



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

CARRERA DE INGENIERÍA EN SOFTWARE

TEMA:

“Aplicación Web Para El Análisis De Contenido Usando Servicios Cognitivos De Inteligencia Artificial En La Plataforma Web Educativa Jovenesweb”.

AUTORES: BETANCOURT LUDEÑA, ELVIS FERNANDO
PANATA CASTILLO, EDGAR SANTIAGO

DIRECTOR: ING. VELASCO SÁNCHEZ, DIEGO ISMAEL
LATACUNGA, 2022



Contenido

Introducción

Planteamiento del problema

Fundamentos Conceptuales y Referenciales

Desarrollo del sistema

Validación y Análisis de los resultados

Conclusiones y Recomendaciones



Introducción



Introducción

Formulación del marco teórico asociado al análisis de contenido mediante el uso de servicios cognitivos de Inteligencia Artificial

Conceptualización de la propuesta

Implementación de la propuesta

Validación de la propuesta



Planteamiento del problema



Planteamiento del problema



- Constante crecimiento del contenido generado por usuarios
- El contenido ofensivo provoca problemas emocionales o de salud mental
- El trabajo como moderador de contenido puede llegar a ser tedioso
- Alto gasto de recursos económicos para la moderación de contenido
- Tiempos de revisión de contenido muy largos
- Normalización de la violencia



Formulación del problema

En consecuencia, a la problemática descrita se obtiene el siguiente problema:

¿Cómo optimizar el proceso del análisis de contenido que publican los usuarios de la plataforma web educativa Jovenesweb?



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Objetivo General

Desarrollar una aplicación web para optimizar el análisis de contenido que publican los usuarios de la plataforma web educativa Jovenesweb, usando servicios cognitivos de Inteligencia Artificial.



Objetivos Específicos

Elaborar el marco teórico con los conceptos fundamentales necesarios para el desarrollo de una aplicación web que analice el contenido de la plataforma educativa Jovenesweb a través del uso de servicios cognitivos de Microsoft Azure.

Configurar e integrar los servicios cognitivos a través de la API Rest de Microsoft Azure

Desarrollar el aplicativo web aplicando la marco de trabajo SCRUM

Implementar el aplicativo web en la plataforma educativa Jovenesweb ubicada en la ciudad de Loja, provincia de Loja

Validar el aplicativo web mediante pruebas de funcionalidad y rendimiento.



Hipótesis

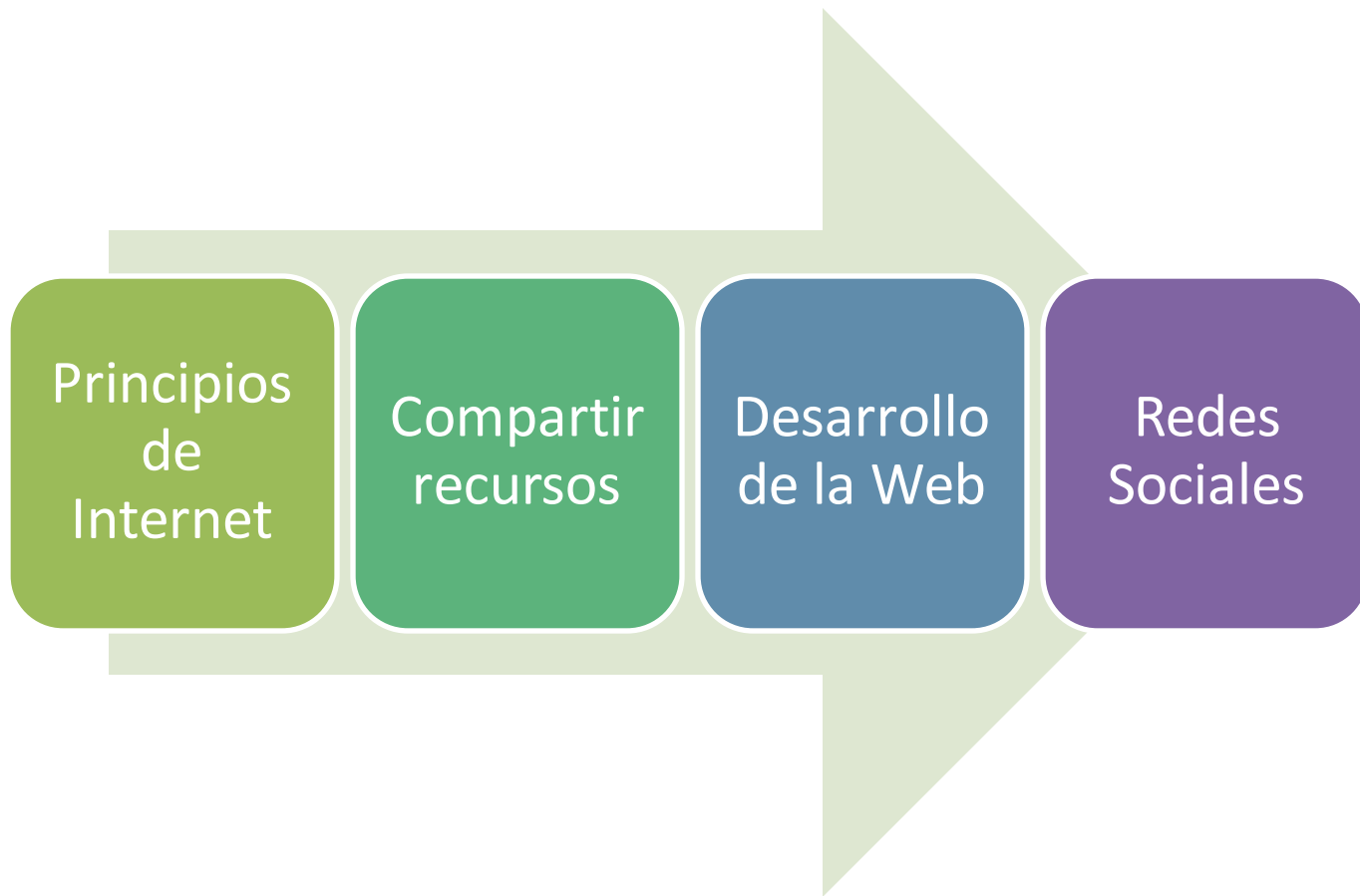
¿Si se desarrolla una aplicación web para el análisis de contenido usando servicios cognitivos de inteligencia artificial entonces se optimizan los tiempo y calidad de revisión en la plataforma web educativa Jovenesweb?



Fundamentos Conceptuales y Referenciales



Antecedentes históricos



Marco de trabajo SCRUM

Para proyectos en entornos complejos donde los requisitos cambian

Permite resultados rápidos donde la innovación, la flexibilidad y la productividad son esenciales y quieren trabajar en un ambiente ágil.

