa, la asignación de clase es igual en la mayoría del espacio y distinto en la región cerca al límite de las clases.

Figura 1

Varios clasificadores



Nota. En la figura N°1, los datos en el plano cartesiano están representados por la figura de círculo color azul y triángulo de color rojo, estas figuras representan dos clases y cada figura tienen dos variables. En la gráfica también se observa 3 hiperplanos posibles que separan las clases.

Figura 2*Máximo margen*

Nota. En la figura N°2 se observa un hiperplano que separa a los datos (círculos y triángulos) según sus respectivas clases. La leyenda de la derecha muestra la representación numérica utilizada para la clasificación de los datos.

Las máquinas de soporte vectorial buscan el hiperplano separador que maximice el margen entre las clases, lo cual es equivalente a buscar una cinta de ancho máximo que separe los datos.

2.1.2 Ventajas de Máquinas de Soporte Vectorial

Las ventajas de las máquinas de soporte vectorial, al encontrar este clasificador, son varias sobre otros clasificadores

Encontrar este clasificador tiene ventajas sobre los otros estudiados ya que:

- Si los datos de cada clase se originan en una distribución latente, es probable que los nuevos datos obtenidos de la misma distribución estén próximos a los datos observados inicialmente. Esto permite que la máquina de soporte vectorial tenga excelentes propiedades de generalización.
- Como se explicará más adelante, el clasificador de margen máximo se define solo por ciertos datos. Esto soluciona de inmediato el problema de los desbalances de clase o de las formas diferentes de las clases.

- En consecuencia, la solución de margen máximo permanece estable al añadir datos fuera del margen. Los datos que determinan el margen se denominan Vectores de Soporte, y su función es limitar la rotación y expansión del margen.

2.1.3 *Formulación del Problema de Máquinas de Soporte Vectorial.*

Sea un conjunto de datos de entrenamiento $\{x_i\}_{i=1}^N$ linealmente separable, con dos clases $\{1, -1\}$, se define un hiperplano separador (Vapnik, 1998).

$$\{x \in \mathbb{R}^n \mid \omega^\top x + b = 0\} \quad (1)$$

Donde:

Si $\omega^\top x + b > 0$, a la clase de x se le asignara 1

Si $\omega^\top x + b < 0$, a la clase de x se le asignara -1

El par (ω, b) forma parte del hiperplano que separa y no es único, entonces $(\mu\omega, \mu b)$ también formara parte del hiperplano.

Sean los vectores de soporte de cada clase x_+ y x_- , Se pone restricciones a los datos sobre el borde del margen, para que no haya escalamiento y los datos permanecerán en su misma clase.

De la siguiente forma:

$$\omega^\top x + b = 1 \quad (2)$$

$$\omega^\top x + b = -1 \quad (3)$$

Se pone las restricciones sin conocer cuáles son los Vectores de Soporte, que pueden ser varios.

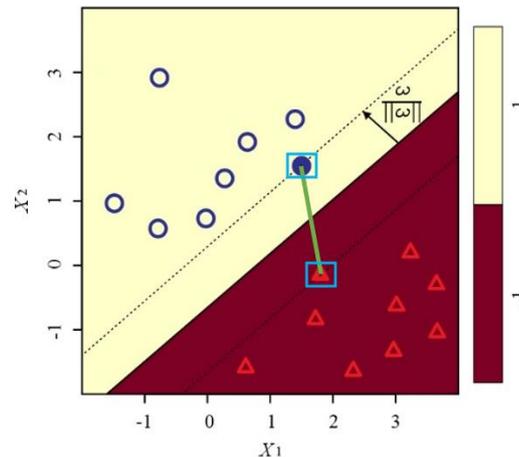
2.1.4 *Ancho de Margen*

Las restricciones anteriores definen hiperplanos paralelos a la región de decisión ya que tienen el mismo parámetro ω .

Por otra parte, el margen (M) se define como la distancia entre la región de decisión y cualquiera de las clases (Vapnik, 1998).

Figura 3

Clasificación del máximo margen.



Nota. En la figura N°3, que ilustra la clasificación de máximo margen, se observa dos clases señaladas por círculos azules y triángulos rojos. La región de decisión está indicada por una línea continua y los bordes del margen se muestran mediante una línea punteada. Los vectores de soporte están indicados con rectángulos de color celeste. La línea verde simboliza el vector diferencia ($x_+ - x_-$), y la flecha negra el vector unitario $\frac{\omega}{\|\omega\|}$.

En la figura se observa que el ancho del margen puede ser calculado como una proyección sobre la dirección normal del hiperplano, es decir. Que el ancho de margen M , es calculado como una proyección sobre la dirección normal del hiperplano.

$$M = \frac{1}{2} \|\text{proy}_{\omega}(x_+ - x_-)\| \quad (4)$$

$$= \frac{1}{2} \|(x_+ - x_-)\| \cos(\theta)$$

$$= \frac{1}{2} \|(x_+ - x_-)\| \left(\frac{\omega^\top (\|(x_+ - x_-)\|)}{\|\omega\| \cdot \|(x_+ - x_-)\|} \right)$$

$$= \frac{1}{2\|\omega\|} \omega^\top (x_+ - x_-) \quad (5)$$

Donde el $\cos(\angle(x, y)) = \frac{\langle x, y \rangle}{\|x\| \|y\|}$

2.1.5 Optimización

Uno de los problemas de optimización en el ancho de margen M , no depende explícitamente de los vectores soporte (Vapnik, 1998).

$$M = \frac{1}{2\|\omega\|} ((\omega^\top x_+) - (\omega^\top x_-)) \quad (6)$$

$$M = \frac{1}{2\|\omega\|} ((1 - b) - (-1 - b))$$

$$M = \frac{1}{\|\omega\|} \quad (7)$$

Se considera las siguientes codificaciones para cada clase.

$$y_i = +1 \leftrightarrow \omega^\top x_i + b \geq +1 \quad (8)$$

Optimización de la clasificación de máximo margen.

$$\begin{aligned} \max_{\omega, b} \frac{1}{\|\omega\|}, \\ \text{s.a}^1 y_i(\omega^\top x_i + b) \geq +1, i = 1, \dots, N \end{aligned} \quad (9)$$

Donde las restricciones exigen que todas las muestras estén correctamente clasificadas.

2.1.6 Formulación del Problema Primal y Dual

Para evitar problemas de diferenciabilidad, se considerará la siguiente formulación equivalente del problema anterior.

De la ecuación (9) anterior tenemos lo siguiente.

$$\begin{aligned} (P) \min_{\omega, b} \frac{1}{2} \|\omega\|^2 \\ \text{s. a } y_i(\omega^\top x_i + b) \geq +1, i = 1, \dots, N. \end{aligned} \quad (10)$$

La dualidad Lagrangiana, nos permite resolver el problema de optimización con restricciones (Vapnik, 1998).

¹ Expresión utilizada en matemáticas para indicar que una función está sujeta a otra

$$L(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 + \sum_{i=1}^N \alpha_i (1 - y_i (\omega^\top x_i + b)) \quad (11)$$

El Lagrangiano dual del problema está dado por.

$$\theta(\alpha) = \inf_{\omega, b} L(\omega, b, \alpha) \quad (12)$$

con L convexo.

Aplicando la condición de primer orden se tiene.

$$\frac{\partial L}{\partial \omega} = \omega^\top - \sum_{i=1}^N \alpha_i y_i x_i^\top = 0 \rightarrow \bar{\omega} = \sum_{i=1}^N \alpha_i y_i x_i \quad (13)$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^N \alpha_i y_i = 0 \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \quad (14)$$

Luego se sustituye en la ecuación L (11), la ecuación Lagrangiano dual se obtiene.

$$\theta(\alpha) = L(\bar{\omega}, \bar{b}, \alpha) \quad (15)$$

$$\begin{aligned} &= \frac{1}{2} \left\langle \sum_{i=1}^N \alpha_i y_i x_i, \sum_{j=1}^N \alpha_j y_j x_j \right\rangle + \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \alpha_i y_i \left\langle \sum_{j=1}^N \alpha_j y_j x_j, x_i \right\rangle - b \sum_{i=1}^N \alpha_i y_i \\ &= \frac{1}{2} \sum_{i,j=1}^N \alpha_i y_i \alpha_j y_j \langle x_i, x_j \rangle + \sum_{i=1}^N \alpha_i - \sum_{i,j=1}^N \alpha_i y_i \alpha_j y_j \langle x_j, x_i \rangle \\ &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \end{aligned} \quad (16)$$

Finalmente, el problema dual se centra en maximizar $\theta(\alpha)$ con la condición que

$\alpha \geq 0$, es decir:

$$\begin{aligned} (D) \max_{\alpha} & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\ \text{s. a} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & \alpha_i \geq 0 \end{aligned} \quad (17)$$

Donde la primera restricción se deriva de la condición de optimalidad impuesta sobre L , al calcular $\theta(\alpha)$.

2.1.7 Observaciones de la Ecuación

Para resolver máquinas de soporte vectorial es mediante la solución de su problema dual.

La forma usual para resolver una máquina de soporte vectorial es mediante la resolución de su problema dual. Para este problema se tienen las siguientes observaciones:

- La función objetivo es una forma cuadrática definida negativa.
- Lo anterior implica que el problema tiene un único máximo.
- Después de resolver formulación dual (es decir., se han determinado los valores óptimos para α), la predicción de un nuevo punto x_m se expresa de la siguiente manera.

$$\hat{y}(x_m) = \text{sgn}(\bar{\omega}^\top x_m + b) \quad (18)$$

$$= \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i \langle x_i, x_m \rangle + b\right) \quad (19)$$

- Según el teorema de holgura complementaria, para α óptimo se cumple que.

$$\alpha_i (1 - y_i (\bar{\omega}^\top x_i + b)) = 0, \forall i \in \{1, \dots, N\}. \quad (20)$$

Por lo que $\alpha_i = 0$, $\forall x_i$ fuera del margen y , consecuentemente x_i fuera del margen no aporta en la predicción \hat{y} .

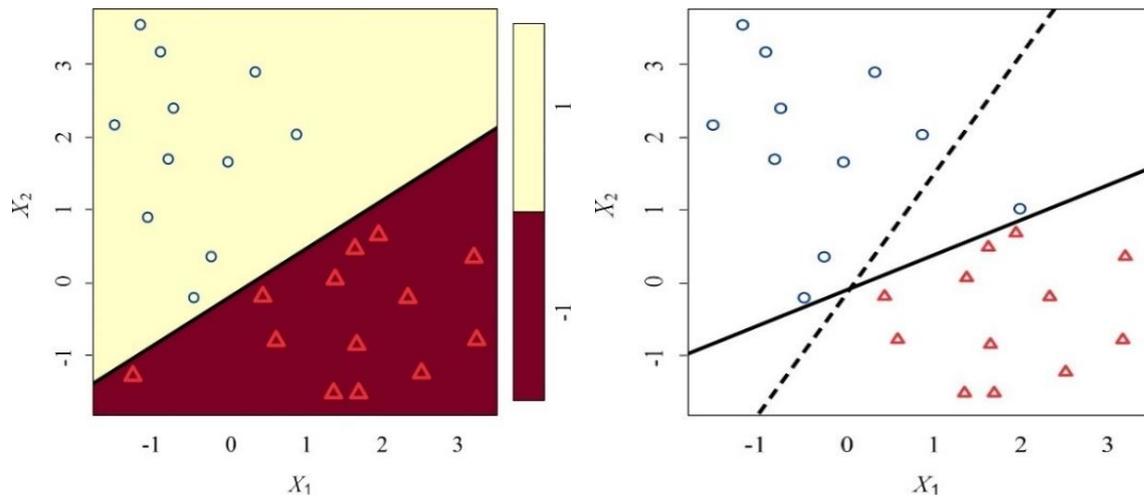
2.1.8 Problemas de la Ecuación

El planteamiento de la ecuación N°19 tiene dos debilidades.

- Los datos no siempre serán separables por lo que el problema puede ser infactible.
- Aunque los datos son linealmente separables, el clasificador puede mostrar una alta sensibilidad a nuevos datos. Esto se observa en la siguiente figura,

Figura 4

Sensibilidad del máximo margen.



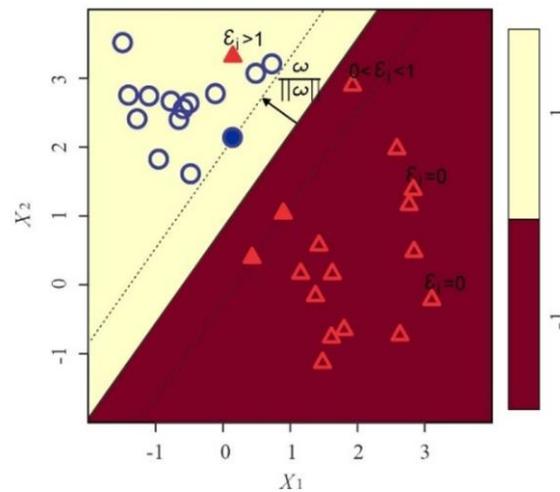
Nota. En la figura N°4, a la izquierda se observa el caso sin outliers y a la derecha el caso con outliers, La solución previamente mencionada se representa mediante una línea discontinua.

2.1.9 Margen Suave y Variable Holgura.

Para solucionar los problemas de mala clasificación tenemos las variables de holgura. Estas variables de holgura permiten al clasificador admitir algunos datos, de las clases que estén mal clasificados (Cortes y Vladimir, 1995).

Figura 5

Valores de la variable holgura.



Nota. En la gráfica N°5 se observa un hiperplano junto con los vectores de soporte, también se muestra los valores de la variable holgura ϵ según su ubicación.

La variable holgura $\{\epsilon_i\}_{i=1}^N$, perdona ciertos datos mal clasificados.

$$(P) \min_{\omega, b, \epsilon} \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^N \epsilon_i \quad (21)$$

$$s. a \ y_i(\omega^\top \cdot x_i + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0$$

Considerando $c > 0$, es un hiperparámetro. En la ecuación de costo, cuando se pone el término $c \sum_{i=1}^N \epsilon_i$, se interpreta como regularización en mínimos cuadrados regularizados. Los ϵ indica el error de clasificación de un dato.

- Si $\epsilon_i = 0$, entonces el dato de x_i se encuentra en el lado correcto del plano.
- Si $0 < \epsilon_i < 1$, entonces el dato x_i se encuentra en el lado correcto del plano, pero se encuentra dentro del margen.
- Si $\epsilon_i > 1$, entonces el dato x_i se encuentra en el lado incorrecto del plano.

2.1.10 Margen Suave Formulación.

La formulación del margen suave, es de la misma forma que la anterior, el problema dual es el siguiente.

$$\begin{aligned}
 (D) \max_{\alpha} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \\
 \text{s. a} \quad & \sum_{i=1}^N \alpha_i y_i = 0, \\
 & 0 \leq \alpha_i \leq c
 \end{aligned} \tag{22}$$

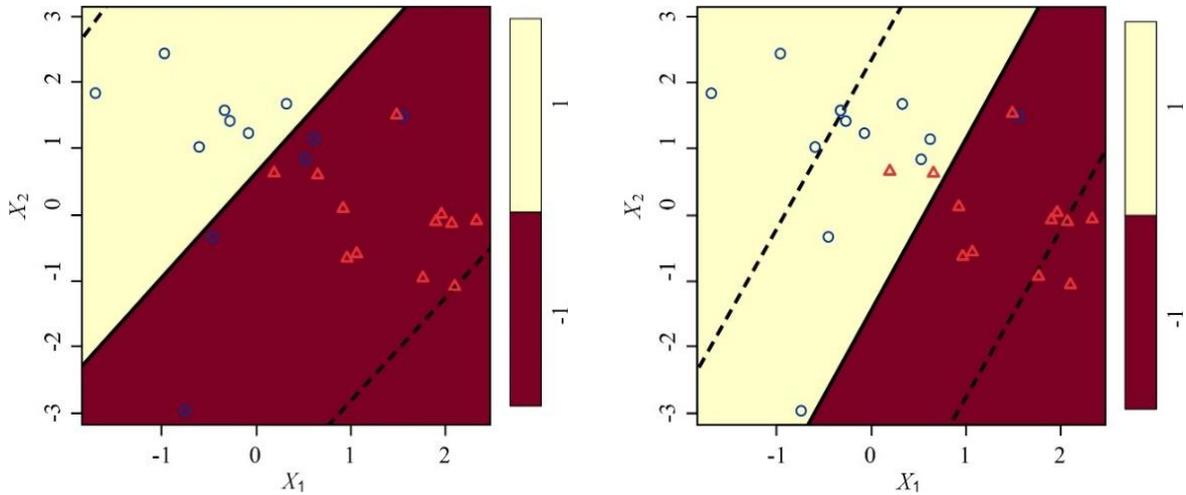
El hiperparámetro (C) en las Maquinas de soporte vectorial actúa como un coeficiente inverso de regularización, ya que su valor determina el equilibrio entre la maximización del margen y el número de muestras incorrectamente clasificadas.

- Si C tiende a infinito, se obtiene nuevamente la formulación del margen duro y su solución de máximo margen es correcta (si el problema es linealmente separable), ya que el problema solo se resolverá si todas las variables de holgura son cero.
- Si C es pequeño, el margen adquiere mayor relevancia que la cantidad de datos incorrectamente clasificados. Como consecuencia, se obtiene un amplio margen que incluye diversas muestras clasificadas incorrectamente. Se prevé que un modelo con un margen amplio, representado por un C pequeño, exhiba una mayor habilidad para adaptarse a datos inéditos, reduciendo así la variabilidad.
- En caso de que el valor de C sea elevado (con un margen amplio), existe la posibilidad de que se ajuste demasiado a los datos disponibles, lo cual puede llevar a un rendimiento deficiente cuando se aplica a datos nuevos. Por esta

razón, se compara a C con un coeficiente inverso de regularización: cuanto menor sea el valor de C , más regularizada será la solución.

Figura 6

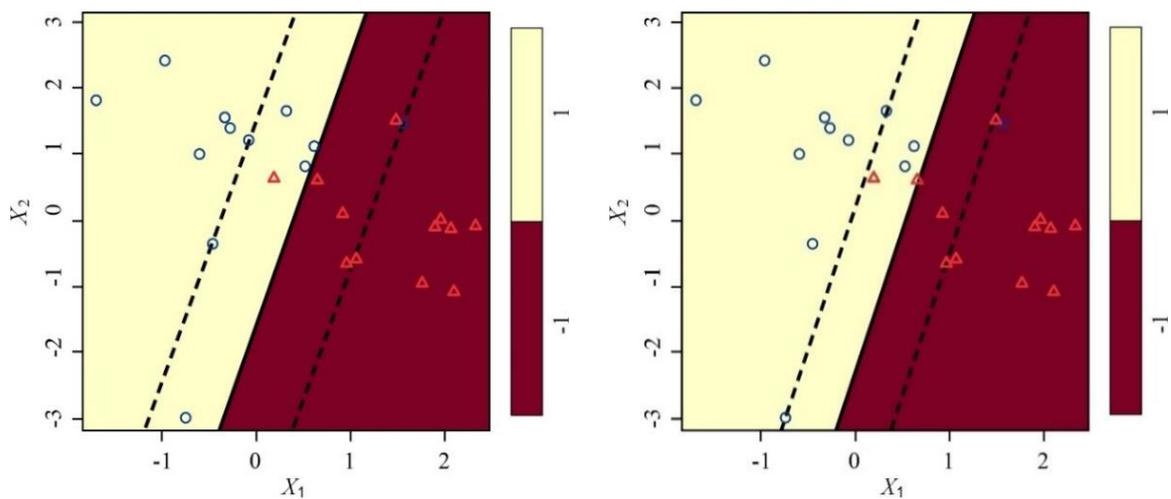
Hiperplano con el costo C aumentado.



Nota. En la figura N°6 se observa un clasificador de vectores de soporte usando diferentes valores para C , cuando C es grande hay un consentimiento para los datos que están en el lado incorrecto del margen y el margen es grande.

Figura 7

Hiperplano con el costo C disminuido.



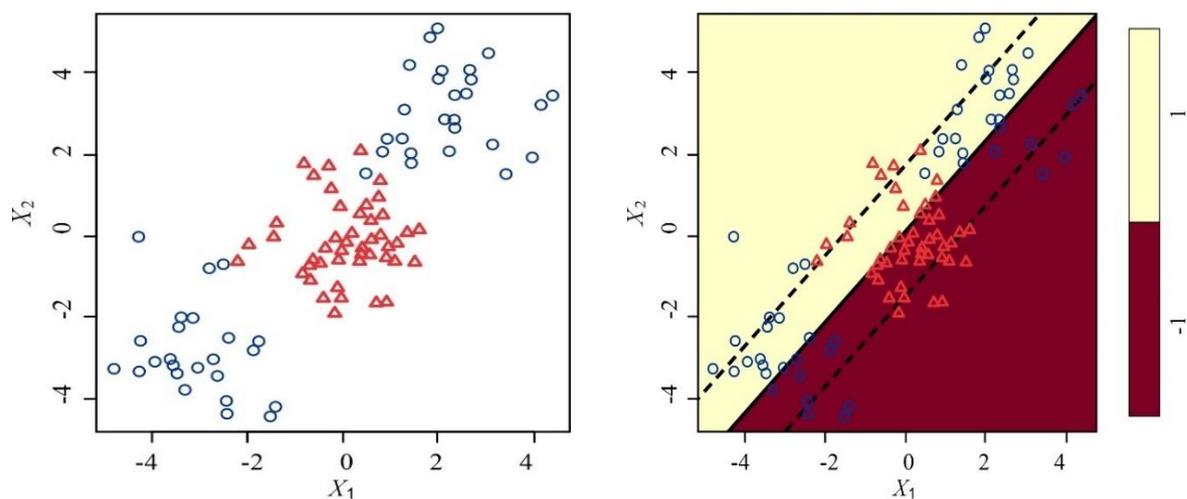
Nota. En la figura N°7 se observa un clasificador de vectores de soporte usando diferentes valores para C . Cuando C es pequeño, hay menor consentimiento para los datos que se encuentran en el lado incorrecto del margen y el margen es más estrecha.

Observaciones:

- La diferencia entre ambas formulaciones radica en que los multiplicadores de Lagrange están ahora limitados por el hiperparámetro C .
- C representa la relevancia asignada a la suma de las variables de holgura en comparación con el ancho del margen.
- Cuando C tiende al infinito, se recupera la formulación original, conocida como margen duro.
- La elección de este hiperparámetro puede realizarse mediante validación cruzada.

Figura 8

Clasificación no lineal



Nota. En la figura N°8, a la izquierda se observa los datos de dos clases y en la figura a la derecha se observa un clasificador de vectores de soporte con el hiperplano, sin embargo, no tienen un buen rendimiento al clasificar los datos.

