

Proyecto del uso de IA y patrones de “feedback”

Índice:

I. Objetivo del proyecto

Qué se busca y porque abordaremos el análisis de datos, junto con una introducción de los términos que se usaran a lo largo de este documento.

II. Descripción del Problema

Descripción de los problemas que este análisis de datos busca resolver, junto a explicaciones de estos problemas en el contexto del uso de la Inteligencia Artificial.

III. Recursos disponibles

Una introducción a las herramientas que se usaran para este análisis de datos; se describe cuáles son, que son y cuál es su propósito en este análisis.

IV. Hipótesis Iniciales

Las aseveraciones iniciales, predicciones sobre el posible resultado de este análisis de datos, a modo de prolegómeno, listadas y explicadas cada una.

V. Definición de “stakeholders” clave (público al que se dirige el producto)

El público al que este análisis de datos podría beneficiar, se describe que tipo de público es, y las diferentes formas en las que el resultado de este análisis podría beneficiarlos.

VI. Preguntas Clave

Otras incógnitas acerca de los detalles extra que rodean este análisis de datos, con el fin de introducir el lector a la perspectiva de este análisis de datos.

VII. Fuentes de Datos Identificadas

Todas las fuentes de datos en la que este análisis de datos está basado, da un vistazo el origen de los datos, junto al sitio web del que se consiguieron.

VIII. Justificación del Proyecto

La justificación del porqué del proyecto, explicando la razón de ser de este análisis de datos, fungiendo como una conclusión y resumén de todo lo previamente visto en este documento.

IX. ¿Cuántos datos y qué tipo son?

Un listado de todos los tipos de datos y su explicación, a modo de introducir cual es el material de trabajo de este análisis.

Objetivo del proyecto:

El objetivo principal de este proyecto es la creación de un pequeño historial *representativo* del empleo por parte del usuario de los “chatbots” a través del análisis de diferentes corpúsculos de datos, tales como el índice de satisfacción, el tiempo de uso, rangos de uso continuo, motivo de uso y extensión de la conversación con el objeto de conocer cuáles son las mejores maneras de mantener la atención de los usuarios en el “chatbot”. Dos logros particulares se desprenden de reunir tal historial representativo: 1) aumentar la eficacia de cualquier usuarios de “chatbots” en lo que concierne a su negocio o negocios; 2) reducción de gastos, es decir, ahorro para los usuarios particulares; 3) garantizar a largo plazo las inversiones del usuario; 4) dar un objetivo e intención clara a la integración de “chatbots” a un negocio.

Este proyecto también podrá dirigir a cualquier tipo de usuario, avanzado o casual, en el mundo de los “chatbots”, Inteligencias Artificiales (IA) que están basadas en el uso de datos en forma de texto para poder “escribir sus propios textos”, siguiendo los textos humanos con los cuales alimentamos a la IA previamente.

La IA es una forma de inteligencia que imita a la inteligencia humana. A diferencia de la inteligencia animal, esta no es el resultado de un proceso biológico o evolutivo, sino meramente un producto humano, una herramienta tecnológica que pretende marcar un antes y un después en la historia; si bien ambiciones respecto al uso de IA han existido desde décadas atrás, sólo a partir del lanzamiento de Chat GPT el interés por la IA ha experimentado un gran y repentino despegue, creando una nueva oportunidad para mejorar negocios y emprendimientos.

Desde un punto de vista global, la IA involucra la toma de decisiones, la creación y modificación de todo tipo de media, por ejemplo, desde crear imágenes altamente realistas (al margen de si pueden ser consideradas “falsas”) hasta texto. En este sentido, se puede decir que lo que diferencia a los “chatbots” de la IA es que, esta vez, se tiene el objetivo con su uso de usar el lenguaje de una forma tan natural que la diferencia entre lo humano y lo no humano resulta indiferenciable, así como de crear textos, como si fuera escritos por un hombre común y corriente.