Integra las mejores prácticas y el trabajo en equipo de colaboración para lograr mejores resultados



Inteligencia Artificial

Capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano

Es una rama de la informática que pretende responder afirmativamente a la pregunta que planteo Alan Turing



Antecedentes Conceptuales y Referenciales

Servicios Cognitivos de Inteligencia Artificial

Proveedor	Imágenes	Videos	Texto
Amazon Web Services	x	x	x
Google	x	x	x
IBM	x	-	x
Microsoft Azure	x	x	x



Antecedentes Conceptuales y Referenciales

Servicios cognitivos de Microsoft Azure

Productos de Inteligencia artificial y aprendizaje automático, cálculo.

Desarrollar sistemas inteligentes que le permitan ver, escuchar, hablar, buscar.

Modelos desarrollados con estrictos estándares éticos para un uso responsable



Antecedentes Conceptuales y Referenciales

Moderación de contenido en redes sociales

Determinar si cumplen o no los estándares y reglas de la comunidad

Se genera información falsa que se propaga a través de publicaciones en redes sociales

Importante considerar el contexto del contenido



Base conceptual

- **Contenido web:** En la actualidad gran parte del contenido se genera de manera dinámica, un caso común son las redes sociales que proporcionan la facilidad de subir contenido de los mismos usuarios y permite que interactúen entre sí.
- **Análisis del contenido:** Es interpretar los datos de manera cualitativa y cuantitativa de un aplicativo web con diversos fines y permite determinar el comportamiento del usuario o de un grupo de usuarios

Base conceptual

- **Contenido ofensivo:** Según la RAE se define como "“humillar o herir el amor propio o la dignidad de alguien, o ponerlo en evidencia con palabras o con hechos”".
- **Sentimientos:** Según la RAE se define como el “estado efectivo del ánimo”, existen tres sentimientos que predominan: positivo, neutral y negativo.
- **Entidades:** Las entidades corresponden a la información específica de personas, lugares, ubicaciones, expresiones numéricas o temas en general que poseen un reconocimiento propio o único.

Antecedentes contextuales



jovenesweb



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

Desarrollo de la propuesta



Visión del proyecto

- La plataforma web educativa Jovenesweb maneja grandes cantidades de contenido generado por sus usuarios cada día y que debe ser moderado para garantizar la calidad del contenido. Para lo cual se desarrolló Content Moderator, que es un aplicativo web que mediante el consumo de Servicios Cognitivos de Inteligencia Artificial analiza y clasifica dicho contenido en:
 - Sentimientos (positivo, negativo neutral)
 - Entidades (información específica de lugares,
 - Ubicaciones, expresiones numéricas, personajes influyentes ...)
 - Contenido ofensivo

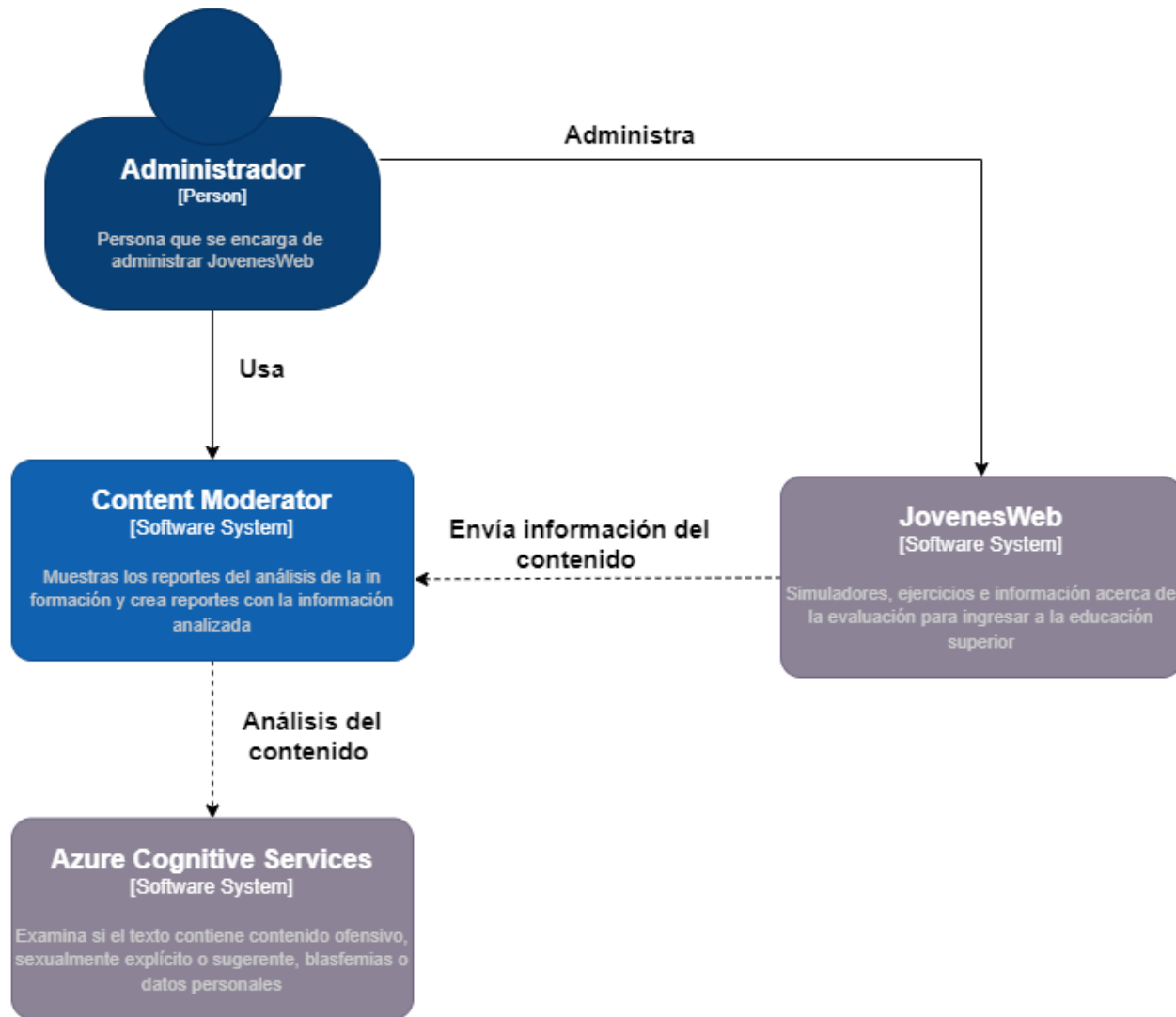
Equipo SCRUM

Core	
Product Owner	Elvis Betancourt
Scrum Master	Santiago Panata
Development Team	Elvis Betancourt Santiago Panata

