



Proyecto Integrador

Avance 3. Baseline

Prof. Titulares

Dra. Grettel Barceló Alonso

Dr. Luis Eduardo Falcón Morales

Prof.a Tutora

Mtra. Verónica Sandra Guzmán de Valle

Integrantes Equipo 3:

Julián David Góngora Forero - A01793036

José Javier Granados Hernández - A00566717

Karel Alejandro Cuevas Hernández - A00301492

11 de febrero de 2024

- ¿Qué algoritmo se puede utilizar como baseline para predecir las variables objetivo?

El presente proyecto se divide en dos partes: la primera es la analítica de las solicitudes por correo electrónico, información que se presentó en la entrega anterior. La segunda consiste en proponer un chatbot que automatice la atención al cliente. Para la construcción de chatbot y por efectos del alcance del proyecto se va a utilizar la herramienta de Watson Assistant de IBM. Para el avance 3 Baseline está centrado no en evaluar el desempeño de dicha herramienta, sino en evaluar el desempeño del modelo de bot que se ha definido acorde al análisis previo.

Cabe aclarar que Watson Assistant no es un motor de aprendizaje automático en sí mismo, sino una plataforma de creación y gestión de asistentes virtuales. No tiene una funcionalidad integrada para entrenar modelos de aprendizaje automático o realizar predicciones utilizando algoritmos específicos.

En su primera versión, el chatbot se definió con 8 intenciones, alimentadas con un banco de frases que oscilaban entre 123 y 196 frases de entrenamiento. El proceso de alimentar intenciones con frases de entrenamiento en Watson Assistant implica la creación de un conjunto de datos estructurado que incluye las intenciones que el chatbot debe identificar y las frases de entrenamiento asociadas a cada una de ellas.

En el análisis previo se recopilaron 1352 frases para este conjunto de datos. Se cargaron en la plataforma de Watson Assistant para entrenar al chatbot, que utiliza técnicas de procesamiento del lenguaje natural para aprender a reconocer y clasificar las intenciones basadas en las frases de entrenamiento proporcionadas. Este proceso de entrenamiento ayuda al chatbot a comprender y responder de manera efectiva a una amplia variedad de consultas de los usuarios con precisión y relevancia.

La métrica utilizada para este **Definir el indicador según Watson.**

- ¿Se puede determinar la importancia de las características para el modelo generado? Recuerden que incluir características irrelevantes puede afectar negativamente el rendimiento del modelo y aumentar la complejidad sin beneficios sustanciales.

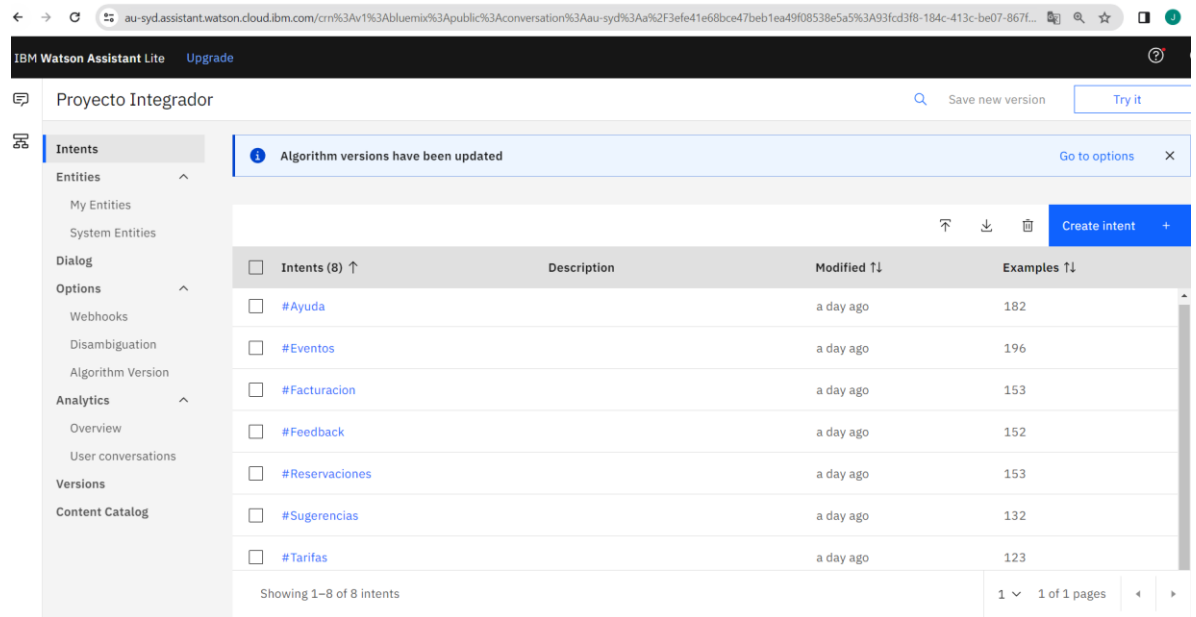
Para el proyecto en cuestión, que implica la construcción de un chatbot utilizando Watson Assistant la determinación de la importancia de características podría no aplicar directamente. En este tipo de aplicaciones, las "características" a menudo se refieren a las palabras o elementos gramaticales utilizados en las interacciones de los usuarios, y no a características numéricas o categóricas típicas en otros problemas de aprendizaje automático. También se resalta que este análisis no tiene como objetivo evaluar el desempeño de dicha herramienta. Por lo tanto, se considera que la evaluación de la importancia de características no es aplicable para este proyecto en comparación con otros casos donde se utilizan características numéricas o categóricas más explícitas.