Etimológicamente, “Chatbots” proviene de las palabras en inglés, “chat” y “bot”, que se traducen al español como “conversaciones” y “robots”, es decir, algo que se refiere a, la Inteligencia Artificial.

Estas IAs tienen la capacidad de replicar diversos estilos de escritura humana si se alimentan con los datos adecuados, y al ser capaces de evaluar el contexto (que se define con las instrucciones que se le den), pueden interactuar con el usuario en tiempo real, siempre manteniendo cierto tono y estructura, a todas horas, dando una consistencia impresionante a la hora de asistir. Lo que da pauta al foco principal de nuestro análisis.

Descripción del Problema:

Los problemas principales que conciernen al uso apropiado de “chatbots” es:

- ❖ La IA tiende a ser usada de manera errónea, se usa fuera de contexto o de manera innecesaria.
El proyecto busca exponer en que tipos de situaciones la IA exhibe calificaciones más altas de parte del usuario y, por ende, saber qué tipo de negocios realmente se van a beneficiar más del uso de la IA y “chatbots”.
- ❖ La IA no es una herramienta barata en largo plazo, y presenta cambios impredecibles en costos; cada mensaje implica costo, tanto el desaprovechamiento y alta demanda causan gastos.
El proyecto busca demostrar que temas tienen más alta frecuencia de mensajes y poder ayudar a saber que esperar según nuestro negocio o emprendimiento, y encontrar estabilidad al usar esta herramienta.

Para dar más contexto, la IA, en esta era, está siendo usada, hasta cierto punto, de forma excesiva, y, a veces inútil y mal supervisada, lo que la despoja de reputación a ojos del público, y potencia de la misma, pero, también provoca casos de alucinación en nuestro “chatbot”. En pocas palabras, significa un resultado muchas veces sin sentido, y, en algunas ocasiones, desastrosamente polémico que puede llegar a causar inconveniencias reales, lo que, lógicamente causa desconfianza y duda sobre el uso de “chatbots” en proyectos grandes, desperdiando potencial que de otra

manera, hubiera dado gran beneficio; por ende es imperativo identificar como evitar tales problemas para triunfar donde otros han fracasado.

Recursos disponibles:

En cuanto a las herramientas y tecnologías que usaremos para llevar nuestro proyecto acabo, junto a la presentación, son:

Python, un lenguaje de programación que hace posible este análisis. Para empezar en la misma página, un lenguaje de programación son los ladrillos que construyen este análisis de datos; Python, en específico, siendo fácil de usar, intuitivo y moderno, lo que lo hace muy flexible a la hora de aplicar a cualquier otro proyecto moderno; volviendo al tema anterior este lenguaje de programación es crucial para analizar datos, este análisis de datos se divide en procesos de interpolación y comparación, todo aplicado a los datos con los que se estará trabajando, a modo de encontrar patrones y respuestas.

Pandas, es la segunda herramienta importante de este análisis; a modo de extender aún más las capacidades de nuestro lenguaje de programación, Python, Pandas es una librería perfecta para análisis de datos; en resumen, esta librería se podría considerar una extensión y/o mejora que se puede añadir a Python. Pandas ayuda al ser, de cierta forma, lo que permitió que este análisis se lleve a cabo, abriendo la capacidad de análisis de datos y otorgando las herramientas para hacerlo; y, por ende, vital.

“Matplotlib”: Se usara esta herramienta digital para una representación inicial de los datos; “Matplotlib” es una librería de Python, similar a Pandas, por lo que también se puede considerar una extensión a las capacidades de Python, esta vez, ofreciendo la capacidad de graficar y representar los datos y los resultados que nos den, de una forma visual, rápida, y, sobre todo, fácil de entender sin requerir conocimientos previos sobre análisis de datos, lo que lo hace una herramienta importante para comunicar los datos al público especializado y general.

“Plotly”: Similar a Matplotlib, es un sitio web que permite crear graficas flexibles que pueden traer un mayor contraste de datos, esta vez, usando las graficas para encontrar patrones diferentes y lograr encontrar una perspectiva diferente a los datos recolectados para un mejor resultado general a la hora exponer y presentar las gráficas.

Trabajamos con 10000 datos, en 11 columnas, la base principal de este análisis de datos. A modo de presentar qué tipo de datos vamos a presentar, vamos a exponer nuestros datos de una manera resumida y general:

- ❖ ID de la sesión: Determina el numero de la sesión del estudiante, que va de un rango de 00001 a 10000.
- ❖ Nivel del Estudiante: Determina el nivel del estudiante al interactuar con la IA, consta de preparatoria, universitario o graduado; esto nos ayuda a saber que

esperar del estudiante y ponerlos en perspectiva junto a los demás datos de cada sesión.

- ❖ Disciplina: Determina la línea de estudios/tema del estudiante, estas materias son: Ciencias de la computación; psicología; biología; matemáticas; historia e ingeniería. Esto nos ayuda enormemente a saber la frecuencia de materias y los detalles que tienen cada materia en concreto, relativo a la Inteligencia Artificial.
- ❖ Fecha de la sesión: Como indica su nombre es la fecha en la que la sesión se realizó, estas fechas empiezan desde finales de 2024 hasta mediados de 2025, y nos ayudan a saber lo cerca que está una sesión con la otra, para calcular tráfico o frecuencia.
- ❖ Duración de la sesión: Determina que tan larga fue cada sesión en minutos, algunas siendo tan cortas como 10 minutos, hasta una hora (60 minutos); estos datos determinan cuánto tiempo una persona está activamente interactuando con la Inteligencia Artificial, y nos ayuda a saber más si lo contrastamos con otros datos, como disciplina o satisfacción.
- ❖ Prompts totales: En otras palabras, mensajes totales, nos da la cantidad total de mensajes que el estudiante tuvo con la Inteligencia Artificial, cantidad variando enormemente; similar a la duración de la sesión, por sí solo no nos dice mucho, pero, contrastado con otros datos nos puede ayudar a poner en mejor perspectiva la eficacia y, sobre todo, enganche de nuestra IA, lo que hace que nuestra IA sea atractiva para el usuario y ellos mismos quieran interactuar con esta, lo que, se traduce a demanda de nuestra Inteligencia Artificial.
- ❖ Tipo de tarea: El tipo de tarea que el estudiante requirió que la Inteligencia Artificial asistiera con, en total, todos los temas son: Estudiar; programación y código; lluvia de ideas; asistencia con tarea e investigación. Con la disciplina, nos ayuda a saber cuál fue el tema de la conversación y que hizo la Inteligencia Artificial, ayudando a saber qué temas se pidieron junto a los demás datos, lo que abre la puerta a comparaciones interesantes.
- ❖ Nivel de asistencia de la Inteligencia Artificial: Lo que indica el rendimiento aplicado y real que nuestra IA tuvo con el usuario, desde errores hasta alucinaciones, esto determina que tan comunes o raros fueron dependiendo del nivel de asistencia, que va de un rango de 1 a 5, ayudándonos a saber si la IA hizo el trabajo que se le pidió, si es una calificación alta; o, si hubo problemas de por medio si es una calificación baja.
- ❖ Resultado Final: Estos datos nos dan un rango de 4 resultados, estos dependiendo del resultado final que tuvo el usuario con la Inteligencia Artificial, los rangos siendo: Asignación completada; confusión, idea postergada/archivada o fallo. Esto nos ayuda a saber que resultado final tuvo nuestro usuario con la IA, y, principalmente, saber si la IA fue de alguna ayuda o si fue todo lo contrario, que, contrastado con los datos anteriores, nos ayudaría a saber qué hace la IA de manera idónea, o que no.