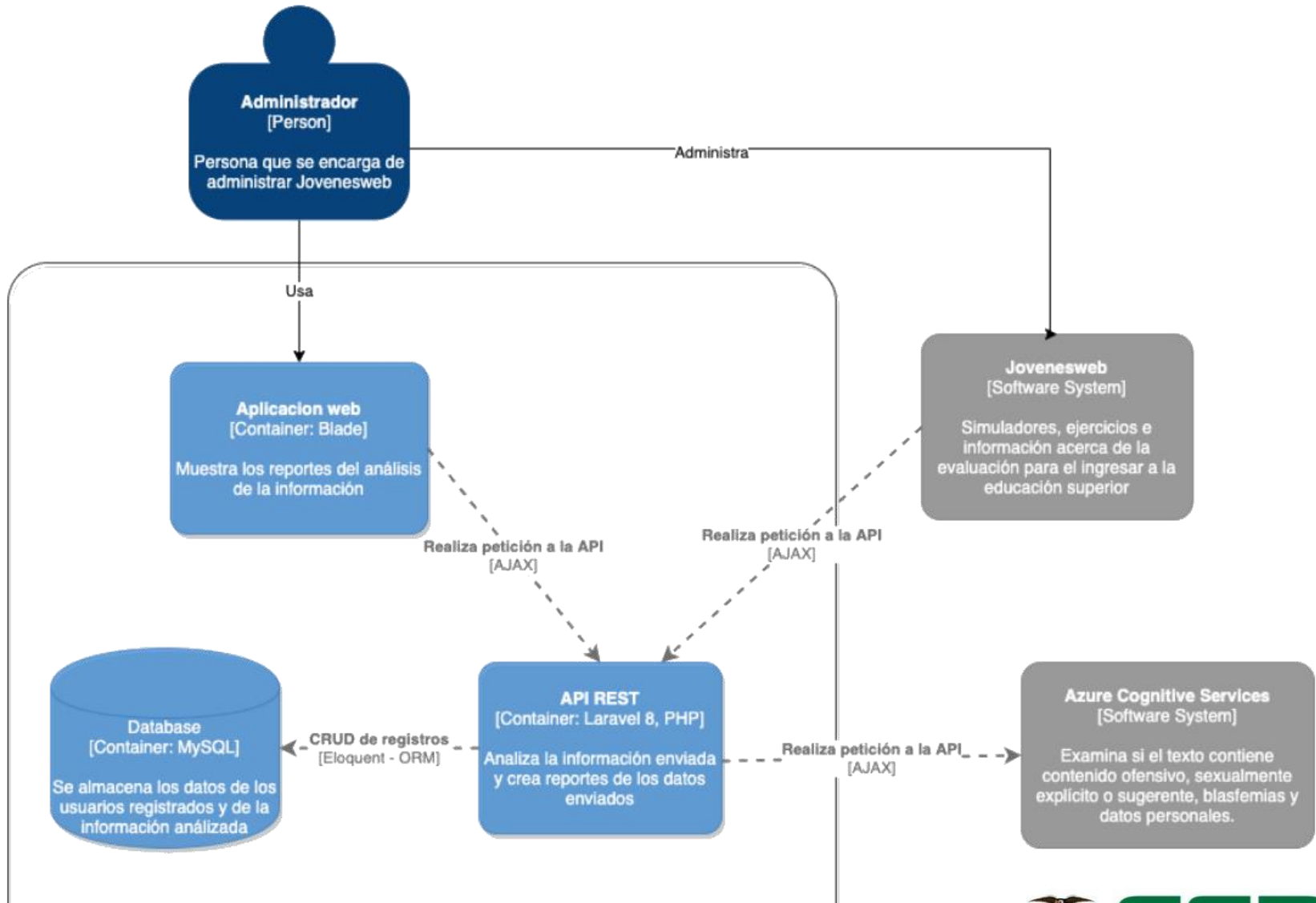
Non - Core	
Stakeholders	Jovenesweb
Scrum Guidance Body	Ing. Diego Velasco



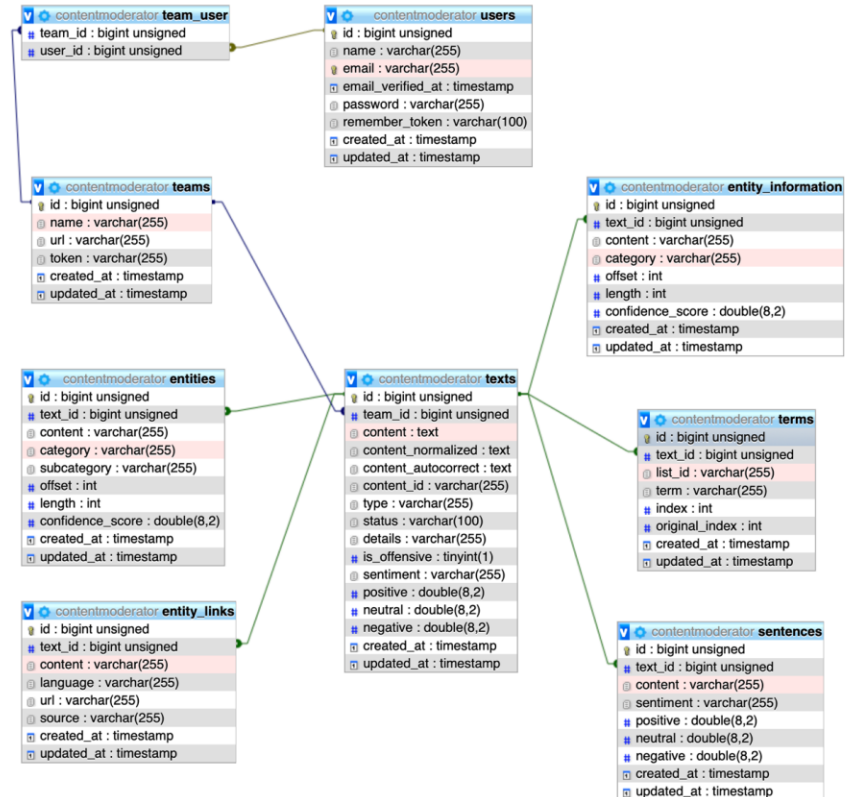
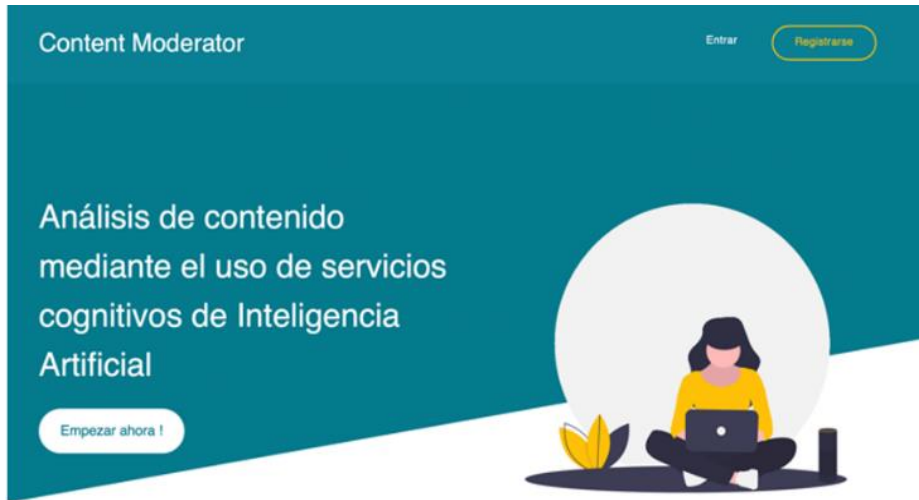
Arquitectura del proyecto: Nivel 1



Arquitectura del proyecto: Nivel 2



Sprint 1



Sprint 1

ContentModerator

Crear una nueva cuenta

Nombre completo

Email

Contraseña

Confirmar la contraseña

[+ Registrarse](#)

[Ya tengo una cuenta](#)

ContentModerator

Crear una nueva cuenta

Nombre completo El campo Nombres completos es obligatorio.