En cambio, el enfoque principal en el que se centra este documento es en la selección de un modelo adecuado, la optimización de los parámetros del modelo y la calidad y cantidad de los datos de entrenamiento para lograr un rendimiento óptimo del chatbot.

- ¿El modelo está sub/sobreajustando los datos de entrenamiento?

Se establece que el foco de este documento es evaluar el desempeño del modelo de bot que se ha definido acorde al análisis previo, más no Watson Assistance. Con lo anterior, para determinar si un modelo está subajustando o sobreajustando en relación con los datos de entrenamiento es crucial para evaluar su rendimiento y hacer ajustes necesarios. Se establecen los siguientes pasos:

Desempeño en datos de entrenamiento: Una forma de evaluar el ajuste del modelo es observar su rendimiento en los datos de entrenamiento. Si el modelo tiene un desempeño deficiente incluso en los datos de entrenamiento, es probable que esté subajustando. Se establece un chatbot con 8 intenciones y un rango variable de frases de entrenamiento: en un rango de 123 – 196 por intención.



<input type="checkbox"/>	Intents (8) ↑	Description	Modified ↑↓	Examples ↑↓
<input type="checkbox"/>	#Ayuda		a day ago	182
<input type="checkbox"/>	#Eventos		a day ago	196
<input type="checkbox"/>	#Facturacion		a day ago	153
<input type="checkbox"/>	#Feedback		a day ago	152
<input type="checkbox"/>	#Reservaciones		a day ago	153
<input type="checkbox"/>	#Sugerencias		a day ago	132
<input type="checkbox"/>	#Tarifas		a day ago	123

Showing 1–8 of 8 intents

1 of 1 pages

Figura 1. Asistente de Watson. Intenciones con frases de entrenamiento.

Desempeño en datos de validación: Divida sus datos en un conjunto de entrenamiento y un conjunto de validación. Entrena el modelo en el conjunto de entrenamiento y evalúa su desempeño en el conjunto de validación. Si el desempeño en el conjunto de validación es significativamente peor que en el conjunto de entrenamiento, es una señal de posible sobreajuste.

Descripción de proceso:

El proceso de evaluación del desempeño del chatbot en datos de validación se llevó a cabo utilizando un conjunto de 40 frases con el propósito de determinar si el modelo estaba sub/sobreajustado. Para realizar esta evaluación, se utilizó el simulador de Watson Assistant, que proporciona una respuesta y muestra las probabilidades de las 3 intenciones más cercanas a la que corresponde la frase de entrada. Los resultados de esta evaluación se registraron en un cuadro que muestra la intención predicha por el chatbot para cada frase, junto con una

indicación de si la predicción fue exitosa o no. De esta manera, se pudo analizar si el modelo estaba clasificando correctamente las frases de validación y si había señales de subajuste o sobreajuste en el rendimiento del chatbot. Este proceso permitió verificar la capacidad del modelo para generalizar y realizar predicciones precisas en datos no vistos, lo que es fundamental para garantizar su efectividad en situaciones del mundo real.

