

UNIVERSIDAD PERUANA UNIÓN
ESCUELA DE POSGRADO
Unidad de Posgrado de Ingeniería y Arquitectura



Una Institución Adventista

Modelo predictivo basado en Machine Learning para el pronóstico de deserción a la vacunación contra el virus del papiloma humano (VPH)

Trabajo de investigación para obtener el Grado Académico de Maestro en Ingeniería de Sistemas con Mención en Ingeniería de Software

Autor:

Bach. Urlish Kleyber Marroquin Marroquin

Asesor:

Mg. Nemias Saboya Rios

Co-Asesor

Mg. Abel Angel Sullon Macalupu

Lima, febrero 2021

DECLARACIÓN JURADA DE AUTORÍA DE TRABAJO DE INVESTIGACIÓN

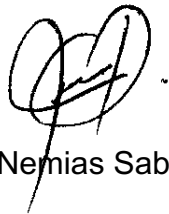
Mg. Nemias Saboya Rios, de la Escuela de Posgrado, Unidad de Posgrado de Ingeniería y Arquitectura, de la Universidad Peruana Unión.

DECLARO:

Que la presente investigación titulada: **“Modelo predictivo basado en Machine Learning para el pronóstico de deserción a la vacunación contra el virus del papiloma humano (VPH)”** constituye la memoria que presenta el bachiller Urlish Kleyber Marroquin Marroquin para aspirar al Grado Académico de Maestro en Ingeniería de Sistemas con Mención en Ingeniería de Software, cuyo trabajo de investigación ha sido realizada en la Universidad Peruana Unión bajo mi dirección.

Las opiniones y declaraciones en este informe son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente declaración en la ciudad de Lima, a los 15 días del mes de marzo del año 2021.



Mg. Nemias Saboya Rios

ACTA DE SUSTENTACIÓN DE TRABAJO DE INVESTIGACIÓN DE MAESTRO(A)

En Lima, Ñaña, Villa Unión, a 22 días del mes de febrero del año 2021, siendo las 8:00 a. m, se reunieron en la modalidad online sincrónica, bajo la dirección del Señor Presidente del Jurado: Dr. Josué Edison Turpo Chaparro el secretario: Mg. Geraldine Verónica Alvizuri Llerena los demás miembros: Mg. Herminio Paucar Curasma y el M.Sc. Fredy Abel Huanca Torres y el asesor: Mg. Nemias Saboya Rios con el propósito de administrar el acto académico de sustentación de trabajo de investigación de Maestro(a) titulada: Modelo predictivo basado en Machine Learning para el pronóstico de deserción a la vacunación contra el virus del papiloma humano (VPH). del Bachiller/Licenciado(a) Urlish Kleyber Marroquin Marroquin

Conducente a la obtención del Grado Académico de Maestro(a) en: Ingeniería de Sistemas (Nomenclatura del Grado Académico) Ingeniería de Software con Mención en

El Presidente inició el acto académico de sustentación invitando al candidato hacer uso del tiempo determinado para su exposición. Concluida la exposición, el Presidente invitó a los demás miembros del Jurado a efectuar las preguntas, cuestionamientos y aclaraciones pertinentes, los cuales fueron absueltos por el candidato. Luego se produjo un receso para las deliberaciones y la emisión del dictamen del Jurado. Posteriormente, el Jurado procedió a dejar constancia escrita sobre la evaluación en la presente acta, con el dictamen siguiente:

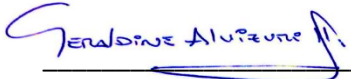
Bachiller/Licenciado (a): Urlish Kleyber Marroquin Marroquin

CALIFICACIÓN	ESCALAS			Mérito
	Vigesimal	Literal	Cualitativa	
Aprobado	18	A-	Con nominación muy bueno	Sobresaliente

(*) Ver parte posterior

Finalmente, el Presidente del Jurado invitó al candidato a ponerse de pie, para recibir la evaluación final. Además, el Presidente del Jurado concluyó el acto académico de sustentación, procediéndose a registrar las firmas respectivas.