Email El campo Email es obligatorio.

Contraseña El campo Contraseña es obligatorio.

Confirmar la contraseña

[+ Registrarse](#)

[Ya tengo una cuenta](#)

ContentModerator

Autenticarse para iniciar sesión

Email

Contraseña

Recordarme [+ Acceder](#)

[Crear una nueva cuenta](#)

ContentModerator

Autenticarse para iniciar sesión

Email El campo Email es obligatorio.


Contraseña El campo Contraseña es obligatorio.





Recordarme [+ Acceder](#)

[Crear una nueva cuenta](#)



Sprint 2

 ContentModerator

-  Inicio
-  Crear grupo
-  Grupos** ▾
-  Jovenesweb ▾
- > Resumen
- > Textos**
- > Textos ofensivos
- > Entidades** ▾
- > Categorías
- > Información Personal
- > Enlaces

menu.pages



Textos analizados



Id	contenido	Ofensivo
27	La respue respuest.	No ofensivo
78	Es: Desci	Ofensivo
124	eso esta tien...	Ofensivo
211	Ya entien	Ofensivo
227	Bueno, e	No ofensivo

Textos ofensivos en Jovenesweb



Buscar

Contenido	Ofensivo
Blanditiis cupiditate ex quia dolore porro.	No ofensivo
Consequatur nihil ut ut dolorum saepe.	Ofensivo
Delectus esse corporis quia quo aut sed.	No ofensivo
Dolores ab saepe repellendus recusandae in.	No ofensivo
Dolores error a consectetur.	Ofensivo
Eligendi a voluptas autem suscipit odit.	Ofensivo
Et sapiente voluptate ullam.	Ofensivo
Ex natus amet incidunt nam.	Ofensivo
Exercitationem non aut officiis nobis et doloribus consequatur.	Ofensivo
Explicabo illo laboriosam fugiat quis.	No ofensivo

Mostrar página 1 de 2

Mostrar por página

Anterios **1** 2 Siguiente



Sprint 2

Search

- Introduction
- Authenticating requests
- Análisis de Entidades
 - Analizar Entidades**
 - Analizar Entidades
 - Analizar Entidades con Enlaces
 - Analizar Entidades con Enlaces
 - Analizar Entidades Info Personal
 - Analizar Entidades Info Personal
- Análisis de Sentimientos
- Contenido ofensivo

Documentation powered by Scribe 📖

Last updated: August 19 2021

Analizar Entidades

requires authentication

Esta URL te permite analizar las entidades existentes en un texto

Se retorna un objeto con el análisis del texto

Request [Try it out](#)

POST `api/analyze/text/entity`

Body Parameters

text string

Texto a ser analizado

text_id integer *optional*

Identificador del texto analizado

bash javascript

Example request:

```
const url = new URL(
  "http://content-moderator.test/api/analyze/text/entity"
);

const headers = {
  "Content-Type": "application/json",
  "Accept": "application/json",
};

let body = {
  "text": "nulla",
  "text_id": 14
}

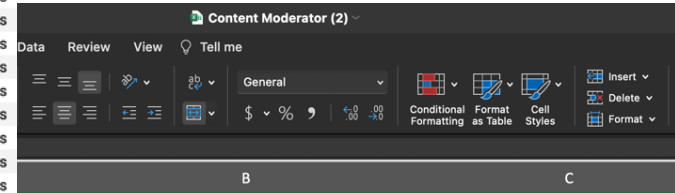
fetch(url, {
  method: "POST",
  headers,
  body: JSON.stringify(body),
}).then(response => response.json());
```



Sprint 3

Content Moderator

Id contenido	Contenido	Tipo	Sentimiento	Estado
2	Ponen la pregunta mas dificil justo, la que no hay tutoriales de youtube ni en Google. Ni Juan ni Al...	-	N/A	success
12	Ponen la pregunta mas dificil justo, la que no hay tutoriales de youtube ni en Google	-	N/A	success
19	Que falta de respeto con esa grafía cancerígena caballero.	comment	N/A	success
27	La respuesta esta mal!! yo la resporte pero no para que cambien el orden de las opciones La respuest...	comment	N/A	success
29	Tienes razón la solución ha sido borrada por falta de grafía	comment	N/A	success
33	15385 x.....100 x= 180 180.....80 x.....100 x= 225 225.....	comment	N/A	success
46	La excreción es un proceso fisiológico, que le permite al organismo eliminar sustancias de desecho t...	solution	neutral	success
49	Una pregunta porque al 80 le resta la edad de Ana ?	comment	N/A	success
50	$x-25x/100-15x/100=975$ $100x-25x-15x/100=975$ $60x/100=975$ $3x/5=975$ $3x=(5)975$ $x=4875/3$ $x=1625//$	comment	N/A	success
56	Le restamos 6 cm al ancho y al largo de las medidas externas del porta retrato porque son 3 cm que o...	comment	N/A	success
57	se me olvidó mensionar que 120 es mayor q 100 por lo tanto se le puede aser el descuento q ofresia e...	comment	N/A	success
72	Tienes razón sergio la respuesta ha sido corregida	comment	positive	success
76	La respuesta es conclusivo porque el parrafo subrayado con negrita comienza con un conector de concl...	comment	N/A	success
78	Es: Descanso es a trabajo La respuesta ya ha sido corregida.	comment	neutral	success
79	esta bien? no debería ser melodía:oido	comment	N/A	success
84	N se de donde sacas esa solución pues a mí me salió 1024 y no fue la correcta no comprendo	comment	negative	success
85	En la solución esta como se resuelve Y como sacaron 1024?	comment	neutral	success
86	por que? si me dicen razón no se trata d una progresión aritmética? si es como nos están indicando e...	comment	N/A	success
92	jajaj listo, esta esta mejor, la solucion es $2^{(10)}$	comment	N/A	success
100	$x + 15/100x = 207$ $100x + 15x/100 = 207$ $115x = 207 \cdot 100$ $x = 20700/115$ $x = 180$	comment	N/A	success
101	Como aclaración hay que tener en cuenta que dos cantidades pueden compararse de dos maneras: Hallan...	comment	N/A	success
114	2- Medidas de superficie Sirven para medir superficies cuadradas, es decir, en dos dimensiones: larg...	comment	N/A	success
119	Se nota que hay triangulos , esferas y cuadrados: el cuadrado gira en diagonal y vuelve a su posició...	comment	neutral	success
124	eso esta mal por que la respuesta en las dos ultimas opciones es la misma disacárido de glucosa tien...	comment	N/A	success