Ejemplo de la evaluación de una frase:

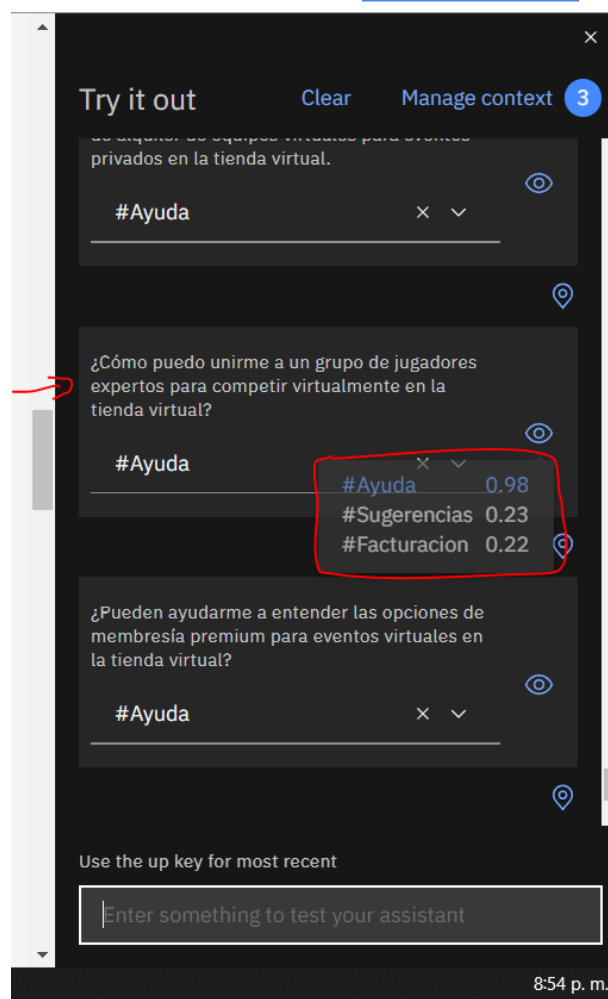


Figura 2. Evaluación de frases de test. Se evidencia las 3 intenciones más probables.

Resultados:

Frases de prueba	Resultado_1	Resultado_2	Resultado_3	Exitoso
¿Cuál es la dirección y el horario de Topgolf Monterrey?	#Ubicacion 0.90	#Trifas 0.27	#Ubicacion 0.22	Si
¿Me podrías dar la dirección y las opciones de transporte de bicicletas para llegar a Topgolf?	#Ubicacion 0.94	#Sugerencias 0.23	#Facturacion 0.22	Si
¿Dónde está ubicado Topgolf en relación con la estación de policía más cercana?	#Ubicacion 0.99	#Tarifas 0.23	#Sugerencias 0.22	Si
¿Cuál es la dirección específica de Topgolf para compartir en aplicaciones de mensajes de texto?	#Ubicacion 0.94	#Sugerencias 0.26	#Eventos 0.23	Si
¿Podrías darme la dirección de Topgolf para agregarla a mi lista de lugares favoritos?	#Ubicacion 0.98	#Sugerencias 0.22	#Facturacion 0.22	Si
¿Ofrecen eventos para celebrar el Día Mundial de los Refugiados?	#Eventos 0.98	#Ubicacion 0.25	#Tarifas 0.24	Si
¿Cuáles son las opciones de bebidas para eventos de lanzamiento de productos sin alcohol?	#Eventos 0.97	#Sugerencias 0.255	#Ubicacion 0.23	Si
¿Puedo organizar un evento para promover la inclusión de personas con discapacidades auditivas?	#Eventos 1.00	0	0	Si
¿Ofrecen eventos para grupos de amantes de la literatura?	#Eventos 0.97	#Reservaciones 0.23	#Feedback 0.22	Si
¿Cuáles son las opciones de entretenimiento para eventos de realidad extendida?	#Eventos 0.98	#Ayuda 0.23	#Sugerencias 0.23	Si
¿Cuánto cuesta el alquiler de espacios adicionales para eventos de lanzamiento de productos de belleza?	#Tarifas 0.98	#Reservaciones 0.23	#Ubicacion 0.22	Si
¿Existen tarifas especiales para eventos de lanzamiento de productos de tecnología y electrónica?	#Tarifas 0.94	#Sugerencias 0.23	#Reservas 0.22	Si
¿Ofrecen descuentos para reservaciones de eventos de lanzamiento de productos de moda y accesorios?	#Tarifas 0.99	#Sugerencias 0.24	#Sugerencias 0.23	Si
¿Cuál es el costo adicional por el uso de servicios de fotografía y videografía para eventos de lanzamiento?	#Tarifas 0.99	#Ubicacion 0.24	#Sugerencias 0.22	Si
¿La factura se puede enviar por mensaje de texto en WhatsApp?	#Facturacion 0.99	#Sugerencias 0.23	#Ayuda 0.23	Si
¿Cómo puedo obtener una factura para una fiesta de fin de año?	#Facturacion 0.98	#Ubicacion 0.24	#Eventos 0.23	Si
¿La factura se emite automáticamente al adquirir un paquete para celebraciones?	#Facturacion 0.97	#Sugerencias 0.23	#Ubicacion 0.22	Si
¿La factura se puede descargar en formato digital firmado digitalmente?	#Facturacion 0.99	#Ayuda 0.22	#Ubicacion 0.26	Si
¿Cómo puedo obtener una factura para una reunión de amigos y familiares?	#Facturacion 0.98	#Eventos 0.27	#Ubicacion 0.20	Si
¿Ofrecen reservas para eventos de lanzamiento de productos sostenibles?	#Reservaciones 0.94	#Eventos 0.25	#Sugerencias 0.23	Si
¿Tienen opciones de reserva para eventos de presentación de nuevas experiencias ecológicas?	#Reservaciones 0.94	#Sugerencias 0.22	#Feedback 0.21	Si
¿Cómo funcionan las reservas para eventos corporativos con actividades de networking ecológicas?	#Reservaciones 0.91	#Ayuda 0.25	#Sugerencias 0.22	Si