2.1.11 Método del Kernel

El método de las máquinas de soporte vectorial es apropiado cuando las clases son linealmente separables, pero cuando los datos no son separables linealmente se aplica el método de Kernel, donde se diseña una característica para que los datos sean linealmente separables (Cortes y Vladimir, 1995).

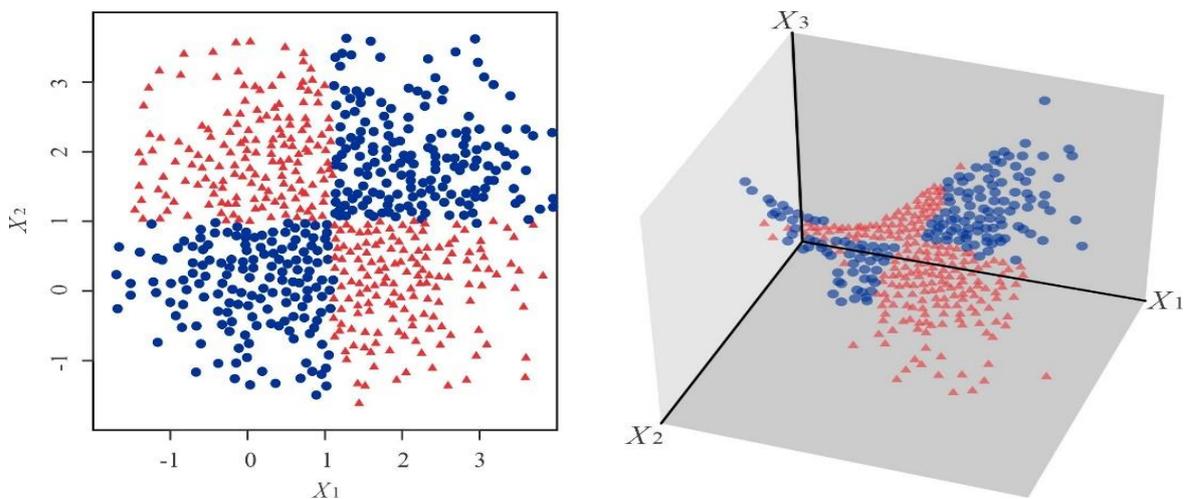
Sea desde \mathbb{R}^2 a \mathbb{R}^3 de la siguiente forma.

$$\phi: [x_1, x_2]^\top \rightarrow [x_1, x_2, x_1x_2]^\top \quad (23)$$

La nueva representación permite clasificar de forma lineal las clases con un plano $Z = 0$ en \mathbb{R}^3 .

Figura 9

Método de Kernel.



Nota. En la figura N°9, a la izquierda, se observa los datos no son linealmente separables. En la figura de la derecha, los datos han sido mapeados a través de ϕ . En este nuevo espacio, el plano $z = 0$ puede separar los datos rojos de los azules. Donde $x_3 = x_1x_2$.

En el caso general no es claro cuál debe ser el buen ϕ . A pesar de esto, notemos que en la formulación de máquina de soporte vectorial solo se requiere poder calcular los productos internos entre las características de cada entrada, es decir, si se considera un ϕ

arbitrario para el problema de clasificación, solo se necesitaría calcular los productos internos de la forma.

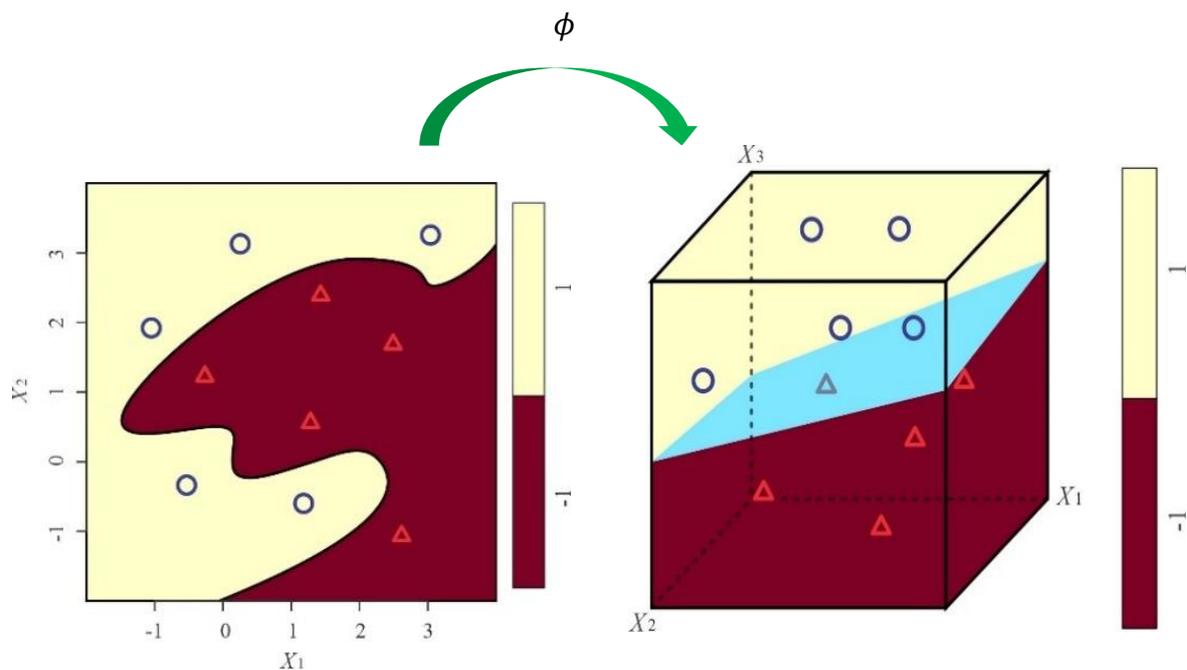
$$\langle \phi(x_i), \phi(x_j) \rangle \quad (24)$$

Por lo tanto, si bien no siempre se puede encontrar el ϕ exclusivo de cada problema, es posible utilizar un modelo muy general, con la esperanza de que alguno de los términos añadidos sea el que realmente separe las clases.

De esta forma, se puede considerar un mapeo de alta dimensión que incorpore varias combinaciones entre las componentes.

Figura 10

Separación en un espacio de dimensión mayor.



Nota. En la figura N°10, a la izquierda se observa el espacio de entrada y a la derecha se representa el espacio de características, donde se aplica la transformación ϕ .

Para encontrar estos ϕ generales, veamos la siguiente definición.

2.1.11.1. Definición del Método Mercer Kernel.

Este método Mercer kernel es una función continua $k: X \times X \rightarrow \mathbb{R}$, tal que:

- Es simétrica $K(x_1, x_2) = K(x_2, x_1)$
- Está definida positivamente (Vapnik, 1998), quiere decir.

$$\int_{x^2} K(x_1, x_2)g(x_1)g(x_2)dx_1dx_2 \geq 0. \quad (25)$$

Donde \forall función $g: X \rightarrow \mathbb{R}$, continua.

La segunda propiedad recibe su nombre debido a su similitud con las matrices definidas positivas: si g se mira como vector de \mathbb{R}^X , entonces la expresión anterior se representa como, $g^\top K g \leq 0$.

De esta forma, se tiene el siguiente teorema de análisis funcional que sustenta la utilización de kernel en los algoritmos de aprendizaje automático:

2.1.11.2. Teorema de Mercer Kernel Simplificado.

Sea un $k: X \times X \rightarrow \mathbb{R}$, un Mercer Kernel, entonces existe un espacio de Hilbert $(\mathcal{H}, \langle, \rangle)$ y una función $\phi: X \rightarrow \mathcal{H}$. (Vapnik, 1998), tal que.

$$K(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle \quad (26)$$

Quiere decir que existe un mapa de las características de $\phi / K(x_1, x_2)$ representa el producto interno en algún espacio de estas características de x_1 y x_2 , este espacio no necesariamente tiene que ser de dimensión finita.

2.1.11.3. Kernel Polinomial.

Se define de la siguiente forma.

$$K_{pol}(x, y) = (c + x^\top y)^d \quad (27)$$

Donde:

$c \geq 0$, es el parámetro libre.

$d \in \mathbb{N}$, es el orden polinomial.

Si $d = 2$, los términos se pueden reagrupar, para observar que el mapa de las características que incita este Kernel.

$$\begin{aligned} \phi_{pol}(x) & \\ &= [x_1^2, \dots, x_m^2, \sqrt{2}x_1x_2, \dots, \sqrt{2}x_mx_{m-1}, \sqrt{2c}x_1, \dots, \sqrt{2c}x_m, c] \end{aligned} \quad (28)$$

Cuando se utiliza el kernel polinomial, se usa un mapa de características que tiene los monomios de grado hasta $d = 2$, en general esta propiedad cumple para cualquier valor $d \in \mathbb{N}$.

2.1.11.4. Función Base Radial Kernel.

A esta función base radial Kernel (RBF Kernel) se le conoce como Kernel gaussiano o exponencial cuadrático, está dado por la siguiente ecuación.

$$K_{RBF}(x, y) = \sigma^2 \exp\left(-\frac{\|x - y\|^2}{2l^2}\right) \quad (29)$$

Es de dimensión infinita el mapa de las características que se induce y sus fronteras son suaves e infinitamente diferenciables.

2.1.11.5. Kernel Periódico.

Está definido con la siguiente ecuación:

$$K_{per}(x, y) = \sigma^2 \exp\left(-\frac{2\text{sen}^2\left(\frac{\pi|x - y|}{p}\right)}{l^2}\right) \quad (30)$$

El Kernel periódico puede extraer características periódicas de los datos, y está regulado por el parámetro p .

2.1.11.6. Truco de Kernel.

Es importante la introducción del método de Kernel en las máquinas de soporte vectorial, porque:

- El teorema de Mercer simplificado indica, que cuando se tiene cualquier función k simetría y positiva, $K(x_1, x_2)$ representara un producto interno en un espacio de características.
- Las máquinas de soporte vectorial simplifican su funcionamiento al calcular únicamente los productos internos, lo que posibilita la creación de kernel en estas máquinas. Esto implica la necesidad de definir el producto interno en el proceso de optimización a través del kernel.
- El truco de kernel se puede aplicar a cualquier algoritmo, en donde las entradas solamente aparezcan en forma de producto punto, a este proceso se le conoce como Kernelización.

2.1.11.7. Kernel y Máquinas de Soporte Vectorial.

Sea un mapa de características ϕ , al formular las máquinas de soporte vectorial se pueden reemplazar las entradas por las características inducidas por ϕ . Entonces, su ecuación primal será la siguiente.

$$(P) \quad \min_{\omega, b} \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^N \varepsilon_i \quad (31)$$

$$s. a \quad y_i(\omega^\top \phi(x_i) + b) \geq 1 - \varepsilon_i, i = 1, \dots, N$$

Su ecuación dual es:

$$(D) \quad \max_{\alpha} \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \quad (32)$$

$$s. a \quad \sum_{i=1}^N \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq c$$

Como el problema de optimización solamente usa los datos mediante producto internos, se utilizará el truco de kernel para parametrizar directamente dicho producto interno. $\langle \phi(x_i), \phi(x_j) \rangle$ Mediante $K(x_i, x_j)$.

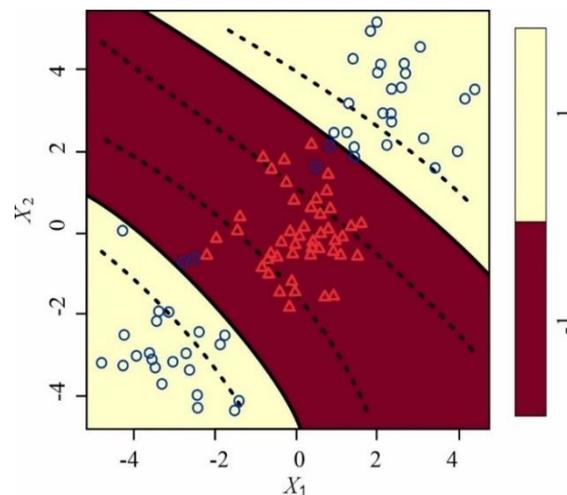
En la ecuación se muestra el problema de optimización, con el dual.

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s. a} \quad & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq c \end{aligned} \quad (33)$$

En la figura (11) se muestra la implementación del kernel en las máquinas de soporte vectorial utilizando dos tipos de kernels.

Figura 11

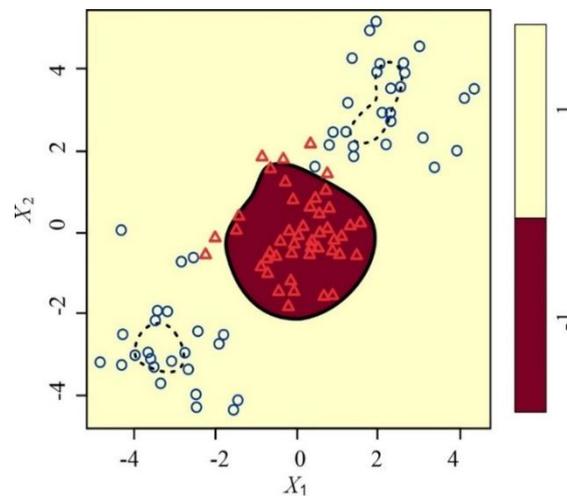
Clasificación con el método de las Máquinas de Soporte Vectorial con núcleo polinomial.



Nota. En la figura N°11 se observa la clasificación utilizando el método de las máquinas de soporte vectorial con núcleo polinomial de grado $d = 3$, aplicado a los datos no lineales.

Figura 12

Clasificación con Máquinas de Soporte Vectorial con núcleo radial.



Nota. En la figura N°12 se observa la clasificación utilizando el método de las máquinas de soporte vectorial con núcleo radial.

2.2 Bases Teóricas de los Factores de Embarazo en Adolescencia.

2.2.1 La Adolescencia.

De acuerdo con la Organización Mundial de la Salud (2024), la adolescencia abarca desde los 10 hasta los 19 años, siendo un periodo crucial de crecimiento tanto físico como psicosocial en términos sociales y reproductivos.

Durante la adolescencia, las jóvenes experimentan cambios físicos, sociales y emocionales mientras transitan de la niñez a la adultez, consolidando su independencia social y económica. Este período se caracteriza por la capacidad reproductiva y la evolución de patrones psicológicos.

Las adolescentes mujeres experimentan cambios hormonales, crecimiento físico, desarrollo cognitivo y forman su identidad, es un momento crucial para adquirir habilidades de autonomía, toma de decisiones y responsabilidades.

La OMS enfatiza la calidad de la educación sexual, acceso a métodos anticonceptivos y apoyo emocional para las adolescentes mujeres y varones.

La palabra adolescencia proviene de adolece donde el significado es carencia o falta de madurez.

En el Perú, el Ministerio de Salud (MINSA), indica que la etapa adolescente percibe entre los 12 años hasta los 17 años y 11 meses, MINSA ha establecido normas para la atención integral de salud adolescentes, considerando sus necesidades específicas, en la Resolución Ministerial N° 1001-2019/MINSA con el nombre, Norma técnica para la atención integral de Salud de Adolescentes (Ministerio de Salud, 2019).

Etapas de la adolescencia:

- Adolescencia temprana. Comprende entre la edad de 11 a 13 años, grandes cambios funcionales (menarquía), pierde interés por los padres, tiene amistades con personas del mismo sexo, no controlan impulsos y se preocupan por su apariencia física.
- Adolescencia media. Entre los 14 y los 16 años, se alcanza la plenitud del crecimiento físico y emocional. Durante este periodo, se intensifica la conexión con los amigos, surgen desafíos con los padres, se inicia la exploración sexual y se pueden manifestar comportamientos desafiantes y arriesgados.
- Adolescencia tardía. En el rango de edades de 17 a 19 años, los jóvenes empiezan a sentirse cómodos con su apariencia física, fortalecen la conexión con sus padres y se distancian un poco de sus amistades.

2.2.2 Embarazo en la Adolescencia

La maternidad temprana representa un reto a nivel global, con repercusiones importantes en la salud, la comunidad y en el desarrollo económico.

Según la Organización Mundial de la Salud (2024), alrededor de 21 millones de jóvenes mujeres entre 15 y 19 años se embarazan anualmente en áreas en desarrollo, y

aproximadamente 12 millones de ellas llegan al parto. Asimismo, señala que, durante el año 2019, un 55% de los embarazos no planificados en jóvenes culminan en interrupción voluntaria del embarazo.

En el año 2022, en Perú, según los datos de la Encuesta Demografía y de Salud (ENDES) 2022, el 9.2% de jóvenes de entre 15 y 19 años experimentaron la maternidad o el embarazo, lo que representa una ligera disminución en comparación con los años anteriores. De acuerdo a las áreas urbanas y rurales. En las zonas urbanas se mantiene en 6.8% (Instituto Nacional de Estadística e Informática, 2022)

Para el año 2022 según MINSA ha reportado 30,000 embarazos en adolescentes, esta cifra supera a los demás años, las regiones con mayor embarazo son Lima, Loreto, Piura, La Libertad y Cajamarca.

Las madres adolescentes enfrentan un mayor riesgo de complicaciones como eclampsia, endometritis puerperal e infecciones sistémicas. Los neonatos están en riesgo de presentar bajo peso al nacer, ser prematuros y desarrollar condiciones neonatales severas.

Las adolescentes con menos educación o bajo estatus económico tienden a tener tasas más altas de embarazo, también el acceso limitado a anticonceptivos, las tasas de embarazo adolescente varían menormente entre países y en sus regiones.

Una forma de prevenir son los programas de embarazo adolescente en los centros educativos, el cual minimizo el porcentaje de adolescentes embarazadas, pero no todos los adolescentes tienen acceso a estos programas por diferentes motivos geográficos, nivel económico, social y cultural. Por lo que explica un alto porcentaje de adolescentes embarazadas.

Cuando las jóvenes se embarazan, muchas veces ya han finalizado su educación básica o avanzada y se ven en la necesidad de buscar empleo para mantenerse a sí mismas,

lo cual tiene un impacto significativo a largo plazo en su familia y en la comunidad en la que viven. Las jóvenes que son madres tienen una mayor probabilidad de caer en la pobreza.

El embarazo adolescente contribuye a la perpetuación de ciclos de pobreza, exclusión y marginación, en los cuales los hijos e hijas tienen a enfrentar condiciones de pobreza, deficiente salud y a experimentar embarazos precoces.

La tasa de embarazo a nivel mundial muestra que las adolescentes sin educación tienen mayores probabilidades de quedar embarazadas en comparación con las adolescentes con educación secundaria. Las adolescentes de familias en el quintil de riqueza más bajo también tienen más probabilidades de ser madres adolescentes que aquellos de familias en el quintil de riqueza más alto del país.

En las adolescentes embarazadas afecta gravemente su salud, dificulta su desarrollo psicosocial y aumenta el riesgo de muerte. En nuestra nación, se presentan múltiples desafíos durante la gestación y el alumbramiento que constituyen la causa principal de defunción en el rango de edades de 15 a 19 años. Las jóvenes madres que dan a luz antes de cumplir quince años enfrentan un mayor peligro de morir durante el parto. En naciones con ingresos moderados y bajos, las madres menores de quince años enfrentan un peligro de fallecimiento durante el parto que es el doble que el de las madres mayores, mientras que las madres de 20 años tienen un menor riesgo de morir durante el parto. Cada año, aproximadamente 16 millones de mujeres jóvenes de 15 a 19 años y dos millones de jóvenes menores de quince años experimentan un embarazo.

La adolescencia y el embarazo forman una combinación arriesgada que refleja desequilibrios causados por la falta de desarrollo físico y emocional.

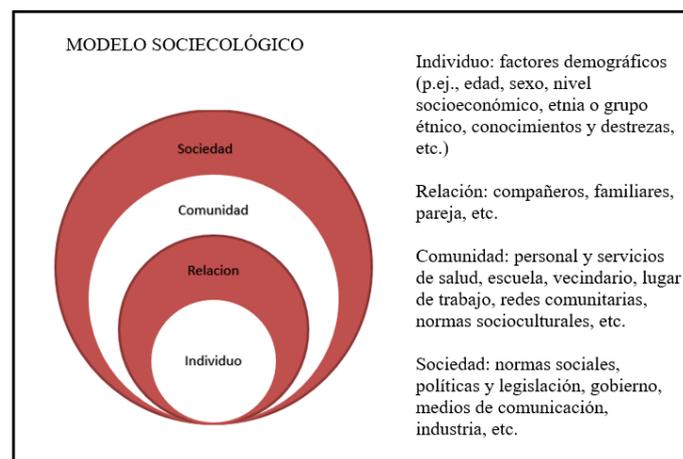
En esta etapa de la adolescencia es especialmente complicada porque corresponde a un periodo de madurez, donde corren el riesgo de fracasar o triunfar y esto marcará el rumbo de su futura vida.

2.2.3 Factores para el Embarazo Adolescente.

Según el informe sobre Violencia Mundial y Salud Global, se han descubierto diversos elementos que influyen en el embarazo en la adolescencia, conocidos como factores de género. Estos factores, también llamados causas fundamentales de la desigualdad entre géneros, incrementan la vulnerabilidad y la probabilidad de embarazo durante la adolescencia (Krug y otros, 2002) . Estos factores se encuentran en el ámbito personal, interpersonal y comunitario, y pueden perdurar en medidas institucionales o sistémicas, como pautas de atención, leyes y normativas.

Figura 13

Forma esquemática de los factores institucionales y de género relacionados con el embarazo en adolescentes.



Nota. En la figura N°13 se observa el modelo sociológico según el informe mundial sobre la salud y violencia en el mundo realizado en Ginebra (Krug y otros, 2002).

a) Factor Individuo:

- Edad
- Estado civil
- Educación
- Motivo de embarazo

- Etnia o grupo étnico
 - Conocimientos
 - destrezas
- b) Factores relaciones sociales
- Presión de Compañeros
 - Presión Familiares
 - Presión de pareja
 - Comunicación con adultos
 - Apoyo familiar
 - Violación sexual
- c) Factores culturales
- Educación sexual
 - Servicios de salud
 - Apoyo comunitario
 - Norma y valores
- d) Factor sociedad
- Oportunidad Laboral
 - Violencia de genero
- e) Factor económico
- Ingreso económico
 - Trabajo del padre
 - Trabajo de la madre
 - Ahorro económico
 - deuda
 - Nivel económico

- Logros laborales
- Negocio
- Agua
- Luz
- Internet

2.3 Antecedentes de la Investigación

2.3.1 Antecedentes Nacionales

Medina (2018), realizó el trabajo de Factores que condicionan el embarazo de adolescentes en Miraflores y Ramon Castilla, distrito de San Miguel en el año 2017, Universidad Nacional De Huancavelica, Perú. cuyo objetivo fue analizar e identificar los factores que condicionaron el embarazo adolescente en Miraflores y Ramón Castilla, durante 2017, para ello el estudio es de tipo interpretativo, cualitativo y no se estableció una relación de causa y efecto. realizan un estudio de los diferentes factores en el embarazo en adolescente como son económicos, sociales, culturales, actitudes y conocimientos, así como también sobre salud sexual y reproductiva, la investigación se realizó a los adolescentes mujeres embarazadas entre las edades de 15 a 17 años, la mayoría vive con sus padres y se iniciaron a su actividad sexual a los 14 años, obteniendo las siguientes conclusiones, lo que más influyen al embarazo adolescentes son los factores socio culturales, la preparación y las prácticas en salud sexual, mientras los que no influye son el factor económico.