Content Moderator		
Total Entidades	Categorías	
5	benigna, no bravo; f	1 Quantity
6	a multiplicar con un	10 Quantity
7		3 Person
8		24 Skill
9		26 Skill
10		3 Quantity
11		1 Quantity
12		8 Quantity
13		1 Quantity
14		3 PersonType
15		1 DateTime
16		2 Product
17		4 Quantity
18		2 Product
19		8 Skill
20		3 Quantity
21		4 Quantity
22		5 Quantity
23		1 Quantity
24		5 DateTime
25		5 Quantity

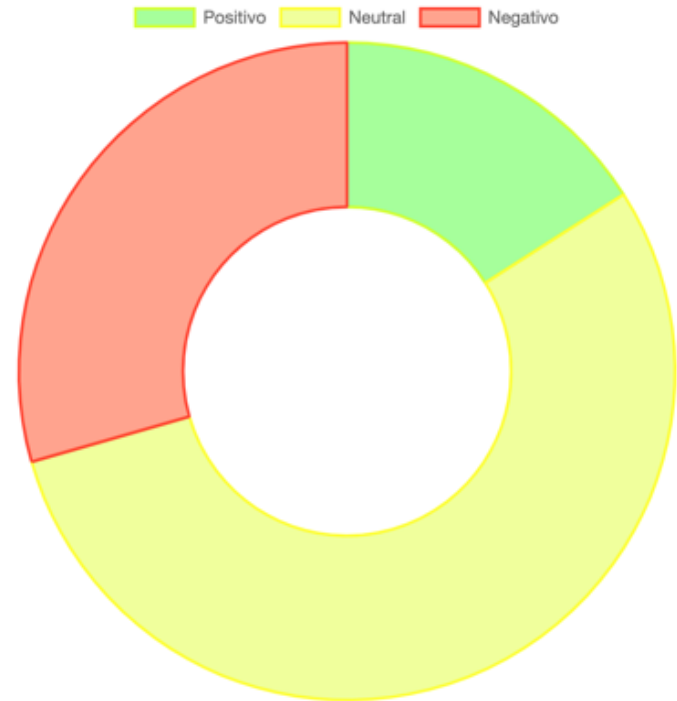


Sprint 3

Cantidad de textos por sentimientos

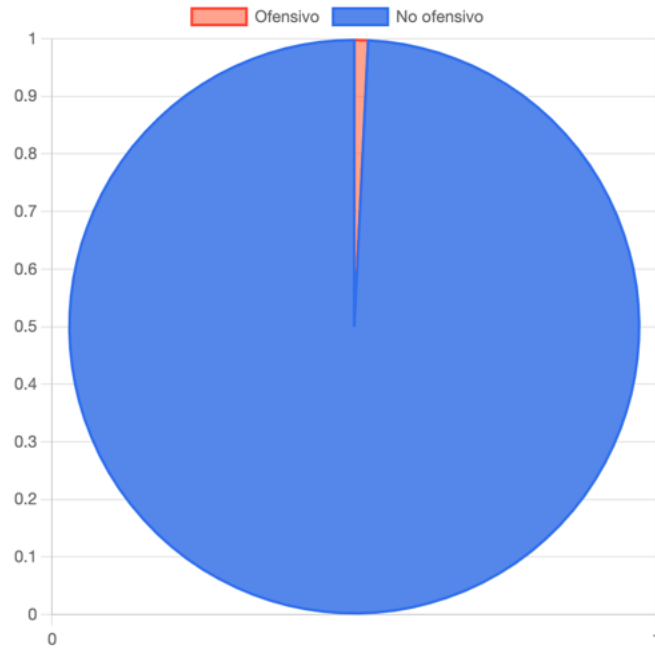


Promedio percentual de sentimientos



Sprint 4

Porcentaje contenido ofensivo



Palabras ofensivas repetidas



Entidades más repetidas



Sprint 4

Entidades identificadas en Jovenesweb



Buscar

Texto	Total Entidades	Categorías	Acción
<p>$A = b \cdot h$ $6x^2 + 22x + 20 = (3x+5) \cdot (???)$ Multiplicamos solo los términos independientes $20 = 5 \cdot ?$ $20 = 5 \cdot 4$ el +4 es el único que cumple que respuestas tiene +4?? la B) y la C). Multiplicamos con la base y tiene que dar el área Se aplica la propiedad distributiva $\downarrow B) (3x+5)(3x+4) = 9x^2 + 12x + 15x + 20 = 9x^2 + 27x + 20$ (No es el área) C) $(3x+5)(2x+4) = 6x^2 + 12x + 10x + 20 = 6x^2 + 22x + 20$ (SI ES EL ÁREA) Respuesta C)</p>	4	<ul style="list-style-type: none">QuantityQuantitySkillQuantity	Ver
<p>$45 \operatorname{Sen} \theta = 45\sqrt{3} \cdot \operatorname{Sen} 30^\circ$ (El 45 pasa a dividir del otro lado, así que eso igual se lo elimina) $\operatorname{Sen} \theta = \sqrt{3} \cdot 1/2 \leftarrow$ (Esto sale del valor de $\operatorname{Sen} 30^\circ$) $\operatorname{Sen} \theta = \sqrt{3}/2 \theta = 60^\circ$ Ya que al ubicarnos en la tabla de ángulos, el $\sqrt{3}/2$ se encuentra en el $\operatorname{Sen} 60^\circ$)</p>	7	<ul style="list-style-type: none">QuantityQuantityQuantityQuantity...	Ver
<p>$(C^3 \cdot C^2)^2$ La que es con la misma base que es "C" La regla es: cuando tiene la misma base se suman los exponentes $3+2=5$ Y el 5 lo multiplicamos por el 2 del exponente de afuera que viene hacer 10 La respuesta es C^{10}</p>	5	<ul style="list-style-type: none">QuantityQuantityDateTimeQuantity...	Ver



Validación de la propuesta



Obtención de los datos

Los datos obtenidos para el análisis estadístico son de fuentes primarias, en este caso provienen directamente del análisis de contenido del plataforma web educativa Jovenesweb, los mismos que no han sido analizados

Se debe considerar que por políticas de la empresa Jovenesweb se tuvo un acceso a una muestra de mil textos generados por el usuario de la plataforma Jovenesweb: 500 soluciones y 500 comentarios



Obtención de los datos

Frecuencia esperada y frecuencia observada del indicador 1

	Frecuencia esperada Jovenesweb	Frecuencia observada Content Moderator
Positivo	16,28	15,96
Negativo	33,24	29,45
Neutral	50,48	54,62