¿Ofrecen descuentos para reservas realizadas durante eventos ecológicos?	#Reservaciones 0.93	#Sugerencias 0.23	#Feedback 0.21	Si
¿Se pueden hacer reservas para eventos de presentación de nuevas soluciones ecológicas?	#Reservaciones 0.96	#Sugerencias 0.25	#Ayuda 0.23	Si
La espera para obtener una cancha de golf fue breve pero inesperada.	#Feedback 0.97	#Ubicacion 0.24	#Facturacion 0.22	Si
Necesito informar sobre un problema con la política de cancelación de reservas en línea.	#Feedback 0.90	#Ayuda 0.23	#Reservas 0.23	Si
La presentación de los platos no fue tan atractiva como esperaba.	#Feedback 0.96	#Sugerencias 0.24	#Facturacion 0.21	Si
Quiero expresar mi descontento con la organización de eventos en general.	#Feedback 0.95	#Eventos 0.25	#Sugerencias 0.22	Si
Experimenté un problema con la disponibilidad de mesas en la zona de juegos.	#Feedback 0.98	#Sugerencias 0.27	#Ayuda 0.21	Si
Quiero expresar mi satisfacción con la atención al cliente en el área de juegos infantiles.	#Feedback 0.93	#Sugerencias 0.23	#Ayuda 0.22	Si
¿Han pensado en introducir noches de música en vivo con artistas reconocidos?	#Sugerencias 0.99	#Feedback 0.32	#Ubicacion 0.22	Si
Me gustaría que hubiera más opciones de entretenimiento virtual en la zona de juegos.	#Sugerencias 0.94	#Feedback 0.24	#Ayuda 0.23	Si
Sugiero mejorar la variedad de opciones de cócteles sin alcohol en el menú.	#Sugerencias 0.98	#Feedback 0.23	#Ayuda 0.22	Si
Sería genial tener más opciones de asientos en la terraza.	#Sugerencias 0.99	#Feedback 0.22	#Ubicacion 0.22	Si
¿Podrían agregar más opciones de juegos para adultos en la terraza?	#Sugerencias 0.91	#Ubicacion 0.22	#Tarifas 0.21	Si
No puedo acceder a la información sobre las políticas de cancelación en la tienda virtual en línea.	#Ayuda 0.99	#Feedback 0.22	#Facturacion 0.21	Si
¿Qué debo hacer si mi reserva está asociada con un código de promoción vencido en la tienda virtual?	#Ayuda 0.99	#Ubicacion 0.24	#Sugerencias 0.23	Si
Necesito asistencia para entender las opciones de alquiler de equipos virtuales para eventos privados en la tienda virtual.	#Ayuda 0.99	#Sugerencias 0.24	#Factura 0.21	Si
¿Cómo puedo unirme a un grupo de jugadores expertos para competir virtualmente en la tienda virtual?	#Ayuda 0.98	#Sugerencias 0.23	#Facturacion 0.22	Si
¿Pueden ayudarme a entender las opciones de membresía premium para eventos virtuales en la tienda virtual?	#Ayuda 0.99	#Sugerencias 0.24	#Eventos 0.22	Si