Presidente



Secretario

Asesor

Miembro

Miembro

Bachiller/Licenciado(a)

Modelo predictivo basado en *Machine Learning* para el pronóstico de deserción a la vacunación contra el virus del papiloma humano (VPH)

Urlish Marroquin

UPG Ingeniería y Arquitectura,
Escuela de Posgrado,
Universidad Peruana Unión,
Carretera Central Km 19.5
Lurigancho, Lima, Perú
+51 943204706

urlishmarroquin@upeu.edu.pe

Nemias Saboya

UPG Ingeniería y Arquitectura,
Escuela de Posgrado,
Universidad Peruana Unión,
Carretera Central Km 19.5
Lurigancho, Lima, Perú
+51 989946356

saboya@upeu.edu.pe

A. Angel Sullon

Escuela Profesional de Ingeniería de
Sistemas,
Universidad Peruana Unión,
Filial Juliaca,
Carretera Salida a Arequipa Km 6
Chullunquiani, Juliaca, Perú
+51 944184691

angeli@upeu.edu.pe

RESUMEN

En la actualidad, una de las enfermedades que está causando gran cantidad de muertes en Perú, es el cáncer de cuello uterino ocasionado por el virus del papiloma humano (VPH), sin embargo, gracias a la aplicación de la vacuna contra esta enfermedad se puede prevenir y proteger por ciertas cepas del VPH. El estudio consistió en la elaboración de un modelo predictivo utilizando *Machine Learning* para el pronóstico de deserción en la vacunación contra el VPH en niñas del grupo etario de 9 a 13 años de edad. Los datos que se utilizaron proceden del “Sistema de vacunación contra el VPH” del Ministerio de Salud del Perú (Minsa). La metodología consistió en desarrollar cuatro modelos de aprendizaje supervisado: *Decision Tree Classifier*, *Random Forest Classifier*, *Extra Trees Classifier* y *Extreme Gradient Boosting* con la intención de comparar los resultados y elegir al modelo que presente mejor desempeño para su respectiva calibración y poder ser utilizado a través de una interfaz gráfica. Los resultados demostraron que el mejor modelo de aprendizaje fue *Random Forest Classifier*, con un *Accuracy Score* 63.6140%, AUC de 63.6183%, precisión de 64%, *Recall* de 63% y *F1-score* de 63%; esto indica que el modelo clasifica el 64% de los casos como niñas que desertan a la vacunación contra el VPH.

CCS Concepts

- Computing methodologies ~ Machine learning ~ Learning paradigms ~ Supervised learning ~ Supervised learning by classification
- Computing methodologies ~ Machine learning ~ Machine learning approaches ~ Classification and regression trees
- Theory of computation ~ Logic ~ Automated reasoning

Keywords

Human papillomavirus (HPV); Machine learning; Random Forest Classifier; Modelo Predictivo.

1. INTRODUCTION

A lo largo de la historia de la humanidad, el ser humano se ha visto afectado por diferentes tipos de enfermedades, estos han impactado negativamente en la salud de las personas, en consecuencia, para evitar estas afecciones se han desarrollado diversos procedimientos para la prevención y cuidado de la salud. La OMS recomienda que se debe utilizar la vacuna como una medida de prevención y protección de la población según las posibles enfermedades que existen en la zona, ante esta situación la OMS estableció “La Agenda de Inmunización (AI) 2030” [1]. En el caso de Perú la vacunación está gestionada por el Ministerio de Salud (MINSA) [2] en la “Norma Técnica de Salud que establece el Esquema Nacional de Vacunación” [3].

Una de las enfermedades que requiere prevención oportuna es el virus del papiloma humano (VPH) [4], ya que afecta a la población susceptible de niñas en el grupo etario de 9 a 13 años de edad, que requieren someterse a 2 dosis obligatorias para garantizar su prevención. El MINSA para dar un control adecuado contra el VPH utiliza una herramienta informática llamada “Sistema de vacunación contra el VPH” [5]; cuyas variables están utilizadas y validadas por los expertos del área de salud del MINSA. Esta institución almacena gran cantidad de datos y de diferentes años. Por esta razón, los datos que se utilizó en el estudio fueron obtenidos de esta base de datos.