Frecuencia esperada y frecuencia observada del indicador 2

	Frecuencia esperada Jovenesweb	Frecuencia observada Content Moderator
% Ofensivo	9,4	0,77
% No ofensivo	90,6	99,23

Frecuencia esperada y frecuencia observada del indicador 3

	Frecuencia esperada Jovenesweb	Frecuencia observada Content Moderator
Tiempo análisis sentimientos	810,24	0,0319444
Tiempo análisis ofensivo	810,24	0,0333333
Tiempo análisis entidades	810,24	0,0655556



Obtención de los datos

Frecuencia esperada y frecuencia observada del indicador 4

	Frecuencia esperada Jovenesweb	Frecuencia observada Content Moderator
Porcentaje de contenido analizado de manera exitosa	90	90
Porcentaje de error de contenido analizado	10	10

Frecuencia esperada y frecuencia observada del indicador 5

	Frecuencia esperada Jovenesweb	Frecuencia observada Content Moderator
Cantidad Categoría	2817	3626
Cantidad Información personal	356	527
Cantidad enlaces	597	692

Frecuencia esperada y frecuencia observada del indicador 6

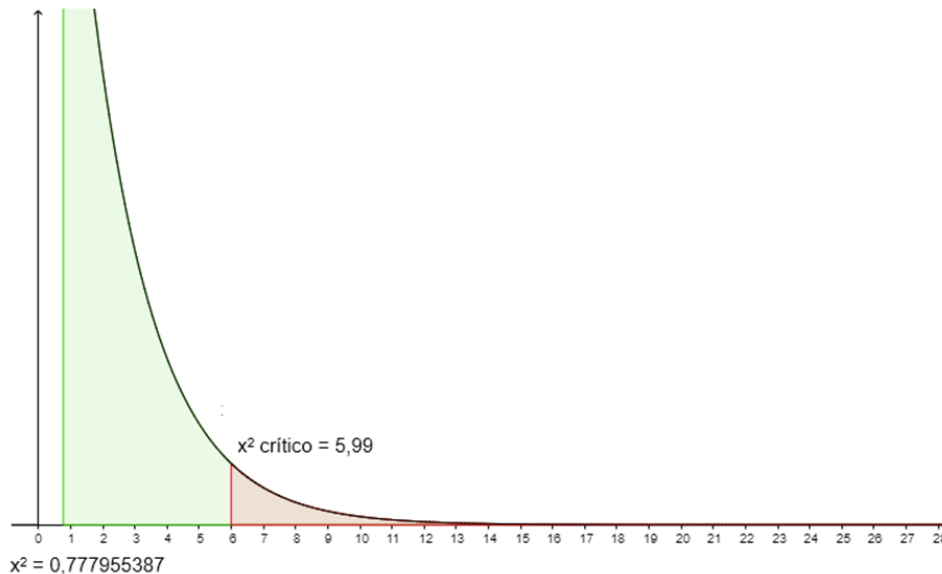
	Frecuencia esperada Jovenesweb	Frecuencia observada Content Moderator
Positivo	132	120
Negativo	219	246
Neutral	564	549



Validación para el indicador 1

Formulación de hipótesis para el indicador: porcentaje de sentimientos positivos, negativos y neutrales de los textos analizados

- **H0:** El aplicativo web se ajusta a la distribución teórica del porcentaje de contenido analizado en los sentimientos: positivos, negativos y neutrales
- **H1:** El aplicativo web no se ajusta a la distribución teórica de porcentaje de contenido analizado en los sentimientos: positivos, negativos y neutrales

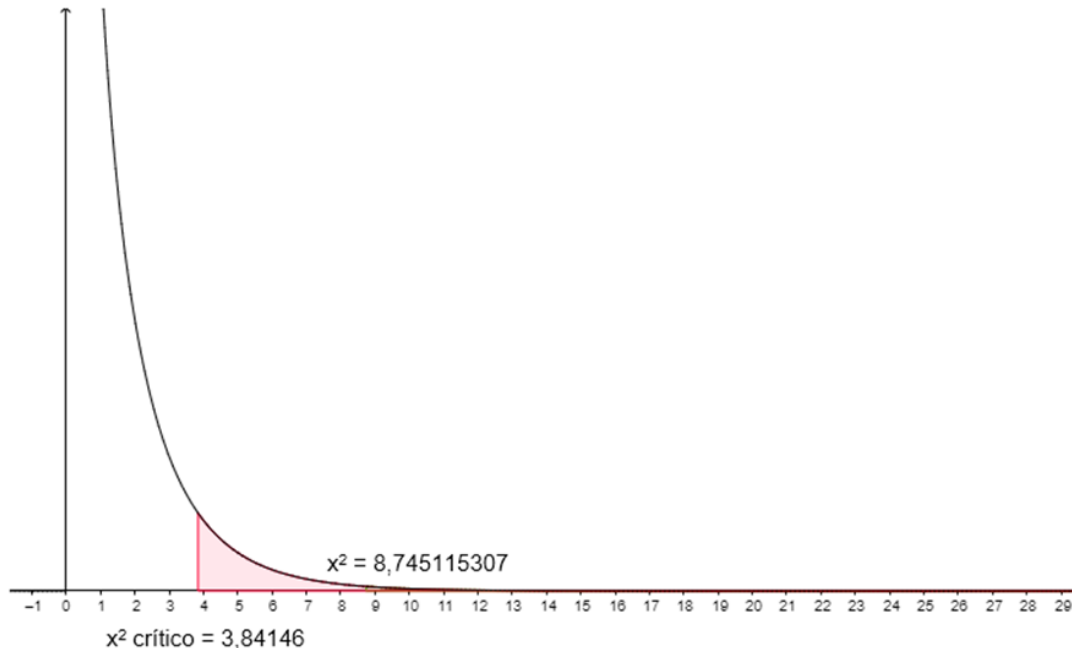


Se acepta la hipótesis nula (H0) al ser el valor chi-cuadrado calculado menor al valor crítico, se rechaza la hipótesis alternativa (H1).