Rojas (2019), realizó el trabajo de Factores de riesgo asociados con el embarazo en adolescentes del hospital nacional Hipólito Unanue, 2018, Universidad Ricardo Palma, Perú, cuyo objetivo fue Explorar los factores asociados de riesgo al embarazo adolescente en el año 2018 en el centro Hospitalario Nacional Hipólito Unanue. La metodología en el estudio fue de tipo observacional, analítico y correlacional, con un diseño de caos y controles

de corte transversal. Para ello El tipo de estudio en los casos de jóvenes embarazadas y controles de jóvenes no embarazadas. La investigación se realizó con una muestra compuesta por 260 jóvenes, de los cuales 86 jóvenes en porcentajes 33,1% respondieron a los casos y 174 jóvenes en porcentajes 66,9% al grupo control. Obteniendo las siguientes conclusiones el factor del estudio resulta ser un factor de peligro individual para las adolescentes embarazadas y el factor familiar donde es el abandono de ambos o un padre y también el grado de educación del padre. La menstruación y el debut sexual, presión de grupo, antecedentes familiares y el nivel socioeconómico no son factores de riesgo significativos.

Prado (2020), realizó una investigación Factores sociales y familiares que predisponen el embarazo en adolescentes en el Centro de Salud Los Licenciados. Ayacucho, en el año 2019 en Universidad Nacional del Callao Perú, con el objetivo de identificar la relación entre los factores sociales y familiares asociados al embarazo en adolescente en el Centro de Salud los Licenciados, ubicado en la región de Ayacucho, Perú. El enfoque utilizado en la investigación es cuantitativo, con un diseño no experimental y correlacional. La presente investigación doctoral se lleva a cabo con una muestra de 70 adolescentes, que reciben atención en el Centro de Salud Los Licenciados, utilizando un método de selección aleatoria simple. Se empleó un cuestionario con 12 preguntas para explorar los aspectos familiares y sociales. Los datos fueron analizados utilizando Excel y SPSS. De acuerdo con los datos recopilados, se encontró que el 42.8% de los adolescentes experimentan una situación familiar regular, mientras que el 27.1% de ellos se encuentran en estado de embarazo. Por otro lado, el 35.7% de los jóvenes tienen un entorno familiar desfavorable, de los cuales el 21.4% están embarazadas. Además, se identificó que el 21.4% de los adolescentes disfrutaban de un ambiente familiar positivo, de los cuales solo el 1.4% están embarazadas. Se examina la influencia del entorno social, revelando que el 42.9% disfruta

de un entorno social favorable, mientras que el 30% no está en estado de gestación. Por otro lado, el 34.3% experimenta un entorno social regular, de los cuales el 18.6% están embarazadas y el 15.7% no lo están. En contraste, el 22.9% enfrenta un entorno social desfavorable, con un 18.6% de embarazadas y un 4.3% que no lo están. Los resultados mostraron que los parientes y circunstancias sociales tienen un impacto importante en el embarazo de las adolescentes.

Ángeles (2022), realizó la investigación de Detección de carcinoma basocelular utilizando red neuronal convolucional y máquinas de soporte vectorial, en la Universidad Nacional Mayor de San Marcos, con el propósito de crear un sistema automático para detectar carcinoma basocelular, se empleó un enfoque innovador que combinó redes neuronal convolucional y máquinas de soporte vectorial. Este estudio utilizó metodologías cuantitativas y experimentales, centrándose su análisis en la clasificación de imágenes dermatológicas de carcinoma basocelular. El estudio destacó el uso de las redes neuronales convolucionales para extraer características relevantes de las imágenes de dermatoscopia, las máquinas de soporte vectorial fueron empleadas para clasificar las imágenes en categorías de benignidad o malignidad. El uso de máquinas de soporte vectorial permitió una clasificación más precisa, optimizando el margen entre las clases para mejorar la exactitud del diagnóstico. Tras llegar a estas conclusiones, se ha conseguido poner en marcha un innovador modelo que combina redes neuronales convolucionales y máquinas de soporte vectorial, el cual clasifica de manera eficaz las imágenes del carcinoma basocelular, lo que facilita al profesional médico cargar una imagen de dermatoscopia en el sistema y obtener un diagnóstico automático de esta afección. Lo que contribuye a mejorar la precisión y la velocidad del proceso diagnóstico.

Cárdenas (2023), realizó la investigación con título Factores familiares y adolescentes embarazadas que fueron atendidas en la Microred de Salud del distrito de San

Jerónimo, 2021, en la Universidad Nacional de San Antonio Abad del Cusco. Facultad de Ciencias de la Salud, Perú. Se buscó descubrir la conexión entre los factores familiares y el embarazo en la adolescencia. Este estudio es transversal, correlacional y no experimental. En el proceso de análisis, se emplean tablas de contingencia que contienen frecuencias y valores, así como la prueba estadística de Wilcoxon. Los datos revelaron que el promedio de edad de las adolescentes embarazadas es de 15 a 17 años en un 60%, el 46.7% tienen estudios de secundaria incompletos, y el 60% provienen de áreas rurales. El coeficiente de Wilcoxon demostró una fuerte correlación negativa entre el embarazo en adolescentes y el factor familiar, con un valor de $r = -0.790$ y $p = 0.0$, siendo significativamente menor que 0.05. Se observa una conexión poderosa y decreciente en ambos sentidos, con un nivel de importancia de 0.0. Al analizar la relación entre la educación sexual y el embarazo en adolescentes, se revela una conexión moderada y negativa $r = -0.62$, $p = 0.0 \leq 0.05$. lo que sugiere una fuerte relación inversa entre la comunicación de padres y adolescentes, con un coeficiente de correlación de -0.75 y un nivel de significancia de $p = 0.0$, el cual es menor o igual a 0.05. en conclusión, el estudio determinó que los factores familiares están significativamente relacionados con el embarazo adolescente en el distrito de San Jerónimo, lo que resalta la importancia de los elementos familiares en la prevención de este fenómeno

2.3.2 Antecedentes Internacionales

Pinzón, Ruiz, Aguilera y Abril (2018), realizó la investigación tiene el título Factores asociados al inicio de vida sexual y el embarazo en adolescentes en la ciudad de Colombia. En la Universidad del Rosario, Bogotá. Cuyo objetivo de la investigación fue analizar los factores de embarazo adolescente en la ciudad de Colombia, durante el año 2008 y 2014 donde indica que el porcentaje de 23.4% de los nacimientos fueron embarazos de adolescentes con importantes consecuencias sociales, emocionales y biológicas tanto para la adolescente madre como para el niño. Se utilizó la metodología es de corte transversal y

componente analítico dónde tomaron una muestra de 13,313 mujeres entre las edades de 13 y 19 años las cuales fueron encuestadas por la Encuesta Nacional de Demografía y Salud 2010. Donde se realizaron análisis de regresión logística multinomial y bivariados, para encontrar los factores individuales, familiares y de comportamiento que influyen en las adolescentes embarazadas. Llegando a la conclusión que a través de un análisis multivariado encontró que los factores relacionados con un mayor riesgo de embarazo adolescente incluyen tener mayor edad, convivir en pareja, violencia parental y el factor violencia sexual también indica asistir al colegio es un factor protector para ambos eventos. Los factores de pobreza, falta de educación y el hacinamiento están relacionados con el embarazo adolescente.

Poovizhi (2019), realizó la investigación con título *An Efficient Skin Cancer Diagnostic System Using Bendlet Transform and Support Vector Machine* En esta investigación desarrollar un sistema de clasificación del cáncer de piel con la abreviatura SCC utilizando imágenes dermatoscópicas. Se considera un problema de clasificación con la ayuda de Bendlet Transform con la abreviatura BT como características y máquinas de soporte vectorial como clasificador. La metodología de la investigación fue cuantitativa, con un diseño experimental y descriptivo-correlacional en esta investigación se emplea un sistema de extracción de características basado en representación direccional que clasifica con precisión la curvatura, la ubicación y la orientación de las imágenes dermastoscópicas. Luego, diseñan dos clasificadores con el método de las máquinas de soporte vectorial para la clasificación de los datos. Llegando a la conclusión la muestra que el sistema de clasificación de cáncer con la abreviatura SCC que utiliza el clasificador BT y el método de las máquinas de soporte vectorial provee mayor precisión, especificidad y sensibilidad en comparación con otras técnicas como las wavelets y Shearlets. También, el sistema realizado

no requiere ninguna intervención manual para el diagnóstico, lo que mejora la eficiencia y la precisión del proceso.

Galindo, Perdomo y Figueroa (2020), realizó la investigación con el título Estudio comparativo de las máquinas de soporte vectorial multiclase con el sistema de inferencia neuro-difuso autoorganizado y redes neuronales para los problemas de clasificación. El objetivo es contextualizar un sistema neuro difuso autoorganizado con la abreviatura de SONFIS² y comparar su desempeño en problemas de clasificación con otros clasificadores universales como las redes neuronales artificiales y las máquinas de soporte vectorial, la metodología fue experimental con datos aleatorios. Cuando se aplica el algoritmo SONFIS a 3 problemas de clasificación: iris de Fisher, cáncer de mama y actividad humana, para consecutivamente comparar los datos conseguidos con los clasificadores universales en los problemas de clasificación como son las redes neuronales artificiales y el método de las máquinas de soporte vectorial con multiclase. Donde obtienen que SONFIS posee rendimiento similar y en algunos casos mejor que las redes neuronales artificiales y el método de máquina de vectores de soporte en problemas de clasificación, y ponen condiciones para entender el problema estructuralmente. Llegando a la conclusión, que los SONFIS nace como un método alternativo para analizar los problemas de clasificación, y este tiene un desempeño equivalente y en otros casos superior que otras metodologías, es más efectiva y más utilizada para este tipo de problemas, también tiene la ventaja de permitir la abstracción de discernimiento acerca de los problemas. El costo computacional SONFIS utiliza el mismo tiempo que los otros algoritmos en otros casos en menor tiempo, obteniendo resultados análogos en términos de los indicadores estimados. La prueba mostrada en los casos de estudio mostrados es placentera y muestra el potencial de SONFIS, El rendimiento

² El método SONFIS, Sistema de Inferencia Neuro-Difuso Auto Organizado (Figueroa, 2015) busca generar reglas para un sistema difuso usando redes neuronales

en tareas de clasificación es importante. Las Técnicas en los algoritmos genéticos, o el método de inteligencia de enjambres se pueden combinar con SONFIS para luego seleccionar los parámetros más altos adecuados para cada tipo de problema.

Lopera, Zuluaga y Jaramillo (2021) realizó la investigación con el título Máquinas de vectores de soporte en el descubrimiento de biomarcadores en experimentos *in vitro* e *in vivo* de exposición a organoclorados, en el Instituto Tecnológico Metropolitano, Colombia, la finalidad de este estudio fue investigar la eficacia del enfoque de las máquinas de soporte vectorial para categorizar metabolitos, así como para descubrir compuestos relevantes en análisis de exposición a organoclorados. Antes de llevar a cabo dos experimentos, se analiza inicialmente la presencia de organoclorados en las células. Posteriormente, analizan la presencia de pesticidas en muestras de suero de individuos que trabajan en la agricultura. Se contrastó la eficacia de las máquinas de soporte vectorial con la de los análisis discriminantes mínimos cuadrados parciales. Se realizaron pruebas en cuatro funciones Kernel en las máquinas de soporte vectorial, y la precisión de los dos métodos se midió a través de una validación cruzada con k iteraciones. Al llegar a la conclusión final del estudio, se evidenció que tanto el enfoque de las máquinas de soporte vectorial como el discriminante de mínimos cuadrados parciales resultan ser herramientas idóneas para examinar la información obtenida en experimentos *in vitro* bajo condiciones controladas. Sin embargo, se observó que el discriminante de mínimos cuadrados parciales enfrentó dificultades al analizar los datos de estudios *in vivo* con situaciones no controladas y exposiciones no lineales a organoclorados. En lo que respecta a predecir clases en datos de estudios en tiempo real, se encontró que el método de las máquinas de soporte vectorial fue más efectivo en su capacidad predictiva que el discriminante de mínimos cuadrados. Asimismo, el núcleo muestra un desempeño determinado mediante el enfoque de las máquinas de soporte vectorial, el cual puede emplearse para obtener un conjunto óptimo de estas características

más destacadas en relación con la exposición a pesticidas. Asimismo, la carga computacional de las máquinas de soporte vectorial es reducida y se está convirtiendo en una herramienta valiosa para descubrir biomarcadores, incluso en situaciones con un número limitado de muestras. Además, se puede ver como un enfoque sólido para examinar información obtenida de experimentos en directo y estudios en laboratorio.

CAPITULO III

HIPOTESIS Y VARIABLES

3.1 Hipótesis

3.1.1 *Hipótesis General*

El método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022, es el polinómico.

3.1.2 *Hipótesis Específicas*

- a) Los factores económico y educativo están asociados significativamente al embarazo adolescentes.
- b) El método de máquinas vectoriales con Kernel polinómico presenta mayor sensibilidad y especifica con respectos a máquinas de base lineal y radial.

3.2 Identificación de Variables e Indicadores

Variable independiente: Factores asociados al embarazo adolescente.

Variable dependiente: Embarazo adolescente.

Dimensiones:

- Factores personales
- Factor socio económico
- Factor familiar
- Factor condición económico
- Factores residencia
- Factor de atención Salud

3.3 Operacionalización de Variables

Tabla 1

Operacionalización de las variables.

Variable	Definición Conceptual	Definición Operacional	Indicadores
Embarazo adolescente	Es el período en el cual un feto se desarrolla en el vientre o útero de una mujer adolescente	Se establece en base a análisis de orina	Presencia y ausencia

Fuente: Elaboración propia

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores
Factores asociados al embarazo adolescente	Las condicionantes asociadas al embarazo	El embarazo adolescente, está en función de los factores personales, socio económico, condición económica, familiares, , residencia y atención de salud.	Factores personales	Edad
				Peso
				Talla
				IMC
			Factor socio económico	Planificación
				Tipo de gestión
				Nivel Educativo
				Grado de instrucción
				Fase Regular
				Quintil Pobreza
Factor condición económico	Pobreza Monetaria			
	Apoderado			
Factor familiar	Distrito			
	Migración de distrito			
Factor residencia	Microredes			
	Categoría centro de salud por lugar			
Factor de atención salud				

Fuente: Elaboración propia

CAPITULO IV

METODOLOGÍA

4.1 **Ámbito de Estudio: Localización Política y Geográfica**

La población es de 11200 adolescentes mujeres afiliadas al SIS (Seguro Integral de Salud) de la Red de servicios de Salud La Convención.

La capital de la región de Santa Ana es Quillabamba y es una de las quince regiones que conforman la provincia de La Convención en el departamento de Cuzco al sur del Perú. Predomina la Iglesia católica que es parte del Vicariato apostólico de Puerto Maldonado.

El distrito de Santa Ana su creación es en la fecha, 21 de junio de 1825 con el decreto dado por el Libertador de América Simón Bolívar.

Santa Ana se encuentra a una altitud de 1 050 msnm, su superficie es de 359.4 km², con coordenadas 12°52'38"S 72°42'01"O.

Tiene una población total según los datos del INEI 2009 de 33,230 habitantes y una densidad de 77,9 habitantes/km², su abigeo es 080901.

La provincia La Convención está conformada por 18 distritos los cuales son:

Tabla 2*Distritos de la provincia de La Convención 2022.*

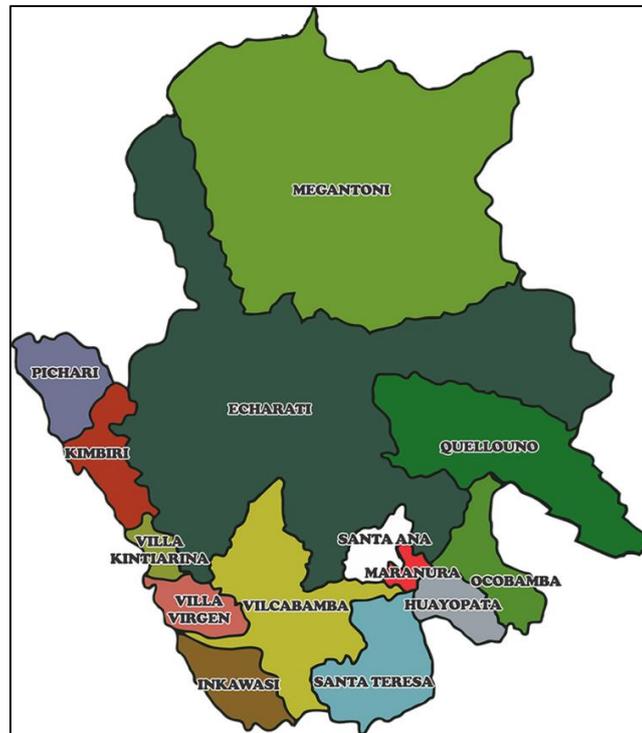
N°	Distrito	Capital
1	Santa Ana	Quillabamba
2	Echarate	Echarate
3	Huayopata	Huyro
4	Maranura	Maranura
5	Ocobamba	Kelcaybamba
6	Quellouno	Quellouno
7	Kimbiri	Kimbiri
8	Santa Teresa	Santa Teresa
9	Vilcabamba	Lucma
10	Pichari	Pichari
11	Inkawasi	Amaybamba
12	Villa Virgen	Villa Virgen
13	Villa Kintiarina	Villa Kintiarina
14	Megantoni	Comunidad Camisea
15	Kumpirushiato	Kepashiato
16	Cielopunco	Chirumpiari
17	Manitea	Tawantinsuyo
18	Unión Asháninca	Mantaro

Fuente: INEI

Nota. En la tabla N°2 se observa los distritos de la provincia de La Convención según el directorio nacional de gobiernos regionales, municipales provinciales y distritales (2022).

Figura 14

Mapa de la provincia La Convención con 14 distritos documento de gestión 2017



Nota. En la figura N°14 se observa los 14 distritos de la provincia La Convención según el documento de gestión de la municipalidad Provincial La Convención del año (Municipalidad Provincial de La Convención, 2017).

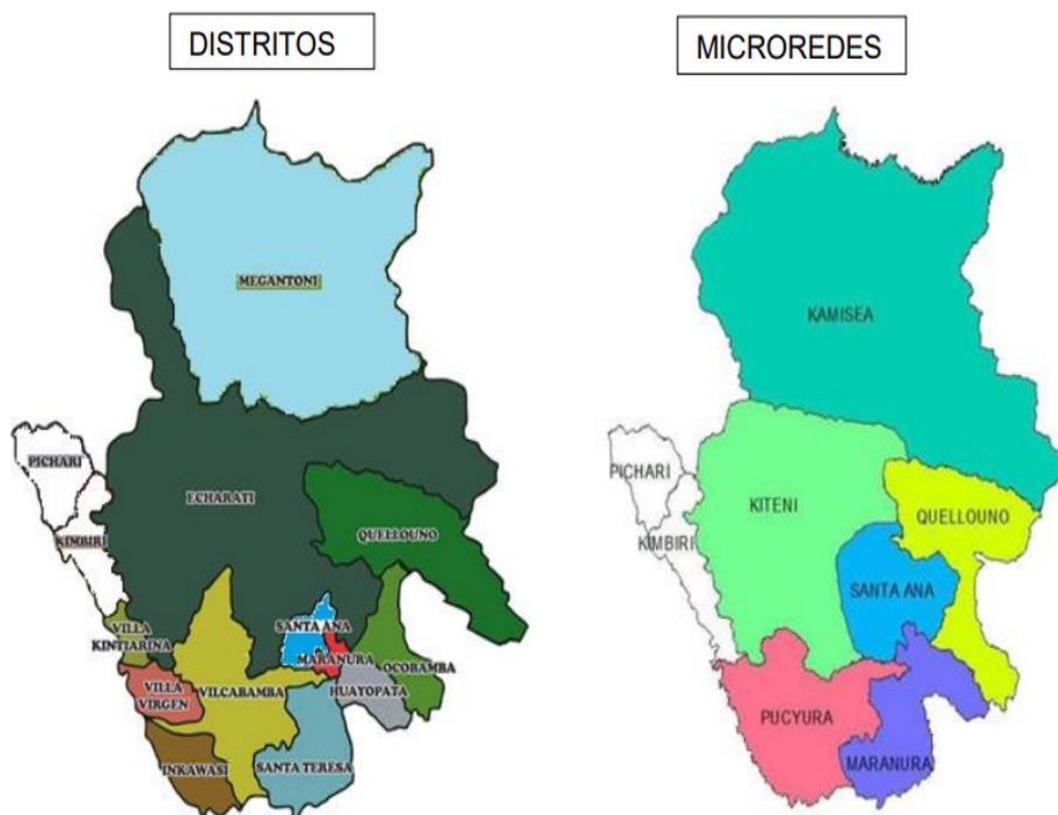
Figura 15*Mapa de la provincia de La Convención*

Nota. En la figura N°15 se observa la ubicación de la provincia La Convención según el documento de gestión de la Municipalidad Provincial de La Convención (Municipalidad Provincial de La Convención, 2017).

La Red de Servicios de Salud de La Convención está conformada por 7 Microredes de servicio de Salud las cuales son: Santa Ana, Maranura, Pucyura, Quellouno, Kamisea, Kiteni, Kimbiri-Pichari.

Figura 16

Mapa de la provincia de La Convención por distritos y Microredes de Salud.



Nota. En la figura N°16 se observa las Microredes de Salud y distritos de la provincia La Convención según el Plan operativo institucional 2021 (Red Salud La Convención, 2021).

4.2 Tipo y Nivel de Investigación.

El tipo de investigación, según el propósito de la investigación es básica y es una investigación no experimental.

La investigación corresponde a un nivel explicativo puesto que se pretende explicar las causas de embarazo en adolescentes.

El diseño de investigación es no experimental, transversal debido a que no se manipulan ninguna variable solo se registra esta tal como se presenta en la realidad.