En este contexto, muchas organizaciones tienen almacenados grandes cantidades de datos que no son explorados y el sector de salud no es ajeno ante esta situación. Hoy en día, para manejar grandes cantidades de datos existen muchos modelos, métodos, herramientas y entre otros. Por ejemplo, el *Machine Learning* [6]. Estos son aplicados en beneficio de las organizaciones u otros según conveniencia y aún más para el sector de salud, porque es una alternativa favorable como herramienta para luchar contra la prevención contra el VPH [7].

Esta investigación otorga oportunidades favorables a la gestión del monitoreo y control del VPH utilizando Machine Learning [6] con la finalidad de mejorar el proceso de seguimiento del cumplimiento de las dosis y evitar así el abandono del tratamiento [8]; ya que en la actualidad se realizan campañas donde se requieren de una gran cantidad de recursos económicos y humanos [9], los cuales no están bien asignados, asimismo, existen reportes donde se evidencia una alta tasa de deserción de niñas que tuvieron acceso a la vacunación contra el VPH y este problema se refleja hoy en día, en las personas adultas, porque en Perú se presentan una alta tasa en mujeres adultas con VPH [8].

De acuerdo a los antecedentes obtenidos en diferentes investigaciones realizadas en el Perú respecto al VPH se encontró que se desarrollaron estudios como la relación entre los determinantes que condicionan la deserción y el cumplimiento del esquema de vacuna contra el VPH [10] y referente a la percepción de las madres sobre la vacuna contra el VPH [11]. También, se encontraron en la investigación temas relacionados a la prevención y lucha contra el VPH utilizando *Machine Learning* [6] como el referente a predecir el riesgo de cáncer de cuello uterino que es una consecuencia por no prevenir adecuadamente el VPH [12], otro, la extensión extracapsular en el carcinoma de células escamosas orofaríngeo [13] y el último en la selección del tratamiento de verrugas [14].

El MINSA tiene la necesidad de identificar aquellas niñas que no completan las dosis de su vacunación contra el VPH; sin embargo, hasta el momento no se ha trabajado ni estudiado temas relacionados al manejo de la gestión de datos para predecir con una exactitud (alta) las acciones de las niñas con el propósito de que los resultados puedan aplicarse en el Perú.

Ante la necesidad y problemática mencionada anteriormente el estudio tiene el objetivo de elaborar un modelo predictivo utilizando *Machine Learning* para el pronóstico de deserción en la vacunación contra el virus del papiloma humano (VPH).

1.1 Norma Técnica de Salud

La "Norma Técnica de Salud que establece el Esquema Nacional de Vacunación" [3]; aprobado por Resolución Ministerial N° 719-2018/MINSA el 01 de agosto de 2018; tiene el objetivo de establecer el esquema ordenado y cronológico de vacunación obligatoria en todo el Perú; el cual describe que la vacuna contra el VPH ofrece protección contra la infección generada por los principales genotipos oncogénicos del VPH, los cuales están asociados con la aparición del cáncer cervical y con las verrugas genitales; además de ofrecer un esquema de vacunación para niñas y adolescentes del 5° grado de primaria regular de instituciones educativas públicas y privadas, o que tengan desde 9 años hasta 13 años 11 meses y 29 días que por alguna razón no estén estudiando [3].

1.2 Virus del Papiloma Humano

Los virus del papiloma humano (VPH) [4] son un grupo de más de 200 virus relacionados y algunos de estos se transmiten por las relaciones sexuales con una persona infectada; el VPH se divide en 2 grupos "Los VPH de riesgo bajo que casi no causan enfermedades; sin embargo, algunos tipos de VPH de riesgo bajo producen verrugas en el área de los genitales, el ano, la boca o la garganta", "Los VPH de riesgo alto que causan varios tipos de cáncer; hay alrededor de 14 tipos de VPH de riesgo alto, dos de estos son el VPH16 y el VPH18 que causan la mayoría de los cánceres relacionados con el VPH, de los cuales destaca el cáncer de cuello uterino" [4].

1.3 Algoritmos de Machine Learning

1.3.1 Decision Tree Classifier

Divide los datos en conjuntos homogéneos, cuya división se basa en la variable de entrada que posea el diferenciador más significativo; identifica la variable más significativa a través de su valor, el cual brinda los mejores grupos de datos homogéneos [15].

1.3.2 Random Forest Classifier

Formado por un grupo de *Decision Trees*; introduce una aleatoriedad adicional al cultivar árboles, en lugar de buscar la mejor característica al dividir un nodo; busca la mejor característica entre un subconjunto aleatorio de características [15].

1.3.3 Extra Trees Classifier

Identifica las características predictivas de cada árbol y un valor de corte para dividir cada nodo, del cual solo evalúa un subconjunto aleatorio de los predictores seleccionados; cada muestra de entrenamiento se selecciona sin reemplazo [15].

1.3.4 Extreme Gradient Boosting

XGBoost es una biblioteca optimizada de aumento de gradiente distribuida que funciona impulsando árboles con el objetivo de ser altamente rápida, escalable y portátil; implementa algoritmos de aprendizaje automático en el marco de Gradient Boosting [15].

1.4 Optimización del Modelo

1.4.1 Balanceo de Datos

El balanceo de datos genera instancias artificiales para equilibrar la muestra de datos basado en la regla del vecino más cercano. La generación se realiza extrapolando nuevas instancias [16].

1.4.2 Calibración del Modelo

Optimización de Parámetros: Proceso conocido como *Grid Search*, intenta encontrar la combinación ideal de parámetros que proporcione mayor precisión al modelo elegido [17].

2. METODOLOGÍA

El proceso metodológico corresponde al estudio de un modelo predictivo [18], este sirvió como guía para el desarrollo del método utilizado y tiene cinco fases ver Figura 1: "Comprensión de datos (obtener datos, exploración de datos y pre procesamiento de datos (limpieza)), Preparación de datos, Modelado, Evaluación y despliegue.

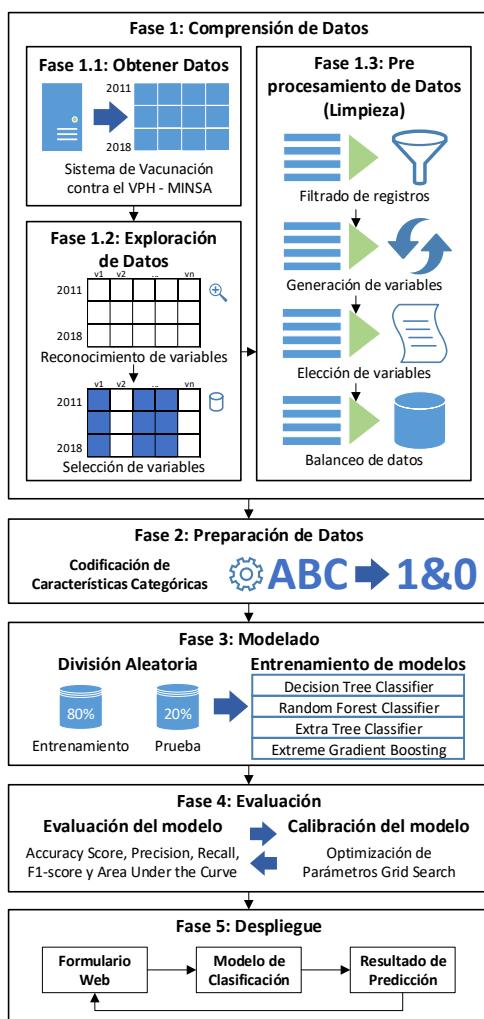


Figura 1. Proceso metodológico del modelo predictivo

2.1 Comprensión de Datos

2.1.1 Obtener datos

Esta fase consistió en la obtención de los datos estructurados que se encuentran en el sistema web “Sistema de vacunación contra el VPH” [5], perteneciente al Ministerio de Salud del Perú (MINSA) [2] y que sus variables están validadas por personal experto en el área de la salud. Los registros fueron obtenidos en un archivo Excel con consentimiento por el Ministerio de Salud. Estos registros contienen los datos de la vacunación contra el virus de papiloma humano (VPH) [4] de niñas del grupo etario de 9 a 13 años de edad de la ciudad de Lima, Perú, entre los años del 2011 al 2018, con un total 540658 registros.

2.1.2 Exploración de datos

2.1.2.1 Reconocimiento de variables

Esta fase consistió en reconocer aquellas variables obtenidas de los registros originales; dando como resultado un total de 41 variables, las cuales se detallan en la Tabla 1.

Tabla 1. Identificación de variables agrupadas por descripción

Nro.	Descripción
1-6	Oficinas responsables de los registros.

7-10	Establecimientos de salud.
11-17	Colegios de las niñas vacunadas.
18-30	Datos personales de las niñas vacunadas.
31-33	Dosis, lugar y fecha de vacunación.
34-35	Datos personales del vacunador.
36-37	Datos personales del supervisor.
38-39	Datos personales del registrador.
40-41	Número y fecha del registro

2.1.2.2 Selección de variables

Esta fase consistió en la selección inicial de variables identificadas como pertinentes al modelo; de las cuales únicamente se eligieron 8 variables, las cuales se detallan en la Tabla 2.

Tabla 2. Selección de variables referentes a la deserción a la vacunación contra el VPH

Variable	Tipo	Descripción
TIPO DOC	Cualitativa nominal	Tipo de documento de Identidad
DNI	Cualitativa nominal	Número de documento de Identidad
EDAD	Cuantitativa discreta	Edad de la niña vacunada
UBIGEO RESIDENCIA HAB	Cualitativa nominal	Ubigeo residencial habitacional
GRADO	Cualitativa ordinal	Grado del colegio de la niña
DOSIS	Cuantitativa discreta	Número de la dosis aplicada
LUGAR VACUNACION	Cualitativa nominal	Lugar donde se realizó la vacunación
FECHA VACUNACION	Cuantitativa continua	Fecha de la vacunación

2.1.3 Pre procesamiento de datos (Limpieza)

2.1.3.1 Filtrado de registros

Esta fase consistió en la selección de los registros de Lima y Callao entre la fecha de vacunación de los años 2011 hasta 2018, reduciendo la cantidad de registros a 538613.

2.1.3.2 Generación de variables

Esta fase consistió en generar nuevas variables en base de las existentes y se detallan a continuación:

- 1) Variable DESERCIÓN obtenida a partir de la variable DOSIS que estaba en relación de las niñas que no completaban su tratamiento.
- 2) Variables ESTRATO_SOCIAL y AREA_ZONA a partir de la variable UBIGEO RESIDENCIA HAB.
- 3) Variable MES_ACCESO en función a la fecha de vacunación (FECHA VACUNACION).

La generación obtuvo un total de 228735 registros y 8 variables, las cuales se detallan en la Tabla 3.

Tabla 3. Distribución de las variables luego de la transformación

Original	Nuevo	Tipo	Descripción
----------	-------	------	-------------

TIPO DOC	IDENTIDAD NACIONAL	Cualitativa nominal	Nacional o Extranjero
EDAD	EDAD_INICIO	Cuantitativa discreta	Edad de la niña vacunada
UBIGEO RESIDENCIA HAB	ESTRATO_SOCIAL	Cualitativa ordinal	Estrato social distrital
	AREA_ZONA	Cualitativa nominal	Área Zona
GRADO	ESTUDIANTE ACTIVO	Cualitativa nominal	Niña sí o no estudia
DOSIS	DESERCIÓN	Cualitativa nominal	Deserción
LUGAR VACUNACIÓN	LUGAR_VACUNACION	Cualitativa nominal	Lugar donde se realizó la vacunación
FECHA VACUNACIÓN	MES_ACCESO	Cuantitativa discreta	Mes de Acceso a la primera dosis

2.1.3.3 Elección de variables

Esta fase consistió en la realización de la correlación de Pearson por pares de todas las variables, las cuales se detallan en la Tabla 4.

Tabla 4. Resultado de la correlación de Pearson en relación a la variable Deserción

Variable	Correlación	Selección
DESERCIÓN	1.000000	Si
IDENTIDAD_NACIONAL	0.242327	Si
MES_ACCESO	0.149886	Si
LUGAR_VACUNACION	0.124278	Si
EDAD_INICIO	0.122800	Si
AREA_ZONA	0.052202	Si
ESTUDIANTE_ACTIVO	0.006419	No
ESTRATO_SOCIAL	0.005713	No

Después del análisis correspondiente se eliminaron las variables con una correlación menor al 5%, quedando 6 variables para la construcción del modelo.

2.1.3.4 Balanceo de datos

Esta fase consistió en balancear los datos; el estudio identificó 80287 registros de niñas desertoras y 148448 registros de niñas no desertoras, existiendo una brecha entre unas y las otras, ante este problema fue necesario la aplicación de la técnica de balanceo de datos “SMOTETomek”, esto ayudó al fortalecimiento del modelo.

2.2 Preparación de datos

Esta fase consistió en realizar una modificación en las opciones de respuesta de las variables categóricas, siendo transformadas aquellas que tenían 3 opciones de respuestas a 2 variables (LUGAR_VACUNACION), así mismo aquellas variables categóricas que tenían 2 opciones de respuesta a 1 variable (IDENTIDAD_NACIONAL y AREA_ZONA); todas estas variables generadas contaban con valores 1 y 0.

2.3 Modelado

Esta fase consistió en distribuir los datos en dos grupos aleatorios, el primero del 80% para entrenamiento y el segundo del 20% para pruebas [15]; se utilizó Python versión 3.8.5 como lenguaje de

programación, a través de la interfaz Spyder 4.1.5 perteneciente a la distribución de Anaconda Navigator 1.10.0. Se diseñaron cuatro modelos de aprendizaje supervisado de *Machine Learning*: “*Decision Tree Classifier, Random Forest Classifier, Extra Tree Classifier y Extreme Gradient Boosting*”, ya que son capaces de encontrar relaciones complejas no lineales en los datos [15]. Se utilizaron las bibliotecas: “*Scikit-learn, Pandas, Numpy y Matplotlib*”.

2.4 Evaluación

Esta fase consistió en ver cuál de los modelos mencionados anteriormente tuvo mejor *Accuracy Score*, para luego elegir el mejor y posteriormente evaluar su *Precision, Recall, F1-score y Area Under the Curve (AUC)*. Luego se continuó con la calibración del modelo seleccionado a través de *Grid Search*, para identificar los mejores parámetros y luego rediseñar el modelo para volver a realizar su evaluación.

2.5 Despliegue

Esta fase consistió en la puesta en ejecución del modelo predictivo a través de un formulario web desarrollado en Python con Flask, almacenada en GitHub y publicada en Heroku, el cual se conecta con el modelo de clasificación y este da un resultado de predicción el cual es devuelto y mostrado en el formulario web.

3. Resultados

3.1 Modelos

Los resultados de los modelos se obtuvieron a través del uso del puntaje de validación cruzada con 10 interacciones de los datos de entrenamiento; se configuró con un *n_splits=10* y *random state = 0*. Los resultados de esta validación se detallan en la Tabla 5.

Tabla 5. Resultado de la validación cruzada

Modelo	Accuracy
Decision Tree Classifier	0.635182
Random Forest Classifier	0.635225
Extra Tree Classifier	0.635191
Extreme Gradient Boosting	0.635208

3.2 Evaluación del mejor modelo

El mejor modelo fue el *Random Forest Classifier*, con un puntaje de exactitud de 0.635225, que es equivalente al 63.5225% (Ver tabla 5). De este modelo se pasó a realizar un análisis más detallado, requiriendo una calibración con *Grid Search*. Los resultados se muestran en la Tabla 6 y Figura 2.

Tabla 6. Resultados del modelo Random Forest Classifier

		Precision	Recall	F1-score
Random Forest Score = 0.636140 AUC = 0.636183	NO	0.63	0.65	0.64
	SI	0.64	0.63	0.63

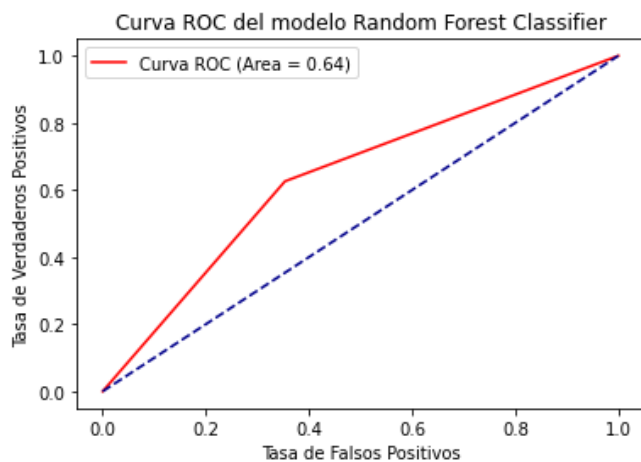


Figura 2. Curva ROC del modelo Random Forest Classifier

El modelo Random Forest Classifier obtuvo los siguientes resultados: Accuracy Score de 63.6140%, el cual muestra un nivel medio alto de predicciones que el modelo realizó correctamente; Precisión de 64%, el cual muestra un nivel medio alto en proporción a los positivos que son verdaderos positivos; Recall de 63%, el cual muestra un nivel medio alto en proporción de positivos reales correctamente clasificados; F1-score de 63%, el cual muestra un nivel medio alto en precisión y exhaustividad; Curve ROC muestra un nivel medio alto de verdaderos positivos contra los falsos positivos; y Area Under the Curve (AUC) de 63.6183%, el cual muestra un nivel medio alto en qué tan bien se clasifican las predicciones.

3.3 Despliegue

El despliegue de la aplicación fue elaborado en la plataforma web gratuita de Heroku. El proceso inicia con la visualización del formulario web, luego se debe ingresar los valores de las variables: Lugar Vacunación, Identidad Nacional, Edad Inicio, Área Zona y Mes Acceso, para posteriormente presionar el botón Predecir, el cual realiza el proceso de predicción de deserción a la vacunación contra el VPH y devuelve el resultado obtenido al formulario web, como se muestra en la Figura 3.

Figura 3. Formulario deserción a la vacunación contra el VPH
Demo en <https://hvp-vaccination.herokuapp.com/>

4. CONCLUSIONES

El modelo predictivo utilizando *Machine Learning* para el pronóstico de deserción en la vacunación contra el virus del papiloma humano (VPH) utiliza el algoritmo de *Random Forest Classifier*, ya que con este algoritmo el modelo obtuvo la mayor precisión con los datos utilizados; el cual género como resultado un *Accuracy Score* 63.6140%, AUC de 63.6183%, *Precision* de 64%, *Recall* de 63% y F1-score de 63%; lo cual indica que el modelo clasifica el 64% de los casos como niñas que desertan a la vacunación contra el VPH. Este modelo podrá ser utilizado para el desarrollo de otros modelos predictivos en temas relacionados a la deserción de la vacunación contra el VPH y además servir como base para construir herramientas para la gestión del monitoreo y control del VPH gracias a su interfaz gráfica de fácil utilización.

5. BIBLIOGRAFÍAS

- [1] W. H. Organization. "Immunization Agenda 2030: A Global Strategy to Leave No One Behind," October 25, 2020; https://www.who.int/immunization/immunization_agenda_2030/en/.
- [2] P. D. Ú. d. E. Peruano. "Ministerio de Salud," October 25, 2020; <https://www.gob.pe/minsa>.
- [3] M. d. S. d. Perú, "Norma Técnica de Salud que establece el Esquema Nacional de Vacunación," 2018, p. 105.
- [4] I. N. d. C. EE.UU. "El virus del papiloma humano (VPH) y el cáncer," October 25, 2020; <https://www.cancer.gov/espanol/cancer/causas-prevencion/riesgo/germenes-infecciosos/vph-y-cancer>.
- [5] M. d. S. O. G. d. E. e. Informática. "Registro Nominal de Vacunación Contra VPH 2014," October 25, 2020; <http://intranet5.minsa.gob.pe/vph2014/account/login.aspx>.
- [6] IBM. "¿Qué es Machine Learning?," October 25, 2020; <https://www.ibm.com/pe-es/analytics/machine-learning>.
- [7] N. I. o. H. (NIH). "Using Artificial Intelligence to Detect Cervical Cancer," December 01, 2020; <https://directorsblog.nih.gov/2019/01/17/using-artificial-intelligence-to-detect-cervical-cancer/>.
- [8] M. d. S. d. Perú, "Coberturas de Vacunación Perú 2018," 2018, p. 13.
- [9] M. d. S. d. Perú. "MINSA Biblioteca Virtual en Salud," October 25, 2020; <http://bvs.minsa.gob.pe/prevencion/>.
- [10] P. M. Cuzcano Pachas, "Relación entre los determinantes que condicionan la deserción y el cumplimiento del esquema de vacuna contra el virus del papiloma humano (VPH) Centro de Salud Surquillo 2016," *Obstetricia y Enfermería*, Universidad de San Martín de Porres, 2016.
- [11] S. M. E. Estefany, G. L. A. Paola, and C. B. Y. Lily, "Percepción de las madres sobre la vacuna contra el Virus del Papiloma Humano en una Institución Educativa Pública, 2016," *ACC CIETNA Para el cuidado de la salud*, vol. 6, no. 2, pp. 83-93, November 18, 2016.
- [12] S. K. Suman, and N. Hooda, "Predicting risk of Cervical Cancer : A case study of machine learning," *Journal of Statistics & Management Systems*, vol. 22, no. 4, pp. 689-696, June 25, 2019. DOI= <https://doi.org/10.1080/09720510.2019.1611227>

- [13] K. M. Hatten, J. Amin, and A. Isaiah, "Machine Learning Prediction of Extracapsular Extension in Human Papillomavirus–Associated Oropharyngeal Squamous Cell Carcinoma," *Otolaryngology-Head and Neck Surgery*, pp. 8, May 13, 2020. DOI=<https://doi.org/10.1177/0194599820935446>
- [14] A. Mishra, and U. S. Reddy, "Machine learning approach for wart treatment selection: prominence on performance assessment," *Network Modeling and Analysis in Health Informatics and Bioinformatics*, vol. 9, no. 1, pp. 14, June 3, 2020. DOI=<https://doi.org/10.1007/s13721-020-00246-7>
- [15] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, Second Edition ed., p.^pp. 510, 1005 Gravenstein Highway North, Sebastopol, CA 95472: O'Reilly Media, 2019.
- [16] F. N. G. Lemaitre, D. Oliveira, C. Aridas. "SMOTETomek," December 01, 2020; <https://imbalanced-learn.readthedocs.io/en/stable/generated/imblearn.combine.SMOTETomek.html>.
- [17] scikit-learn. "Tuning the *hyper*-parameters of an estimator," December 01, 2020; https://scikit-learn.org/stable/modules/grid_search.html.
- [18] N. Saboya, A. A. Sullon, and O. L. Loaiza, "Predictive Model Based on Machine Learning for the Detection of Physically Mistreated Women in the Peruvian Scope," in 3rd International Conference on Advances in Artificial Intelligence, ICAAI 2019, Turkey, 2019, pp. 18-23. DOI=<https://doi.org/10.1145/3369114.3369143>