Validación para el indicador 2

Formulación de hipótesis para el indicador: porcentaje de contenido ofensivo detectado

- **H0:** El aplicativo web se ajusta a la distribución teórica del porcentaje de contenido ofensivo detectado
- **H1:** El aplicativo web no se ajusta a la distribución teórica del porcentaje de contenido ofensivo detectado



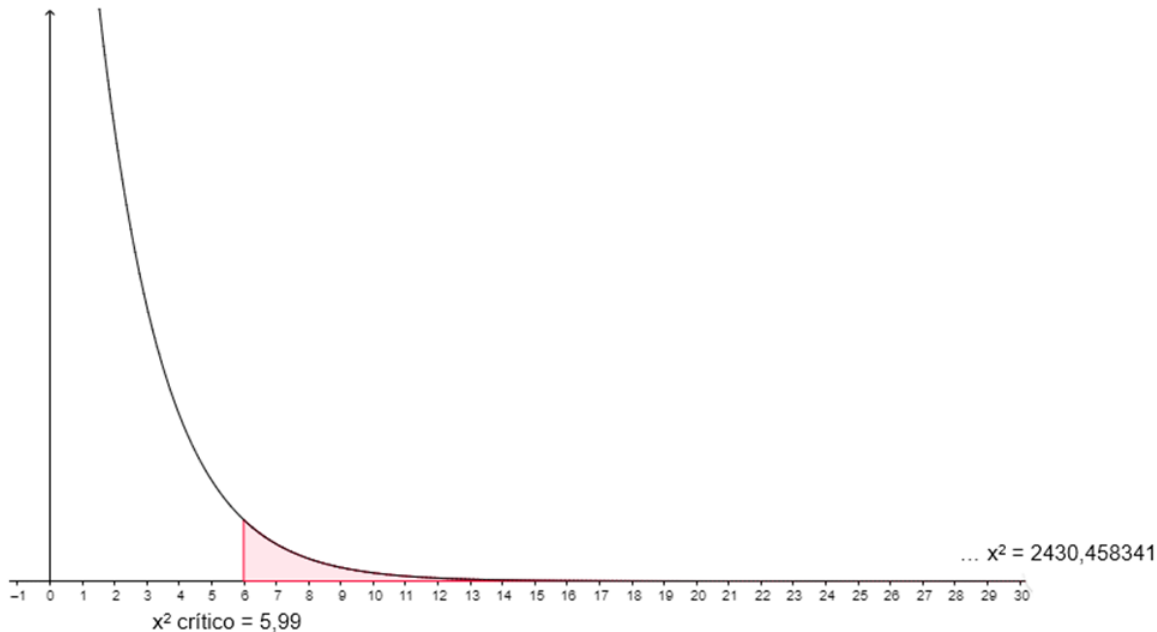
Se rechaza la hipótesis nula (H0) al ser el valor chi-cuadrado calculado mayor al valor crítico, se acepta la hipótesis alternativa (H1).



Validación para el indicador 3

Formulación de hipótesis para el indicador: tiempo en el análisis de contenido

- **H0:** El tiempo de análisis del aplicativo web es mayor a al tiempo de análisis de Jovenesweb
- **H1:** El tiempo de análisis del aplicativo web es menor a al tiempo de análisis de Jovenesweb



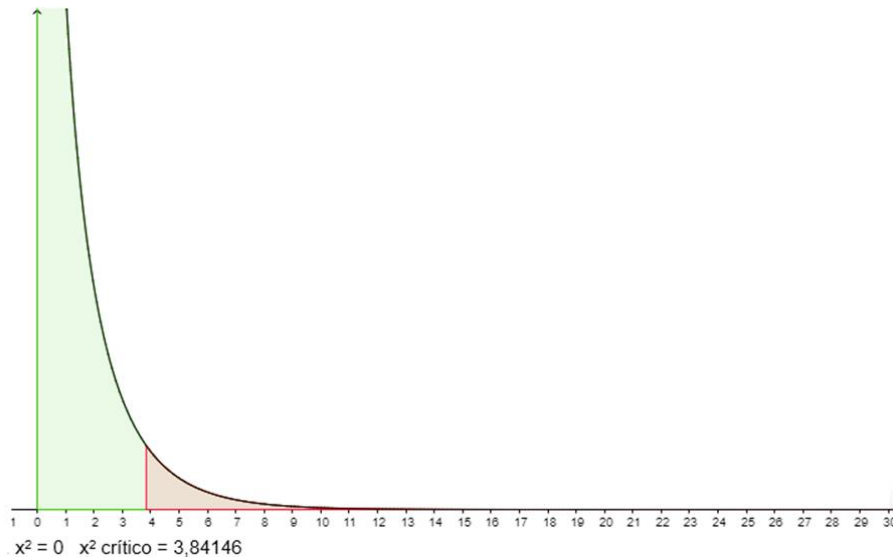
Se rechaza la hipótesis nula (H0) al ser el valor chi-cuadrado calculado mayor al valor crítico, se acepta la hipótesis alternativa (H1).



Validación para el indicador 4

Formulación de hipótesis para el indicador: porcentaje del contenido analizado

- **H0:** El aplicativo web se ajusta a la distribución teórica en el porcentaje de contenido analizado.
- **H1:** El aplicativo web no se ajusta a la distribución teórica en el porcentaje de contenido analizado



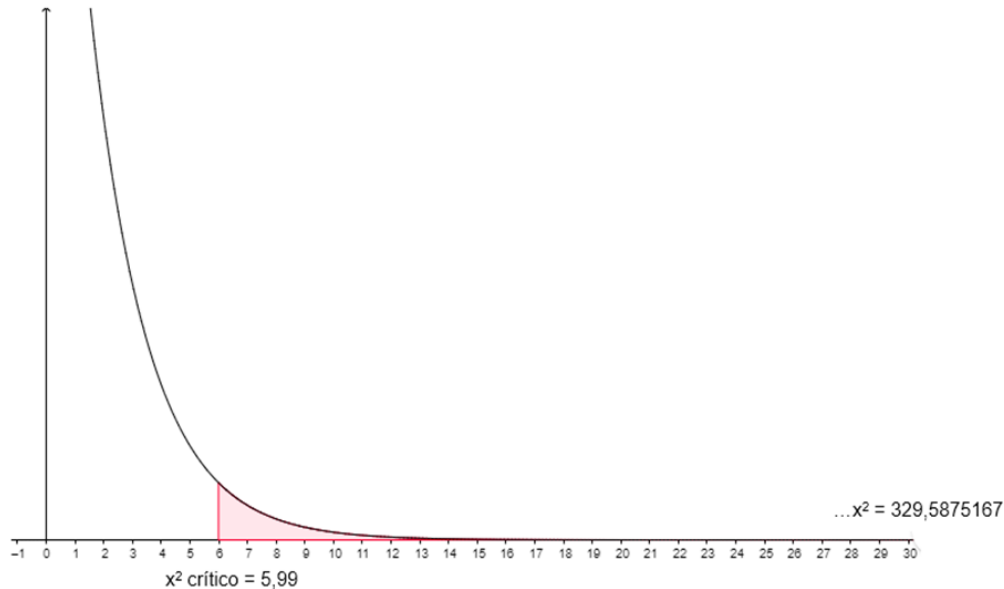
Se acepta la hipótesis nula (H0) al ser el valor chi-cuadrado calculado menor al valor crítico, se rechaza la hipótesis alternativa (H1).