- ❖ ¿Uso de nuevo?: En un rango de falso o verdadero, nos ayuda a saber si el estudiante uso nuestra IA otra vez, o, si de lo contrario, no la volvió a usar. Por sí solo, estos datos nos informan bastante, pues, nos dicen si el usuario se interesó lo suficiente para volver otra vez, o, no. Y, contrastado con otros datos nos ayuda a saber aún más de lo bien que haría la IA en ciertas tareas que otros, y saber dónde la Inteligencia Artificial es más atractiva para cierto tipo de estudiantes, por ejemplo.
- ❖ Índice de satisfacción: Finalmente, esto nos indica en una escala de decimales del 1 al 5 cuál fue la experiencia general que el usuario consideró que tuvo con la Inteligencia Artificial, siendo unos de los datos que otorgan más información en general, pues, saber la opinión que el usuario tuvo, contrastado y comparado con el resto de los datos nos da una idea bastante clara de lo que está ocurriendo, y ayuda bastante a completar todo lo que se propone en este análisis de datos.

Hipótesis Iniciales:

Saber qué buscamos dará una mejor perspectiva de nuestras ideas y opiniones, junto a qué esperar y qué no:

- ❖ **Hipótesis 1:** Una menor cantidad de mensajes se correlaciona a insatisfacción y/o fallo al realizar la tarea objetivo.
- ❖ **Hipótesis 2:** Llevar a cabo una tarea de manera óptima está directamente correlacionado con el usuario volviendo a usar el “chatbot”.
- ❖ **Hipótesis 3:** Estudiantes de preparatoria para arriba exhiben un mayor uso de los “chatbots”.
- ❖ **Hipótesis 4:** Los temas relacionados a matemáticas y programación exhiben el mayor uso de “chatbots”.

Definición de “stakeholders” clave (público al que se orienta el producto):

Siendo específicos, podemos tener en cuenta:

- Usuarios / interesados en el uso de Inteligencia Artificial: Aquellos que quieran saber, ya sea a modo de guía o para aumentar sus conocimientos acerca de las capacidades y/o debilidades de esta, que pueda aumentar la eficiencia personal y ayudar a comprender las nuevas herramientas con las que se está trabajando con a la fecha de escritura de este documento.
- Dueños de negocios pequeños o emprendimientos: La IA puede ayudar a acortar trabajos enormemente, y agilizar trabajos considerados triviales para el nivel de habilidad o conocimiento del usuario; esto puede ayudar enormemente

a acortar los tiempos que hay entre procesos, mientras que hace trabajos grandes más fáciles de soportar y llevar a cabo, ayudado enormemente a aquellos que no tienen un equipo grande para respaldarlos, o quieren llevar un pequeño proyecto a una escala mayor.

- Asistencia al cliente: Es uno de los muchos campos en la que la inteligencia artificial es más apta de aplicar, mientras se ejecute de manera correcta, claro está; dejar una maquina como la cara de la compañía, aquel que interactúa con los usuarios y tiene contacto directo, ya sea ayudando y/o respondiendo, es una responsabilidad grande en sí misma, y por ello es algo que ha generado opiniones varias, negativas y positivas, pues: un fallo podría causar una alucinación por respuesta (la alucinación, como se mencionó antes, es una respuesta sin sentido o incluso inapropiada), lo cual es bastante malo, y podría dañar la reputación de aquellos a los que la IA pertenezca. En el caso contrario, puede ayudar mucho a las compañías de asistencia al cliente, bajando costes y aumentando la accesibilidad y constancia de esta, pues, una IA nunca se cansa, haciendo posible una cobertura del servicio total las 24 horas del día, los 7 días de la semana. El determinante del triunfo es la aplicación correcta de la misma, por lo que este análisis puede dar una asistencia valiosa para evitar los problemas antes mencionados.
- Marketing y publicidad: El poder entender que vendemos y con que tratamos da una ventaja a la hora de promocionar un cierto servicio. Este análisis ayuda a entender justamente eso, el poder detectar áreas fuertes y los porcentajes de que tipo de personas las usa más, y, con que propósito; dando una clara visión sobre con qué tipo de público se está tratando y que buscan.

Preguntas Clave:

Preguntas objetivo:

1. ¿Cuál es la correlación entre nivel educativo y/o si trabaja, con los temas que los usuarios eligen?
2. ¿Menos cantidad de mensajes (prompts) se traduce a fallar en realizar la tarea?
3. ¿La inteligencia artificial muestra más efectividad (número de trabajos completados) en trabajos relacionados a escritura que en programación? ¿O son iguales?
4. ¿Estudiantes de ramas de ciencias computacionales exhiben el mayor uso del resto de materias/disciplinas?
5. ¿Un uso más prolongado de la IA se traduce a un menor índice de efectividad general de la IA?

6. ¿Mientras menor sea el nivel educativo del usuario, se presentan más errores y más insatisfacción al terminar de usar la IA?
7. ¿Cuál es la materia/disciplina que tiene la mayor cantidad de insatisfacción con la IA? ¿Cuál es el contrario de este?
8. ¿Las sesiones de estudio con IA brindan buenos resultados y satisfacción?
9. ¿Los usuarios que volvieron a usar la IA exhiben un mayor índice de satisfacción?
10. ¿De todos los casos de uso de IA para tareas, en su mayoría son exitosas?

Fuentes de Datos Identificadas:

Kaggle: “AI Assistant Usage in Student Life”, que se traduce a el uso de los asistentes potenciados por Inteligencia Artificial en la vida estudiantil, es la fuente de los datos que hace este análisis de datos realidad, y la principal razón de su existencia. Contiene todos los datos antes mencionados. Viene de Kaggle, un sitio web que se especializa en bases de datos, y que, además de esta base de datos contiene muchas más de todo tipo de temas diferentes. Esta misma base de datos tiene sus orígenes en la necesidad de saber el estado y efecto de la IA en la vida estudiantil, basado en:

- ❖ (“EDA and visualization practice”) - Practica de visualización y EDA.
- ❖ (“Machine learning modeling”) - Modelado de Aprendizaje automático.
- ❖ (“Feature engineering workflows”) - Flujos de trabajo de Ingeniería orientada a características.
- ❖ (“Educational data science exploration”) - Ciencia de datos educacional y exploración.

Todo esto se resume a modelos de predicción (es el uso de la Inteligencia Artificial para predecir datos o eventos gracias al uso de bases de datos y otros procesos que les enseñan que tipo de datos se están manejando junto al contexto para crear predicciones aproximadas de cuáles serán los siguientes datos o eventos por suceder) y análisis de datos, como este, que ayudan a obtener los datos para entrenar el modelo predictivo y obtener los resultados deseados de manera satisfactoria; este es uno de los muchos usos que se les pueden dar a bases de datos como esta en conjunto con la Inteligencia Artificial.

Justificación del Proyecto:

Buscar como aprovechar al máximo la IA, junto con nuevas formas de saber cómo optimizar y acelerar los procesos de trabajo dentro de ciertos límites, es el enfoque general de este análisis.

Imaginar un mundo donde la Inteligencia Artificial sea un complemento del humano, como a su vez este es un complemento de la Inteligencia Artificial es una de las

razones por la que este análisis es, pues, aunque sea contradictorio, la IA es una herramienta que también depende en el rendimiento y “feedback” del usuario, en este caso, el alumno; el alumno tiene que dar un “prompt” (mensaje) de calidad para obtener un resultado final de calidad de parte de la Inteligencia Artificial.

Entender que áreas pueden aventajarse más con el uso adecuado de la IA, que sector de la población lo usa más y para qué, junto a otros datos, nos puede ayudar a entender donde la IA se puede aplicar mejor, y cuál es el enfoque más estable que se puede tomar para sacar provecho de esta, no solo desde la perspectiva de una empresa o emprendimiento, sino a nivel local y personal; todo esto, es el objetivo de este análisis, y los resultados que se esperan de ella.