La investigación se lleva a cabo de manera transeccional o transversal, recopilando información en un instante específico, en un momento singular (Hernandez Sampiery et al.,

2014). Busca detallar las variables y examinar cómo se afectan y se relacionan entre sí en un instante específico.

4.3 Unidad de Análisis

Para la presente investigación la unidad de análisis está conformada por los adolescentes mujeres de edad 12 a 17 años de edad afiliados en el SIS de la Red de Servicios de Salud La Convención del año 2022.

4.4 Población de Estudio

La población son todos los adolescentes mujeres afiliados en el SIS de la Red de Servicios de Salud La Convención, según el reporte de estadística se tiene un total de 11200 registros de estas 8537 cumplen con los criterios de integridad de información, Mas del 90% de los indicadores completos respecto a peso y talla.

4.5 Tamaño de Muestra

Muestra

$$n = \frac{Z^2 * p * (1 - p) * N}{\epsilon^2 * (N - 1) + Z^2 * p * (1 - p)} \quad (34)$$

N=8537

Nivel de Confianza 95%, (Z=1.96)

Margen de Error 5% ($\epsilon = 0.05$)

P=Proporción de Éxito

P=0.50

$$n = \frac{1.96^2 * 0.50 * (1 - 0.500) * 8537}{0.05^2 * (8537 - 1) + 1.96^2 * 0.50 * (1 - 0.50)}$$

$$n = \frac{8198.9348}{22.3004} = 367.7$$

El tamaño de muestra es $n \geq 367.7$

Se consideró una muestra 2993 mayor a lo establecido en el tamaño de muestra para garantizar la confiabilidad de los resultados y además estas técnicas de máquinas de soporte vectorial para su modelamiento requieren muestras de tamaño grande.

4.6 Técnicas de Selección de Muestra

La muestra se seleccionó mediante una muestra no probabilística por conveniencias y voluntaria considerándose un total 2993 adolescentes registrados en sistema de afiliados SIS en la Red de Servicios de Salud La Convención. La técnica de máquinas de soporte vectorial requiere una gran cantidad para su aplicación.

4.6.1 Técnica

El estudio se basa en un análisis documental referido a las historias clínicas y una ficha de información sociodemográfica de cada adolescente, por tanto, la técnica de investigación es el documental.

4.6.2 Instrumento

El instrumento más adecuado para el registro de información que se ha diseñado es la historia clínica que considera también información sociodemográfica.

Recopilar información implica crear un minucioso plan de acciones que nos guíen en la obtención de datos con un objetivo concreto.

4.7 Validez y Confiabilidad de Instrumentos

Para la validación del instrumento se aplicó la técnica de juicio de expertos esta se basa en opiniones de especialistas con conocimientos en el área de investigación.

Tabla 3*Validez y confiabilidad de Instrumento*

Nro.	Expertos	% de Validación
1	Dra. APARICIO ARENAS, Karla Zelmira	90%
2	Mtro. ZUÑIGA BLANCO, Arturo	92%
3	Mag. LAZO CHUQUIHUAYTA, Henry	91%
Promedio		91%

Fuente: Elaboración propia

Se aplicó prueba piloto de 30 adolescentes, obteniéndose un alfa de Cronbach de 0.87 lo cual indica que el instrumento presenta alta confiabilidad.

Tabla 4*Confiabilidad.*

Variable	Alfa de Cronbach
Factores en el embarazo adolescente	0.87

Fuente: Elaboración propia

En la tabla N°4 los instrumentos presentan confiabilidad alta.

4.8 Técnicas de Análisis e Interpretación de la Información.

Los datos que se obtuvieron tras la aplicación de los instrumentos de recolección son organizados en Excel y luego en el programa RStudio versión 4.3.2, antes de utilizar el modelo Maquinas de Soporte Vectorial los datos se procesa de la siguiente forma para asegurar que el análisis sea confiable.

Se realizó la limpieza de datos durante la recopilación de información donde se encontraron datos faltantes o inconsistencias, luego la imputación reemplazando los valores faltantes con la mediana o moda dependiendo de la variable es numérica o categoría, de esta manera se evitó perder información y sesgos en el modelo. También se identificó valores

atípicos esto se eliminó cuidadosamente. Se normalizaron los datos para que sus valores estén dentro del mismo rango.

Se particiona los datos de entrenamiento y prueba. Se analizó con el modelo de Máquinas de soporte vectorial que es una técnica robusta para clasificar los datos al encontrar un hiperplano que mejor clasifique a las adolescentes embarazadas y no embarazadas, se utiliza el Kernel lineal, kernel radial y kernel polinómico, luego se elige el kernel más adecuado mediante un proceso de prueba y error, evaluando el desempeño en varias configuraciones, se utiliza el parámetro C para controlar la penalización por errores de clasificación, se aplicó la validación cruzada dividiendo en varias partes, entrenando el modelo en algunas y probándola en las restantes para evaluar cómo funciona el modelo en datos nuevos y así reduciendo el riesgo de sobreajuste,

Se mide la eficacia de las máquinas de soporte vectorial con kernel utilizando las métricas como la precisión sensibilidad y especificidad, luego se evaluó el balance entre falsos adolescentes embarazadas y falsos adolescentes no embarazadas con la curva AUC-ROC, donde ROC representa gráficamente la tasa de verdadero adolescentes embarazadas frente a la tasa frente a la tasa de falsos adolescentes embarazadas y AUC (área bajo la curva) es la medida del rendimiento general de un clasificador representado por la curva ROC , luego identificamos que factores están asociados al embarazo adolescente y finalmente se muestra en tablas y graficas los resultados obtenidos.

4.9 Técnicas para Demostrar la Verdad o Falsedad de la Hipótesis Planteada

Tabla 5

Plan de análisis de datos.

Hipótesis de investigación	El método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022, es el polinómico.
Hipótesis nula	Los factores económico y educativo están asociados significativamente al embarazo adolescentes El método de máquinas vectoriales con Kernel polinómico presenta mayor sensibilidad y especifica con respecto a máquinas de base lineal y radial
Hipótesis alterna	Los factores económico y educativo no están asociados significativamente al embarazo adolescentes El método de máquinas vectoriales con Kernel polinómico no presenta mayor sensibilidad y especifica con respecto a máquinas de base lineal y radial
Nivel de significancia	5%
Prueba estadística	Máquinas de soporte vectorial
Regla de decisión	Si $p \leq 0.05$ se acepta la hipótesis alterna Si $p > 0.05$ se acepta la hipótesis nula

Fuente: Elaboración propia

CAPITULO V

RESULTADOS Y DISCUSIÓN

5.1 Procesamiento, Análisis, Interpretación y Discusión de Resultados

Con la finalidad de describir el método de máquina de soporte vectorial que presenta mayor exactitud en los factores de embarazo de adolescentes en el distrito de Santa Ana 2022 se consideró una muestra no probabilística por conveniencias y voluntaria considerándose un total 2993 adolescentes registrados en sistema de afiliados SIS en la Red de Servicios de Salud La Convención.

Resumen de la base de datos.

Tabla 6

Tipo de Gestión.

Tipo de Gestión	Frecuencia	Porcentaje
Privado	54	2%
Publico	2749	92%
NA	190	6%
Total	2993	

Fuente: Elaboración propia

En la tabla N°6 se observa que, el 92% de las adolescentes mujeres se encuentran vinculados al tipo de gestión público, solo el 2% de las adolescentes mujeres al tipo de gestión privado y el 6% de las adolescentes mujeres no reportan el tipo de gestión.

Tabla 7*Nivel Educativo*

Nivel Educativo	Frecuencia	Porcentaje
Primaria	709	24%
Secundaria	2072	69%
Básica Alternativa Avanzado	22	1%
NA	190	6%
Total	2993	

Fuente: Elaboración propia

En la tabla N°7 se observa que, el 69% de las adolescentes mujeres se encuentran vinculado al nivel educativo secundaria, solo el 1% de las adolescentes mujeres al nivel educativa básica alternativa avanzado y el 6% de las adolescentes mujeres no reportan el nivel educativo.

Tabla 8*Grado de Instrucción*

Grado de Instrucción	Frecuencia	Porcentaje
Primaria(3ro a 6to)	709	24%
1ro secundaria	536	18%
2do secundaria	502	17%
3ro secundaria	480	16%
4to secundaria	437	15%
5to secundaria	139	5%
NA	190	6%
Total	2993	

Fuente: Elaboración propia

En la tabla N°8 se observa que, el 24% de las adolescentes mujeres se encuentran vinculados al grado de instrucción primaria (3ro y 6to), solo el 5% de las adolescentes mujeres se encuentran vinculados al grado de instrucción 5to grado de secundaria y el 6% de las adolescentes mujeres no reportan su grado de instrucción.

Tabla 9*Quintil de Pobreza.*

Quintil Pobreza	Frecuencia	Porcentaje
Extrema Pobreza	2366	79%
Muy Pobre	354	12%
Pobre	273	9%
Total	2993	

Fuente: Elaboración propia

En la tabla N°9 se observa que, el 79% de las adolescentes mujeres se encuentran vinculados al quintil pobreza de extrema pobreza y solo el 9% de adolescentes mujeres al quintil de pobreza pobre.

Tabla 10*Pobreza Monetaria.*

Pobreza Monetaria	
Min	8.3
Max	45.45
NA s	179

Fuente: Elaboración propia

En la tabla N°10 se observa que el indicador de pobreza monetaria tiene un rango de valores entre un mínimo de 8.3% y un máximo de 45.45%, estos porcentajes son la proporción de las adolescentes que viven por debajo de la línea de pobreza monetaria y 179 valores faltantes en pobreza monetaria

Tabla 11*Planificación Familiar.*

Planificación	Frecuencia	Porcentaje
No	1732	58%
SI	1261	42%
Total	2993	

Fuente: Elaboración propia

En la tabla N°11 se observa que, el 58% de las adolescentes mujeres no realizaron su planificación y solo el 42% si realizaron su planificación.

Tabla 12*Distrito.*

Distrito	Frecuencia	Porcentaje
Santa Ana	273	9%
otros	2720	91%
Total	2993	

Fuente: Elaboración propia

En la tabla N°12 se observa que, el 91% de las adolescentes mujeres se encuentran vinculados a otros distritos de la provincia La Convención y solo el 9% de las adolescentes mujeres se encuentran vinculado al distrito Santa Ana.

Tabla 13*Migración.*

Migración	Frecuencia	Porcentaje
No	2373	79%
Si	430	14%
NA	190	6%
Total	2993	

Fuente: Elaboración propia

En la tabla N°13 se observa que, el 79% de las adolescentes mujeres no realizaron migración, solo el 14% de las adolescentes mujeres si realizaron migración y el 6% de las adolescentes mujeres no reportan su migración.

Tabla 14

Microred de Salud.

Microred	Frecuencia	Porcentaje
Santa Ana	700	23%
Otros	2293	77%
Total	2993	

Fuente: Elaboración propia

En la tabla N°14 se observa que, el 77% de las adolescentes mujeres se encuentran vinculado a otro Microred de Salud de la provincia La Convención y solo 23% de adolescentes mujeres a Microred de Salud Santa Ana.

Tabla 15

Categoría de salud por lugar.

Categoría Salud Lugar	Frecuencia	Porcentaje
I-1	814	27%
I-2	660	22%
I-3	463	15%
I-4	1056	35%
Total	2993	

Fuente: Elaboración propia

En tabla N°15 se observa que, el 35% de las adolescentes mujeres se encuentran vinculado a la categoría de Salud por lugar I-4 y solo el 15% de las adolescentes mujeres a categoría de Salud lugar I-3.

Tabla 16*Embarazo.*

Embarazo	Frecuencia	Porcentaje
No	2889	97%
Si	104	3%
Total	2993	

Fuente: Elaboración propia

En la tabla N°16 se observa que, el 97% de adolescentes mujeres no están embarazadas y el 3% si están embarazadas.

Tabla 17*Valores faltantes según variable*

Valores Perdidos según variable				
Tipo Gestión	Nivel Educativo	Grado de Instrucción	Pobreza Monetaria	Migración
1	2	3	5	8

Fuente: Elaboración propia

En la tabla N°17 se observa que, existen variables que presentan datos perdidos como por ejemplo tipo de gestión, nivel educativo, grado de instrucción, pobreza monetaria y si provienen de la migración.

Tabla 18*Valores faltantes en las variables.*

Variable	Frecuencia	Porcentaje
Tipo de Gestión	190	20%
Nivel Educativo	190	20%
Grado de Instrucción	190	20%
Pobreza Monetaria	179	19%
Migración	190	20%
Total	939	

Fuente: Elaboración propia

En la tabla N°18, se observa que, el 20% de valores faltantes de cada variable tipo de gestión, nivel educativo, grado de instrucción y migración, la pobreza monetaria tiene un 19% de valores faltantes y en total de datos faltantes entre estas variables es de 939 datos.

Tabla 19

Combinación de valores Faltantes.

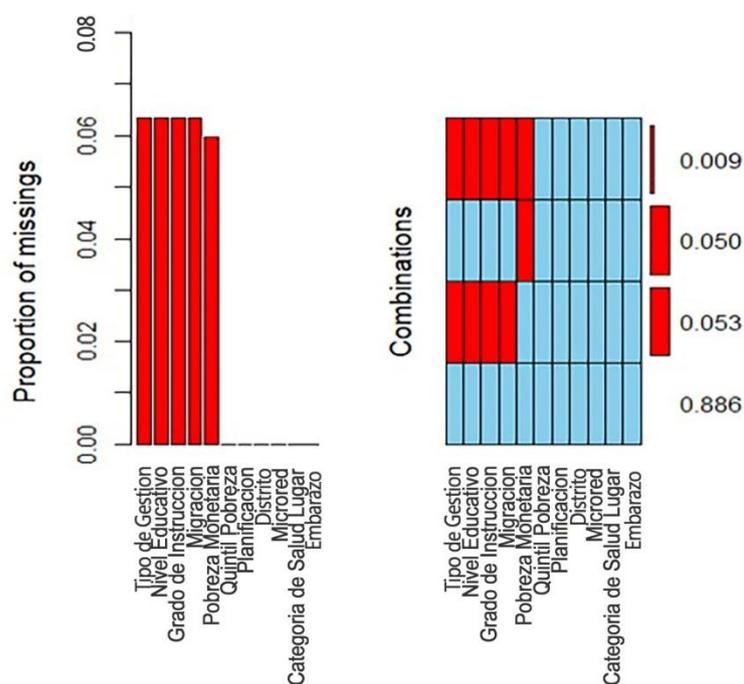
Combinación	Frecuencia	Porcentaje
0	2653	89%
1	150	5%
4	161	5%
5	29	1%
Total	2993	

Fuente: Elaboración propia

En la tabla N°19, el 89% de los registros no tienen valores faltantes, esto indica que la base de datos es predominantemente limpia, el 11% restante tiene combinaciones de valores faltantes, 5% (combinación 1) faltantes en una sola variable, 5% (combinación 4) faltantes en múltiples variables y 1% (combinación 5) es el de menor proporción.

Figura 17

Proporción de valores faltantes y combinación



En la figura N° 17 se observa, en la gráfica de la izquierda la proporción de los valores faltantes con respecto a los valores totales. El 20% de los valores faltantes son de las variables tipo de gestión, nivel educativo, grado de instrucción y migración, y el 19% de los datos faltantes son de la variable pobreza monetaria. en la gráfica de la derecha se observa una matriz de combinaciones de valores faltantes donde las áreas de color rojo representan combinaciones donde faltan valores en las variables analizadas y las áreas en color azul corresponde a los valores completos. El 88.6% de los registros no tienen valores faltantes, representado en color azul, el 5.3% de valores faltantes en múltiples variables área de color rojo combinadas, el 5% de valores faltantes distribuidos en patrones específicos y 0.9% de valores faltantes en menor cantidad.

Tabla 20

Variables ordenadas por valores faltantes.

Variables	Frecuencia	Porcentaje
Tipo de Gestión	190	6.35%
Nivel Educativo	190	6.35%
Grado de Instrucción	190	6.35%
Migración	190	6.35%
Pobreza Monetaria	179	5.98%
Quintil Pobreza	0	0.00%
Planificación	0	0.00%
Distrito	0	0.00%
Microred	0	0.00%
Categoría de Salud Lugar	0	0.00%
Embarazo	0	0.00%

Fuente: Elaboración propia

En la tabla N°20, el 6.35% de los valores faltantes correspondientes a la variable tipo de gestión, nivel educativo, grado de instrucción, migración y el 5.98% de los datos perdidos corresponde a pobreza monetaria.

Después de realizado la imputación en base a las medidas de tendencia central se tiene los datos completos para poder utilizar el método de máquinas de soporte vectorial y poder predecir cuales son las variables que están asociadas a determinar si una adolescente este embarazada o no.

5.1.1 Separación de los Datos en Entrenamiento y Prueba.

A continuación, se separa la data para el entrenamiento y para la prueba, considerando un 70% de datos para el entrenamiento y el restante para la prueba. Quedando la distribución de estos datos de la siguiente forma.

Tabla 21

Separación de los datos en Entrenamiento y Prueba.

Embarazo	Datos de Entrenamiento		Datos Prueba	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
NO	2023	97%	866	97%
SI	73	3%	31	3%
Total	2096		897	

Fuente: Elaboración propia

En la tabla N°21 se observa dos clases, mujeres adolescentes embarazadas y no embarazadas, el 97% de las mujeres adolescentes no están embarazadas del total de datos de entrenamiento y solo el 3% si están embarazadas.

5.1.2 Análisis de Máquinas de Soporte Vectorial Lineal

Para poder proporcionar los resultados se puede indicar que se usó 2096 muestras. Con 8 predictores y el variable objetivo embarazo con 2 clases: no y sí.

Dentro del preprocesamiento los datos fueron centrados y escalados antes de entrenar el modelo, así como validación cruzada de 10 fólderes para evaluar el modelo. El modelo fue ajustado utilizando diferentes valores del parámetro de regularización C , Los resultados de la precisión y el índice Kappa para cada valor de C .

Tabla 22

Máquinas de Soporte Vectorial con Kernel Lineal

Máquinas de soporte vectorial con Kernel Lineal	
Muestra	2096
Predictor	8
Clases : NO , SI	2

Fuente: Elaboración propia

En la tabla N°22 se observa que, 2096 datos de muestra, 8 predictores (variables) y 2 clases (si embarazo y no embarazo) para aplicar el método de máquinas de soporte vectorial con Kernel lineal.

Tabla 23

Matriz de confusión y estadísticas

Predicción	NO	Si
NO	850	14
SI	16	17
Accuracy	0.9665	
95% CI	(0.9513 , 0.9764)	
No información Rate	0.9654	
P-Value(Acc > NIR)	0.5476	
McNemar's Test P-Value	7.12e-08	
Sensitivity	0.86	
Specificity	0.92	
Prevalence	0.9654	
Detection Rate	0.9654	
Detection Prevalence	1	
Balanced Accuracy	0.5	
'Positive' Class	No	

Fuente: Elaboración propia

En la tabla N°23 se observa con respecto a la precisión del modelo es alta (0.9665), indicando que predice correctamente la mayoría de los casos.

La prueba de McNemar tiene un p-valor de 7.12×10^{-8} , lo que indica una diferencia significativa entre las clasificaciones falsos negativos y falsos positivos.

En el contexto de la predicción del embarazo, una sensibilidad del 86% indica que la prueba es bastante buena para detectar si una adolescente puede estar embarazada cuando realmente lo está. Esto es importante en escenarios donde no se quiere perder ningún caso

de embarazo, por ejemplo, en contextos médicos donde un diagnóstico temprano es crucial y más si son adolescentes que pueden estar siendo acosadas o violentadas.

Así mismo la especificidad fue del 92%, la prueba es muy efectiva para asegurar que las personas que no están embarazadas sean correctamente identificadas como tales. Esto es importante para evitar falsos positivos que pueden causar ansiedad y llevar a pruebas adicionales innecesarias.

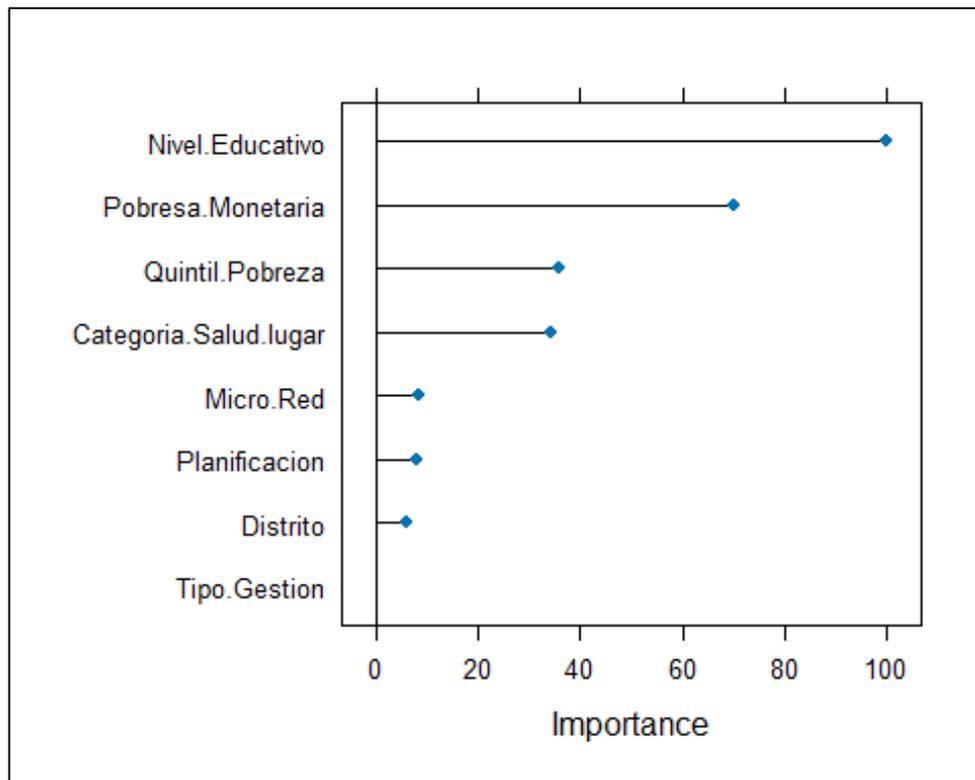
Tabla 24

ROC Curva de importancia de Variable.

Variables	Importancia(%)
Nivel Educativo	100
Pobreza Monetaria	70
Quintil Pobreza	35.692
Categoría de Salud de Lugar	34.177
Microred	8.379
Planificación	7.877
Distrito	5.899
Tipo de Gestión	0

Fuente: Elaboración propia

El nivel educativo es la variable más influyente con un 100% de importancia en la clasificación de los datos relacionados al embarazo adolescente seguido de pobreza monetaria con un 70%, las variables como quintil de pobreza y categoría de salud por lugar tienen importancias moderadas y el tipo de gestión no muestra relevancia con 0% indicando que no influye en los resultados del modelo

Figura 18*Variables más importantes*

En la figura N°18 se observa que las variables más importantes para predecir las adolescentes embarazadas, es el nivel educativo seguido de pobreza monetaria y la menos importante es la variable tipo de gestión.

De la tabla N°24 y figura N°18 Las variables más importantes que ayudan a predecir o caracterizar a las adolescentes que están embarazadas y así en lo posterior prevenir embarazos no deseados son: el nivel educativo es la variable más importante para la predicción, seguida por pobreza monetaria y quintil de pobreza, además de la categoría de establecimiento de salud y hasta la planificación.

5.1.3 Análisis de Máquinas de Soporte Vectorial Radial

Para este análisis se trabajó con 8 predictores y la variable objetivo embarazo con 2 clases: no y sí.

Dentro del preprocesamiento los datos fueron centrados y escalados antes de entrenar el modelo, así como validación cruzada de 10 fóldees para evaluar el modelo. El modelo fue ajustado utilizando diferentes valores del parámetro de regularización C , además se observa los resultados de la precisión y el índice Kappa para cada valor de C , obteniendo los mejores indicadores con $\sigma = 0.07570959$ y $C = 1$.

Tabla 25

Máquinas de Soporte Vectorial con Kernel Radial.

Máquinas de Soporte Vectorial con Kernel Radial	
Muestra	2096
Predictor	8
Clases : NO , SI	2

Fuente: Elaboración propia

En la tabla N°25 se observa que, 2096 datos de muestra, 8 predictores (variables) y 2 clases (si embarazo y no embarazo) para aplicar el método de máquinas de soporte vectorial con Kernel Radial.

Tabla 26*Matriz de confusión y estadísticas.*

Matriz de confusión y estadísticas		
Predicción	NO	Si
NO	850	8
SI	26	23
Accuracy	0.9602	
95% CI	(0.9513, 0.9764)	
No información Rate	0.9654	
P-Value (Acc > NIR)	0.5476	
Mcnemar's Test P-Value	7.118e-08	
Sensitivity	0.7401	
Specificity	0.9364	
Prevalence	0.9654	
Detection Rate	0.9654	
Detection Prevalence	1	
Balanced Accuracy	0.5	
'Positive' Class	No	

Fuente: Elaboración propia

En la tabla N°26 respecto a la precisión del modelo es alta (0.9602), lo cual indica que predice correctamente la mayoría de los casos de embarazo y no embarazo.

La prueba de McNemar tiene un p-valor de 7.11×10^{-8} , lo que indica una diferencia significativa entre las clasificaciones falsos negativos y falsos positivos.

En el contexto de la predicción del embarazo, una sensibilidad del 74% indica que la prueba es buena en cierto nivel para detectar si una adolescente puede estar embarazada cuando realmente lo está. Esto es importante en escenarios donde no se quiere perder ningún caso de embarazo, por ejemplo, en contextos médicos donde un diagnóstico temprano es crucial y más si son adolescentes que pueden estar siendo acosadas o violentadas.

Así mismo la especificidad fue del 93.64%, la prueba es muy efectiva para asegurar que las personas que no están embarazadas sean correctamente identificadas como tales. Esto es importante para evitar falsos positivos que pueden causar ansiedad y llevar a pruebas adicionales innecesarias.

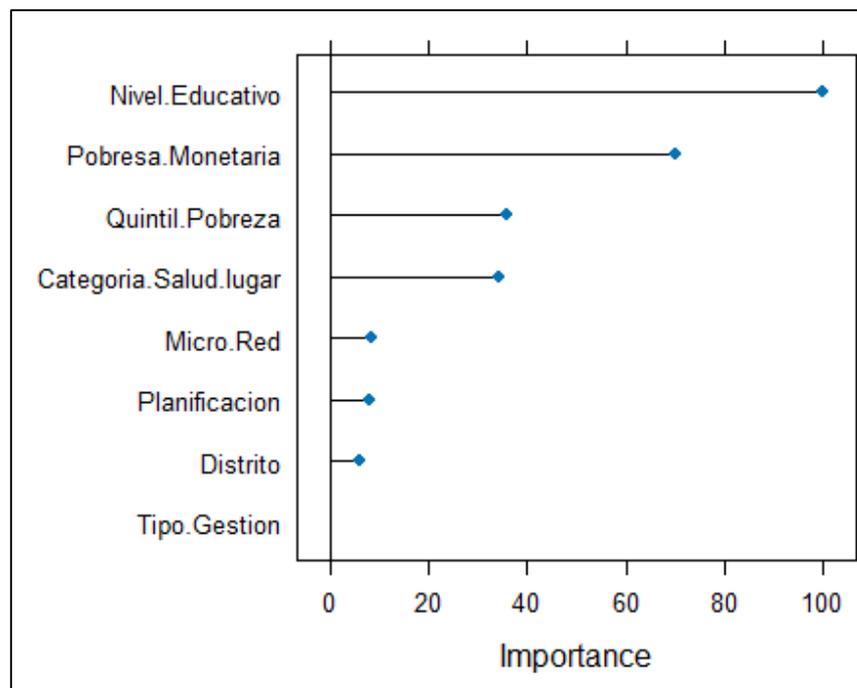
Tabla 27

ROC Curva de importancia de Variable.

Variables	Importancia (%)
Nivel Educativo	100
Pobreza Monetaria	70.028
Quintil Pobreza	35.692
Categoría de Salud de Lugar	34.177
Microred	8.379
Planificación	7.877
Distrito	5.899
Tipo de Gestión	0

Fuente: Elaboración propia

El nivel educativo es la variable más influyente con un 100% de importancia en la clasificación de los datos relacionados al embarazo adolescente seguido de pobreza monetaria con un 70.028%, las variables como quintil de pobreza y categoría de salud por lugar tienen importancias moderadas y el tipo de gestión no muestra relevancia con 0% indicando que no influye en los resultados del modelo

Figura 19*Variables más importantes*

En la figura N°19 se observa las variables más importantes para predecir el embarazo adolescente, la variable más importante es el nivel educativo seguido de pobreza monetaria y el menos importante es el tipo de gestión.

De la tabla N°27 y figura N°19 las variables más importantes que ayudan a predecir o caracterizar a las adolescentes que están embarazadas y así en lo posterior prevenir embarazos no deseados son: el nivel educativo es la variable más importante para la predicción, seguida por pobreza monetaria y quintil de pobreza, además de la categoría de establecimiento de salud y hasta la planificación.

5.1.4 Análisis de Máquinas de Soporte Vectorial Polinómica

Para este análisis se trabajó también con 8 predictores y la variable objetivo embarazo con 2 clases: no y sí.

Dentro del preprocesamiento los datos fueron centrados y escalados antes de entrenar el modelo, así como validación cruzada de 10 fólderes para evaluar el modelo. El modelo

fue ajustado utilizando diferentes valores del parámetro de regularización C , además se observa los resultados de la precisión y el índice Kappa para cada valor de C , Los valores finales utilizados para el modelo fueron de grado 2, escala 0.1 y $C = 0.5$.

Tabla 28

Máquinas de Soporte Vectorial con Kernel Polinomial.

Máquinas de Soporte Vectorial con Kernel Polinomial	
Muestra	2096
Predictor	8
Clases : NO , SI	2

Fuente: Elaboración propia

En la tabla N°28 se observa que, 2096 datos de muestra, 8 predictores (variables) y 2 clases (si embarazo y no embarazo) para aplicar el método de máquinas de soporte vectorial con Kernel Polinomial.

Tabla 29*Matriz de confusión y estadísticas.*

Predicción	NO	Si
NO	854	7
SI	12	24
Accuracy	0.9788	
95% CI	(0.9513 , 0.9764)	
No informacion Rate	0.9654	
P-Value(Acc > NIR)	0.5476	
McNemar's Test P-Value	6.12E-08	
Sensitivity	0.7742	
Specificity	0.9409	
Prevalence	0.9654	
Detection Rate	0.9654	
Detection Prevalence	1	
Balanced Accuracy	0.5	
'Positive' Class	No	

Fuente: Elaboración propia

En la tabla N°29 se puede observar de la salida anterior del programa que la precisión del modelo SVM polinómico es alta (0.9788), lo cual indica que el modelo de soporte vectorial radial predice correctamente la mayoría de los casos de embarazo y no embarazo, recalcar que este cálculo se está llevando a comprobar en la data de prueba.

La prueba de McNemar tiene un p-valor de 6.12×10^{-8} , lo que indica una diferencia significativa entre las clasificaciones falsos negativos y falsos positivos.

En el contexto de la predicción del embarazo, una sensibilidad del 77.42% indica que la prueba es buena en cierto nivel para detectar si una adolescente puede estar embarazada cuando realmente lo está. Esto es importante a la hora de dar un diagnóstico.

Así mismo la especificidad fue del 94.09%, la prueba es muy efectiva para asegurar que la adolescente no están embarazada correctamente identificada como tal. Esto es importante para evitar falsos positivos que pueden llevar a pruebas adicionales innecesarias.

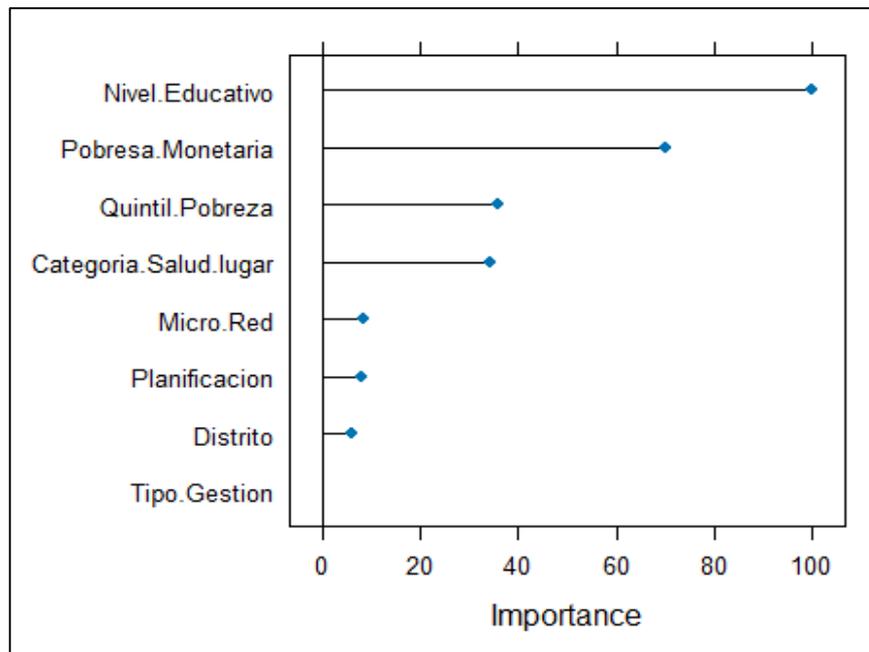
Tabla 30

ROC Curva de importancia de Variable.

Variables	Importancia (%)
Nivel Educativo	100
Pobreza Monetaria	70.028
Quintil Pobreza	35.692
Categoría de Salud de Lugar	34.177
Microred	8.379
Planificación	7.877
Distrito	5.899
Tipo de Gestión	0

Fuente: Elaboración propia

El nivel educativo es la variable más influyente con un 100% de importancia en la clasificación de los datos relacionados al embarazo adolescente seguido de pobreza monetaria con un 70.028%, las variables como quintil de pobreza y categoría de salud por lugar tienen importancias moderadas y el tipo de gestión no muestra relevancia con 0% indicando que no influye en los resultados del modelo

Figura 20*Variables más importantes*

En la figura N°20 se observa que, la variable más importante es el nivel educativo seguido de pobreza monetaria y el menos importante es el tipo de gestión.

De la tabla N°30 y figura N°20 las variables más importantes que ayudan a predecir o caracterizar a las adolescentes que están embarazadas y así en lo posterior prevenir embarazos no deseados son: el nivel educativo es la variable más importante para la predicción, seguida por pobreza monetaria y quintil de pobreza, además de la categoría de establecimiento de salud y hasta la planificación.

5.1.5 Comparando los Indicadores de Precisión.

Tabla 31

Comparación de los indicadores de precisión.

Comparando los Indicadores de Precisión	
svmLineal AUC-ROC	0.78
svmPoly AUC-ROC	0.84
svmRadial AUC-ROC	0.82

Fuente: Elaboración propia

En la tabla N°31 se observa la comparación de los indicadores de precisión de tres modelos basados en máquinas de soporte vectorial con el AUC-ROC.

SVMPolynomial (AUC-ROC = 0.84): Este es el modelo con el mejor rendimiento general, lo que sugiere que es el más efectivo para predecir la situación (si y no) de embarazo. Es probable que ofrezca el mejor equilibrio entre sensibilidad y especificidad, lo que significa que minimizará tanto los falsos negativos como los falsos positivos.

SVMRadial (AUC-ROC = 0.82): Este modelo también es muy bueno, solo un poco inferior al modelo polinómico. Puede ser una buena opción si se prefiere la simplicidad del Kernel radial y si el rendimiento adicional del Kernel polinómico no justifica el aumento en la complejidad.

SVMLineal (AUC-ROC = 0.78): Aunque es el modelo con el menor rendimiento relativo, sigue siendo una opción válida y podría ser preferido en situaciones donde la interpretabilidad y la simplicidad del modelo son más importantes que el rendimiento máximo.

Tabla 32

Comparación de los indicadores de precisión.

Indicador	SVM Lineal	SVM Radial	SVM Polinómico
Accuracy	0.9665	0.9602	0.9788
Sensibilidad	0.86	0.7401	0.7742
especificidad	0.92	0.9364	0.9409
AUC-ROC	0.78	0.82	0.84

Fuente: Elaboración propia.

En la tabla N°32 se observa que, respecto a Accuracy (Precisión global). el modelo con mayor precisión es de SVM Polinómico es 0.9788, lo que indica que el modelo clasifica correctamente el 97.88% de las observaciones, el cual es una excelente precisión, seguido de SVM Lineal con 96.65% y el radial con 96.02% las diferencias en precisión entre los modelos son muy pequeñas, pero estadísticamente podría ser relevantes.

Respecto a la Sensibilidad, el modelo SVM Lineal tiene una sensibilidad de 0.86, indica que el modelo es capaz de identificar correctamente el 86% de los casos positivos reales, seguido de SVM Polinómico y SVM Radial.

Respecto a la especificidad, el modelo SVM Polinómico tiene una especificidad de 0.9409 indicando que el modelo clasifica correctamente el 94.09% de los casos negativos reales, seguido de SVM Radial y SVM Lineal.

Con respecto a AUC-ROC, el modelo SVM Polinómico tiene un AUC- ROC de 0.84, indica que el modelo tiene una buena capacidad discriminativa para distinguir entre las dos clase positiva y negativa en el problema de clasificación, este indicador refuerza la ventaja del modelo polinómico en términos de balance entre sensibilidad y especificidad.

El método de máquinas vectoriales con Kernel polinómico presenta mayor sensibilidad y especifica con respectos a máquinas de base lineal y radial.

5.2 **Discusión de Resultados.**

Con respecto a los factores de embarazo adolescente Medina (2018), realizo el trabajo de Factores que condicionan el embarazo de adolescentes en Miraflores y Ramon Castilla, distrito de San Miguel en el año 2017, Universidad Nacional de Huancavelica obteniendo como conclusiones que lo que más influyen al embarazo adolescentes son los factores socio culturales , la preparación y las prácticas en salud sexual y los que no influye son el factor económico, por otra parte Rojas (2019), realizo el trabajo de Factores de riesgo asociados con el embarazo en adolescentes del hospital nacional Hipólito Unanue ,2018, Universidad Ricardo Palma del Perú, llego a la conclusión de que el factor del estudio se convierte en un riesgo personal para las jóvenes embarazadas, junto con el factor familiar que incluye el abandono de ambos padres o de uno de ellos, así como el nivel educativo del padre, estos hallazgos presentan similitud con los resultados de nuestra investigación, que señalan que el factor educativo es el que más influye en el embarazo en adolescentes.

Con respecto a los factores de embarazo adolescente. Prado (2020), es su investigación Factores sociales y familiares que predisponen el embarazo en adolescentes en el Centro de Salud Los Licenciados. Ayacucho, en el año 2019 en la Universidad Nacional del Callao del Perú, obteniendo como conclusión que indica que existe una relación significativa de los familiares y factores sociales en el embarazo adolescentes, este hallazgo presenta similitud con respecto a la componente educativa establecida como factor predisponente en nuestro estudio para el embarazo.

Con respecto a la aplicación de las técnicas de máquinas de soporte vectorial, Ángeles (2022), realizo la investigación de Detección de carcinoma basocelular utilizando red neuronal convolucional y máquinas de soporte vectorial, en la Universidad Nacional Mayor de San Marcos, obteniendo como conclusión que se logró implementar un modelo híbrido de redes neuronales convolucional y máquinas de soporte vectorial, el cual clasifica

eficientemente el carcinoma basocelular el cual permite al clínico cargar una imagen dermatoscopia al sistema para conseguir una determinación automática del carcinoma basocelular, los resultados de eficiencia de este estudio presentan similitud en virtud de que en nuestro estudio se estableció altos niveles de exactitud del método de Máquinas de Soporte Vectorial.

Con respecto a los factores de embarazo adolescente. Cárdenas (2023), investigación con título Factores familiares y adolescentes embarazadas que fueron atendidas en la Microred de Salud del distrito de San Jerónimo en el año 2021, llevada a cabo en la facultada de Ciencias de la Salud de la Universidad Nacional de San Antonio Abad del Cusco, Perú. Obteniendo como conclusión, que los factores familiares y el embarazo en la adolescencia están significativamente relacionados. Estos estudios difieren a los obtenidos en nuestro estudio, donde se estableció que el factor económico es que más condiciona el embrazo en adolescentes.

Con respecto a los factores de embarazo adolescente. Pinzón, Ruiz, Aguilera y Abril (2018), realizo la investigación titulada Factores asociados al inicio de vida sexual y el embarazo en adolescentes en la ciudad de Colombia. En la Universidad del Rosario, Bogotá. Obteniendo como conclusión, que a través de un análisis multivariado encontró los factores como de tener mayor edad, convivir en pareja, violencia parental y violencia sexual son determinantes. también indica asistir al colegio es un factor protector para ambos eventos. Los factores de pobreza, falta de educación y el hacinamiento están relacionados con el embarazo adolescente. Este hallazgo presenta similitudes con respecto a la componente educativa establecida como factor predisponente en nuestro estudio para el embarazo.

Con respecto a la aplicación de las máquinas de soporte vectorial. Poovizhi (2019), realizo la investigación titulada An Efficient Skin Cancer Diagnostic System Using Bendlet Transform and Support Vector Machine. En esta investigación, desarrollo un sistema de

clasificación del cáncer de piel utilizando imágenes dermatoscópicas. Dando la conclusión, que los resultados muestran que el sistema de clasificación de cáncer que utiliza el clasificador Transformada Bendlet y el método de máquinas de soporte vectorial provee mayor precisión, especificidad y sensibilidad en comparación con otras técnicas. Por otra parte, Galindo, Perdomo y Figueroa (2020), realizó la investigación con el título Estudio comparativo de las máquinas de soporte vectorial multiclase con el sistema de inferencia neuro-difuso autoorganizado y redes neuronales para los problemas de clasificación. Obteniendo como conclusión que los SONFIS es un método alternativo para analizar los problemas de clasificación y este tiene un desempeño equivalente o superior que otras metodologías, El costo computacional SONFIS utiliza el mismo tiempo que los otros algoritmos en otros casos en menor tiempo, obteniendo resultados análogos en términos de los indicadores estimados. La prueba mostrada en los casos de estudio mostrados es placentera y muestra el potencial de SONFIS, El rendimiento en tareas de clasificación es importante. Las Técnicas en los algoritmos genéticos, o el método de inteligencia de enjambres se pueden combinar con SONFIS para luego seleccione los parámetros más altos adecuados para cada tipo de problema. Por otra parte, Lopera, Zuluaga y Jaramillo (2021) realizó la investigación con el título Máquinas de vectores de soporte en el descubrimiento de biomarcadores en experimentos en vitro e vivo de exposición a organoclorados, en el Instituto Tecnológico Metropolitano, Colombia, obteniendo como conclusión que en este estudio, se demostró que el método de las máquinas de soporte vectorial son métodos oportunos para analizar los datos de estudios en vitro con situaciones controladas, pero los mínimos cuadrados parciales presentó conflictos con los datos de estudios en vivo con situaciones no controladas y datos no lineales en la exposición a organoclorados. En cuanto al pronóstico de clases en datos de estudios en vivo, el método de las máquinas de soporte vectorial reveló una mayor eficiencia predictivo que mínimos cuadrados parciales. También

el Kernel tiene un rendimiento identificado por el método de máquinas de soporte vectorial se puede utilizar para obtener una lista adecuada de estas características más selectas en el contexto de la exhibición a pesticidas. Además, el coste computacional de las máquinas de soporte vectorial es menor y se está convirtiendo en un instrumento útil para la identificación de biomarcadores, inclusive cuando tenemos pocas muestras. También se puede considerar un método robusto para analizar datos derivados de experimentos en vivo y estudios en vitro. Estos resultados presentan similitud en virtud de que en nuestro estudio se estableció altos niveles de exactitud del método de máquinas de soporte vectorial.

5.3 Contrastación de Resultados

En presente trabajo se formuló la hipótesis general:

El método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022, es el polinómico.

De la tabla N°32 se concluye que la máquina de soporte vectorial con kernel polinómico de grado 2 presenta una exactitud de 97.88%, superior a la de los otros modelos de máquinas de soporte vectorial; es decir, este modelo predice en base a la información correcta el 98% de los embarazos en los adolescentes. Esta información es trascendental para las políticas de salud, puesto que se puede prevenir mediante programas e identificar a posibles adolescentes antes de un embarazo. La sensibilidad de este modelo 0.7742 es decir predice correctamente los casos positivos de embarazo en un 77.42% y la especificidad es 0.9409 lo que significa que predice correctamente los casos negativos de embarazo adolescente en un 94.09%.

El modelo lineal es el segundo que presenta los mejores indicadores de precisión con una exactitud de 96.65% superior a la del otro modelo de máquinas de soporte vectorial; es decir, este modelo predice correctamente el 96.65% de los embarazos en las adolescentes. La sensibilidad de este modelo 0.86, es decir, predice correctamente los casos positivos de embarazo en un 86% y la especificidad es 0.92, lo que significa que predice correctamente los casos negativos de embarazo adolescente en un 92%.