Validación para el indicador 5

Formulación de hipótesis para el indicador: cantidad de entidades detectadas y clasificadas

- **H0:** El aplicativo web se ajusta a la distribución teórica en las cantidades de entidades detectadas y las clasifica en: categoría, información personal, cantidad de enlaces
- **H1:** El aplicativo web no permite mejorar las cantidades de entidades detectadas y las clasifica en: categoría, información personal, cantidad de enlaces



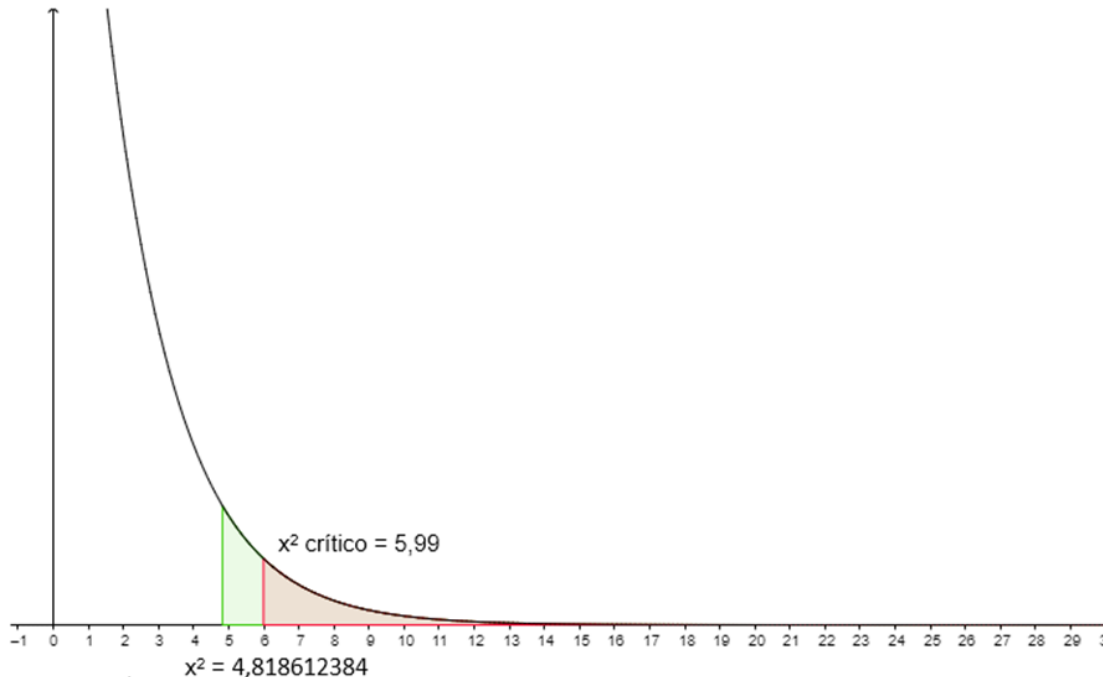
Se rechaza la hipótesis nula (H0) al ser el valor chi-cuadrado calculado mayor al valor crítico, se acepta la hipótesis alternativa (H1).



Validación para el indicador 6

Formulación de hipótesis para el indicador: Cantidad de contenido con sentimientos identificados y clasificado (positivo, negativo y neutro)

- **H0:** El aplicativo web se ajusta a la distribución teórica en las cantidades de sentimientos detectadas y los clasifica en: positivo, negativo y neutral
- **H1:** El uso del aplicativo web no permite mejorar las cantidades de sentimientos detectadas y los clasifica en: positivo, negativo y neutral



Se acepta la hipótesis nula (H0) al ser el valor chi-cuadrado calculado menor al valor crítico, se rechaza la hipótesis alternativa (H1).

Conclusiones y Recomendaciones



Conclusiones

- Se cumplió con el objetivo de desarrollar un aplicativo web que permita optimizar el análisis de contenido publicado por los usuarios de la plataforma web educativa Jovenesweb con el uso de servicios cognitivos de Inteligencia artificial de Microsoft Azure
- La elaboración del marco teórico permitió el entendimiento de los conceptos fundamentales para el desarrollo de un aplicativo web que permita implementar los servicios cognitivos de Inteligencia Artificial ofrecidos por Microsoft Azure para realizar el reconocimiento de: entidades, sentimientos y contenido ofensivo a fin de automatizar el moderamiento de contenido en la plataforma Jovenesweb

Conclusiones

- A través de la integración del aplicativo web con los servicios cognitivos de Inteligencia Artificial de Microsoft Azure optimizó el análisis del contenido generado por los usuarios de Jovenesweb.
- El desarrollo del aplicativo web por medio del marco de trabajo Scrum permitió obtener un dinamismo de trabajo sostenible, priorizando las funcionalidades por su valor agregado, estimando el esfuerzo requerido en cada una de ellas y respondiendo con agilidad a los cambios.

Conclusiones

- Para el desarrollo del aplicativo web se utilizó el framework Laravel 8 de PHP en conjunto con el servicio de Laravel Forge, el cual permite desplegar y configurar el aplicativo en un servidor de DigitalOcean para que sea accesible a través de internet
- Se implementó de manera satisfactoria el aplicativo web en conjunto con el equipo de desarrollo de Jovenesweb y en base al cronograma de entregas establecido en la etapa del inicio del proyecto

Conclusiones

- Para la validación de los indicadores establecidos se utilizó la técnica estadística del Chi cuadrado por medio de bondad de ajuste con un nivel de significancia del 5% lo que comprobó la hipótesis.
- Se debe tener en cuenta que el análisis de contenido usando los servicios cognitivos de Inteligencia Artificial de Microsoft Azure está sujeto a limitaciones que podrían provocar resultados no deseados como; el mal uso de las reglas gramaticales el cual distorsiona el contexto de la oración y el uso de abreviaturas que dificulta el reconocimiento de contenido ofensivo e identidades

Recomendaciones

- Se recomienda trabajar con la versión de pago de los servicios cognitivos de Inteligencia Artificial de Microsoft Azure, ya que en la versión gratuita existen limitaciones sobre la cantidad de textos a analizar al mes y limitaciones en la cantidad de textos a analizar por minuto los cuales podrían aumentar el tiempo de análisis.
- Se recomienda el uso del marco de trabajo SCRUM para facilitar el desarrollo del proyecto dada su flexibilidad y adaptación.

Recomendaciones

- Se recomienda antes de analizar el contenido validar que las reglas gramaticales sean las correctas y que no existen abreviaturas en palabras para evitar resultados inesperados en el análisis.
- Se recomienda normalizar los textos antes de ser analizados, eliminando código HTML, letras duplicadas, espacios en blanco duplicados para evitar resultados inesperados en el análisis de entidades, sentimientos o contenido ofensivo mediante la aplicación de expresiones regulares y funciones de texto propias del lenguaje de programación.

Recomendaciones

- Se sugiere leer la documentación de los servicios cognitivos de Inteligencia Artificial de Microsoft Azure para poder comprender el funcionamiento de la API y sus limitaciones.
- Para futuras investigaciones con temáticas similares se recomienda integrar el análisis de moderación de contenido para imágenes cuya integración es similar a la moderación de contenido ofensivo en textos y que permitirá evaluar la eficacia del análisis.

Gracias por su atención



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA