

Universidad de Lima
Facultad de Ingeniería y Arquitectura
Carrera de Ingeniería de Sistemas



**CHATBOT PEDIÁTRICO PARA LA
ORIENTACIÓN SOBRE APENDICITIS AGUDA
BASADO EN NLP Y MODELOS DE
CLASIFICACIÓN SUPERVISADA**

Tesis para optar el Título Profesional de Ingeniero de Sistemas

Juan Domingo Benate Mendoza

Código 20140151

Asesor

Hernan Alejandro Quintana Cruz

Lima – Perú

Setiembre de 2020

Chatbot Pediátrico para la orientación sobre Apendicitis Aguda basado en NLP y modelos de clasificación supervisada

Juan Domingo Benate Mendoza

20140151@aloe.ulima.edu.pe

Universidad de Lima

Resumen: Esta investigación demuestra la viabilidad de utilizar un Chatbot Pediátrico para asesorar a padres de familias sobre una enfermedad común en niños: Apendicitis Aguda. Este sistema utiliza técnicas de *Natural Language Processing* (también conocido en español como Procesamiento de Lenguaje Natural o por sus siglas en inglés NLP), que mediante un modelo de *Naive Bayes* clasifica en intenciones los mensajes enviados por el usuario. También se utiliza un modelo de *Random Forest*, para poder brindar un prediagnóstico sobre si un menor tiene altas sospechas de padecer esta enfermedad o no y según el resultado brinda recomendaciones/orientaciones a los padres. El prediagnóstico ha obtenido más del 95% de precisión a nivel de clasificación de un niño enfermo. Por el otro lado, la precisión del modelo de clasificación de intenciones fue 80%, por lo que su aprendizaje debería ser continuo ya que todos tenemos diferentes formas de expresarnos.

Palabras Clave: Chatbot pediátrico, Naive Bayes para NLP, Clasificación de Intenciones, Pediatría, Prediagnóstico médico, Árbol de Decisiones, Random Forest.

Abstract: This research demonstrates the viability of using a Pediatric Chatbot to advise parents about a very common disease in children: Acute Appendicitis. This system uses Natural Language Processing techniques (also known as NLP), that use a Naive Bayes model to classify the user's messages into intents. It also provides a pre-diagnosis on whether a child has high suspicions of suffering from Acute Appendicitis or not through a Random Forest classifier, and according to the result, the Chatbot would send a recommendation to parents. Pre-diagnosis has obtained more than 95% accuracy in classification of a sick child. On the other hand, the accuracy of Intent Classification model was 80%, it's learning should be continuous because everybody has different ways of expressing ourselves.

Keywords: Pediatric Chatbot, Naive Bayes for NLP, Intent Classification, Pediatrics, Medical pre-diagnostic, Decision Tree, Random Forest.

1. INTRODUCCIÓN

En el Perú, existe una ineficiencia en el servicio de Salud Pública, el cual se ha evidenciado en esta época de pandemia por el COVID-19, Sausa (2018) informa que el tiempo de espera por una cita médica es de dos semanas y para ser atendido se debe esperar poco más de dos horas. Una de las causas de este problema se puede identificar gracias a unas estadísticas publicadas por ESSALUD a finales del 2018, por ejemplo: El total de asegurados era de aproximadamente 11,5 millones de personas, que representan el 35.7% de la población total peruana. A nivel nacional hay 12,596 médicos laborando en el sector público, esto nos indica que existe una gran escasez de personal calificado ya que por cada diez mil asegurados hay 10.96 médicos asignados para atenderlos. Uno de los segmentos más perjudicados por el prolongado tiempo de espera son los niños ya que si la atención no es oportuna puede generar varias complicaciones que aumentan la morbilidad y mortalidad (Rodríguez, Becerra y Rueda, 2016). Estos datos presentados fueron publicados antes del inicio de la pandemia, este año 2020 los hospitales se encuentran colapsados por personas necesitadas de atención debido al COVID-19, por eso solo se debería acudir a emergencias en casos extremos ya que existe el riesgo de contraer este virus en el camino. Se escogió la Apendicitis Aguda ya que según el Instituto Nacional del Niño (2018) es considerada como uno de los procesos quirúrgicos de emergencia en niños que con mayor frecuencia enfrentan los cirujanos y cuya identificación es más confusa en pacientes de edad pediátrica, en consecuencia un retardo en su diagnóstico incrementa la frecuencia de las complicaciones ocasionando mayor morbilidad y mortalidad, también se reporta que en la mayor parte de casos las complicaciones se presentan antes de una evaluación por parte del especialista, cuya tasa de aparición es del 20 al 30%, y en niños menores de 5 años esto puede llegar hasta el 57%. Luego de esta explicación se da a entender que es necesario una herramienta que sirva de apoyo a los padres de familia para que puedan identificar oportunamente altas sospechas de esta enfermedad y así acudir inmediatamente a una sala de emergencias.

La Inteligencia Artificial (IA) es software capaz de emular a la mente humana con el objetivo de que las máquinas tengan la capacidad de realizar tareas humanas. Esta investigación está enfocada en una tecnología que ya comenzó

a ser utilizada por una gran cantidad de organizaciones a nivel nacional e internacional, estos son los Chatbots, que son agentes de chat capaces de simular una conversación humana al interactuar con una persona, por lo que su popularidad en las organizaciones es debido a su capacidad de reducir costos de servicios al cliente y atender a varios usuarios a la vez las 24 horas del día, los 365 días del año (Ranoliya, Raghuwanshi y Singh, 2017). Un Chatbot especializado en pediatría que pueda contestar las consultas más frecuentes, sea capaz de realizar un prediagnóstico y en base a eso brindar recomendaciones personalizadas servirá para que los padres o apoderados de un niño puedan obtener información confiable desde casa, facilitaría la detección temprana de Apendicitis Aguda, se evitaría que asistan a salas de emergencias por sospechas erróneas de esta enfermedad y agilizaría la atención a personas que si requieran atención urgente.

En este artículo se comenzará analizando investigaciones relacionadas que sirven de base para usarlas como referencia en la realización de este trabajo. Luego se revisarán los antecedentes que brindarán una base teórica sobre las áreas del conocimiento involucradas en la investigación. Después se presentarán los métodos utilizados para llevar a cabo la investigación, se describirán los resultados obtenidos y se mostrará las interpretaciones sobre estos. Por último, se brindará un análisis crítico y objetivo de los resultados obtenidos en este trabajo.

2. ESTADO DEL ARTE

Ahmad et al. (2018) menciona que los Chatbots son usualmente usados para varios propósitos como atención al cliente y abastecimiento de información. Algunas implementaciones realizadas en el Perú son: la cadena de pizzerías Papa Jhon's desarrolló su Chatbot llamado Papa Bot cuya función principal es reemplazar el canal telefónico para ordenar los pedidos que serán entregados vía delivery a los clientes. Otro ejemplo es el Bot de la AFP Habitat, llamado Habi, cuya función es resolver las dudas o consultas frecuentes sobre las AFP que tengan los clientes y realizar otras funciones (Semana Económica, 2018). Los Chatbots también pueden ser útiles en otros campos como la medicina, a continuación, se muestran investigaciones previas donde se implementaron estas soluciones en dicho campo:

Tabla 1. Chatbots en el ámbito de la medicina

Autor	Año	Enfoque	Características	Tecnología	Validación
Amato, Marrone, Moscato, Piantadosi, Picariello, y Sansone	2017	Prevención	Brinda recomendaciones para prevenir enfermedades.	Python NLP: IBM Watson	-
Shah y Philip	2019	Diagnóstico Médico	Diagnostica y predice enfermedades.	Python NLP: RN (TensorFlow)	No se realizaron pruebas. Se menciona la prueba de Turing.
Ahmad, Sanusi, Wahab, Mustapha, Sayadi, y Saringat	2018	Farmacia	Da consejos sobre medicamentos.	NLP: IBM Watson	Muestra: 50 usuarios. Prueba de funcionalidad, Prueba de aceptación del usuario: Seguridad, funcionalidad, interfaz y satisfacción del usuario.
Comendador, Francisco, Medenilla y Mae	2015	Pediatría	Prescribe, sugiere y brinda información sobre medicamentos genéricos.	C# Base de datos: MS Access	Muestra: 14 estudiantes de farmacia y 4 pediatras. Se evaluó (del 1 al 5): facilidad de uso, adecuación, consistencia y velocidad de respuesta.

Autor	Año	Enfoque	Características	Tecnología	Validación
Vaira, Bochicchio, Conte, Casaluci, y Melpignano	2018	Maternidad	Brinda rápidos y confiables consejos sobre maternidad.	Microsoft Bot Builder V4 NLP: Microsoft LUIS	-

Amato et al. (2017) explica que para que un Chatbot pueda entender los mensajes de una persona utilizaron una técnica de Procesamiento de Lenguaje Natural llamada Clasificación de Intenciones, en su investigación utilizó un servicio de IBM llamado Watson, también explicó que los mensajes enviados (input o expresión) por los usuarios son procesados por este servicio que devuelve una intención (output) y en base a la intención se define el flujo correcto en el diálogo. Algunas de las intenciones entrenadas en su modelo de NLP son: Saludar, reservar una cita y consultar información general.

Comendador et al. (2015) diseñó un Chatbot capaz de prescribir, sugerir y brindar información sobre medicamentos genéricos para niños cuyo público objetivo fueron padres de familia confundidos sobre el uso de medicamentos genéricos para el tratamiento de los menores. Este Bot ejecuta un dialogo donde se le pregunta al usuario datos sobre el paciente dando como resultado una prescripción de medicamentos genéricos e indicaciones sobre cómo deben ser suministrados, las precauciones que deben seguir, etc. Se validó este sistema con estudiantes de farmacia y pediatras, donde llenaron una encuesta que evaluaba 4 criterios: facilidad de uso, adecuación, consistencia y velocidad de respuesta, a cada variable se le asignaba un valor del 1 al 5, finalmente en los resultados se mostró que cada criterio evaluado tuvo un puntaje aproximado de 4, es decir si hubo una buena aceptación por parte de los especialistas.

Vaira et al. (2018) también utilizó clasificación de intenciones en su Chatbot, pero con una herramienta diferente llamada LUIS o Language Understanding Intelligent System que es un servicio cognitivo de Microsoft. Para el entrenamiento de su modelo de NLP utilizó un Dataset de 422 expresiones divididas en 11 intenciones y durante su experimentación los investigadores se dieron cuenta que a mayor número de intenciones menor es la precisión del modelo de clasificación debido a la posibilidad de utilizar palabras similares en contextos diferentes.

Shah y Philip (2019) desarrollaron un modelo de clasificación de intenciones utilizando una red neuronal que fue construida con librerías de Tensorflow en Python. Antes de iniciar el entrenamiento realizaron un preprocesamiento a las expresiones mediante las siguientes técnicas: splitting, tokenización, stemming y vectorización.

Para evaluar el desempeño de un Chatbot se desarrollaron prototipos y se realizaron pruebas con especialistas de la rama en la que está enfocada esta aplicación, además se deberían recopilar las recomendaciones de los usuarios (Gabrielli, Marie, y Corte, 2018). Ahmad et al. (2018), propuso realizar una prueba de funcionalidad para el testeo del Bot cuya finalidad es comprobar que cada entrada tenga asignada correctamente su salida, por último, se debe tener en cuenta los siguientes puntos en la prueba de aceptación del cliente: seguridad y funcionalidad.

Con respecto a la identificación de pacientes con Apendicitis Aguda existen varias investigaciones sobre clasificación de pacientes enfermos y sanos usando técnicas de Machine Learning donde se emplean diferentes técnicas, atributos y cuyos resultados son variados.

Aydin et al. (2020) hicieron pruebas con varios modelos lineales de Machine Learning para detectar Apendicitis Aguda y diferenciar un caso complicado y no complicado en niños. Para entrenar su modelo recolectaron data de casos admitidos entre el año 2010 y 2016, en total son 7244 pacientes menores de 18 años. Cada registro incluía datos demográficos, análisis de sangre y un diagnóstico postoperatorio basado en opiniones clínicas y patológicas. La mayoría de los modelos arrojaron resultados similares en las pruebas, pero los investigadores decidieron que el modelo de Árbol de Decisiones es el mejor debido a su sencilla interpretación, este modelo obtuvo un 94.69% de precisión. Por último, los autores mencionan que el Aprendizaje Automático es un enfoque novedoso para prevenir intervenciones quirúrgicas innecesarias y así disminuir la carga tanto a pacientes como a sistemas de salud.