El modelo radial es el tercero que presenta menores indicadores de precisión con una exactitud el 96.02% inferior a la de los otros modelos de máquinas de soporte vectorial; es decir, este modelo predice correctamente el 96% de los embarazos en adolescentes. La sensibilidad de este modelo 0.7401, es decir, predice correctamente los casos positivos de

embarazo en un 74.01%, y la especificidad es 0.9364, lo que significa que predice correctamente los casos negativos de embarazo adolescente en un 93.64%.

De los resultados anteriores se verifica la hipótesis formulada.

CONCLUSIONES

- Primero. El modelo de máquinas de soporte con Kernel polinómico de grado 2 presenta mayor exactitud en la predicción de embarazo adolescente con una precisión de 84% y exactitud de 97.88 %
- Segundo. El principal factor asociado al embarazo en adolescentes según el modelo de máquinas de soporte vectorial es el nivel educativo seguido de pobreza monetaria, condiciones salud y la planificación.
- Tercero. El modelo de máquinas de soporte vectorial polinómico de grado 2 presenta mejor comportamiento en los indicadores de sensibilidad 0.7742 y especificidad 0.9409. con respecto a los otros modelos de máquinas de soporte vectorial lineal y radial.

RECOMENDACIONES

- Primero. Se recomienda a las autoridades de salud implementar modelos de aprendizaje automático para pronosticar y prevenir embarazo en adolescentes.
- Segundo. A los docentes de la escuela profesional de matemáticas mención estadística se les recomienda fomentar la aplicación de los modelos de aprendizaje automático como máquinas de soporte vectorial, para que los egresados fomenten la aplicación de estas técnicas en las diferentes áreas.

BIBLIOGRÁFIAS

- Angeles Rojas, J. A. (2022). *Detección de carcinoma basocelular utilizando red neuronal convolucional y Support Vector Machine* [Tesis de grado, Universidad Nacional Mayor de San Marcos de Lima]. ALICIA, Peru.
<https://hdl.handle.net/20.500.12672/18365>
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer.
<https://link.springer.com/book/9780387310732>
- Campo Leon, E. (s.f.). *Introducción a las máquinas de vector (SVM) en aprendizaje supervisado* [Tesis de grado, Universidad de Zaragoza].
- Cardenas Macote, G. R. (2023). *Factores familiares y adolescentes embarazadas que fueron atendidas en la Microred de Salud del distrito de San Jerónimo, 2021* [Tesis de licenciatura, Universidad Nacional de San Antonio Abad del Cusco]. *Obstetra*. RENATI, Cusco, Peru. <http://purl.org/pe-repo/ocde/ford#3.02.02>
- Carmona suarez, E. (2016). Tutorial sobre máquinas de vectores soporte (SVM). *Cartagena99*, 25.
https://www.cartagena99.com/recursos/alumnos/apuntes/Tema8._Maquinas_de_Vectores_Soporte.pdf
- Cortes, C., & Vladimir, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/https://doi.org/10.1007/BF00994018>
- Cruz Huacac, R. D. (2019). *Uso de máquina de soporte vectorial para predicción de resistencia a la compresión de concreto* [Tesis de maestría, Universidad Nacional de San Agustín de Arequipa]. RENATI, Arequipa.
<http://repositorio.unsa.edu.pe/handle/UNSA/9092>
- Fernández Casal, R., Costa Bouzas, J., & Oviedo de la Fuente, M. (2021). *Aprendizaje Estadístico*. https://rubenfcasal.github.io/aprendizaje_estadistico.

- Fondo de las Naciones Unidas para la Infancia. (2014). Vivencias y relatos sobre el embarazo en adolescentes. UNICEF.
<https://www.unicef.org/lac/informes/vivencias-y-relatos-sobre-el-embarazo-en-adolescentes>
- Galindo, E., Perdomo, J., & Figueroa-Garcia, J. (2020). Estudio comparativo entre máquinas de soporte vectorial multiclase, redes neuronales artificiales y sistema de inferencia neuro-difuso auto organizado para problemas de clasificación. (42, Ed.) *Revista de Inteligencia Artificial*, 42(1), 1-16. <https://doi.org/10.4067/S0718-07642020000100273>
- Gil Martinez, C. (2018). *Maquinas de Soporte Vectorial*. R Pubs.
https://rpubs.com/Cristina_Gil/SVM
- Hernandez Sampier, R., Fernandez Collado, C., & Batista Lucio, P. (2014). *Metodología de la Investigación*. (6ta, Ed.) Mexico: McGraw Hill.
<https://archive.org/details/hernandez-sampieri>
- Instituto Nacional de Estadística e Informática (INEI). (2022). *Directorio Nacional de Municipalidades Provinciales, Distritales y de Centros Poblados 2022*.
<https://www.gob.pe/institucion/inei/informes-publicaciones/3313436-directorio-nacional-de-municipalidades-provinciales-distritales-y-de-centros-poblados-2022>
- Instituto Nacional de Estadística e Informática. (2022). *Encuesta Demográfica y de Salud Familiar (ENDES) 2022 - Nacional y Departamental*. INEI.
<https://encuestas.inei.gob.pe/endes/?cod=01>
- Instituto Nacional de Estadística e Informática. (2022). *Encuesta Nacional Demográfica y de Salud Familiar*. INEI.
https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitaes/Est/Lib1898/libro.pdf

- Ipanaqué Alama, W. (2021). *Clasificación de hojas de tomate con plagas o enfermedades usando una máquina de soporte vectorial (SVM)* [Tesis, Universidad de Piura]. RENATI. <https://hdl.handle.net/11042/4940>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning with applications in R*. (2, Ed.) Springer Science, Business Media. https://doi.org/10.1007/978-1-0716-1418-1_1
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning with applications in Python*. Springer. <https://link.springer.com/book/10.1007/978-3-031-38747-0>
- Krug, E., Dahlberg, L., Mercy, J., Zwi, A., & Lozano, R. (2002). Informe mundial sobre la violencia y la salud. *World Health Organization*. <https://iris.who.int/handle/10665/42495>
- Lopera-Rodríguez, J., Zuluaga, M., & Jaramillo-Garzón, J. (2021). Máquinas de vectores de soporte para detección de biomarcadores en experimentos in vitro e in vivo de exposición a organoclorados. *Tecnologías*, 24(52), 2088. <https://doi.org/10.22430/22565337.2088>
- Medina Roca, L. N. (2018). *Factores que condicionan el embarazo de adolescentes en los barrios Miraflores y Ramón Castilla, distrito de San Miguel . 2017* [Tesis de especialidad, Universidad Nacional de Huancavelica]. ALICIA. <http://repositorio.unh.edu.pe/handle/UNH/1711>
- Mesa de Concentración para la Lucha contra la Pobreza. (2023). Prevenir el embarazo en niñas y adolescentes y reducir las muertes maternas. *Mesa de Concentración*. <https://www.mesadeconcentacion.org.pe/>
- Ministerio de Salud. (2019). *Normas legales 1001-2019-MINSA*. Gobierno del Perú: <https://www.gob.pe/institucion/minsa/normas-legales/344744-1001-2019-minsa>

- Municipalidad Provincial de La Convención. (2017). *Mapa de la provincia de La Convención*. Gobierno del Perú.:
<https://www.gob.pe/institucion/munilaconvencion/informes-publicaciones/1999831-mapa-de-la-provincia-de-la-convencion>).
- Organizacion Mundial de la Salud. (22 de Febrero de 2023). *Mortalida Materna*.
<https://www.who.int/es/news-room/fact-sheets/detail/maternal-mortality>
- Organizacion Mundial de la Salud. (10 de 04 de 2024). *Embarazo en la adolescencia*.
<https://www.who.int/es/news-room/fact-sheets/detail/adolescent-pregnancy#:~:text=1.%20En%202019%20se%20registraban>
- Pinzon-Rondon, A. M., Ruiz-Sternberg, Á. M., Aguilera-Otalvaro, P. A., & Abril-Basto, P. D. (2018). Factores asociados al inicio de vida sexual y al embarazo adolescente en Colombia. Estudio de corte transversal. *Revista chilena de obstetricia y ginecología*, 83, 487-499. <https://doi.org/10.4067/s0717-75262018000500487>
- Poovizhi, s. (2019). An Efficient Skin Cancer Diagnostic System Using Bendlet Transform and Support Vector Machine. *International Journal of Medical Imaging*, 12(3), 45-46. <https://doi.org/10.1590/0001-3765202020190554>
- Prado Juscamaita, J. I., & Reves Huapaya, E. S. (2018). *El Embarazo Adolescente en Perú: Estrategias para el afrontamiento para el bienestar emocional y la inclusion educativa* [Investigacion y Postgrado, Universidad Nacional Hermilio Valdizan,Peru]. Peru.
<https://dialnet.unirioja.es/servlet/articulo?codigo=6736285#:~:text=Este%20trabajo%20estudia%20el%20embarazo%20adolescente%20en%20Per%C3%BA,social%20C%20negativa%20para%20el%20desarrollo%20de%20la%20mujer>.
- Prado Martinez, C. (2020). *Factores familiares y sociales que predisponen el embarazo en adolescentes en el Centro de Salud Los Licenciados. Ayacucho, 2019* [Tesis

doctoral, Universidad Nacional del Callao]. Renati, Lima.

<http://hdl.handle.net/20.500.12952/5621>

Red Salud La Convención. (2021). *Plan operativo institucional 2021*.

<https://redsaludlaconvencion.online-e.com/wp-content/uploads/2024/09/PLAN-OPERATIVO-INSTITUCIONAL-2021.pdf>

Renato, S. (25 de agosto de 2024). *Aumenta casos de embarazo en menores de edad*.

Infobae: <https://www.infobae.com/peru/2024/04/14/aumentan-casos-de-embarazo-en-menores-de-edad-lima-loreto-y-la-libertad-registran-3086-partos/>

Rodríguez Bernate, I., & Tobar, F. (2022). Consecuencia socioeconómicas del embarazo y la maternidad adolescente en Perú. UNFRA.

<https://peru.unfpa.org/es/video/%C2%BFcu%C3%A1-es-el-impacto-del-embarzoadolescente%E2%80%8B-en-el-per%C3%BA>

Rojas Humán, E. B. (2019). *Factores de riesgo asociados con el embarazo en adolescentes en el Hospital Nacional Hipólito Unanue en el año 2018* [Tesis de grado, Universidad Ricardo Palma]. Repositorio URP, Lima, Peru.

<https://hdl.handle.net/20.500.14138/1816>

Sanchez Vargas, P. d. (2020). *Apoyo social y resiliencia en adolescentes embarazadas Una revision Teórica* [Tesis, Universidad Señor de Sipán]. Univesidad Señor de Sipán, Lambayeque, Perú.

UNFPA. (20 de setiembre de 2023). *Embarazo adolescente en Perú, 2023*. Fondo de Población de las Naciones Unidas:

<https://peru.unfpa.org/es/news/embarzoadolescenteperu2023>

Vapnik, V. (1998). *Statistical learning theory*. Simon & Schuster.

<https://www.amazon.com/Statistical-Learning-Theory-Vladimir-Vapnik/dp/0471030031>

ANEXOS

Anexo 1: Matriz de Consistencia

TITULO: Máquina de soporte vectorial en el estudio de factores de embarazo en adolescentes, del distrito de Santa Ana, 2022.

PROBLEMA GENERAL	OBJETIVO GENERAL	HIPOTESIS GENERAL	VARIABLE INDEPENDIENTE:
¿Cuál de los métodos de máquinas de soporte vectorial presenta mejores niveles de predicción en los factores de embarazo de adolescentes en el distrito de Santa Ana 2022?	Determinar el método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022.	El método de máquina de soporte vectorial que presenta mayor exactitud en la predicción de los factores de embarazo de adolescentes, en el distrito de Santa Ana 2022, es el polinómico.	<ul style="list-style-type: none"> • Edad • Peso • Talla • IMC • Planificación • Tipo de gestión • Nivel Educativo • Grado de instrucción • Fase Regular • Quintil Pobreza • Pobreza Monetaria • Apoderado • Distrito • Migración de distrito
PROBLEMAS ESPECÍFICOS	OBJETIVOS ESPECIFICOS	HIPÓTESIS ESPECÍFICOS	
a) ¿Cuáles son los factores asociados en el embarazo adolescente en el distrito de Santa Ana 2022, obtenidos de la aplicación del método máquina de soporte vectorial?	a) Establecer los factores asociados al embarazo de adolescentes en el distrito de Santa Ana 2022, obtenidos de la aplicación del método máquina de soporte vectorial.	a) Los factores económico y educativo están asociados significativamente al embarazo adolescentes.	

			<ul style="list-style-type: none"> • Microredes • Categoría centro de salud por lugar
<p>b) ¿Cuál es la sensibilidad y la especificidad de las máquinas de soporte vectorial en la predicción de factores de embarazo de adolescentes en el distrito de Santa Ana 2022?</p>	<p>b) Determinar la sensibilidad y la especificidad de las máquinas de soporte vectorial en la predicción de factores de embarazo de adolescentes en el distrito de Santa Ana 2022.</p>	<p>b) El método de máquinas vectoriales con Kernel polinómico presenta mayor sensibilidad y especifica con respecto a máquinas de base lineal y radial.</p>	<p>VARIABLE DEPENDIENTE:</p> <ul style="list-style-type: none"> • Embarazo adolescente

JUSTIFICACION	METODOLOGIA	UNIVERSO POBLACION Y MUESTRA
<p>El estudio es importante porque permitirá conocer los factores de embarazo adolescentes esto permitirá tomar mejores decisiones para la población adolescentes</p>	<p>El tipo de investigación es básica.</p> <p>El Alcance o nivel de investigación es descriptiva</p> <p>Diseño de investigación no experimental, transversal.</p>	<p>POBLACION: Todas las adolescentes afiliadas en el SIS de la Red de Servicios de Salud La Convención.</p> <p>MUESTRA: Adolescentes de 12 a 17 años de edad afiliadas en el SIS embarazadas y no embarazadas de la Red de Servicios de Salud La Convención 2022.</p> <p>METODO ESTADÍSTICO</p> <p>Aplicación máquina de soporte vectorial.</p>

Anexo 2: Instrumento de Recolección de Información

INSTRUMENTOS DE RECOLECCIÓN DE DATOS

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
ESCUELA DE POSGRADO MAESTRÍA ESTADÍSTICA

ENCUESTA PARA EVALUAR EL EMBARAZO ADOLESCENTE EN EL DISTRITO DE SANTA ANA 2022

Objetivo del instrumento: El presente instrumento tiene la finalidad de describir los factores que se asocian al embarazo en adolescentes del distrito de Santa Ana.

- Se solicita su participación ya que es de mucha importancia.
- Por favor responder con toda sinceridad, ya que los resultados permitirán tener un conocimiento objetivo sobre el particular.
- Toda información recolectada mediante el presente cuestionario es confidencial.
- Agradezco de antemano su valiosa colaboración.

Datos generales

1. Edad:
2. Peso:
3. Talla:
4. IMC:
5. ¿Estado civil?:
 - Soltera
 - Conviviente

Factores Socio Económico

6. ¿Cuál es el tipo de Institución Educativa donde estudia o estudio este año?
 - Privado
 - Publico
7. ¿Cuál es su nivel educativo?
 - Primaria
 - Secundaria
 - Básica Alternativa Avanzado
8. ¿Cuál es su grado de instrucción?
 - No tiene estudios
 - primaria
 - Primero de Secundaria
 - Segundo de Secundaria
 - Tercero de Secundaria
 - Cuarto de Secundaria
 - Quinto de Secundaria
9. ¿Cuál es su situación de estudio este año?
 - Matriculado
 - Promoción guiada
 - Promovido
 - Requiere Recuperación Pedagógica
 - Retirado

Factores Condición Económico:

10. ¿Cuál es su situación económica?
 - Extrema pobreza

Ficha de encuesta para evaluar el embarazo adolescente en el distrito de Santa Ana 2022.

- Muy pobre
- Pobre

11. ¿Cuánto es su ingreso económico?

Factores Familiares:

12. ¿Quién es su apoderado?

- Madre
- Padre
- Abuelos
- Hermanos
- Tíos
- Otros

Factores Residencia:

13. ¿Cuál es el lugar donde vive actualmente?

- Santa Ana
- Otros

14. ¿Actualmente vive en el distrito donde nació?

- Si
- No

Factores de Atención de Salud:

15. ¿Cuál es la Micro Red donde se realiza su atención?

- Santa Ana
- Otros

Gracias por su colaboración

VALIDACIÓN DE INSTRUMENTOS

I. DATOS GENERALES
 I.1. Título Del Trabajo De Investigación : Máquina de soporte vectorial en el estudio de factores de embarazo en adolescentes, Distrito de Santa Ana 2022
 I.2. Investigador : Lic. Bárcena Carrasco Pascual
 II. DATOS DEL EXPERTO:
 II.1. Nombres y apellidos : KARLA ZELMIRA APARCICIO ARENAS
 II.2. Especialidad : ESTADÍSTICA
 II.3. Lugar y Fecha :
 II.4. Cargo e institución donde labora : DOCENTE - UNIVSUDC

Componente	Indicadores	Criterios	Deficiente 0-20%	Regular 21-40%	Bueno 41-60%	Muy bueno 61-80%	Excelente 81-100%
Forma	Redacción	Los indicadores e ítems están redactados considerando los elementos necesarios.				X	
	Claridad	Están formulado con un lenguaje apropiado.					X
	Objetividad	Esta expresado en conducta observable.					X
Contenido	Actualidad	Esta adecuado al avance de la ciencia y la tecnología.				X	
	Suficiencia	Los ítems son adecuados en cantidad y claridad.				X	
	Intencionalidad	El instrumento mide pertinentemente las variables de investigación.				X	
Estructura	Organización	Existe una organización lógica.					X
	Consistencia	Se basa en aspectos teóricos científicos de la investigación educativa.				X	
	Coherencia	Existe coherencia entre los ítems, indicadores, dimensiones y variables.					X
	Metodología	La estrategia responde al propósito del diagnóstico.					X

- Opinión de aplicabilidad : APLICABLE.
- Promedio de Valoración : 90%
- Luego de revisado el instrumento : (X)
- Procede a su aplicación : ()
- Debe corregirse : ()

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO
 FACULTAD DE CIENCIAS

Dra. Karla Zelmira Aparicio Arenas
 COMAP 496

Firma y sello del Experto

DNI:

Validación de instrumentos

FICHA DE CALIFICACIÓN PARA LA VALIDACIÓN DEL INSTRUMENTO

	Preguntas	Escala De Validación				
		1	2	3	4	5
1	¿Considera Ud. ¿Que los ítems de los instrumentos miden lo que pretende medir?					
2	¿Considera Ud. ¿Que la cantidad de ítems registrados en esta versión son suficientes para tener una comprensión de la materia en estudio?					
3	¿Considera Ud. ¿Que los ítems contenidos en este instrumento son una muestra representativa del universo materia de estudio?					
4	¿Considera Ud. ¿Que los conceptos utilizados en estos instrumentos son todos y cada uno de ellos propios de las variables de estudio?					
5	¿Considera Ud. ¿Que si aplicamos en reiteradas oportunidades este instrumento a muestras similares obtendríamos datos también similares?					
6	¿Considera Ud. ¿Que todos y cada uno de los ítems contenidos en este instrumento tienen los mismos objetivos?					
7	¿Considera Ud. ¿Que el lenguaje utilizado en el presente instrumento es lo sencillo y no da lugar a diferentes interpretaciones?					
8	¿Estima Ud. ¿Que las escalas de medida utilizadas son pertinentes a los objetos materia de estudio?					

¿Qué aspectos habría que modificar o que aspectos tendrían que incrementarse para mejorar el instrumento?

.....

.....

.....

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAO DEL CUSCO
 FACULTAD DE CIENCIAS
 DNI. 80810720
 DNI. 80810720
 DNI. 80810720

Ficha de calificación para la validación de instrumento

VALIDACIÓN DE INSTRUMENTOS

I. DATOS GENERALES

I.1. Título Del Trabajo De Investigación : Máquina de soporte vectorial en el estudio de factores de embarazo en adolescentes, Distrito de Santa Ana 2022
 I.2. Investigador : Lic. Bárcena Carrasco Pascual

II. DATOS DEL EXPERTO:

II.1. Nombres y apellidos : Arturo Zuñiga Blanca
 II.2. Especialidad : Mtro. en estadística e investigación científica
 II.3. Lugar y Fecha : Cusco
 II.4. Cargo e institución donde labora : Docente nombrado en UNSAAC

Componente	Indicadores	Criterios	Deficiente 0-20%	Regular 21-40%	Bueno 41-60%	Muy bueno 61-80%	Excelente 81-100%
Forma	Redacción	Los indicadores e ítems están redactados considerando los elementos necesarios.				X	
	Claridad	Están formulado con un lenguaje apropiado.					X
	Objetividad	Esta expresado en conducta observable.					X
Contenido	Actualidad	Esta adecuado al avance de la ciencia y la tecnología.				X	
	Suficiencia	Los ítems son adecuados en cantidad y claridad.				X	
	Intencionalidad	El instrumento mide pertinentemente las variables de investigación.				X	
Estructura	Organización	Existe una organización lógica.					X
	Consistencia	Se basa en aspectos teóricos científicos de la investigación educativa.				X	
	Coherencia	Existe coherencia entre los ítems, indicadores, dimensiones y variables.					X
	Metodología	La estrategia responde al propósito del diagnóstico.					X

- Opinión de aplicabilidad : Aplicable
- Promedio de Valoración : 90%
- Luego de revisado el instrumento :
- Procede a su aplicación : (X)
- Debe corregirse : ()


 MTRD. Arturo Zuñiga Blanca
 DNI: 469728024

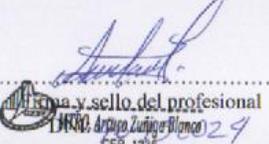
Validación de instrumento

FICHA DE CALIFICACIÓN PARA LA VALIDACIÓN DEL INSTRUMENTO

Preguntas		Escala De Validación				
1	¿Considera Ud. ¿Que los ítems de los instrumentos miden lo que pretende medir?	1	2	3	4	5
2	¿Considera Ud. ¿Que la cantidad de ítems registrados en esta versión son suficientes para tener una comprensión de la materia en estudio?	1	2	3	4	5
3	¿Considera Ud. ¿Que los ítems contenidos en este instrumento son una muestra representativa del universo materia de estudio?	1	2	3	4	5
4	¿Considera Ud. ¿Que los conceptos utilizados en estos instrumentos son todos y cada uno de ellos propios de las variables de estudio?	1	2	3	4	5
5	¿Considera Ud. ¿Que si aplicamos en reiteradas oportunidades este instrumento a muestras similares obtendríamos datos también similares?	1	2	3	4	5
6	¿Considera Ud. ¿Que todos y cada uno de los ítems contenidos en este instrumento tienen los mismos objetivos?	1	2	3	4	5
7	¿Considera Ud. ¿Que el lenguaje utilizado en el presente instrumento es lo sencillo y no da lugar a diferentes interpretaciones?	1	2	3	4	5
8	¿Estima Ud. ¿Que las escalas de medida utilizadas son pertinentes a los objetos materia de estudio?	1	2	3	4	5

¿Qué aspectos habría que modificar o que aspectos tendrían que incrementarse para mejorar el instrumento?