Probabilidad de éxito	96.59%
-----------------------	--------

Análisis de resultados:

- Con base en la evaluación de desempeño realizada en los datos de validación del chatbot, utilizando un conjunto de 40 frases, se evidencia un rendimiento notable. La tasa de reconocimiento exitoso del 100% de los intentos sugiere una capacidad de clasificación precisa y efectiva del modelo. Además, el promedio de probabilidad de reconocimiento del 96.59% denota una alta confianza en las predicciones efectuadas por el chatbot.
- Estos resultados reflejan un modelo bien ajustado y validado, capaz de comprender y responder adecuadamente a una variedad de consultas de los usuarios. No obstante, se recomienda continuar monitoreando el rendimiento del chatbot y realizar evaluaciones

periódicas para garantizar su consistencia y eficacia a medida que se enfrenta a nuevos datos y escenarios.

- Los resultados exitosos obtenidos en la evaluación de desempeño del chatbot pueden atribuirse la calidad del conjunto de datos de entrenamiento: Es probable que el chatbot haya sido entrenado con un conjunto de datos de alta calidad y representativo, que abarque una amplia gama de posibles consultas de los usuarios. Esto permite al modelo aprender patrones relevantes y realizar clasificaciones precisas.
- ¿Cuál es la métrica adecuada para este problema de negocio?

La métrica asignada para definir el éxito es la precisión global del chatbot. Esta se define como el número de predicciones correctas realizadas por el chatbot sobre el total de predicciones realizadas. Se calcula como:

$$\text{Precisión global} = \frac{\text{Número de predicciones correctas}}{\text{Total de predicciones}} \times 100\%$$

Para este problema de negocio, donde el chatbot debe clasificar las intenciones de los usuarios, la precisión global proporciona una medida de cuántas veces el chatbot proporciona la respuesta correcta en comparación con el total de interacciones.

El porcentaje de precisión global que se considera adecuado puede variar dependiendo de varios factores, como la complejidad del problema, la disponibilidad de datos de entrenamiento y las expectativas del negocio. Sin embargo, en general, una precisión global del 90% o superior suele considerarse buena para muchos casos de uso de chatbots. Un porcentaje inferior puede indicar que el chatbot necesita mejoras en su capacidad para comprender y responder efectivamente a las consultas de los usuarios.

- ¿Cuál debería ser el desempeño mínimo a obtener?

El desempeño mínimo aceptable para un chatbot puede variar según varios factores, incluidos los objetivos del negocio, las expectativas de los usuarios y la complejidad del problema que el chatbot está diseñado para abordar. Sin embargo, proporcionaré algunas pautas generales sobre qué considerar al establecer un desempeño mínimo aceptable:

1. **Precisión:** La precisión del chatbot, medida por la cantidad de respuestas correctas proporcionadas en comparación con el total de interacciones, del 90% explicado en punto anterior. Un desempeño mínimo aceptable debe garantizar una precisión razonable que demuestre la capacidad del chatbot para realizar clasificaciones precisas de las intenciones del usuario. Peso para la evaluación del proyecto 50%

2. **Experiencia del usuario:** El desempeño mínimo aceptable también debe tener en cuenta la experiencia del usuario. El chatbot debe ser capaz de manejar consultas de manera efectiva, brindar respuestas claras y comprensibles, y ofrecer una experiencia de usuario satisfactoria en general. Se establece una calificación aceptable superior a 70% en una encuesta NPS en el chatbot con frecuencia de 1 vez cada quince días por usuario. Peso para la evaluación del proyecto 30%

3. **Expectativas del negocio:** Las expectativas del negocio juegan un papel importante en la determinación del desempeño mínimo aceptable. El chatbot debe ser capaz de

cumplir con los objetivos y requisitos específicos del negocio, como la reducción de costos, la mejora de la eficiencia operativa o la satisfacción del cliente. Se establece unos criterios calificación Peso para la evaluación del proyecto 20%

El desempeño mínimo aceptable para un chatbot podría ser una precisión global del 80% o superior, junto con una experiencia de usuario satisfactoria y una capacidad para cumplir con los objetivos del negocio. Es importante realizar pruebas y evaluaciones continuas para ajustar y mejorar el desempeño del chatbot a medida que se implementa y se utiliza en el mundo real.