Khumrin et al. (2018) desarrollaron modelos de clasificación multinomial y binomial que fueron entrenados con 208 registros de pacientes que registraron dolor abdominal, dicho conjunto de datos tiene un total de 75 atributos relacionados a la historia clínica, exámenes físicos y resultados de laboratorio. Este modelo es capaz de clasificar en 5 clases (Apendicitis, Gastroenteritis, Infección Urinaria, Embarazo Ectópico y Enfermedad Pélvica Inflamatoria). Los modelos con los mejores resultados fueron Naive Bayes y LogitBoost ya que cada uno obtuvo una precisión mayor al 80% y los resultados del segundo eran ligeramente superiores al primero, pero los investigadores decidieron escoger Naive Bayes como el mejor modelo debido a que sus predicciones encajan más con los juicios clínicos reales.

Akmese et al. (2020) realizaron un estudio para desarrollar un fácil, rápido y preciso modelo binario de Machine Learning para el diagnóstico temprano de Apendicitis Aguda, para esto hicieron la comparación de diversos modelos entrenados con un conjunto de datos históricos como edad, género y datos de laboratorio de 595 pacientes que presentaron dolor abdominal, utilizando el 70% para entrenamiento y lo restante para pruebas. Los modelos que obtuvieron los mejores resultados fueron un Árbol de Decisiones con Gradiente Potenciada y Random Forest con un nivel de predicción de 95.31% y 92.96% respectivamente. Finalmente, para determinar el mejor modelo los autores tomaron como criterio más importante el nivel de precisión.

La mayoría de los autores que han realizado estudios sobre Chatbot coincidieron en ampliar los conocimientos de la aplicación para que puedan procesar un mayor número de consultas. Otra oportunidad de mejora es que los diálogos sean más personalizados, es decir se necesita recopilar más información sobre el usuario como por ejemplo su edad, peso, ritmo cardiaco, etc. También tomar en cuenta aspectos de seguridad de la información en los Chatbot, por ejemplo: encriptación de mensajes, seguridad en las bases de datos que están conectadas al Bot, etc. (Vaira et al., 2018). Por último, la interacción en diferentes idiomas es un aspecto que se debe tener en cuenta para que la aplicación esté disponible en varios países y facilitar el acceso a la información a nivel mundial (Amato et al., 2017).

Los Chatbots no reemplazaran a las personas, pero son un canal innovador para brindarles información útil y personalizada a través de una conversación evadiendo así el tiempo de espera por una consulta tradicional. Si la aplicación está bien diseñada y elaborada podrá aumentar el auto empoderamiento del usuario, brindaría una mejor experiencia y reduciría costos en el sistema de salud ya que el número de consultas innecesarias disminuiría (Vaira et al., 2018).

3. MARCO TEÓRICO

3.1. Visión general sobre un Chatbot

Khan y Das (2018) definen un Chatbot como un programa de computadora que procesa una entrada de lenguaje natural enviada por el usuario y le responde con una respuesta inteligente y relativa, la interacción con el usuario es mediante texto solamente, pero se puede usar voz o imágenes.

3.1.1. Características de un Chatbot

Según Toledo estas son las características que diferencian a estos sistemas (como se citó en Cobos, 2013; Inglada, 2012):

Tabla 2. Características de un Chatbot

Característica	Descripción
Autonomía	Puede actuar sin la necesidad de que intervenga un ser humano.
Sociabilidad	Puede comunicarse con otros agentes u otras entidades.
Racionalidad	Sabe que respuesta enviar según el mensaje anterior enviado por el usuario.
Reactividad	Percibe el entorno y su comportamiento depende de lo percibido.
Adaptabilidad	Puede adaptar su comportamiento mediante el aprendizaje.
Movilidad	Puede trasladarse a cualquier dispositivo gracias al internet.
Veracidad	Comunica información proveniente de fuentes confiables.
Personalidad	Cada Chatbot es único, el programador es quien define su personalidad.

3.2. Arquitectura y Diseño

A continuación, se explicará a nivel conceptual los componentes que conforman un Chatbot, por ejemplo: Técnicas de Procesamiento de Lenguaje Natural, modelos de Machine Learning utilizados en NLP y clasificación, por último, aspectos generales de un gestor de diálogos.

3.2.1. Procesamiento de Lenguaje Natural

El Lenguaje Natural es aquel medio que las personas usan cotidianamente para establecer una comunicación entre ellos, esto puede ser mediante escritura o el habla (Vásquez, Quispe y Huayna, 2009). Asimismo, el Procesamiento de Lenguaje Natural (NLP) es la capacidad de la máquina para la conversión del texto en data estructurada, comprender su significado y determinar la acción adecuada (Fernandes, 2018).

La Clasificación de Intenciones es la capacidad de la máquina para interpretar lo que desea el usuario desde sus mensajes enviados (Tur y Deng, 2011). Esta técnica es muy usada en sistemas conversacionales para poder determinar la siguiente acción del sistema como respuesta al último mensaje enviado por el usuario (Yang et al., 2017).

Las aplicaciones de NLP requieren realizar un procesamiento de los textos mediante las siguientes técnicas:

- a. Stopwords: Son palabras comunes que añaden un mínimo o nulo significado a comparación de otras palabras claves. En NLP se elimina estas palabras comunes para poder enfocarse en las keywords (Kulkarni y Shivananda, 2019).
- b. Splitting y tokenización: Estas técnicas van de la mano, tokenización se refiere a separar o hacer un split a una cadena de texto en unidades mínimas con significado, es decir obtendríamos el listado de palabras que conforman esta cadena (Kulkarni y Shivananda, 2019).
- c. Stemming: Es el proceso de extraer la palabra raíz a un conjunto de palabras (Kulkarni y Shivananda, 2019).
- d. Vectorización: dicha técnica consiste en la transformación de las palabras en números para facilitar el procesamiento. Existen diversas técnicas para esto, una de ellas es: TF-IDF, en la cual se le asigna un puntaje a cada palabra en la expresión según con la frecuencia en la que se repite en todo el Dataset (Manaswi, 2018).

3.2.2. Modelos de Machine Learning

- a. Naive Bayes

Es un clasificador basado en el teorema de Bayes capaz de distinguir entre más de dos clases, dicho algoritmo es uno de los más utilizados para clasificar textos debido a su fácil construcción y alta precisión. Para determinar si un texto x pertenece a una clase y , dicho modelo utiliza la siguiente fórmula:

$$p(x|y) = \prod_{k=1}^n p(w^k|y) \quad (1)$$

Donde n es el número de palabras dentro del texto x y w es la k -ésima palabra en el texto x . Cada palabra en el texto es evaluada de manera independiente. Se haya cada posibilidad que tiene un texto para pertenecer a cada clase y es seleccionada aquella clase que tuvo la mayor posibilidad (Hastie, Tibshirani y Friedman, 2017).

El modelo multinomial es una técnica que le asigna su frecuencia a cada una de las palabras que conforman un texto, y toma en consideración dicha frecuencia para predecir la probabilidad de que un texto x pertenezca a una clase y , a diferencia de otros tipos de modelos ya no se toma solo en cuenta que la aparición o no de la palabra (Aggarwal y Zhai, 2012).

- b. Árbol de decisiones

Barrientos et al. (2009) señala que un Árbol de Decisiones es un modelo de clasificación utilizado en la Inteligencia Artificial que se basa en el aprendizaje inductivo a partir de observaciones y construcciones lógicas y que el proceso de dicho tipo de aprendizaje se representa mediante un árbol formado por un conjunto de nodos, hojas y ramas. También explica que el nodo raíz es el atributo con el que se inicia el proceso de clasificación, los siguientes nodos representan otros atributos particulares del problema, un nodo hijo representa una posible respuesta a los cuestionamientos que se representan, cada rama esta etiquetada con los posibles valores que puede tomar un atributo y los nodos hoja representan una decisión que coincide con una de las variables clase del problema a resolver.

El algoritmo CART es utilizado para clasificación y regresión, construye árboles binarios utilizando la característica y el umbral que producen la mayor ganancia de información en cada nodo, donde las variables de entrada pueden ser valores discretos o continuos (Pedregosa et al., 2011).

- c. Random Forest

Este algoritmo de Machine Learning es uno de los robustos que existen, se construye en base a un conjunto de Árboles de Decisiones que son entrenados en subconjuntos aleatorios de características y se promedia sus predicciones. Su uso es más conveniente y efectivo que un Árbol de Decisiones (Géron, 2017).

d. Datos no balanceados

Han, Wang y Mao (2005) afirman que un Dataset se puede considerar desbalanceado cuando una clase tiene mayor número de registros (clase mayoritaria) que la otra clase (clase minoritaria). También mencionan que algunas técnicas en las cuales se altera la distribución original del Dataset para combatir dicho fenómeno son:

Random Oversampling: Equilibra el Dataset al duplicar de manera aleatoria registros pertenecientes a la clase minoritaria, esta técnica puede provocar Overfitting ya que puede hacer la región de decisión del modelo más pequeño y específico.

SMOTE Oversampling: También conocido en inglés como Synthetic Minority Oversampling Technique, genera registros sintéticos pertenecientes a la clase minoritaria. Primero se ubican los vecinos más cercanos a los registros minoritarios y aleatoriamente se generan nuevos registros a lo largo de la línea entre los registros minoritarios y sus vecinos más cercanos.

e. Gestión del diálogo

Según Jack Cahn (2017), los desarrolladores de Bot deben definir los siguientes aspectos sobre la gestión del diálogo:

Estrategias de comunicación: Primero se debe determinar quién guía la conversación: el usuario, el Bot o ambos. En el primer caso, el usuario inicia el dialogo y el Bot responde. En el segundo, el Bot dirige el dialogo y el usuario responde a las consultas del Bot. Y en el caso mixto, el Bot y el usuario guían la conversación en distintos contextos.

Trucos de lenguaje: Para las respuestas del Bot durante el flujo conversacional que tengan poca probabilidad de ser las correctas se deben considerar los siguientes trucos: (1) Cambiar de tema, (2) Hacer preguntas abiertas, es decir en vez de responder “No lo sé”, debe responder, “No entiendo la respuesta, ¿podría enviar otra consulta?”, así incentiva al usuario de continuar con la conversación. (3) Contar un chiste. (4) Solicitar más información, si el Bot no está seguro de su respuesta puede solicitar más información para tener un contexto más preciso.

Principios de diseño de un diálogo: (1) Desambiguación, el Bot debe aclarar las entradas del usuario cuando este tenga diversos significados. (2) Relajación, capacidad de eliminar restricciones para manejar adecuadamente las entradas del usuario. (3) Confirmación, chequear los detalles del usuario antes de culminar el diálogo. (4) Terminación, solicitando los detalles necesarios que se dejaron fuera en un input de usuario.

Estrategias de imitación de humanos: Estas estrategias permitirán al Bot a aparentar ser más humano: (1) Desarrollo de personalidad, asignarle un nombre y una personalidad para hacerlos más humanos. (2) Dirigir la conversación, el Bot deberá hacer las preguntas principalmente, para controlar la participación del usuario y minimizar errores. (3) Charla, los desarrolladores incorporan una funcionalidad de conversación trivial que crea una relación con el humano y evita el silencio (si el humano no está iniciando el diálogo). (4) Errar como humanos, incluir disfluencias en el diálogo hace que los Bots parezcan humanos.

3.3. Implementación

Khan y Das (2018) mencionan que los Chatbots son aplicaciones recientes por lo que todavía no existe una guía que nos indique los estándares que debemos seguir para su desarrollo, pero poco a poco se van definiendo estándares que buscan brindarle una buena experiencia al usuario. A continuación, se explicará algunos principios basados en UX que se deben tomar en cuenta para su construcción:

Tabla 3. Buenas prácticas para la construcción de un Chatbot

Buena práctica	Descripción
Ser corto y preciso	Enviar mensajes cortos y así evitar la saturación del usuario con información.
Enviar mensajes ricos en contenido	Utilizar de mensajes enriquecidos en contenido para mejorar la experiencia visual del usuario.
Tomar en cuenta la fuente del mensaje	Tomar en cuenta el canal de origen del mensaje y darles un formato adecuado a los mensajes.
Brindar información para contactar a un agente humano	Brindar una forma de contacto con un agente humano al usuario en caso de que no se pueda responder su consulta evitará que este deje una mala impresión sobre el Bot.

4. DISEÑO E IMPLEMENTACIÓN

4.1 Requerimientos del Sistema

Para esta investigación se desarrolló un prototipo similar al Bot desarrollado en la investigación de Comendador (2015), cambiando el enfoque a la Apendicitis Aguda, en base a esto se definieron los siguientes requerimientos:

Tabla 4. Listado de requerimientos funcionales y no funcionales

Funcionales	No Funcionales
Informar al usuario sobre los siguientes temas relacionados a Apendicitis Aguda: definición, síntomas, tratamiento, prevención y complicaciones.	El sistema deberá estar disponible 24x7.
El Bot deberá ser capaz de clasificar los mensajes de los usuarios en las siguientes intenciones: saludo, despedida, afirmación, negación, solicitar un prediagnóstico y consultar información sobre el primer requerimiento.	El usuario deberá interactuar con el Bot desde una canal web, puede utilizar un computador o dispositivo móvil.
En caso de que el Bot no entienda un mensaje del usuario, se deberá enviar un mensaje informando que no entendió el mensaje del usuario.	El sistema deberá soportar mínimo 1000 usuarios concurrentes.
Recopilar información sobre síntomas que siente el menor de edad.	
Dar un prediagnóstico sobre si el menor tiene o no Apendicitis Aguda en base a los datos recopilados del usuario.	
Luego de responder la consulta del usuario, el Bot deberá consultar al usuario si desea realizar alguna consulta más.	
Los modelos de clasificación para NLP y prediagnóstico deberán ser consumidos desde servicios REST.	

4.2 Metodología para Desarrollo de un Chatbot



Figura 1. Metodología para desarrollo de un Chatbot.

En la Fig. 1, se puede observar la metodología utilizada que consiste en 5 fases principales, a continuación, se explicará cada una de ellas:

4.2.1 Diseño

En base a los requerimientos se definió la siguiente arquitectura de la aplicación:

- a. Arquitectura de la aplicación:

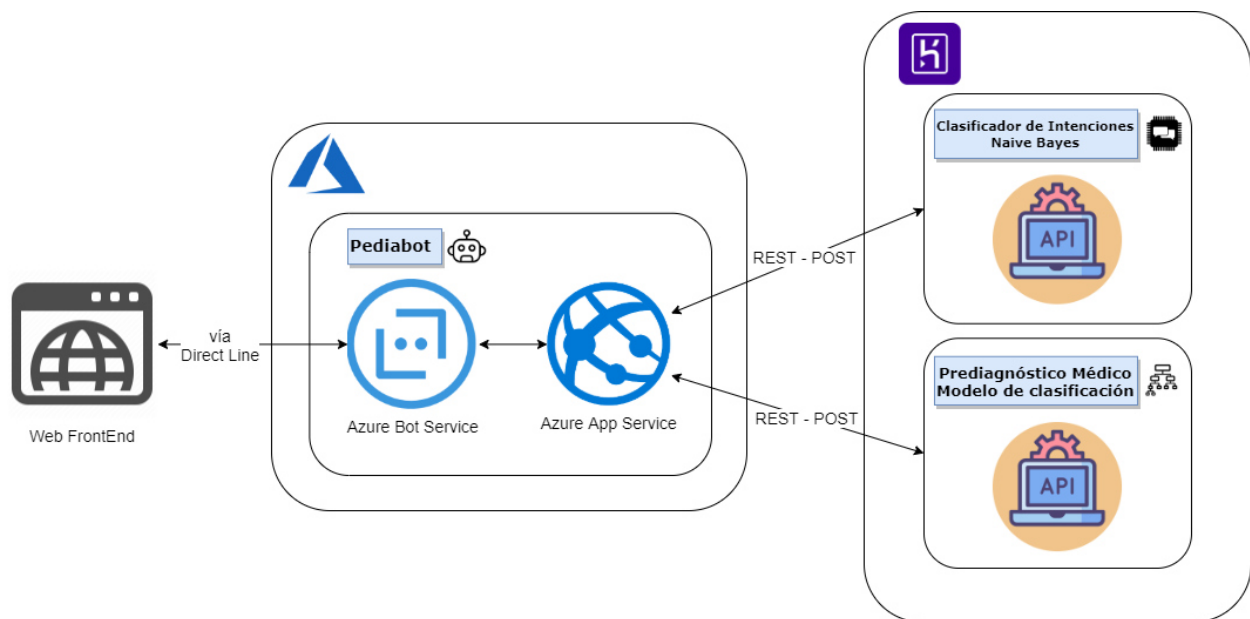


Figura 2. Arquitectura de la solución.

i. Pediabot

- Azure App Service: Se utilizará el framework Microsoft Bot Framework v4 en su versión para C#, que provee un conjunto de herramientas para la construcción, depuración y despliegue de un Bot (Delgado, León y Sánchez, 2017). La aplicación se alojará en la nube de Azure, para esto debe crearse un servicio llamado App Service.
- Azure Bot Service: Servicio que complementa las aplicaciones de Bot en Azure que ofrece los conectores para poder comunicarse con diversos canales de mensajería instantánea. En esta solución se utilizará el canal de Direct Line.

ii. Clasificador de Intenciones

- El Procesamiento de Lenguaje Natural le dará una capa de inteligencia al Chatbot, mediante servicios REST, utilizando el método POST se enviará un archivo JSON que contendrá el mensaje del usuario en formato de texto. Luego el API preprocesará el texto y luego lo clasificará en una intención mediante un modelo entrenado de Naive Bayes de tipo Multinomial y se enviará como respuesta al Bot el resultado de esta clasificación.

iii. Prediagnóstico Médico

- El Chatbot realizará una serie de preguntas al usuario e ira creando un perfil en base a sus respuestas, igual que con el modelo de NLP, luego se conectará mediante servicios REST a un modelo de clasificación que será definido en base a los resultados para determinar si el menor sufre o no de Apendicitis Aguda donde el input será un archivo JSON que contiene el perfil del paciente y la salida será un valor binario (0: No tiene Apendicitis, 1: Tiene Apendicitis).

iv. Web FrontEnd

- Mediante el canal de Direct Line el usuario podrá interactuar con el Bot desde una página web mediante cualquier dispositivo móvil y de escritorio que tenga conexión a internet. El enlace de la plataforma cliente es: <http://pediabotweb.herokuapp.com/>

b. Flujo del diálogo:

Todos los mensajes enviados por el Bot al usuario fueron elaborados junto a un especialista en Cirugía Pediátrica. A continuación, en la Fig. 3 se muestra el diseño del flujo del diálogo del Chatbot. En primer lugar, el diálogo inicia enviándole un saludo al usuario y se le pregunta cuál es su consulta, luego se prepara el mensaje enviado por el usuario y se envía el mensaje al servicio de NLP alojado en Heroku cuya respuesta es la intención del mensaje que será utilizada para determinar la respuesta al usuario, por consiguiente, el flujo puede tomar 4 caminos diferentes. En caso de que inicie el diálogo de prediagnóstico, el usuario deberá responder un cuestionario sobre síntomas que siente el menor, donde se espera una afirmación o negación, luego de cada pregunta se irá guardando la respuesta en un registro,

con el objetivo de enviarlo al servicio de Árbol de Decisiones para la clasificación, de ahí se retorna el resultado al Chatbot y se le brinda una recomendación en base a este. Otro caso es que el usuario consulte información sobre la enfermedad (síntomas, tratamiento, prevención, etc.) y el Bot le enviará la respuesta a su consulta. En caso de que el Bot no entienda el mensaje del usuario este le enviará un mensaje informando esto. Finalmente, el Bot le preguntará al usuario si es que tiene alguna otra consulta para continuar con la conversación.

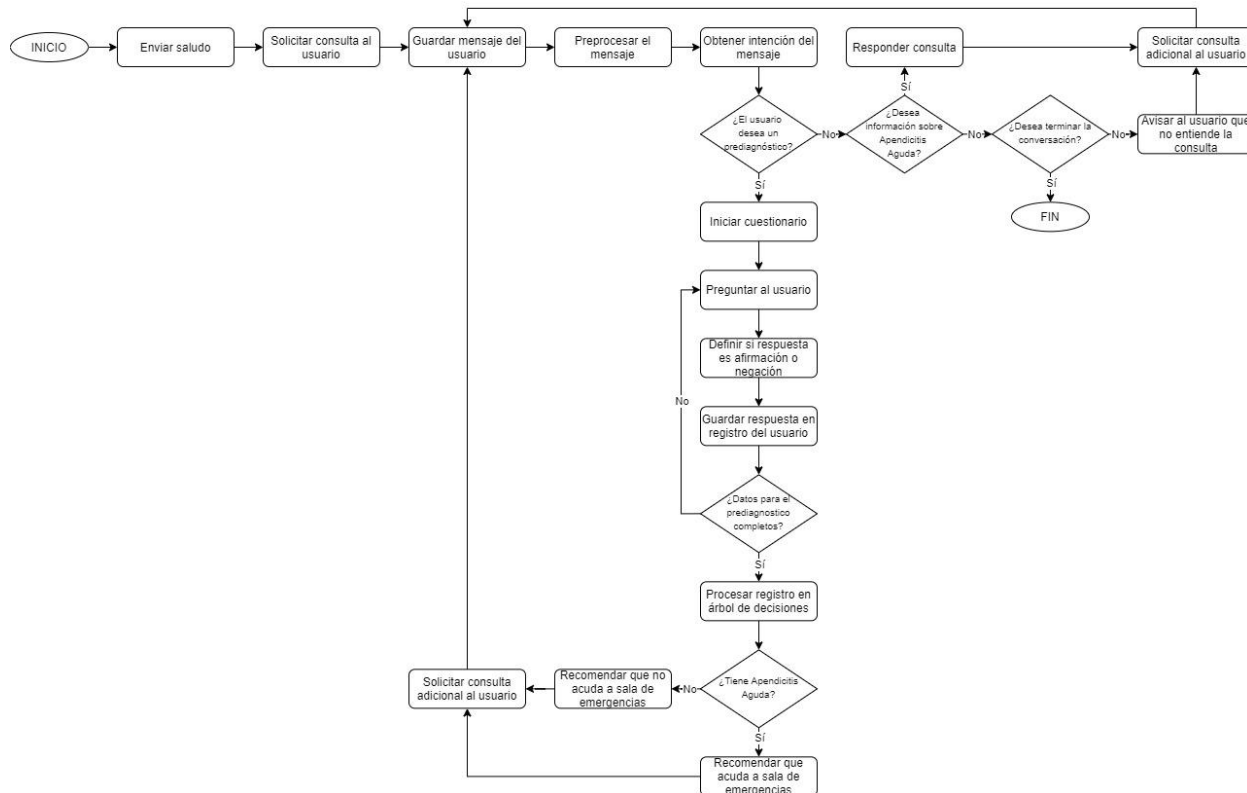


Figura 3. Flujo del diálogo.

4.2.2 Desarrollo

a. Clasificación en Intenciones:

Para darle la capa de inteligencia al Chatbot se debe generar un modelo de clasificación de intenciones. Para el entrenamiento del modelo de NLP se tiene un conjunto de 409 expresiones etiquetadas con sus respectivas intenciones.

En el presente estudio se ha recopilado dichas expresiones mediante una encuesta en línea, donde se le preguntaba por cada intención que escriba las expresiones que iban relacionadas a esa clase, Por ejemplo: en el caso de la intención Consultar tratamiento de apendicitis, se le pidió al usuario que escriba formas de expresión para consultar dicha información. Esta encuesta fue respondida por 25 personas y se ha verificado que cada expresión no se repita. Esto con la finalidad de que se registre diversas formas de expresiones de las personas y el Chatbot no esté sesgado a una sola forma de expresarse.

En la Tabla 5, se muestra el listado de intenciones que podrá identificar el Bot usando Procesamiento de Lenguaje Natural. Dichas intenciones cumplen las siguientes funciones principales: Iniciar y terminar una conversación con el Chatbot (1), Solicitar la respuesta a alguna consulta sobre esta enfermedad (2), Solicitar un prediagnóstico (3), Durante el prediagnóstico negar o afirmar alguna pregunta realizada por el Chatbot (4).

Tabla 5. Listado de intenciones y expresiones registradas

Intención	Descripción	Nº Expresiones etiquetadas
Saludo	Identifica cuando el usuario está saludando.	36

Intención	Descripción	N° Expresiones etiquetadas
Despedida	Identifica cuando el usuario se esté despidiendo.	33
Afirmación	Identifica cuando el usuario este afirmando algo.	36
Negación	Identifica cuando el usuario este negando algo.	43
SolicitarDefinicionEnfermedad	Identifica cuando el usuario quiere saber qué es la Apendicitis Aguda.	29
SolicitarSintomasEnfermedad	Identifica cuando el usuario quiere saber los síntomas de la Apendicitis Aguda.	42
SolicitarTratamientoEnfermedad	Identifica cuando el usuario quiere saber el tratamiento para la Apendicitis Aguda.	44
SolicitarPrevencionEnfermedad	Identifica cuando el usuario quiere saber cómo prevenir la Apendicitis Aguda.	46
SolicitarComplicacionesEnfermedad	Identifica cuando el usuario quiere saber cuáles son las complicaciones de la Apendicitis Aguda.	42
Prediagnostico	Identifica cuando el usuario desea saber si el/la menor tiene altas sospechas de Apendicitis Aguda.	58

Se ha adaptado un código de Python de Github para clasificación de intenciones del usuario Priya997 (2018). Antes de dividir el Dataset para el entrenamiento y las pruebas, primero se realizó el preprocesamiento de las expresiones, dicho procedimiento se puede observar en la Fig. 4, en donde a cada expresión lo primero que se hizo fue remover los signos de puntuación, luego son removidos los stop-words, después de la limpieza viene la tokenización, y finalmente, se realiza la vectorización de las palabras mediante la técnica TF-IDF.

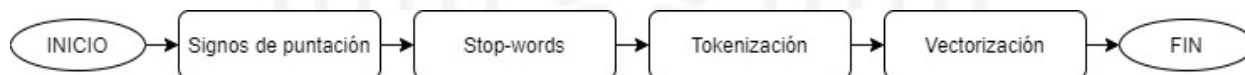


Figura 4. Preprocesamiento de mensajes.

El siguiente paso es construir un modelo de Multinomial Naive Bayes con los parámetros por defecto de la función, en este paso utilizamos las librerías de scikit-learn, después entrenamos el modelo y se realizan las pruebas para medir la precisión. Finalmente, este modelo entrenado se conectará con el Chatbot para que determine la intención de los mensajes enviados por el usuario.

b. Prediagnóstico:

Para el entrenamiento se utilizó un dataset de 172 registros recopilados en una investigación hecha por Sheikh y Latif (2015) sobre el análisis del dolor abdominal agudo y otros síntomas en un departamento de cirugía pediátrica. Este dataset fue solicitado mediante correo electrónico al Dr. Muhammad Ali Sheikh, quien me entregó la data generada en su investigación solamente para uso académico y personal por lo cual no tengo el permiso de compartirlo, la cadena de correos se encuentra adjuntada en los anexos de este documento.

La data se recibió en formato .sav por lo que para el análisis se utilizó la herramienta SPSS Statistics. Este conjunto de datos cuenta con 28 atributos, junto con el asesoramiento de un especialista se descartaron 9 atributos del dataset, el resumen de los 19 atributos restantes se puede observar en la Tabla 6. De dicho conjunto de datos el 80% de los registros será para el entrenamiento y lo restante para pruebas.

Tabla 6. Atributos del Dataset

Dataset (Acute abdominal pain in children: clinical outcomes & values of symptoms and signs in diagnosis)		
Atributos descartados	Atributos para entrenamiento	
Nombre	Nombre	Valores

Dataset (Acute abdominal pain in children: clinical outcomes & values of symptoms and signs in diagnosis)		
Sorethroat: Irritación de la garganta	Age: Edad	Entre 1 a 15 años
Quantity: Veces en la que el paciente ha vomitado	Sex: Género	• 0: Masculino • 1: Femenino
Color: Color del vómito	Duration: Duración del dolor por más de un día	• 0: No • 1: Sí
Bleedingpr: Sangrado por el recto	Location: Dolor abdominal en el cuadrante inferior derecho	• 0: No • 1: Sí
Pulseadmission: Pulso cardiaco en la admisión	Shifting: Propagación del dolor abdominal	• 0: No • 1: Sí
Tempadm: Temperatura corporal en la admisión (°F)	Character: Dolor abdominal tipo cólico	• 0: No • 1: Sí
Hb: Nivel de hemoglobina	Anorexia: Falta de apetito	• 0: No • 1: Sí
TLC: Conteo total de leucocitos	Vomiting: Nauseas o vómitos	• 0: No • 1: Sí
ClinicalDiagnosis: Diagnóstico clínico	Fever: Fiebre	• 0: No • 1: Sí
	Micturition: Disuria	• 0: No • 1: Sí
	Constipation: Estreñimiento	• 0: No • 1: Sí
	Loosemotions: Diarrea	• 0: No • 1: Sí
	Throatcongestion: Congestión nasal / garganta	• 0: No • 1: Sí
	Tenderness: Sensibilidad en el abdomen inferior derecho	• 0: No • 1: Sí
	Guarding: Resistencia abdominal	• 0: No • 1: Sí
	Mass: Masa palpable	• 0: No • 1: Sí
	Reboundtender: Dolor abdominal al rebote	• 0: No • 1: Sí
	Distension: Distensión abdominal	• 0: No • 1: Sí
	Final Diagnosis: Apendicitis	• 0: No • 1: Sí

Durante la construcción de los modelos de clasificación, primero se transformó los datos del atributo Edad en una unidad de desviación estándar utilizando el puntaje Z o Z-Score, esto se realizó para estandarizar dicho atributo en relación con las otras columnas del Dataset.

Una observación importante es que la data está desbalanceada ya que la clase mayoritaria son los menores que no tienen Apendicitis Aguda con 106 registros y la clase minoritaria son aquellos con esta enfermedad con 66 registros, por eso para equilibrar la data se hizo una comparación de dos técnicas: Random Oversampling y SMOTE Oversampling. Esto solo se aplicó a la data de entrenamiento para evitar realizar pruebas con registros creados

artificialmente. En ambos casos solo se está creando registros artificiales a la clase minoritaria antes de entrenar el modelo. Para esta tarea existe una librería de Python llamada imblearn. En la Fig. 5 pueden observar las nuevas distribuciones de las clases del Dataset aplicando dichas técnicas, como puede ver se aumentó los registros de la clase minoritaria para que sean la misma cantidad que la clase mayoritaria.



Figura 5. Distribución de clases en el Dataset según la técnica empleada.

Para cada Dataset se desarrollaron dos modelos y ver cuál es el mejor, Árbol de Decisiones y Random Forest, para el primero se utilizó la librería de scikit-learn, DecisionTreeClassifier cuyo algoritmo utilizado es CART (Pedregosa et al., 2011), y para el segundo se utilizó la librería RandomForestClassifier.

Para obtener la profundidad óptima de cada árbol se utilizó un gráfico de ajuste y cada uno se configuró para utilizar el criterio de entropía. Los modelos de Random Forest utilizan 1000 árboles como estimadores y también usan el criterio de entropía. Luego se emplearon varias técnicas de validación para elegir el mejor modelo que utilizará el Chatbot, la decisión tomada está explicada en la sección de discusiones.

El pseudocódigo 1 explica cómo el Chatbot determina la respuesta a los mensajes que envía el usuario. Cuando se recibe un mensaje lo primero que se hace es guardarlo en una variable de tipo texto. Luego el texto es enviado a una API REST donde será preprocesado antes de ser clasificado por el modelo de NLP entrenado que retornará la intención del mensaje. Finalmente se realiza una secuencia de preguntas lógicas para determinar la respuesta del Bot.

PSEUDOCÓDIGO 1: Determinación de respuesta del Bot en base a la Intención

```

SI Actividad ES Mensaje_Recibido {
  #Capturar mensaje enviado por el usuario
  Mensaje = GuardarMensaje()
  #Realizar preprocesamiento al mensaje
  Mensaje_Limpio = PreprocesarMensaje(Mensaje)
  #Invocar servicio de NLP
  Clasificador_Intenciones = LlamarServicioREST('https://chatbotnlpmodel.herokuapp.com/classify')
  #Clasificar mensaje
  Intencion_Mensaje = Clasificador_Intenciones(Mensaje_Limpio)
  #Definir respuesta en base a la intención
  SI Intencion_Mensaje ES "Saludo" {
    Saludar()
  }
  SINO SI Intencion_Mensaje ES "Despedida" {

```

```

    Despedirse()
}
#Se informa al Bot que se requiere un prediagnóstico
SINO SI Intencion_Mensaje ES "Prediagnostico" {
    IniciarDialogo.Prediagnostico ()
}
#Buscar información sobre apendicitis aguda
SINO SI Intencion_Mensaje ES "SolicitarDefinicionEnfermedad" {
    IniciarDialogo.DefinirApendicitis()
}
}
}

```

En el pseudocódigo 2 se puede observar cómo es el diálogo del prediagnóstico, dicho diálogo consiste en una secuencia de preguntas que el Chatbot le realizará al usuario para poder crearle un perfil que es un arreglo donde se guardan las respuestas del usuario al cuestionario hecho durante el prediagnóstico, este será procesado por el Árbol de Decisiones y así determinará si es sospechoso o no de esta enfermedad. Como se puede ver en el segundo Step (Preguntar_Genero), se captura la respuesta del usuario sobre su edad y se le realiza una normalización escalar antes de ser guardado en el perfil del usuario, luego se le envía la siguiente pregunta donde se adjunta un par de opciones para que escoja el usuario. En los siguientes Step serán preguntas de Sí o No donde se le adjuntará al usuario botones denominados quick replies para que puede responder más rápido, esto con el fin de facilitar más la interacción. Finalmente, cuando se haya completado el cuestionario y todas las respuestas se adjunten el perfil del usuario, dicho objeto será enviado al API REST del Árbol de Decisiones y retornará el resultado de la clasificación. Dependiendo de este se le enviará un mensaje con recomendaciones sobre lo que debería hacer, y se finalizará el diálogo. Finalmente, se regresará al diálogo principal que está representado en el pseudocódigo 1 para determinar la respuesta a las próximas consultas que tenga el usuario.

PSEUDOCÓDIGO 2: Diálogo de Prediagnóstico

```

STEP Preguntar_Edad(
    #Enviar mensaje al usuario para saber su edad
    EnviarMensaje("¿Cuántos años tiene el/la menor?")
)
STEP Preguntar_Genero(
    #Capturar mensaje enviado por el usuario
    Mensaje = GuardarMensaje()
    #Hallar Normalización Escalar de la edad
    Edad = NormEscalar()
    #Guardar edad en el perfil del usuario
    Perfil_Usuario.Edad = Edad
    #Enviar mensaje al usuario para saber su genero
    EnviarMensaje("¿Cuál es el sexo del menor?", opciones = [Masculino, Femenino])
)
STEP Preguntar_Dolor_Abdominal(
    #Capturar mensaje enviado por el usuario
    Mensaje = GuardarMensaje()
    #Guardar edad en el perfil del usuario
    SI Mensaje ES "Masculino" {
        Perfil_Usuario.Genero = 0 #Es niño
    }
    SINO{
        Perfil_Usuario.Genero = 1 #Es niña
    }
    #Enviar mensaje al usuario para saber si tiene dolor abdominal
    EnviarMensaje("¿Siente dolor abdominal?", opciones = [Si, No])
)
STEP Preguntar_Nauseas_Vomitos(
    #Capturar mensaje enviado por el usuario
    Mensaje = GuardarMensaje()

```

```

#Actualizar el perfil del usuario
SI Mensaje ES "Si" {
  Perfil_Usuario.DolorAbdominal= 1
}
SINO SI Mensaje ES "No" {
  Perfil_Usuario.DolorAbdominal = 0
}
#Enviar mensaje al usuario para saber si tiene nauseas o vomitos
EnviarMensaje("¿Tiene nauseas o vomitos?", opciones = [Si, No])
)
.
.
.
STEP HallarPrediagnostico(
  #Invocar servicio de clasificación
  Clasificador = LlamarServicioREST('https://skmodel.herokuapp.com/predict')
  #Clasificar mensaje
  Resultado = Clasificador(Perfil_Usuario)
  SI Resultado ES 0 {
    EnviarMensaje("No tiene apendicitis aguda")
  }
  SINO{
    EnviarMensaje("Sí tiene apendicitis aguda")
  }
  #Terminar dialogo de prediagnóstico y volver al dialogo principal
  FinalizarDialogo()
)

```

Microsoft Azure tiene sus propios servicios cognitivos para NLP para Chatbots llamado LUIS (Language Understanding Intelligent System) pero no hay documentación oficial que indique que modelos se utilizan para la clasificación de intenciones. También tiene su estudio de Machine Learning pero se descartó su uso porque genera costos al requerir diversos servicios de Azure para su construcción. El uso de estas herramientas se aleja del cumplimiento de los objetivos definidos para este trabajo ya que se definió construir modelos de clasificación propios y utilizando herramientas open-source que busca utilizar el Chatbot como un mecanismo de interfaz con el usuario.

4.2.3 Pruebas

En este paso de manera local se realizan pruebas para verificar el correcto funcionamiento del Chatbot para esto se ha utilizado la aplicación de escritorio Bot Framework Emulator que permite probar Bots creados con Microsoft Bot Framework ejecutándose de manera local o de manera remota mediante un túnel. Se está tomando como referencia el trabajo de Ahmad et al. (2018) para definir qué tipo de pruebas realizar, en donde se realizaron pruebas de funcionalidad y aceptación del usuario.

Junto a un médico especialista, se definieron dos tipos de flujos conversacionales para probar toda la funcionalidad del Chatbot, en el primero se probará que el Chatbot responda correctamente a las consultas frecuentes sobre Apendicitis Aguda, y en el segundo flujo se probará el diálogo de prediagnóstico. El especialista verificará que todos los mensajes enviados por el Bot dan información correcta y fácil de entender. Si es que los resultados son satisfactorios se sigue con el siguiente paso, caso contrario se vuelve a la fase de desarrollo para arreglar los errores detectados.

4.2.4 Despliegue

La aplicación ha sido desplegada en la nube de Microsoft Azure porque esta plataforma de nube provee un servicio llamado Azure Bot Service que es compatible con el framework que estamos utilizando para el desarrollo, por lo que se facilita el despliegue de la solución, además dicho servicio provee conectores a diversas plataformas de mensajería instantánea, esta primera versión del prototipo se conectará mediante el canal de Direct Line a una página web.

Por otro lado, la plataforma de Heroku solo se utilizará para el despliegue de las APIs que contienen los modelos de NLP y clasificación, y el frontend para que los usuarios puedan interactuar con el Bot. En esta primera versión no se espera un uso masivo de la aplicación por lo que se decidió utilizar Heroku al tener un plan gratuito de alojamiento y de esta manera no se generarán costos por utilizar plataformas cloud que ofrecen mayor capacidad de cómputo y almacenamiento.

4.2.5 Validación

Se publicó el Chatbot para que los voluntarios puedan interactuar con él y completaron una encuesta cuyos criterios a evaluar se adaptaron de las encuestas de aceptación de usuario de los trabajos de Comendador et al. (2015) y Ahmad et al. (2018), estos criterios son:

- a. Facilidad del uso del Chatbot
- b. Fiabilidad de las respuestas
- c. Velocidad de respuesta
- d. Aporte a la medicina
- e. Experiencia de uso
- f. Recomendaciones y/o comentarios

Para los cinco primeros criterios, los encuestados asignaron un puntaje entre 1 y 5, que representan el puntaje más bajo y alto respectivamente, y para el criterio de recomendaciones y/o comentarios, los usuarios registraron alguna apreciación extra sobre el uso del Bot.

Por el otro lado, para validar el modelo de Árboles de Decisiones, los especialistas recopilaron registros anonimizados de pacientes que acudieron a emergencias debido a sospechas de Apendicitis Aguda para procesarlos, Se está utilizando las siguientes métricas ya que son las ideales para validar un modelo de clasificación según Géron (2017).

- a. F1-score.
- b. Matriz de confusión.
- c. Curva ROC y AUC.
- d. Sensibilidad y Especificidad.

Esta metodología es un ciclo, si existen oportunidades de mejora para el Bot desde la fase de Evaluación se va a la fase de Diseñar para definir y desarrollar nuevas mejoras y mantener al Chatbot en constante entrenamiento.

5. RESULTADOS

5.1 Evaluación del Chatbot

Se validó el Chatbot con 68 personas de las cuales 28 son especialistas en cirugía pediátrica. Estoy considerando como puntaje positivo los valores 4 y 5 en cada criterio de evaluación. Por cada criterio aproximadamente más del 70% de encuestados ha dado un puntaje positivo. A continuación, se muestra los resultados obtenidos en las encuestas:

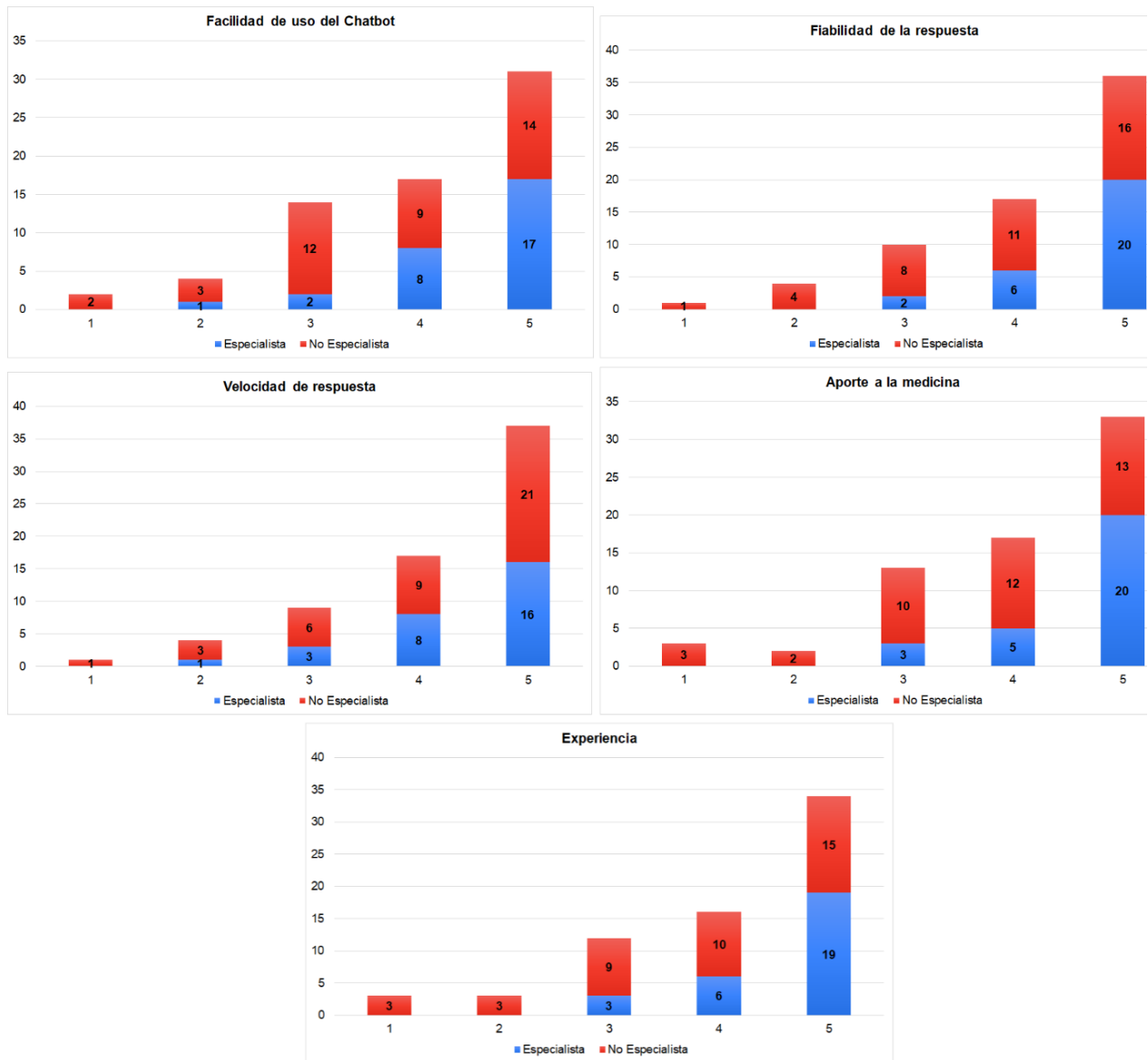


Figura 6. Resultados de la encuesta sobre Pediabot.

5.2 Clasificación de Intenciones

El modelo de Naive Bayes con los datos de prueba obtuvo en la métrica F1-score un resultado de 93.00% con los datos de prueba. Para validar este modelo, luego de integrarlo al Chatbot se enviaron 5 mensajes relacionados a cada intención y se contaron las respuestas correctas enviadas por el Bot, se obtuvo un porcentaje de aciertos de 80.00% (40 aciertos de un total de 50 expresiones). A continuación, se muestra un resumen de los resultados obtenidos:

Tabla 7. Resultados de validación del modelo de NLP

Intención	Aciertos	No aciertos	% Aciertos
Saludo	3	2	60%
Despedida	3	2	60%
Afirmación	3	2	60%
Negación	4	1	80%
SolicitarDefinicionEnfermedad	4	1	80%

Intención	Aciertos	No aciertos	% Aciertos
Solicitar Sintomas Enfermedad	3	2	60%
Solicitar Tratamiento Enfermedad	5	0	100%
Solicitar Prevencion Enfermedad	5	0	100%
Solicitar Complicaciones Enfermedad	5	0	100%
Prediagnostico	5	0	100%
TOTAL	40	10	-
	PROMEDIO		80%

5.3 Prediagnóstico

Los modelos que obtuvieron el mejor resultado en la métrica F1-score fueron Árbol de Decisiones y Random Forest, ambos con el Dataset Desbalanceado, y Random Forest con SMOTE Oversampling (95.45%), a continuación, se mostrarán los resultados obtenidos en las diferentes métricas:

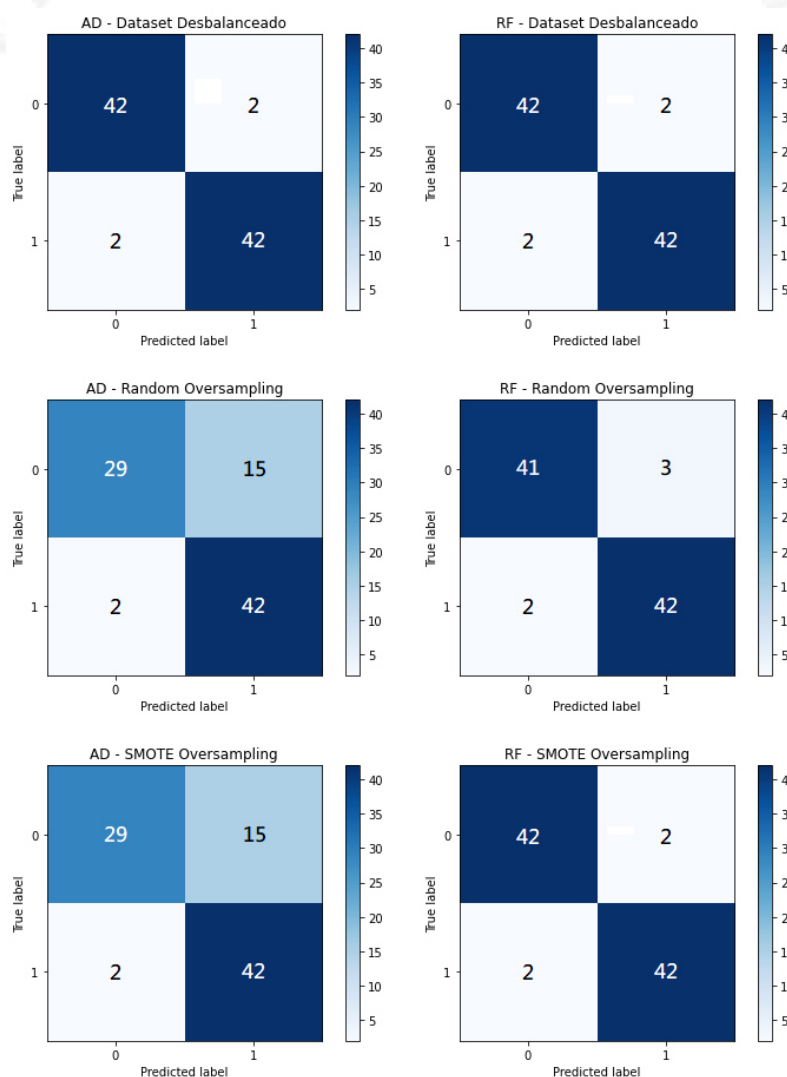


Figura 7. Matriz de confusión.

A partir de la matriz de confusión podemos calcular los índices de Sensibilidad (capacidad para detectar un paciente enfermo con Apendicitis Aguda) y Especificidad (capacidad para detectar un paciente no enfermo de la enfermedad evaluada) para cada modelo, los resultados los podemos observar en la Tabla 8.

Tabla 8. Sensibilidad y Especificidad.

	Modelo	Sensibilidad	Especificidad
Dataset desbalanceado	Árbol de Decisiones	0.95	0.95
	Random Forest	0.95	0.95
Random Oversampling	Árbol de Decisiones	0.95	0.66
	Random Forest	0.95	0.93
SMOTE Oversampling	Árbol de Decisiones	0.95	0.66
	Random Forest	0.95	0.95

En base a los resultados de la matriz de confusión también se graficó la curva ROC, ver Fig. 8.

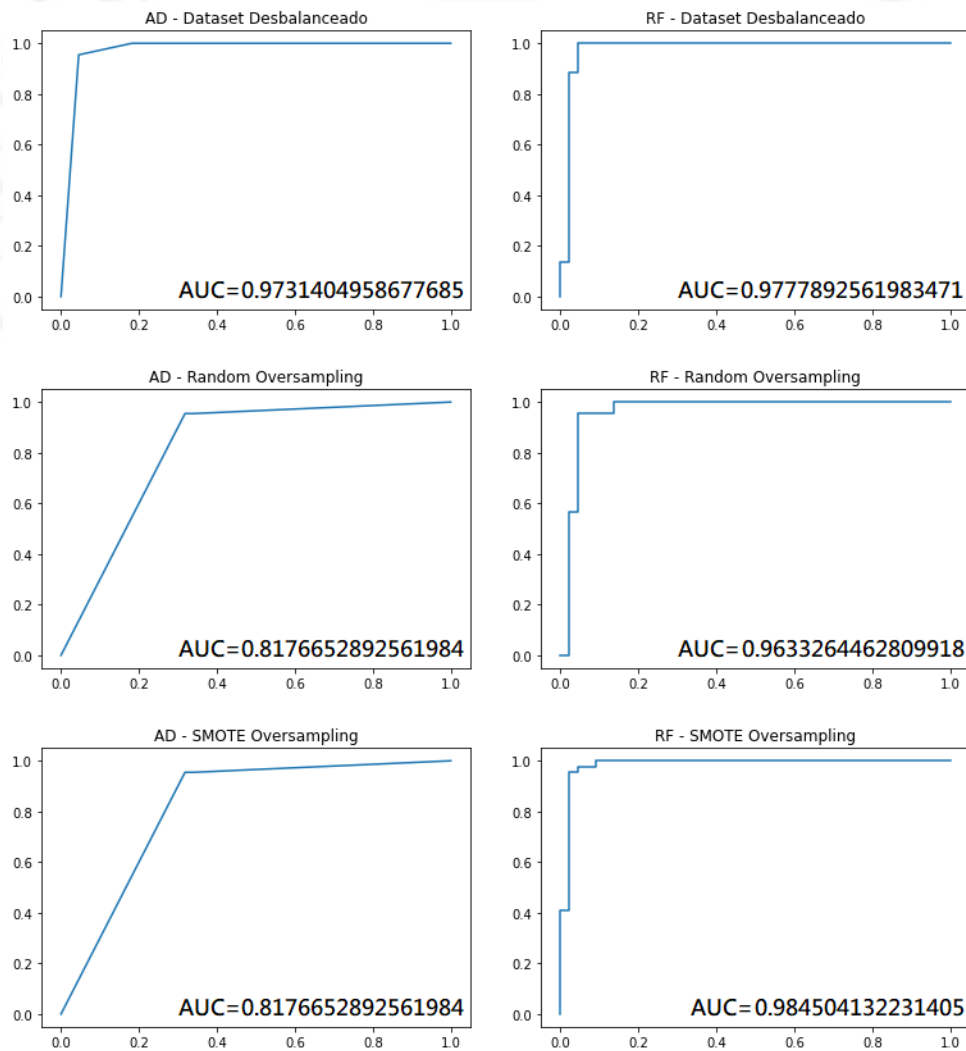


Figura 8. Curva ROC.

6. DISCUSIÓN

Según los resultados de la encuesta los usuarios tuvieron una buena experiencia con el prototipo creado debido a que aproximadamente el 90% de los especialistas dieron una calificación positiva (4 o 5). El criterio que tuvo el mejor puntaje fue velocidad de respuesta con un 79.41%.

Por otro lado, el que tuvo menor puntaje fue facilidad de uso con un 70.59%. Esto se ve reflejado en la sección de comentarios donde la recomendación más frecuente fue mejorar la facilidad de uso, por ejemplo: añadir botones a la interfaz para evitar el tipeo de su consulta y así agilizar la conversación. Otra herramienta para restar la complejidad de interacción que debí incluir es una guía sobre cómo interactuar con el Bot en el mensaje inicial, o también haber agregado una intención de ayuda para que así el Bot envíe un mensaje de instrucciones para su uso.

Con respecto al Procesamiento de Lenguaje Natural se puede decir que una precisión de 80% es aceptable, pero podría mejorar. Braun, Hernandez, Matthes y Langen (2017), realizaron su propio modelo de NLP con solo dos intenciones donde cada una tenía 43 y 57 expresiones registradas respectivamente, en las pruebas obtuvo un puntaje promedio de 0.979. Comparando estos resultados, se puede inferir que la cantidad de expresiones registradas para el entrenamiento en mi modelo de NLP son muy pocas y además que tener más clases hace que la clasificación sea más compleja. Igualmente, el Chatbot demuestra una gran capacidad para responder correctamente a las consultas recibidas que desean iniciar el módulo de Recomendaciones y Prediagnóstico.

Según los resultados de los modelos de clasificación, en la Fig. 7 podemos ver que todos los modelos tienen una gran capacidad para clasificar menores con Apendicitis, respecto a clasificación de menores sin Apendicitis se puede observar que los modelos de árboles de decisiones con Random y SMOTE Oversampling tienen una gran cantidad de falsos positivos, por lo que ambos modelos deben ser descartados ya que confundirán a los padres de familia y llevarían a sus hijos a un centro de salud por una sospecha errónea. Según la Tabla 8, son 3 los modelos que tienen la mejor capacidad para diferenciar pacientes sin y con Apendicitis Aguda, estos son: los dos modelos con el dataset desbalanceado y Random Forest con SMOTE Oversampling. Si observamos la Fig. 8 solo un modelo tiene el mejor resultado, Random Forest con SMOTE Oversampling tiene el valor más alto de AUC, dejando atrás a los modelos con el dataset desbalanceado, quienes obtuvieron también los resultados más altos en las otras métricas.

Se escogió el modelo de Random Forest con SMOTE Oversampling para integrarlo con el Chatbot ya que tiene las mejores métricas y además Random Forest es un algoritmo más robusto que Árbol de Decisiones y al equilibrar las clases antes del entrenamiento mejorará la precisión para determinar si un menor está enfermo o sano, el modelo de Random Oversampling también muestra resultados óptimos, pero se descartó su uso debido a que la creación de nuevos registros aleatorios genera problemas de sobreajuste.

A comparación con los trabajos explicados en el estado del arte, los resultados obtenidos por mi modelo de clasificación son casi similares, en las investigaciones de Aydin et al. (2020) y Akmese et al. (2020) se obtuvo un resultado de 94.96% y 95.31% respectivamente, aunque la diferencia es que estos resultados los obtuvieron con un modelo de Árbol de Decisiones, con respecto a la investigación de Khumrin et al. (2018), se observa un mejor rendimiento de mi modelo de Random Forest con respecto a sus modelos de Naive Bayes y LogitBoost. Finalmente, otra diferencia por destacar es que no hay necesidad de datos de laboratorio (exámenes de orina, sangre, radiografías, etc.) para realizar un prediagnóstico, esto significa que los padres de familia pueden realizarlo desde casa, sin la necesidad de asistir a un centro de salud para obtener este tipo de resultados.

7. CONCLUSIONES

En conclusión, el objetivo principal de construir un prototipo aceptado por especialistas y padres de familia se cumplió ya que según la encuesta los resultados por criterio son mayores al 70%, sin embargo, lo ideal hubiera sido que la precisión del modelo de NLP tenga una precisión mayor al 90% como el modelo de Random Forest para prediagnóstico.

Con respecto a la arquitectura, el framework de desarrollo y los componentes en la nube de Microsoft apoyaron mucho en construcción y despliegue del Bot, aunque faltaron agregar a la arquitectura otros componentes como una base de conocimientos para registrar más preguntas frecuentes sobre Apendicitis Aguda y una base de datos donde se guarde los mensajes no reconocidos por el modelo de NLP y así realizar un entrenamiento activo.

Un Chatbot sería un gran soporte para evitar que padres de familias acudan a un centro de emergencias por falsas sospechas de cualquier enfermedad, debido a la coyuntura actual no se pudo acudir a una sala de emergencias para medir como disminuye la frecuencia de visitas innecesarias, quedaría pendiente para una investigación futura.

Una de las principales dificultades de la investigación fue obtener un Dataset de expresiones adecuado para el entrenamiento del modelo de NLP ya que la mayoría de las expresiones obtenidas mediante encuestas fueron repetidas. Durante la experimentación fue una sorpresa que se obtenga los mismos resultados con el Dataset desbalanceado y el de SMOTE, se esperaba que con las clases equilibradas se obtengan resultados superiores, pero igual se resalta que se haya podido construir un modelo de clasificación muy preciso con una cantidad regular de registros.

El punto para reforzar en el prototipo es la facilidad de uso, por ejemplo: añadiendo botones y otros componentes, también habilitando guías para apoyar al usuario. Finalmente, en base a esta experiencia concluyo que para el desarrollo de este tipo de aplicaciones aparte de un programador se requiere del apoyo de lingüistas para la definición del flujo de conversación y los mensajes que enviará el Bot, y especialistas en UX que definan una interfaz amigable para el usuario y otros componentes para agregar al diálogo como botones u otros elementos enriquecidos.

8. TRABAJOS FUTUROS

Para una futura investigación, sería recomendable agregar una nueva técnica de NLP, que es el Reconocimiento de Entidades, dicha técnica consiste en la detección de palabras clave dentro de los mensajes enviados por el usuario, esto aumentará el entendimiento del mensaje recibido y permitirá al Chatbot entender mejor el contexto de la conversación. En este primer prototipo no se consideró una fuente de persistencia para almacenar los mensajes enviados por el usuario que no fueron entendidos por el Bot y utilizar esta fuente como entrada para reentrenar el modelo de NLP, por lo que se agregaría este componente en una siguiente versión para un aprendizaje continuo. Otra cosa para tener en cuenta es que las aplicaciones de Chat también permiten el envío de audios e imágenes como mensajes, por lo que sería interesante que el Bot sea capaz de procesar dichos mensajes y brindar una respuesta adecuada al usuario. El procesamiento de audios también puede ser realizado con técnicas de NLP, con la conversión de audio a texto. Sobre el procesamiento de imágenes se puede utilizar alguna técnica de Visión Computacional, para buscar algún objeto en específico, clasificación, etc. Otro punto importante a tener en cuenta es la seguridad de la aplicación, en el caso de manejar data sensible en un trabajo futuro se debería añadir controles de seguridad que protejan la aplicación a ciertas amenazas, algunas medidas a tener en cuenta son: configurar la autenticación de entre el frontend cliente y el backend del Chatbot mediante un token con tiempo de expiración, toda comunicación con aplicaciones externas deberá ser mediante protocolo https, configurar un whitelist de canales permitidos para que no puedan utilizar tu Chatbot desde canales que no hayas configurado, si es que se almacena la data sensible en base de datos encriptar la información, etc. Con respecto a la evaluación del Chatbot, se podría implementar algún framework más robusto para evaluar diversos aspectos del Bot como el manejo de los diálogos, la seguridad, la interfaz de usuario, etc. Por último, para el futuro se implementaría las recomendaciones obtenidas mediante las encuestas al prototipo para que más padres de familia utilicen el Chatbot y ver si disminuyen las asistencias a una sala de emergencias de pediatría por una falsa sospecha de Apendicitis Aguda.

REFERENCIAS

- Aggarwal, C. C., & Zhai, C. (2012). A Survey of Text Classification Algorithms [Un estudio de los algoritmos de clasificación de texto]. *Mining Text Data*, 163–222. https://doi.org/10.1007/978-1-4614-3223-4_6
- Ahmad, N. S., Sanusi, M. H., Wahab, M. H. A., Mustapha, A., Sayadi, Z. A., & Saringat, M. Z. (2018). Conversational Bot for Pharmacy: A Natural Language Approach [Bot conversacional para farmacia: un enfoque de lenguaje natural]. *2018 IEEE Conference on Open Systems (ICOS)*, 76-79. IEEE. <https://doi.org/10.1109/ICOS.2018.8632700>
- Akmese, O. F., Dogan, G., Kor, H., Erbay, H., & Demir, E. (2020). The Use of Machine Learning Approaches for the Diagnosis of Acute Appendicitis [El uso de enfoques de aprendizaje automático para el diagnóstico de apendicitis aguda]. *Emergency Medicine International*. <https://doi.org/10.1155/2020/7306435>
- Amato, F., Marrone, S., Moscato, V., Piantadosi, G., Picariello, A., & Sansone, C. (2017). Chatbots Meet eHealth: Automating Healthcare. *WAIHA@ AI* IA*. 40-49. <https://pdfs.semanticscholar.org/b9a3/fe17ee8869c8501c1a41ee9b2f6988bc103.pdf>
- Aydin, E., Türkmen, İ. U., Namli, G., Öztürk, Ç., Esen, A. B., Eray, Y. N., Eroğlu, E., & Akova, F. (2020). A novel and simple machine learning algorithm for preoperative diagnosis of acute appendicitis in children

- [Un algoritmo de machine learning novedoso y simple para el diagnóstico preoperatorio de apendicitis aguda en niños]. *Hernia*, 353, 4-9. <https://doi.org/10.1007/s00383-020-04655-7>
- Barrientos, R., Cruz, N., Acosta, H., Rabatte, I., Gogeoascoechea, M. D. C., Pavón, P., & Blázquez, S. (2009). Árboles de decisión como herramienta en el diagnóstico médico. *Revista médica de la Universidad Veracruzana*, 9(2), 19-24. <https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=27872>
- Braun, D., Hernandez-Mendez, A., Matthes, F., & Langen, M. (2017). Evaluating natural language understanding services for conversational question answering systems. *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, 174-185. <https://www.aclweb.org/anthology/W17-5522>
- Cahn, J. (2017). *CHATBOT: Architecture, Design, & Development* [Tesis de licenciatura, Universidad de Pensilvania]. Academia.
- Cobos Torres, J. C. (2013). *Integración de un Chatbot como habilidad de un robot social con gestor de diálogos* [Tesis de doctorado, Universidad Carlos III de Madrid]. Academia.
- Comendador, B. E. V., Francisco, B. M. B., Medenilla, J. S., & Mae, S. (2015). Pharmabot: a pediatric generic medicine consultant Chatbot. *Journal of Automation and Control Engineering*, 3(2). <http://www.joace.org/uploadfile/2014/0801/20140801025000959.pdf>
- ESSALUD-GCPP-GGI-Sub Gerencia de Estadística (2018). Principales Indicadores de Salud y Series Estadísticas 2000 – 2018. http://www.essalud.gob.pe/downloads/series_Estadisticas_2000_2018.pdf
- Fernandes, A. (15, Agosto, 2018). Definición de NLP, NLU, NLG y cómo funcionan los Chatbots. Planeta Chatbot. <https://planetachatbot.com/definicion-nlp-NLP-nlg-y-c0mo-funcionan-los-chatbots-fcf4b22e065a>
- Gabrielli, S., Marie, K., & Corte, C. D. (2018). SLOWBot (chatbot) Lifestyle Assistant [SLOWBot (chatbot) asistente de estilo de vida]. *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 367-370. ACM. <https://doi.org/10.1145/3240925.3240953>
- Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* [Machine Learning práctico con Scikit-Learn y TensorFlow: conceptos, herramientas y técnicas para construir sistemas inteligentes] (2° ed.). California: O'Reilly
- Han, H., Wang, W. Y., & Mao, B. H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning [Borderline-SMOTE: un nuevo método de oversampling en el aprendizaje de conjuntos de datos desequilibrados]. *International conference on intelligent computing*. 878-887. https://doi.org/10.1007/11538059_91
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. <https://web.stanford.edu/~hastie/ElemStatLearn/>
- Inglada, V. J. J., & Navarro, V. J. B. (2003). Desarrollo de sistemas multi-agente en tiempo real. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, 7(18), 65-80. <https://dialnet.unirioja.es/servlet/articulo?codigo=1023229>
- Instituto Nacional del Niño – San Borja (2018). Guía de Práctica Clínica para el Diagnóstico y Tratamiento de Apendicitis Aguda. <http://www.insnsb.gob.pe/docs-trans/resoluciones/archivopdf.php?pdf=2018/RD-204-2018.pdf>
- Khan, R., & Das, A. (2018). Build Better Chatbots. A Complete Guide to Getting Started with Chatbots [Cree mejores chatbots. Una guía completa para comenzar con los chatbots]. <https://doi.org/10.1007/978-1-4842-3111-1>

- Khumrin, P., Ryan, A., Judd, T., & Verspoor, K. (2018). Diagnostic machine learning models for acute abdominal pain: towards an e-learning tool for medical students [Modelos de machine learning de diagnóstico para el dolor abdominal agudo: hacia una herramienta de e-learning para estudiantes de medicina]. *MEDINFO 2017: Precision Healthcare Through Informatics: Proceedings of the 16th World Congress on Medical and Health Informatics*, 447-451. IOS Press. <https://doi.org/10.3233/978-1-61499-830-3-447>
- Kulkarni, A., & Shivananda, A (2019). Natural Language Processing Recipes: Unlocking text data with machine learning and deep learning using Python [Recetas de Procesamiento de Lenguaje Natural: Desbloqueando data de tipo texto con machine learning y deep learning usando Python]. <https://doi.org/10.1007/978-1-4842-4267-4>
- Ley N° 29733, Ley de protección de datos personales. (03 de julio del 2011). http://www.pcm.gob.pe/transparencia/Resol_ministeriales/2011/ley-29733.pdf
- Manaswi N.K. (2018) Developing Chatbots. In *Deep Learning with Applications Using Python* [Desarrollando Chatbots. En *Deep learning con aplicaciones que utilizan Python*]. https://doi.org/10.1007/978-1-4842-3516-4_11
- Microsoft. (s.f.). *Intelligent bot creation that scales on demand*. <https://azure.microsoft.com/en-us/services/bot-service/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. <http://www.jmlr.org/papers/v12/pedregosa11a>
- Priya997 (31, Julio, 2018). SKF-Chatbot. Github. https://github.com/Priya997/SKF-Chatbot/blob/master/Basic_Approach/notebook/Intent_classifier.ipynb
- Ranoliya, B. R., Raghuwanshi, N., & Singh, S. (2017). Chatbot for university related FAQs [Chatbot para preguntas frecuentes relacionadas con la universidad]. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1525-1530. <https://doi.org/10.1109/ICACCI.2017.8126057>
- Rodríguez, S., Becerra, M., & Rueda, S. A. (2016). Validación de un programa electrónico de historia clínica que orienta el estudio de la apendicitis aguda en pediatría. *Pediatría*, 49(1), 8-16. <https://doi.org/10.1016/j.rcpe.2016.02.003>
- Sausa, M. (17, Febrero, 2018). Ineficiencias del sector salud están afectando a millones de peruanos. Peru21. <https://peru21.pe/peru/situacion-salud-peru-ineficiencias-sector-afectando-millones-peruanos-396225>
- Semana Económica. (31 de mayo del 2018). AFP Habitat lanzó nuevo Chatbot para atender afiliados en Facebook. <https://semanaeconomica.com/que-esta-pasando/articulos/292922-afp-habitat-lanzo-su-nuevo-chatbot-para-atender-afiliados-en-facebook>
- Shah, D., & Philip, T. J. (2019). An Assistive Bot for Healthcare Using Deep Learning: Conversation-as-a-Service [Un bot de asistencia para la atención médica mediante Deep Learning: Conversation-as-a-Service]. *Progress in Advanced Computing and Intelligent Engineering*, 109-118. https://doi.org/10.1007/978-981-13-1708-8_10
- Sheikh, M. A., & Latif, T. (2015). Acute abdominal pain in children: clinical outcomes & values of symptoms and signs in diagnosis. *The Professional Medical Journal*, 22(08), 1080-1086. <http://theprofesional.com/index.php/tpmj/article/download/1161/909>
- Toledo Cambizaca, A. F. (2018). Desarrollo de un chatbot que ayude a responder preguntas frecuentes referentes a becas en la Universidad Técnica Particular de Loja [Tesis de licenciatura, Universidad Técnica Particular de Loja]. Semantic Scholar
- Tur, G., & Deng, L. (2011). Intent determination and spoken utterance classification. *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech* [Determinación de intenciones

y clasificación de expresiones habladas. Comprensión del lenguaje hablado: sistemas para extraer información semántica del habla], 93-118. <https://doi.org/10.1002/9781119992691.ch4>

- Vaira, L., Bochicchio, M. A., Conte, M., Casaluci, F. M., & Melpignano, A. (2018). MamaBot: a System based on ML and NLP for supporting Women and Families during Pregnancy [MamaBot: un sistema basado en ML y NLP para apoyar a mujeres y familias durante el embarazo]. *Proceedings of the 22nd International Database Engineering & Applications Symposium*, 273-277. ACM. <https://doi.org/10.1145/3216122.3216173>
- Vásquez, A. C., Quispe, J. P., & Huayna, A. M. (2009). Entendimiento de lenguaje natural. *Revista de investigación de Sistemas e Informática de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos*, Lima, Perú, 6(2), 45-54. <http://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/5923>
- Yang, X., Chen, Y. N., Hakkani-Tür, D., Crook, P., Li, X., Gao, J., & Deng, L. (2017). End-to-end joint learning of natural language understanding and dialogue manager [Aprendizaje conjunto end-to-end de entendimiento del lenguaje natural y administrador de diálogos]. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5690-5694. IEEE. <https://doi.org/10.1109/ICASSP.2017.7953246>



Anexo 1: Comunicación con el Dr. Muhammad Ali Sheik para solicitar el Dataset de su investigación para el entrenamiento de mi modelo de prediagnóstico.

Access to dataset : ACUTE ABDOMINAL PAIN IN CHILDREN; CLINICAL OUTCOMES & VALUE OF SYMPTOMS AND SIGNS IN DIAGNOSIS 



Juan Domingo Benate Mendoza <20140151@aloe.ulima.edu.pe>
para drali444

24 jun 2019 8:52   

Dear Dr. Ali




My name is Juan Benate, I'm a System Engineering student from University of Lima, Peru. I'm writing this email to ask you for access to dataset of the investigation " ACUTE ABDOMINAL PAIN IN CHILDREN; CLINICAL OUTCOMES & VALUE OF SYMPTOMS AND SIGNS IN DIAGNOSIS ". I'm making a investigation about appendicitis diagnosis using machine learning so I need a dataset to train and test my model. I hope you can answer my email.

Yours faithfully,

--
Juan Domingo Benate Mendoza
Estudiante de Ingeniería de Sistemas
Universidad de Lima



Muhammad Ali <drali444@yahoo.com>
para mí

24 jun 2019 22:05   

 inglés > español [Traducir mensaje](#)

[Desactivar para: inglés](#) 

Dear Juan Benate

I am glad to know about your model. If you read the article, you will find the information there in results section. If you don't have full text of the article, let me know I will mail you the full text. I will be happy to help you.

Dr Muhammad Ali Sheikh

Associate Professor
Dept of Paediatric Surgery
Shaikh Zayed Hospital, Lahore.



Juan Domingo Benate Mendoza <20140151@aloe.ulima.edu.pe>
para Muhammad

mar, 25 jun 2019 13:25   




Dear Dr. Ali


Thank you so much for your answer. I have the full text of your investigation and read it but I need the distribution of the data, do you have, for example, an excel file with the data registered?

Yours faithfully,



Muhammad Ali <drali444@yahoo.com>
para mí

lun, 1 jul 2019 21:02   

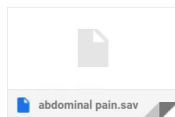
 inglés > español [Traducir mensaje](#)

[Desactivar para: Inglés](#) 

Hello, i am attaching the data set in SPSS format. This is confidential material.You can use it for personal use but do not share it.

Dr Muhammad Ali Sheikh

Associate Professor
Dept of Paediatric Surgery
Shaikh Zayed Hospital, Lahore.



Juan Domingo Benate Mendoza <20140151@aloe.ulima.edu.pe>
para Muhammad

1 jul 2019 21:23   

Dear Dr. Ali

Thank you so much for the dataset. I will use it only for academic use. Don't worry about it.

Yours faithfully,



Anexo 2: Pruebas de Funcionalidad de Pediabot.

Entrada	Respuesta esperada	Respuesta obtenida	Estado
Inicio de conversación	Hola soy PediaBot 🤖, un asistente virtual especializado en Apendicitis Aguda en niños y adolescentes de 0 a 15 años. Te puedo orientar sobre los siguientes puntos: 1) Definición 2) Síntomas 3) Tratamiento 4) Complicaciones 5) Prevención 6) Conocer si su hijo/a tiene altas sospechas de Apendicitis Aguda	Hola soy PediaBot 🤖, un asistente virtual especializado en Apendicitis Aguda en niños y adolescentes de 0 a 15 años. Te puedo orientar sobre los siguientes puntos: 1) Definición 2) Síntomas 3) Tratamiento 4) Complicaciones 5) Prevención 6) Conocer si su hijo/a tiene altas sospechas de Apendicitis Aguda	CORRECTO
Intención: Saludo	Iniciando... ó Bot iniciando, espere por favor... ó Un momento por favor, Pediabot se está conectando	Iniciando... ó Bot iniciando, espere por favor... ó Un momento por favor, Pediabot se está conectando	CORRECTO
Intención: Despedida	¡Gracias por consultar, que tengas un buen día! ó Nos vemos pronto, ¡que tengas un buen día! ó Espero haberte ayudado, ¡nos vemos pronto!	¡Gracias por consultar, que tengas un buen día! ó Nos vemos pronto, ¡que tengas un buen día! ó Espero haberte ayudado, ¡nos vemos pronto!	CORRECTO
Intención: SolicitarDefinicionEnfermedad	El apéndice es un órgano de tamaño reducido conectado al inicio del intestino grueso en el cuadrante inferior derecho del abdomen. La inflamación de este órgano se llama apendicitis. Es una emergencia que luego de las 24 horas de evolución puede complicarse y poner en riesgo la vida de su niño/a.	El apéndice es un órgano de tamaño reducido conectado al inicio del intestino grueso en el cuadrante inferior derecho del abdomen. La inflamación de este órgano se llama apendicitis. Es una emergencia que luego de las 24 horas de evolución puede complicarse y poner en riesgo la vida de su niño/a.	CORRECTO
Intención: SolicitarSíntomasEnfermedad	Existen 3 síntomas que aparecen con el paso de las horas: Generalmente primero se manifiesta el dolor abdominal alrededor del ombligo que migra a la parte inferior derecha y se intensifica con el paso del tiempo. Luego se asocia náuseas, vómitos y disminución del apetito. Finalmente aparece la fiebre que inicialmente es baja y luego se intensifica cuando el cuadro se complica (pasada las 24hrs). Existen otros síntomas pocos frecuentes sobre todo en niños pequeños: diarreas, molestas urinarias.	Existen 3 síntomas que aparecen con el paso de las horas: Generalmente primero se manifiesta el dolor abdominal alrededor del ombligo que migra a la parte inferior derecha y se intensifica con el paso del tiempo. Luego se asocia náuseas, vómitos y disminución del apetito. Finalmente aparece la fiebre que inicialmente es baja y luego se intensifica cuando el cuadro se complica (pasada las 24hrs). Existen otros síntomas pocos frecuentes sobre todo en niños pequeños: diarreas, molestas urinarias.	CORRECTO
Intención: SolicitarTratamientoEnfermedad	El médico hará una operación para extraer el apéndice afectado. Esta operación se llama apendicectomía. Actualmente, los cirujanos infantiles usan una técnica de acceso mínimo con cortes muy pequeños para poder extraer el apéndice del paciente. Por lo general esta cirugía es curativa y cursa sin complicaciones si se hace a tiempo.	El médico hará una operación para extraer el apéndice afectado. Esta operación se llama apendicectomía. Actualmente, los cirujanos infantiles usan una técnica de acceso mínimo con cortes muy pequeños para poder extraer el apéndice del paciente. Por lo general esta cirugía es curativa y cursa sin complicaciones si se hace a tiempo.	CORRECTO
Intención: SolicitarPrevencionEnfermedad	Existen pocas medidas preventivas descritas para Apendicitis, entre ellas se recomiendan: 1) Dieta rica en fibras para evitar el estreñimiento. 2) Evitar infecciones virales respiratorias o digestivas, o infestaciones por parásitos. Lo más importante es conocer la enfermedad para prevenir las complicaciones y acudir prontamente a la evaluación médica.	Existen pocas medidas preventivas descritas para Apendicitis, entre ellas se recomiendan: 1) Dieta rica en fibras para evitar el estreñimiento. 2) Evitar infecciones virales respiratorias o digestivas, o infestaciones por parásitos. Lo más importante es conocer la enfermedad para prevenir las complicaciones y acudir prontamente a la evaluación médica.	CORRECTO
Intención: SolicitarComplicacionesEnfermedad	Las complicaciones aparecen después de las 24 hrs, se asocian a la perforación del apéndice, produciendo peritonitis, que puede ser local o afectar todo el abdomen. En este caso la familia puede observar gran limitación para la movilidad del menor por el dolor intenso (el niño camina lentamente, encorvado, cogiéndose el abdomen), se puede notar la hinchazón progresiva del abdomen debido a la peritonitis, el niño se va tornando en mal estado general, con fiebres altas asociadas al dolor intenso y decaimiento. Otra veces está perforación es bloqueada por los órganos vecinos formando una colección de pus.	Las complicaciones aparecen después de las 24 hrs, se asocian a la perforación del apéndice, produciendo peritonitis, que puede ser local o afectar todo el abdomen. En este caso la familia puede observar gran limitación para la movilidad del menor por el dolor intenso (el niño camina lentamente, encorvado, cogiéndose el abdomen), se puede notar la hinchazón progresiva del abdomen debido a la peritonitis, el niño se va tornando en mal estado general, con fiebres altas asociadas al dolor intenso y decaimiento. Otra veces está perforación es bloqueada por los órganos vecinos formando una colección de pus.	CORRECTO
Intención: Prediagnostico	Inicio de diálogo de prediagnostico	Inicio de diálogo de prediagnostico	CORRECTO
Intención: None	Lo siento pero no entiendo tu mensaje ó Disculpa, no tengo la respuesta a tu consulta	Lo siento pero no entiendo tu mensaje ó Disculpa, no tengo la respuesta a tu consulta	CORRECTO
Prediagnostico: 1	Al parecer el/la menor es sospechoso de apendicitis aguda 🚨. Debería llevarlo inmediatamente a algún centro de salud por emergencia para que sea atendido lo antes posible 🚑.	Al parecer el/la menor es sospechoso de apendicitis aguda 🚨. Debería llevarlo inmediatamente a algún centro de salud por emergencia para que sea atendido lo antes posible 🚑.	CORRECTO
Prediagnostico: 0	Al parecer el/la menor tiene baja sospecha de apendicitis. Sin embargo, si no existe mejoría y los síntomas persisten más de 8 horas debe llevarlo al centro de salud más cercano para la evaluación médica. (Podría ser una apendicitis atípica)	Al parecer el/la menor tiene baja sospecha de apendicitis. Sin embargo, si no existe mejoría y los síntomas persisten más de 8 horas debe llevarlo al centro de salud más cercano para la evaluación médica. (Podría ser una apendicitis atípica)	CORRECTO

Anexo 3: Informe sobre revisión de datos para entrenamiento de modelo de Apendicitis Aguda.

Estimado Juan,

En el presente informe se describe la revisión y análisis hecho al conjunto de datos que deseas utilizar en tu investigación. A continuación, se explica aquellos atributos que tu Chatbot puede utilizar y los que no son necesarios para orientar y realizar el prediagnóstico de Apendicitis Aguda. En el caso de los atributos descartados la principal razón es que son data de laboratorio, requieren la evaluación de un profesional médico o no son relevantes para el diagnóstico del menor, ya que mencionaste que el Chatbot está enfocado en brindar prediagnóstico en el domicilio, sin que los padres de familia tengan la necesidad de asistir a un centro médico o a un laboratorio. Con respecto a los atributos para utilizar, estos contribuyen con la sospecha diagnóstica de esta enfermedad, según múltiples publicaciones de la especialidad.

Atributos por descartar:

- Data de laboratorio:
 - Nivel de hemoglobina (Hb)
 - Conteo total de leucocitos (TLC)
- Evaluación por parte de un profesional médico
 - Pulso cardiaco en la admisión (Pulseadmission)
 - Irritación de la garganta (Sorethroat)
 - Diagnóstico clínico (ClinicalDiagnosis)
- No son relevantes para el diagnóstico:
 - Veces en la que el paciente ha vomitado (Quantity)
 - Color del vómito (Color)
 - Sangrado por el recto (Bleedingpr)
 - Temperatura corporal en la admisión (°F) (Tempadm)

Atributos para utilizar:

- Edad (Age)
- Género (Sex)
- Duración del dolor por más de un día (Duration)
- Dolor abdominal en el cuadrante inferior derecho (Location)
- Propagación del dolor abdominal (Shifting)
- Dolor abdominal tipo cólico (Character)
- Falta de apetito (Anorexia)
- Nauseas o vómitos (Vomiting)
- Fiebre (Fever)
- Disuria (Micturition)
- Estreñimiento (Constipation)
- Diarrea (Loosemotions)
- Congestión nasal / garganta (Throatcongestion)
- Sensibilidad en el abdomen inferior derecho (Tenderness)
- Resistencia abdominal (Guarding)
- Masa palpable (Mass)
- Dolor abdominal al rebote (Reboundtender)
- Distensión abdominal (Distension)
- Diagnostico final (Final Diagnosis)

MINISTERIO DE SALUD
HOSPITAL NACIONAL DOCTOR JOSÉ
Dra. YADIRA GUERRA TEJADA
C.M.P. 3807 - t. 22668
Médico Asistente - Cirujía General