.....


 Firma y sello del profesional
 MSc. Arque Zúñiga Vilmon
 CEP. 1256

Ficha de calificación para la validación del instrumento.

VALIDACIÓN DE INSTRUMENTOS

I. DATOS GENERALES

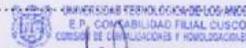
I.1. Título Del Trabajo De Investigación : Máquina de soporte vectorial en el estudio de factores de embarazo en adolescentes, Distrito de Santa Ana 2022
 I.2. Investigador : Lic. Bárcena Carrasco Pascual

II. DATOS DEL EXPERTO:

II.1. Nombres y apellidos : Henry Lazo Chuguihuayta
 II.2. Especialidad : Maestro en Estadística
 II.3. Lugar y Fecha :
 II.4. Cargo e institución donde labora : Docente - Universidad Nacional de San Antonio Abad del Cuzco

Componente	Indicadores	Criterios	Deficiente 0-20%	Regular 21-40%	Bueno 41-60%	Muy bueno 61-80%	Excelente 81-100%
Forma	Redacción	Los indicadores e items están redactados considerando los elementos necesarios.				X	
	Claridad	Están formulado con un lenguaje apropiado.				X	
	Objetividad	Esta expresado en conducta observable.				X	
Contenido	Actualidad	Esta adecuado al avance de la ciencia y la tecnología.				X	
	Suficiencia	Los items son adecuados en cantidad y claridad.			X		
	Intencionalidad	El instrumento mide pertinentemente las variables de investigación.				X	
Estructura	Organización	Existe una organización lógica.				X	
	Consistencia	Se basa en aspectos teóricos científicos de la investigación educativa.				X	
	Coherencia	Existe coherencia entre los items, indicadores, dimensiones y variables.				X	
	Metodología	La estrategia responde al propósito del diagnóstico.				X	

- Opinión de aplicabilidad : Aplicar el instrumento para la investigación
- Promedio de Valoración :
- Luego de revisado el instrumento :
- Procede a su aplicación : (X)
- Debe corregirse : ()



Henry Lazo Chuguihuayta
 Maestro en Estadística

Firma y sello del Experto

DNI: 23963075

Validación de instrumento.

FICHA DE CALIFICACIÓN PARA LA VALIDACIÓN DEL INSTRUMENTO

	Preguntas	Escala De Validación				
		1	2	3	4	5
1	¿Considera Ud. ¿Que los ítems de los instrumentos miden lo que pretende medir?	1	2	3	4	5
2	¿Considera Ud. ¿Que la cantidad de ítems registrados en esta versión son suficientes para tener una comprensión de la materia en estudio?	1	2	3	4	5
3	¿Considera Ud. ¿Que los ítems contenidos en este instrumento son una muestra representativa del universo materia de estudio?	1	2	3	4	5
4	¿Considera Ud. ¿Que los conceptos utilizados en estos instrumentos son todos y cada uno de ellos propios de las variables de estudio?	1	2	3	4	5
5	¿Considera Ud. ¿Que si aplicamos en reiteradas oportunidades este instrumento a muestras similares obtendríamos datos también similares?	1	2	3	4	5
6	¿Considera Ud. ¿Que todos y cada uno de los ítems contenidos en este instrumento tienen los mismos objetivos?	1	2	3	4	5
7	¿Considera Ud. ¿Que el lenguaje utilizado en el presente instrumento es lo sencillo y no da lugar a diferentes interpretaciones?	1	2	3	4	5
8	¿Estima Ud. ¿Que las escalas de medida utilizadas son pertinentes a los objetos materia de estudio?	1	2	3	4	5

¿Qué aspectos habría que modificar o que aspectos tendrían que incrementarse para mejorar el instrumento?

Considerar mas preguntas en los factores donde se tienen pocas



Mg. Henry Lazo Chuquiwaya

Firma y sello del Experto
DNI: 23963075

Ficha de calificación para la validación del instrumento.

Anexo 3: Medios de Verificación


UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAAD DEL CUSCO
ESCUELA DE POSGRADO


Cusco, 28 de agosto de 2023

Oficio N° 1237-2023-EPG-UNSAAC



Dr. CARLOS F. HUESEMBERG FARFÁN
director de RED DE SALUD LA CONVENCION
Presente.-

Ref. Expediente 562562
ASUNTO: SOLICITA ACCESO A INFORMACION

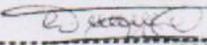
De mi mayor consideración:

Por la presente es grato dirigirme a usted, con la finalidad de hacer de su conocimiento que la Lic. Barcena Carrasco Pascual, egresada de la maestría en Estadística en la UNSAAC, para realizar su tesis "Máquina de soporte vectorial en el estudio de factores de embarazo en adolescentes 2022", solicita tener acceso a información de la base de datos de la Red de Salud de su Dirección.

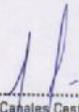
Por lo expuesto solicito se sirva disponer se de acceso a la información requerida, la misma que será manejada con la discreción del caso, siendo responsabilidad absoluta de la tesista cualquier infidencia que se pueda producir sujeta a las acciones legales correspondientes.

Sin otro particular, hago propicia la ocasión para expresarle las muestras de mi mayor consideración.

Atentamente,


Dra. NELLY AYDE CAVERO TORRE
DIRECTORA GENERAL (e)

EPG
OVC/ERZ
Archivo


Jorge Cahales Castro
MEDICO RADICOLOGO
CMP 77946 - RNE 048804



Paraninfo Universitario 2do. Piso - Telefax (051 - 084 - 231751)
 Email: postgrado@unsaac.edu.pe

Autorización para la aplicación del instrumento de investigación del centro de salud Hospital Quillabamba, en la provincia de La Convención, distrito de Santa Ana.


GOBIERNO REGIONAL CUSCO
GERENCIA REGIONAL DE SALUD CUSCO
TRAMITE DOCUMENTARIO
FORMULARIO ÚNICO DE TRAMITE (FUT)

N° _____

1.- SEÑOR GERENTE REGIONAL DE SALUD CUSCO

2.-

APELLIDO PATERNO	APELLIDO MATERNO	NOMBRES
Barcena	Canasco	Pascual
DNI	DOMICILIO	TELÉFONO

3.- Señor Gerente solicito: MARCAR EL CASILLERO CON UN ASPA (X)

<input type="checkbox"/> APROBACIÓN	<input type="checkbox"/> EXONERACIÓN	<input type="checkbox"/> REGISTRO
<input type="checkbox"/> AUTENTICACIÓN	<input type="checkbox"/> EXPEDICIÓN	<input type="checkbox"/> VISACIÓN
<input type="checkbox"/> AUTORIZACIÓN	<input type="checkbox"/> INSCRIPCIÓN	<input type="checkbox"/> RECTIFICACIÓN/MODIFICACIÓN
<input type="checkbox"/> CERTIFICADO	<input type="checkbox"/> INSPECCIÓN	<input type="checkbox"/> SUBSIDIO
<input type="checkbox"/> CONSTANCIA	<input type="checkbox"/> OTORGAMIENTO	<input checked="" type="checkbox"/> OTROS
<input type="checkbox"/> RECONSIDERACIÓN/APELACIÓN		

ESPECIFICAR EL PEDIDO: Solicito datos para realizar Maestría en la LUISAAC, al SES datos de afiliados de la provincia La Convención

4.- ANEXOS si Ud. Desea adjuntar documentos, señale los principales. Si le falta espacio puede escribir en el reverso de este formulario

- Resolución Directoral de inscripción de proyecto de Tesis

- DNI

- Solicito los siguientes datos, Edad, sexo, peso, talla, embarazo, aborto, Nivel educativo, apoyo económico, anticonceptivos planificación, DNI, enfermedades.

Cusco, 17 de enero de 2024

REGIÓN CUSCO
GERENCIA REGIONAL DE SALUD
UNIDAD DE TRAMITE DOCUMENTARIO
17 ENE. 2024
0773

Folio: _____


 Lic. Barcena Canasco Pascual
 DNI: 42477510
 Cel: 925013587
 Correo: pascualbarcena@gmail.com

Autorización de recolección datos de la Red de Servicios de Salud La Convención, en su sede central CUSCO



PERÚ

Ministerio
de Educación

"Decenio de la Igualdad de Oportunidades para Mujeres y Hombres"
"Año del Bicentenario, de la consolidación de nuestra Independencia, y de la conmemoración de las heroicas batallas
de Junín y Ayacucho"

Lima, 28 de febrero de 2024

OFICIO N° 03221-2024-MINEDU/SG-OACIGED

Señor
PASCUAL BARCENA CARRASCO
Presente.-

Asunto : Solicitud de Acceso a la Información Pública

Referencia : MPT2024-EXT-0158369

Me dirijo a usted en atención al expediente de la referencia, a través del cual - en el marco del Texto Único Ordenado de la Ley N° 27806, Ley de Transparencia y Acceso a la Información Pública, aprobado por Decreto Supremo N° 021-2019-JUS (en adelante, TUO de la Ley N° 27806) - solicitó lo siguiente:

1. "SOLICITO ACCESO A LOS DATOS DE LOS ESTUDIANTES DEL AÑO 2022 DE LA PROVINCIA LA CONVENCION, DNI Y GRADO". (*Texto extraído de su solicitud*)

Al respecto, es oportuno señalar que de acuerdo al numeral 5 del artículo 2° de la Constitución Política del Perú establece que toda persona tiene derecho a solicitar sin expresión de causa la información que requiera y a recibirla de cualquier entidad pública, en el plazo legal, con el costo que suponga el pedido, con excepción de aquellas informaciones que afectan la intimidad personal y las que expresamente se excluyan por ley o por razones de seguridad nacional.

Asimismo, el artículo 10° del TUO de la Ley N°27806, establece que las entidades de la Administración Pública tienen la obligación de proveer la información requerida si se refiere a la contenida en documentos escritos, fotografías, grabaciones, soporte magnético o digital, o en cualquier otro formato, siempre que haya sido creada u obtenida por ella o que se encuentre en su posesión o bajo su control.

Por otro lado, mediante **OFICIO N°00351-2024-MINEDU/SPE-OSEE-UE**, la Unidad de Estadística brinda atención a su pedido remitiendo la información solicitada a través de un enlace electrónico, la cual se adjunta al presente.

Cabe mencionar que la información descrita será notificada al correo electrónico consignado en el recurso de reconsideración, debiendo responder con el acuse de recibo correspondiente, en atención a lo establecido en el numeral 20.4 del artículo 20 del TUO de la Ley 27444, Ley del Procedimiento Administrativo General¹.

¹ Texto Único Ordenado de la Ley N° 27444, Ley del Procedimiento Administrativo General, aprobado por Decreto Supremo N° 004-2019-JUS.

Artículo 20. Modalidades de notificación.

20.4 (...) La notificación dirigida a la dirección de correo electrónico señalada por el administrado se entiende válidamente efectuada cuando la entidad reciba la respuesta de recepción de la dirección electrónica señalada por el administrado. La notificación surte efectos el día que sucede a haber sido recibida, conforme lo previsto en el numeral 2 del artículo 20.

EXPEDIENTE: MPT2024-EXT-0158369 CLAVE: AD2733

Esto es una copia auténtica imprimible de un documento electrónico archivado en el Ministerio de Educación, aplicando lo dispuesto por el Art. 25 de D.S. 070-2013-PCM y la Tercera Disposición Complementaria Final del D.S. 028-2016-PCM. Su autenticidad e integridad pueden ser contrastadas a través de la siguiente dirección web:

https://lesinad.minedu.gob.pe/s_inadmed_1/VDD_ConsultaDocumento.aspx



BICENTENARIO
PERÚ
2024



www.gob.pe/minedu

Calle Del Comercio 193
San Borja, Lima 41, Perú
T: (511)615 5800

Autorización al acceso a los datos de los estudiantes del año 2022 de la provincia La
Convención



Sin otro particular.

Atentamente,

FIRMADO DIGITALMENTE

GISELA JANETT CRUZ SILVA
 FUNCIONARIA RESPONSABLE DE ENTREGAR LA INFORMACIÓN PÚBLICA²
 MINISTERIO DE EDUCACIÓN

Documento electrónico firmado digitalmente en el marco de la Ley N° 27269, Ley de Firmas y Certificados Digitales, su Reglamento y modificatorias. La integridad del documento y la autoría de la(s) firma(s) pueden ser verificadas en <https://app.s.firmaperu.gob.pe/web/validador.xhtml/>



CRUZ SILVA Gisela Janett
 FAU 20131D19981 hard
 JEFA DE LA OFICINA DE
 ATENCIÓN AL CIUDADANO
 Y GESTIÓN DOCUMENTAL
 - OIGDGED MINEDU
 Soy el autor del documento
 2024/02/28 15:12:29



PAULO VALDIVIA Jessica
 Enajm FAU 20131D19981
 soft
 ESPECIALISTA LEGAL DE
 ACCESO A LA
 INFORMACIÓN PÚBLICA -
 OIGDGED MINEDU
 Soy V° B°
 MINISTERIO DE EDUCACIÓN 2024/02/28 12:47:47

² Resolución Ministerial N° 005-2019-MINEDU a través de la cual se designa al Jefe de la Oficina de Atención al Ciudadano y Gestión Documental como Funcionario Responsable de Entregar Información; y mediante Resolución Ministerial N° 077-2023-MINEDU se designa a la actual jefa de la Oficina de Atención al Ciudadano y Gestión Documental.

EXPEDIENTE: MPT2024-EXT-0158369 CLAVE: AD2733



Esto es una copia auténtica imprimible de un documento electrónico archivado en el Ministerio de Educación, aplicando lo dispuesto por el Art. 25 de D.S. 070-2013-PCM y la Tercera Disposición Complementaria Final del D.S. 026-2018-PCM. Su autenticidad e integridad pueden ser contrastadas a través de la siguiente dirección web:

https://es.inad.minedu.gob.pe/s_inadmed_1/VDD_ConsultaDocumento.aspx



BICENTENARIO
 PERU
 2024



www.gob.pe/minedu

Calle Del Comercio 193
 San Borja, Lima 41, Perú
 T: (511)815 5800

Autorización al acceso a los datos de los estudiantes del año 2022 de la provincia La
 Convención



"Decenio de la Igualdad de oportunidades para mujeres y hombres"
 "Año del Bicentenario, de la consolidación de nuestra independencia, y de la conmemoración de las heroicas batallas de Junín y Ayacucho"

Lima, 27 de febrero de 2024

OFICIO N° 00351-2024-MINEDU/SPE-OSEE-UE

Señora
GISELA JANETT CRUZ SILVA
 Jefa de la Oficina de Atención al Ciudadano y Gestión Documental
 Presente. -

Asunto : Solicitud de información en el marco de la Ley de Transparencia y Acceso a la Información Pública.

Referencia : Expediente SINAD N° MPT2023-EXT-0158369

De mi consideración:

Es grato dirigirme a usted en atención al expediente SINAD de la referencia, en el que el señor Pascual Barcena Carrasco (DNI N° 42477510) solicita al Ministerio de Educación (Minedu), en el marco de la Ley de transparencia y acceso a la información pública, data innominada del Sistema de Información de Apoyo a la Gestión de la Institución Educativa (SIAGIE), correspondiente a la información de estudiantes de la provincia La Convención durante el año 2022. Cabe señalar que, en dicho expediente, el solicitante mencionó las variables que requiere del SIAGIE, a fin de que el Minedu le brinde la información idónea.

Sobre el particular, informo a usted que el Minedu a partir del año 2011 brinda el aplicativo web del SIAGIE, el cual está administrado por esta Unidad¹, con la finalidad de que las instituciones educativas (I.E.E.) registren la matrícula y evaluación de los estudiantes de las modalidades de Educación Básica Regular (EBR) y, a partir del año 2016 en Educación Básica Especial (EBE). Cabe mencionar que desde el mes de noviembre del año 2020 la modalidad de Educación Básica Alternativa (EBA) se encuentra implementada en el SIAGIE.

En atención a lo solicitado, este despacho, en el marco de sus funciones y la información disponible, procedió a extraer de la base de datos del SIAGIE la data correspondiente, a fin de atender lo requerido por el señor Pascual Barcena Carrasco. Tal información es remitida mediante el siguiente enlace:

- [SINAD_0158369_AIP.rar](#)

Sin otro particular, aprovecho la oportunidad para expresarle las muestras de mi especial consideración.

Atentamente,

CLAUDIA PAOLA LISBOA VÁSQUEZ
 Jefa de la Unidad de Estadística



Firmado digitalmente por:
 LISBOA VÁSQUEZ Claudia
 Paola FAU 20131370998 soft
 Motivo: Soy el autor del
 documento
 Fecha: 27/02/2024 17:15:33-0500

¹ Tal como lo indica el Reglamento de Organización y Funciones (ROF) del MINEDU, aprobado mediante el Decreto Supremo N° 001-2015-MINEDU.

EXPEDIENTE: MPT2024-EXT-0158369 CLAVE: F948E2

Esto es una copia auténtica imprimible de un documento electrónico archivado en el Ministerio de Educación, aplicando lo dispuesto por el Art. 25 de D.S. 070-2013-PCM y la Tercera Disposición Complementaria Final del D.S. 026-2016-PCM. Su autenticidad e integridad pueden ser contrastadas a través de la siguiente dirección web:

https://esinad.minedu.gob.pe/es_sinadmed_#/VDD_ConsultaDocumento.aspx



BICENTENARIO
 PERÚ
 2024



www.gob.pe/minedu Calle Del Comercio 100
 San Borja, Lima 41, Perú
 T: (51)615 5800

Documento electrónico firmado digitalmente en el marco de la Ley N° 27269, Ley de Firmas y Certificados Digitales, su Reglamento y modificatorias. La integridad del documento y la autenticidad de la(s) firma(s) pueden ser verificadas en <https://apps.firmaperu.gob.pe/web/validador.xhtml>



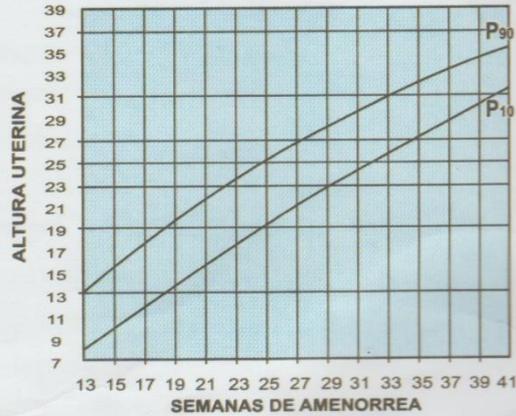
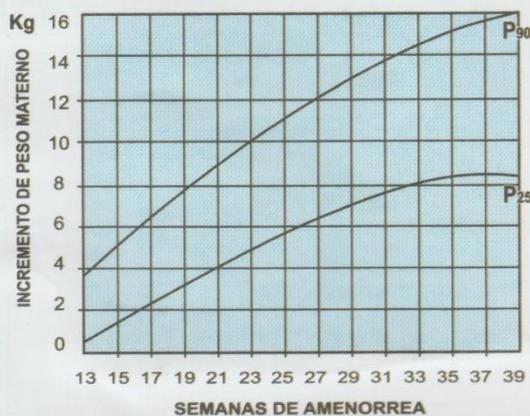
Autorización al acceso a los datos de los estudiantes del año 2022 de la provincia La Convención.