¿Cuántos datos y que tipo son?:

A modo de entrar en detalles, tenemos un acercamiento más atento a los datos que tenemos en la base de datos:

- ID de la sesión: Siendo un total de 10000, como se mencionó, son la principal forma de discernir cada sesión, y son de un rango de 00001 a 10000; siendo un objeto.
- Nivel del Estudiante: 10000 en total, es un objeto, debido a que este contiene la característica de cual es nivel académico de cada sesión de los estudiantes.
- Disciplina: 10000 en total, y como los anteriores, también es un objeto, esta vez conteniendo la materia del estudiante, más específicamente, el tema que se interactuó con la Inteligencia Artificial.
- Fecha de la sesión: 10000 en total, es un objeto, esta vez conteniendo la característica de la fecha en la que el estudiante interactuó con la Inteligencia Artificial.
- Duración de la sesión: 10000, esta vez siendo diferente a las anteriores, al ser un flotante, esta vez, son decimales, pues, contiene los datos en minutos y segundos de cada una de las sesiones, por ende, se usan decimales en este caso.
- “Prompts” totales: Siendo 10000, esta vez siendo un entero, pues, no maneja ningún tipo de decimales, esta vez siendo entero, conteniendo el número de mensajes totales por cada sesión que los estudiantes tienen con la Inteligencia Artificial.
- Tipo de tarea: 10000, es un objeto que determina por cada sesión que los estudiantes tienen con la Inteligencia Artificial, el tipo de tarea realizada entre el estudiante y la Inteligencia Artificial, siendo un objeto porque contiene la característica del tema tratado.
- Nivel de asistencia de la Inteligencia Artificial: 10000 datos en total, siendo el último de los dos datos de clase de enteros, sin decimales, son números enteros en un rango de 1 a 5 que determina la funcionalidad técnica y rendimiento general que la Inteligencia Artificial tuvo en la sesión con el estudiante.

- Resultado Final: 10000 datos, esta vez volviendo a objetos; representando el valor final de la sesión del estudiante con la Inteligencia Artificial, y determina si llegó a una conclusión satisfactoria, pues recordemos que son 4 posibles conclusiones, si la tarea fue exitosa, si el estudiante terminó confundido, si la idea se aplazó o si el estudiante se aplazó con la Inteligencia Artificial; al ser cuatro características/resultados, es la razón por la que este apartado se aplazó.
- ¿Uso de nuevo?: 10000 datos, este siendo el único apartado de toda la base de datos en contener datos booleanos, datos de true y false, pues, recordemos que este apartado de, si el alumno uso de nuevo es, si la Inteligencia Artificial fue capaz de interesar los suficiente al estudiante, o satisfacer sus necesidades, como para que el estudiante vuelva una segunda o más veces para seguir usando la Inteligencia Artificial en busca de resultados similares; por ende, True y False es la mejor manera de describir y clasificar estos datos.
- Índice de satisfacción: 10000 datos, como todos, siendo el último apartado con datos flotantes, decimales, esta vez estableciendo el índice de satisfacción, que abarca de 1 a 5, pero esta vez, con la capacidad de ser cualquier decimal dentro de los rangos establecidos.

La cantidad de datos como se acaba de resumir es en total 10000, de estos 10000, hay un conjunto de números enteros, flotantes y texto, estas asignaciones nos ayudan a saber de qué tipo de dato se trata y diferenciarlos.

Tenemos un total de solamente un booleano (unidad lógica; definida por true y false, verdadero y falso en español, y, simplificado, se podría ver como si y no); dos flotantes (son números, pero, esta vez, decimales, los flotantes son todo tipo de número que siga una estructura de 00.00); dos enteros (números enteros con los cuales todos somos familiares con) y seis objetos (una entidad de uno o más datos, funciones o características, en otras palabras, una característica especial). Con todo esto, ya sabemos con qué estamos tratando y podremos tener una mejor idea con qué tipo de datos estamos trabajando con.