PARTO O ABORTO	Estado HCMP : <input type="checkbox"/> Atención Prenatal <input type="checkbox"/> Aborto <input type="checkbox"/> Parto <input type="checkbox"/>		Producto de la concepción Hijo Único <input type="checkbox"/> Embarazo Múltiple <input type="checkbox"/> Orden <input type="checkbox"/> Aborto <input type="checkbox"/>	
	Ingreso al establecimiento por parto incompleto REFERENCIA AL INGRESO: Si <input type="checkbox"/> No <input type="checkbox"/> No aplica <input type="checkbox"/> FREC. RESP.: <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> TEMP.: <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> SITUACIÓN: Longitudinal <input type="checkbox"/> Transversa <input type="checkbox"/> Izquierda <input type="checkbox"/> POSICIÓN: Derecha <input type="checkbox"/> Izquierda <input type="checkbox"/> TAMAÑO FETAL ACORDE: No <input type="checkbox"/> Si <input type="checkbox"/> INICIO: Espontáneo <input type="checkbox"/> Inducido <input type="checkbox"/> DILATACION: <input type="checkbox"/> <input type="checkbox"/> LIQUIDO AMNÍOTICO: Claro <input type="checkbox"/> Verde claro <input type="checkbox"/> Verde oscuro <input type="checkbox"/> MEMBRANAS: Rotas <input type="checkbox"/> Integras <input type="checkbox"/> FECHA RUPTURA: / /			
Signos y Síntomas de Alerta Anasarca: Si <input type="checkbox"/> No <input type="checkbox"/> Hematuria: Si <input type="checkbox"/> No <input type="checkbox"/> Hemorragia Vaginal: Si <input type="checkbox"/> No <input type="checkbox"/> Hipot. ortostática: Si <input type="checkbox"/> No <input type="checkbox"/> Escotomas: Si <input type="checkbox"/> No <input type="checkbox"/> Ictericia: Si <input type="checkbox"/> No <input type="checkbox"/> Epigastralgia: Si <input type="checkbox"/> No <input type="checkbox"/> Petequias: Si <input type="checkbox"/> No <input type="checkbox"/> Dolor hipocond.der.: Si <input type="checkbox"/> No <input type="checkbox"/> Proteinuria: Si <input type="checkbox"/> No <input type="checkbox"/>		Corticoides antenatales (28-34sem) Completo <input type="checkbox"/> No recibe <input type="checkbox"/> Incompleto <input type="checkbox"/> No Aplica <input type="checkbox"/> SEM. INICIO: <input type="checkbox"/> <input type="checkbox"/>		
Terminación TERMINACIÓN: P. Espontáneo <input type="checkbox"/> Forceps <input type="checkbox"/> Vacuum <input type="checkbox"/> Cesarea electiva <input type="checkbox"/> Cesarea emerg. <input type="checkbox"/> POSIC. GESTANTE: Horizontal <input type="checkbox"/> Vertical <input type="checkbox"/> No Aplica <input type="checkbox"/> PARTO GRAMA: Si <input type="checkbox"/> No <input type="checkbox"/> PARTO CON ACOMPAÑANTE: Si <input type="checkbox"/> No <input type="checkbox"/> DURACIÓN: Normal <input type="checkbox"/> Prolongado <input type="checkbox"/> Precipitado <input type="checkbox"/> No aplica <input type="checkbox"/> MUERTE INTRAUTERINA: No hubo <input type="checkbox"/> Durante embarazo <input type="checkbox"/> Durante parto <input type="checkbox"/> Momento Desconocido <input type="checkbox"/> EPISIOTOMIA: Si <input type="checkbox"/> No <input type="checkbox"/> No Aplica <input type="checkbox"/> DESGARROS: No hubo <input type="checkbox"/> Grado I <input type="checkbox"/> Grado II <input type="checkbox"/> Grado III/IV <input type="checkbox"/> No aplica <input type="checkbox"/> ALUMBRAMIENTO: Activo <input type="checkbox"/> Expontáneo <input type="checkbox"/> Manual <input type="checkbox"/> PLACENTA: Completa <input type="checkbox"/> Incompleta <input type="checkbox"/> Retenida <input type="checkbox"/> LIGADURA CORDÓN: Precoz <input type="checkbox"/> Temprana <input type="checkbox"/> Tardía <input type="checkbox"/>		Tipo Procedimiento CESAREA: Corporal <input type="checkbox"/> Segmentaria <input type="checkbox"/> No aplica <input type="checkbox"/> Medicación en Parto (ver anexo): Sin medicación <input type="checkbox"/> MEDICACION: 1. _____ 2. _____ 3. _____ MEDICAMENTOS: 1. _____ 2. _____ 3. _____		
Indicación Principal Parto Operatorio _____ No hubo <input type="checkbox"/>		Atención NIVEL: FONP <input type="checkbox"/> FONB <input type="checkbox"/> FONE <input type="checkbox"/> FONI <input type="checkbox"/> Domiciliario <input type="checkbox"/> Otro <input type="checkbox"/> PARTO o LEGRADO NEONATO: Médico <input type="checkbox"/> Residente Médico <input type="checkbox"/> Obstetriz <input type="checkbox"/> Interno <input type="checkbox"/> Enfermera <input type="checkbox"/> Téc. de Enfermería <input type="checkbox"/> Estudiante <input type="checkbox"/> Empírica o partera <input type="checkbox"/> Familia <input type="checkbox"/> Otro <input type="checkbox"/> Responsable de la atención del Parto: _____ (Firma y sello) Responsable de la atención del Neonato: _____ (Firma y sello)		
Recién Nacido Sexo: Femenino <input type="checkbox"/> Masculino <input type="checkbox"/> Indeterminado <input type="checkbox"/> Peso: <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> g < 2500g <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> g < 1500g <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> g > 4000g P.Cef. <input type="checkbox"/> mm Temp. <input type="checkbox"/> C Talla <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> mm		HC RN: _____ NOMBRE RN: _____ Patología Recién Nacido: sin patologías <input type="checkbox"/> Fecha: / / 1. _____ 2. _____ 3. _____ Otras (CIE 10): 1: _____ 2: _____		
Edad por Ex. Físico: <input type="checkbox"/> sem < 37 <input type="checkbox"/> > 42 APGAR: 4-6 <input type="checkbox"/> 0-3 <input type="checkbox"/> Peso x Edad Gestacional: Adecuado <input type="checkbox"/> Pequeño <input type="checkbox"/> Grande <input type="checkbox"/>		Reanimación Respiratoria No <input type="checkbox"/> Sí <input type="checkbox"/> Bolsa y Máscara <input type="checkbox"/> Reanim. Avanz. <input type="checkbox"/>		
Exam. Físico Normal <input type="checkbox"/> Anormal <input type="checkbox"/> Hospitaliz. Si <input type="checkbox"/> No <input type="checkbox"/> S.Luética RN: Si <input type="checkbox"/> No <input type="checkbox"/> VDR/LRPR: Si <input type="checkbox"/> No <input type="checkbox"/> Se hizo <input type="checkbox"/> No <input type="checkbox"/>		Medicación En reanim. RN Sí <input type="checkbox"/> No <input type="checkbox"/>		
Exámenes de Laboratorio Glicemia: Si <input type="checkbox"/> No <input type="checkbox"/> Gota Gruesa: Si <input type="checkbox"/> No <input type="checkbox"/> Gota Fresca: Si <input type="checkbox"/> No <input type="checkbox"/>		Vitamina K Si <input type="checkbox"/> No <input type="checkbox"/> Profilaxis Ocular Si <input type="checkbox"/> No <input type="checkbox"/> Ctrl. Puerperio Inmediato Si <input type="checkbox"/> No <input type="checkbox"/>		
Tipos de Sangre antes de Alta Grupo: A <input type="checkbox"/> B <input type="checkbox"/> AB <input type="checkbox"/> O <input type="checkbox"/> No se hizo <input type="checkbox"/> Rh: Rh+ <input type="checkbox"/> Rh- <input type="checkbox"/> No se hizo <input type="checkbox"/>		Reingreso RN Fecha: / / Reingreso: No <input type="checkbox"/> Si <input type="checkbox"/> Diagnóst.: _____ Fecha egreso: / / Egreso: Sano <input type="checkbox"/> Con Patología <input type="checkbox"/> Fallece <input type="checkbox"/> NO Aplica <input type="checkbox"/> Dx.Fallecim.: _____ Dx.Traslado: _____ Establ.Trasl.: _____		
Exámenes de Laboratorio TSH: Si <input type="checkbox"/> No <input type="checkbox"/> Fibrosis Quística: Si <input type="checkbox"/> No <input type="checkbox"/> Fenilceto nuria: Si <input type="checkbox"/> No <input type="checkbox"/> Hiperplasia Suprarrenal: Si <input type="checkbox"/> No <input type="checkbox"/>		Reingreso Materno Fecha: / / Reingreso: No <input type="checkbox"/> Si <input type="checkbox"/> Diagnóst.: _____ Fecha egreso: / / Egreso: Sano <input type="checkbox"/> Con Patología <input type="checkbox"/> Fallece <input type="checkbox"/> NO Aplica <input type="checkbox"/> Dx.Fallecim.: _____ Dx.Traslado: _____ Establ.Trasl.: _____		
Alojamiento Conjunto Si <input type="checkbox"/> No <input type="checkbox"/> Contacto Piel a Piel (LMH): Si <input type="checkbox"/> No <input type="checkbox"/> LME: Si <input type="checkbox"/> No <input type="checkbox"/>		Tamizaje Neonatal TSH: Si <input type="checkbox"/> No <input type="checkbox"/> Fibrosis Quística: Si <input type="checkbox"/> No <input type="checkbox"/> Fenilceto nuria: Si <input type="checkbox"/> No <input type="checkbox"/> Hiperplasia Suprarrenal: Si <input type="checkbox"/> No <input type="checkbox"/>		
Vacunas RN al Alta BCG: Si <input type="checkbox"/> No <input type="checkbox"/> Hepatitis B: Si <input type="checkbox"/> No <input type="checkbox"/>		Anticonceptivos Ligadura Tubaria <input type="checkbox"/> Anticoncec. combinada <input type="checkbox"/> Abstinencia Peridica <input type="checkbox"/> MELA <input type="checkbox"/> Solo Ori/Consej <input type="checkbox"/> Cordón: Progéstag. Inyectables <input type="checkbox"/> Ninguno <input type="checkbox"/> DIU: Orales <input type="checkbox"/> Otro <input type="checkbox"/>		
PUERPERIO INMEDIATO Horas/días postparto o aborto: _____ Temperatura: _____ Pulso (latidos/minuto): _____ Pres. Arterial max/min. (mm. Hg): _____ Involución uterina: _____ Caract. loquios (Sangrado: Leve Moderado Severo): _____ Herida oper. (abdominal/periné): _____ Observaciones: _____		Fecha: / / Control RN: Si <input type="checkbox"/> No <input type="checkbox"/>		

Ficha de registro de afiliación a SIS

ATENCIÓNES PRENATALES	ATENCIÓN 1	ATENCIÓN 2	ATENCIÓN 3	ATENCIÓN 4						
Fecha y hora atención (a/m/d hh:mm)	// : // :	// : // :	// : // :	// : // :	// : // :	// : // :	// : // :	// : // :	// : // :	// : // :
Edad Gest. (Semanas)										
Peso madre (kg)										
Temperatura (°c)										
Tensión Arter. (mm. hg)										
Pulso materno (por min.)										
Altura Uterina (cm)										
Situación (L/TNA)										
Presentación (C/P/NA)										
Posición (D/I/NA)										
F.C.F. (por min./NA)										
Mov. fetal (+/+/+/+/SM/NA)										
Proteinuria Cualitativa (+/+/+/+/NSH)										
Edema (+/+/+/+/se)										
Reflejo Osteotendinoso (+/+/+/+/)										
Examen de Pozón (Formado/NO Form/Sin Exam)										
Indic. Hierro/ AC Fólico (mayor o igual a 14 sem)										
Indic. Calcio (mayor o igual a 20 sem)										
Indic. Ac. Fólico										
Orient. Consej. (PF/TS/Nut/ am/VIH/TBC/ No se hizo /NA)										
E G de Eco. Control. (sem/No se hizo/NA)										
Perfil Biofísico (4,6,8,10 de 10/NSH/NA)										
Cita (a/m/d)										
Visita domicil. Si/NO/NA										
Plan Parto (Control/visita/No se hizo/NA)										
Estab. de la atención										
Responsable de atención										
Nro. Formato SIS										

L = Longitudinal T= Transversa C = Cefálica P= Pélvica D = Derecha I = Izquierda SM = Sin Movimiento SE = Sin Edema NA = No Aplica NSH = No se hizo



Patología Maternas (CIE10) Diagnosticadas

Sin patologías Fecha: / /

Otras patologías (CIE 10):
 1:
 2:

Referencia - Consulta Externa Si No No Aplica Fecha / / Establ. Trasl. :
 Referencia - Emergencia Si No No Aplica Fecha / / Establ. Trasl. :
 Referencia - Apoyo al Diagnóst. Si No No Aplica Fecha / / Establ. Trasl. :

PSICOPROFILAXIS ESTIMULACIÓN PRENATAL PLAN DE PARTO Si No No Aplica ALOJADA EN CASA DE ESPERA Si No No Aplica

Ficha de registro de afiliación a SIS

HISTORIA CLINICA INDIVIDUAL

CODIGO DEL ESTABLECIMIENTO

CATEGORIA: NIVEL II - I

NOMBRES Y APELLIDOS DEL PACIENTE

N° HISTORIA CLINICA

DATOS GENERALES

LUGAR DE NACIMIENTO

FECHA DE NACIMIENTO

EDAD	M	F	DOMICILIO		N° DE TELEFONO		FECHA DE INGRESO	
							/ /	
DOCUMENTO DE IDENTIDAD	DNI.		PASA PORTE		CARNE DE EXTRANJERIA			
N° ESSALUD	N° SIS		SOAT		ESTADO CIVIL			
					S	C	V	D
GRADO DE INSTRUCCION								PROFESION
ANALFABETO	PRIMARIA		SECUNDARIA		SUPERIOR			
OCUPACIÓN	RELIGIÓN			PROCEDENCIA				
Fono del Acompañante	Nombres de la persona acompañante				Dirección de la persona acompañante			

Ficha de registro de la historia clínica individual

Anexo 4: Otros

Recolección de dato en el Hospital de Quillabamba



Verificación de datos según la ficha de registro



Verificación de datos según la ficha de registro



Recolección de la base de datos de la central Red de Salud Cusco

Codificación en R

```
summary(datos) # resumen de La base de datos
```

```
## Tipo.Gestion Nivel.Educativo Grado.Instruccion
## Privado: 54 Primaria : 709 Primaria (3ro a 6to):709
## Publico:2749 Secundaria :2072 1ro secundaria :536
## NA's : 190 Básica altrnativa-Avanzado: 22 2do secundaria :502
## NA's : 190 3ro secundaria :480
## 4to secundaria :437
## 5to secundaria :139
## NA's :190
## Quintil.Pobreza Pobreza.Monetaria Planificacion Distrito
## Extrema pobreza:2366 Min. : 83 No:1732 Santa Ana: 273
## Muy pobre : 354 1st Qu.: 218 Si:1261 Otros :2720
## Pobre : 273 Median :2325
## Mean :1836
## 3rd Qu.:3355
## Max. :4545
## NA's :179
## Migracion Micro.Red Categoria.Salud.lugar Embarazo
## No :2373 Santa Ana: 700 I-1: 814 No:2889
## Si : 430 Otros :2293 I-2: 660 Si: 104
## NA's: 190 I-3: 463
## I-4:1056
##
##
##
```

```
#Para ver que columnas tienen valores perdidos
```

```
which(colSums(is.na(datos))!=0)
```

```
## Tipo.Gestion Nivel.Educativo Grado.Instruccion
```

```
Pobreza.Monetaria
```

```
## 1 2 3
```

```
5
```

```
## Migracion
```

```
##
```

```
a
```

```
##
```

```
## Missings in variables:
```

```
## Variable Count
```

```
## Tipo.Gestion 190
```

```
## Nivel.Educativo 190
```

```
## Grado.Instruccion 190
## Pobreza.Monetaria 179
## Migracion 190
```

```
summary(a)
```

```
##
## Missings per variable:
## Variable Count
## Tipo. Gestión 190
## Nivel.Educativo 190
## Grado.Instruccion 190
## Quintil.Pobreza 0
## Pobreza.Monetaria 179
## Planificación 0
## Distrito 0
## Migración 190
## Micro.Red 0
## Categoria.Salud.lugar 0
## Embarazo 0
##
## Missings in combinations of variables:
## Combinations Count Percent
## 0:0:0:0:0:0:0:0:0:0 2653 88.6401604
## 0:0:0:0:1:0:0:0:0:0 150 5.0116940
## 1:1:1:0:0:0:0:1:0:0 161 5.3792182
## 1:1:1:0:1:0:0:1:0:0 29 0.9689275
```

```
aggr(datos,numbers=T, sortComb=TRUE,
     sortVar=TRUE, only.miss=TRUE)
```

```
##
## Variables sorted by number of missings:
## Variable Count
## Tipo.Gestion 0.06348146
## Nivel.Educativo 0.06348146
## Grado.Instruccion 0.06348146
## Migracion 0.06348146
## Pobreza.Monetaria 0.05980621
## Quintil.Pobreza 0.00000000
## Planificacion 0.00000000
## Distrito 0.00000000
## Micro.Red 0.00000000
## Categoria.Salud.lugar 0.00000000
## Embarazo 0.00000000
```

```
# Imputación
# Usando una medida de Tendencia Central
```

```
datos.c<-centralImputation(datos)
summary(datos.c)
```

```
## Tipo.Gestion Nivel.Educativo Grado.Instruccion
## Privado: 54 Primaria : 709 Primaria (3ro a 6to):899
## Publico:2939 Secundaria :2262 1ro secundaria :536
## Básica altrnativa-Avanzado: 22 2do secundaria :502
## 3ro secundaria :480
## 4to secundaria :437
## 5to secundaria :139
## Quintil.Pobreza Pobreza.Monetaria Planificacion Distrito
## Extrema pobreza:2366 Min. : 83 No:1732 Santa Ana: 273
## Muy pobre : 354 1st Qu.: 218 Si:1261 Otros :2720
## Pobre : 273 Median :2325
## Mean :1866
## 3rd Qu.:3355
## Max. :4545
## Migracion Micro.Red Categoria.Salud.lugar Embarazo
## No:2563 Santa Ana: 700 I-1: 814 No:2889
## Si: 430 Otros :2293 I-2: 660 Si: 104
## I-3: 463
## I-4:1056
```

```
t.ids<-createDataPartition(datos$Embarazo,p=0.7, list=F)
```

```
table(datos.entreno$Embarazo)
```

```
##
## No Si
## 2023 73
```

```
table(datos.test$Embarazo)
```

```
##
## No Si
## 866 31
```

```
# SVM Lineal
```

```
control <- trainControl(method="cv", classProbs=T,savePredictions = T)
hiperparametros <- expand.grid( C = c(0,0.1,0.25,0.5,1, 5,10))
fit.svmLineal <- train(Embarazo~.,
                      data=datos.entreno,
                      method="svmLinear",
                      metric="Accuracy",
                      preProc=c("center", "scale"),
                      tuneGrid = hiperparametros,
```

```

trControl=control,
fit=T)

fit.svmLineal

## Support Vector Machines with Linear Kernel
##
## 2096 samples
## 8 predictor
## 2 classes: 'No', 'Si'
##
## Pre-processing: centered (12), scaled (12)
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1886, 1886, 1887, 1886, 1887, 1887, ...
## Resampling results across tuning parameters:
##
## C Accuracy Kappa
## 0.10 0.9651743 0
## 0.25 0.9651743 0
## 0.50 0.9651743 0
## 1.00 0.9651743 0
## 5.00 0.9651743 0
## 10.00 0.9651743 0
##
## Accuracy was used to select the optimal model using the largest value
.
## The final value used for the model was C = 0.1.

prediC.svmLineal <- predict(fit.svmLineal,datos.test)
confusionMatrix(prediC.svmLineal ,datos.test$Embarazo)

## Confusion Matrix and Statistics
##
## Reference
## Prediction No Si
## No 850 14
## Si 16 17
##
## Accuracy : 0.9665
## 95% CI : (0.9513, 0.9764)
## No Information Rate : 0.9654
## P-Value [Acc > NIR] : 0.5476
##
##
## McNemar's Test P-Value : 7.118e-08
##
## Sensitivity : 0.86
## Specificity : 0.92
## Prevalence : 0.9654
## Detection Rate : 0.9654
## Detection Prevalence : 1.0000
## Balanced Accuracy : 0.5000
##

```

```
##      'Positive' Class : No
##
```

```
impor1=varImp(fit.svmLineal)
impor1

## ROC curve variable importance
##
##              Importance
## Nivel.Educativo      100.000
## Pobreza.Monetaria    70.028
## Quintil.Pobreza     35.692
## Categoria.Salud.lugar 34.177
## Micro.Red            8.379
## Planificacion        7.877
## Distrito             5.899
## Tipo.Gestion         0.000

plot(impor1)
```

```
# SVM radial
set.seed(123)
library(caret)
control <- trainControl(method="cv", classProbs=T,savePredictions = T)
fit.svmRadial <- train(Embarazo~.,
                      data=datos.entreno,
                      method="svmRadial",
                      metric="Accuracy", preProc=c("center", "scale"),
                      trControl=control, fit=FALSE)

fit.svmRadial

## Support Vector Machines with Radial Basis Function Kernel
##
## 2096 samples
##   8 predictor
##   2 classes: 'No', 'Si'
##
## Pre-processing: centered (12), scaled (12)
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1886, 1886, 1887, 1886, 1887, 1887, ...
## Resampling results across tuning parameters:
##
##   C      Accuracy      Kappa
## 0.25  0.9637389  -0.002050781
## 0.50  0.9651720   0.023525357
## 1.00  0.9651743   0.000000000
```

```
##
## Tuning parameter 'sigma' was held constant at a value of 0.07570959
## Accuracy was used to select the optimal model using the largest value
.
## The final values used for the model were sigma = 0.07570959 and C = 1
.

prediC.svmRadial <- predict(fit.svmRadial,datos.test)
confusionMatrix(prediC.svmRadial ,datos.test$Embarazo)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction  No  Si
##           No 840  8
##           Si  26  23
##
##           Accuracy : 0.9602
##           95% CI : (0.9513, 0.9764)
##           No Information Rate : 0.9654
##           P-Value [Acc > NIR] : 0.5476
##
##
##           McNemar's Test P-Value : 7.118e-08
##
##           Sensitivity : 0.7401
##           Specificity : 0.9364
##           Prevalence : 0.9654
##           Detection Rate : 0.9654
##           Detection Prevalence : 1.0000
##           Balanced Accuracy : 0.5000
##           'Positive' Class : No
##
```

```
impor2=varImp(fit.svmRadial)
impor2

## ROC curve variable importance
##
##           Importance
## Nivel.Educativo      100.000
## Pobreza.Monetaria    70.028
## Quintil.Pobreza     35.692
## Categoria.Salud.lugar 34.177
## Micro.Red            8.379
## Planificacion        7.877
## Distrito            5.899
## Tipo.Gestion         0.000

plot(impor2)
```

```

## svm polinomial
set.seed(123)
library(caret)
control <- trainControl(method="cv", classProbs=T,savePredictions = T)
fit.svmpoly <- train(Embarazo~.,
                    data=datos.entreno, method="svmPoly", metric="Accuracy",
                    preProc=c("center", "scale"),
                    trControl=control, fit=FALSE)

fit.svmpoly

## Support Vector Machines with Polynomial Kernel
##
## 2096 samples
## 8 predictor
## 2 classes: 'No', 'Si'
##
## Pre-processing: centered (12), scaled (12)
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 1886, 1886, 1887, 1886, 1887, 1887, ...
## Resampling results across tuning parameters:
## Accuracy was used to select the optimal model using the largest value
.
## The final values used for the model were degree = 2, scale = 0.1 and
C = 0.5.

prediC.svmpoly <- predict(fit.svmpoly,datos.test)
confusionMatrix(prediC.svmpoly ,datos.test$Embarazo)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction No  Si
##           No 854  7
##           Si  12 24
##
##           Accuracy : 0.9788
##           95% CI : (0.9513, 0.9764)
##           No Information Rate : 0.9654
##           P-Value [Acc > NIR] : 0.5476
##
##
##           McNemar's Test P-Value : 6.118e-08
##
##           Sensitivity : 0.7742
##           Specificity : 0.9409
##           Prevalence : 0.9654
##           Detection Rate : 0.9654
##           Detection Prevalence : 1.0000
##           Balanced Accuracy : 0.5000
##
##           'Positive' Class : No
##

```

```

impor3=varImp(fit.svmPoly)
impor3

## ROC curve variable importance
##
##              Importance
## Nivel.Educativo      100.000
## Pobresa.Monetaria    70.028
## Quintil.Pobreza     35.692
## Categoria.Salud.lugar 34.177
## Micro.Red            8.379
## Planificacion        7.877
## Distrito             5.899
## Tipo.Gestion         0.000

plot(impor3)

```

```

#run res and plot ROC
library(MLevel)

modelos<-list(svmLineal=fit.svmLineal, svmPoly=fit.svmPoly, svmRadial=fi
t.svmRadial)

res <- evalm( modelos,gnames=c('svmLineal','svmPoly','svmRadial'),
              plots='roc')

## svmLineal AUC-ROC = 0.78
## svmPoly AUC-ROC = 0.84
## svmRadial AUC-ROC = 0.82

```