

Daniel Esteban Aguilera Figueroa – 202010592
Laura Vanessa Martínez Prieto – 202012624
Cristian Armando Sánchez Ocampo – 202022112

Informe: Proyecto 1 – Etapa 2

1. Introducción

La Analítica de Textos (AT) es un campo de estudio que combina el aprendizaje automático y el procesamiento de lenguaje natural para procesar grandes cantidades de textos y extraer conocimiento útil para la toma de decisiones. Este proyecto en específico se enfoca en fortalecer las competencias necesarias para una aplicación efectiva del proceso de descubrimiento de conocimiento a partir de textos, utilizando una metodología propia del mundo de la analítica. Una de las opciones de problemáticas a abordar en este proyecto es el análisis de sentimientos de películas en español, se tienen comentarios de películas que deben ser clasificados en las categorías de positivo o negativo.

Durante esta nueva etapa se propondrá una aplicación que permita el desarrollo del modelo analítico generado en la etapa anterior, aplicable a la resolución de problemas específicos en organizaciones, empresas o instituciones enfocadas al cine con el fin de ofrecer una perspectiva basada en el sentimiento general obtenido con los resultados del modelo ya mencionado.

2. Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API

Para el proceso de construcción del modelo y automatización del proceso analítico, se optó por la realización de un Pipeline, que en conjunto con el modelo que presentó mejores resultados en la primer parte del proyecto junto con el vectorizador usado fue posible generar este. Como fue visto en la primera parte, el modelo que mejores resultados presentó fue la red neuronal MLPClassifier con el vectorizer CountVectorizer.

Adicionalmente, para la construcción del pipeline ya mencionado, se utilizó la librería Joblib para lograr generar el pipeline y exportar el modelo desarrollado en la primera fase del proyecto junto con el vectorizer utilizado y anteriormente entrenado. Junto a este, se creó una clase TextProcessor, la cual se encarga de la preparación de los datos como fue estipulado en la primer parte.

```
vectorizer = joblib.load('backend/static/logic/datos/CountVectorizer.pkl')

pipeline = Pipeline([
    ('preprocessor', TextProcessor()),
    ('vectorizer', vectorizer),
    ('classifier', MLPClassifier(verbose=True, early_stopping=True, random_state=42))
])
```

Figura 1. Pipeline construido usando el modelo de la primera parte

Ambos elementos fueron importados en el API, para luego ser usados para clasificar los sentimientos de los comentarios en positivos y negativos. También se utilizó la herramienta FastAPI para desarrollar el backend de la aplicación, que consta de diferentes funciones que permiten enviar y recibir datos entregados por diferentes usuarios.

3. Desarrollo de la aplicación y justificación

La aplicación que se desarrolló podría llegar a ser una herramienta muy útil para personas que trabajan en el mundo del cine, como directores, actores y productoras, o llegado al caso también para espectadores. La aplicación les permitirá analizar comentarios y/o reseñas de espectadores en diferentes plataformas para evaluar el éxito de sus producciones y hacer ajustes necesarios. Esto es importante porque les ayuda a tomar decisiones informadas en un mercado competitivo y a entender mejor las necesidades y expectativas del público, lo que puede resultar en producciones más exitosas y rentables.

Para desarrollar esta aplicación se utilizó FastAPI para el backend y Angular para el frontend. FastAPI se eligió debido a que es un marco de desarrollo web moderno y de alto rendimiento, diseñado para crear APIs RESTful capaces de procesar grandes volúmenes de datos de manera eficiente. Por otro lado, Angular es una herramienta ideal para crear interfaces de usuario atractivas y fáciles de usar. La combinación de ambos frameworks permite ofrecer una experiencia de usuario fluida e intuitiva, lo que hace que la aplicación sea muy fácil de usar. La aplicación cuenta con una sección para subir un archivo Excel y una ventana emergente que proporciona al usuario una descripción de los resultados obtenidos por el modelo.



Figura 2. Inicio base de la aplicación realizada

4. Resultados

Como se mencionó anteriormente, la idea es que el usuario ingrese una recopilación de comentarios de la película que desea evaluar. Como resultado, la aplicación mostrará información de los datos obtenidos. En primera instancia, se observa una pequeña muestra de los comentarios y su respectivo sentimiento según el modelo que utilizamos con el fin de tener retrospectiva del proyecto que

se está analizando.

Reseña	Sentimiento
"La saga medieval alemana de Fritz Lang continúa en Die Nibelungen: Kriemhilds Rache (1924) .kriemhild (Margarete Schoen) quiere vengar a su esposo asesinado, Siegfried. Hermano es demasiado débil para traer a los asesinados, Hagen, a Justice.Kriemhild se casa con Etzel, elRey de los hunos (Rudolf Klein-Rogge). Ella da a luz a un niño, e invita a sus hermanos para una fiesta. Tzel y los otros hunos deberían Kil Hagen, pero está protegido por sus hermanos. Vemos una gran batalla de la vida.Y la muerte comienza, y ella pone a todo el lugar en Fire. Incluso Hagen está muerto, y también lo es Kriemhild. Estos películas tratan con grandes temas, como la venganza y el amor eterno. Más tarde, hemos visto algunas mejores películas hechas de esos temas,Pero este fue uno de los primeros."	negative
"Este anime es una visita obligada para los fanáticos de Evangelion. Es un trabajo anterior de Anno Hideaki, pero su estilo sin restricciones, dramático está bastante en su lugar. Además, aquellos que no les gustaron Evangelion pueden encontrar que esta liberación sea un poco más sabrosa. Gunbuster es bastante exclusivo de Sci-Fi anime, ya que en realidad se basa en la ciencia real. De hecho, el espectáculo tiene varias interludios de "lección de ciencia" pequeña que explican la física detrás de algunos de los eventos en la película. Unió de los grandes puntos dramáticos de la película es el paso relativo del tiempo a velocidades cerca de la de la luz. La serie hace un trabajo maravilloso para enfrentar la experiencia imaginablemente traumática de dejar la Tierra en una misión de seis meses que viaja cerca de la velocidad de la luz y regresando a una tierra donde han pasado diez años. El personaje principal sigue siendo 17 o 18 años en toda la serie, mientras que casi todos los otros personajes, edad, considerablemente. Tenga cuidado, este espectáculo es pesado en la savia a veces. También tiene un par de los disparos de pechos más totalmente desahollados que he visto. Me pareció bastante fácil ignorar los uniformes escasos y las escenas de Boo-Hoo, porque la serie es muy buena, pero los espectadores con una tolerancia de baja savia pueden querer mantenerse alejados de este. En una nota interesante, GAINAX, como siempre, logró quedarse sin dinero en la última pareja de episodios. Sin embargo, lograron usar una película en blanco y negro y bocetos de acción para producir una buena resolución de todos modos. El final es un poco tonto, pero me dejó con un buen sentimiento en mi intestino, no pude evitar amarlo. Gunbuster es, en mi opinión, una de las mejores piezas de anime."	positive
"Esta es una de las mejores películas para mascotas para niños.Lloro cada vez que veo la sombra gritando "¡¡espera, espere a Peter!!"A medida que el coche familiar se está alejando,¡Esto es una visita obligada si amas a los animales!La mejor película de la historia!Las líneas en la película son a veces estúpidas.Como cuando Sassy dice al azar,"¡Regla de gato y perros bebé!"Líneas como esta pude hacer sin, pero cuando tenía seis años, me encantó esa línea.La historia puede parecer gancho para algunos, pero me gusta.Shadow como el perro mayor que está preparando la oportunidad de hacerse cargo de él cuando se ha ido realmente se está moviendo cuando lo piensas.Me recordó a mi perro infantil.Creo que todos pueden encontrar una pieza de sí mismos en "atajo hacia el hogar"."	positive
"Cuando se lanzó su DVD, llegué al mercado y lo compré. Y creo que mi dinero estaba de manera correcta, como lo esperaba antes de comprarlo. Movie impresionante ¿Qué más puedo decir para Will Smith, ha sido un actor increíble como siempre, ya sea en las acciones de las acciones o en serio? Siempre le da un rendimiento de frenado récord. Creo que esta es la película después de apresurarse de agosto, lo que hace llorar a una persona mientras lo miraba. La forma en que el director describió la historia fue realmente increíble. Su vida anterior y su nueva vida en la película se elaboraron correctamente a la audiencia. Incluso no pude encontrar ninguna falla en la historia o en la forma en que se disparan. Creo que su DVD debe ser un hogar porque esto será realmente bueno para su colección. No es la película que necesita Pop-Corns para disfrutar, esta es la película que deja que la audiencia aprenda una lección. Ahora, ¿cuál es la lección puedes ver eso mientras miraba? Y aconsejo a aquellas personas que son críticos de la película, vea esto si pudieras encontrar cualquier crítica sobre esta película, por favor, háblame."	positive
"No me he reído tan duro en una película en mucho tiempo.Tengo que ir a una proyección anticipada, y estaba emocionada porque me había estado muriendo por verlo.Tenía lágrimas en mis ojos de la risa a lo largo de una gran parte de la película.Toda la audiencia compartió mi risa, y estaba aplaudiendo y gritando a lo largo de la mayor parte de la película. Kudos a Steve Carrell (a quien ya había sido fanático).Él demuestra en esta película su tremendo talento para la comedia.Él tiene un estilo que no he visto antes.Y Catherine Keener es excelente como siempre.Gracias a Dios allí, no había un cameo de Ferrell (Love Your, pero lo vi demasiado en este verano). Había partes del genio cómico en esta película.En parte, gracias a Carrell, y en parte gracias a la escritura (también Carrell).La escena de depilación y la velocidad del dater con el "problema obvio" fueron absolutamente históricamente. Sin duda, iré a ver '40 años de edad virgen 'cuando se libera.Mi consejo: ve a verlo para enormes risas y una película increíblemente agradable encima de ella."	positive

Figura 3. Display de los resultados del modelo parte 1.

Por otro lado, hay una opción de botón llamada “mostrar resultado” que ilustrará el porcentaje de opiniones positivas y negativas de la película en un diagrama de pastel y un pequeño mensaje con un análisis muy general dependiendo del resultado obtenido, con el fin de proveer información del rendimiento general de la película.

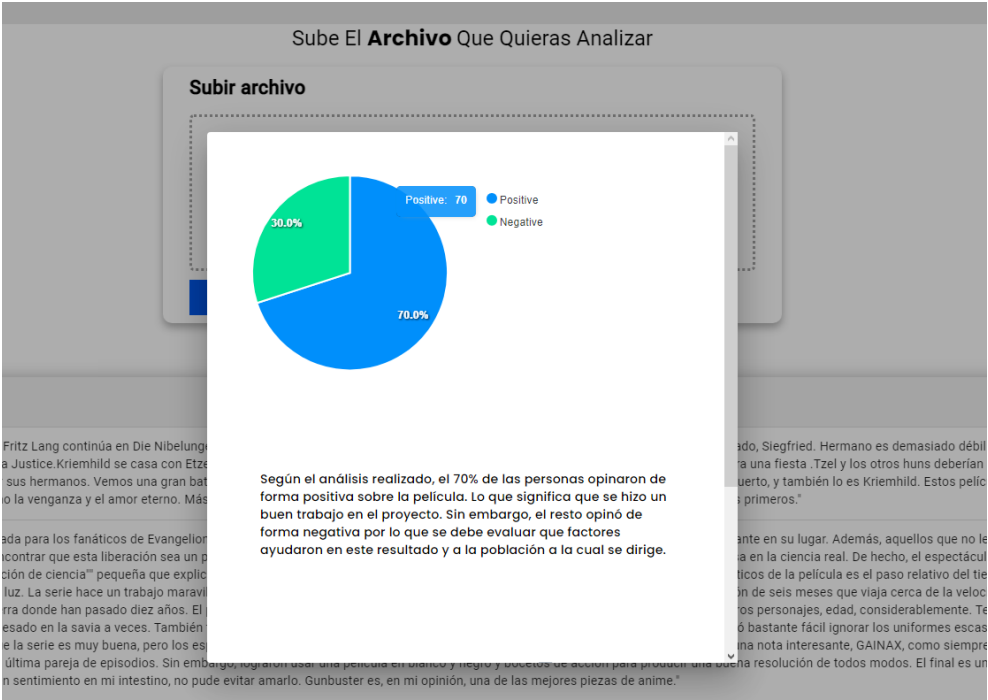


Figura 4. Display de los resultados del modelo parte 2.

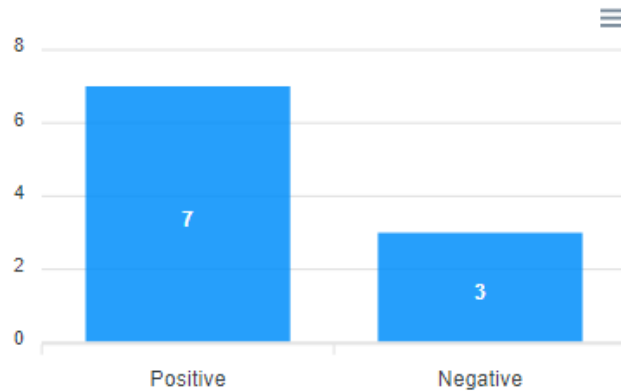


Figura 5. Display de los resultados del modelo parte 3.

En el **video** proporcionado se observará con más detalle las funcionalidades de la aplicación.

Adicionalmente, la aplicación se encuentra desplegada en la nube con ayuda de máquinas virtuales en Google Cloud Platform, en estas se encuentran desplegados ambos servicios, tanto back como frontend.

<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	back-2	us-central1-a	SSH	▼	⋮
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	front	us-central1-a	SSH	▼	⋮

Figura 6. MV en GCP

5. Contribución equipo de estadística

En cuanto a la contribución del grupo de estadística podemos destacar los siguientes aspectos. Cito textualmente la conversación realizada con Amy Nicole Ramírez Rodríguez. *“Pues estaba pensando para que el análisis no sea tan plano, podrían preguntar cosas más específicas como por ejemplo a qué hora ven las películas, dependiendo de lo que digan pueden hacer análisis de autocorrelación o diagramas comportamentales para ver si las películas se ven más en cierto horario o si lo hacen por edad también pueden sacar análisis y ahí si hacer la parte estadística, ¡porque si se los piden si es importante que lo puedan sacar de alguna parte!”*

El grupo nos plantea la importancia de realizar un análisis más detallado y específico en relación con la visualización de películas, con el fin de obtener información valiosa que pueda ayudar a comprender mejor a la audiencia y mejorar la calidad de los productos audiovisuales. Aunque la idea propuesta pueda ser compleja de realizar, se enfatiza en la necesidad de obtener datos

estadísticos más precisos y completos para realizar una evaluación más profunda.

Se destaca la importancia de realizar preguntas específicas, como conocer la hora del día en que se ven las películas, con el fin de hacer análisis de autocorrelación y diagramas comportamentales que permitan determinar si existe un patrón en la audiencia, en función de la edad o de otros factores relevantes. Se sugiere también la posibilidad de filtrar los datos por grupos de edad y aplicar herramientas estadísticas como diagramas de caja y bigote, histogramas y líneas de tendencia para obtener información más precisa.

En resumen, Amy hace énfasis en la importancia de realizar un análisis más profundo y detallado en relación con la visualización de películas, con el fin de obtener información valiosa que permita comprender mejor a la audiencia y mejorar la calidad de los productos audiovisuales. Aunque, teniendo en cuenta todos los aspectos, esto es algo que es complicado de realizar debido a las limitaciones con las que contamos.

Por último, discutimos la población objetivo de nuestra aplicación con ella y llegamos a la conclusión que esta está bien definida dado que es crucial un análisis de este estilo para el desarrollo de una película o su resultado general.

6. Trabajo en equipo

a. Roles desempeñados en el equipo

Rol	Nombre	Código	Usuario Github
Líder de proyecto	Daniel Aguilera	202010592	Daniagui12
Ingeniero de datos	Laura Martinez	202012624	lmartinezp2003
Ingeniero de software responsable del diseño de la aplicación y resultados	Cristian Sánchez	202022112	Panis26
Ingeniero de software responsable de desarrollar la aplicación final	Daniel Aguilera	202010592	Daniagui12

b. Distribución del trabajo

Tema	Nombre
Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API	Todos los integrantes
Desarrollo de la aplicación y justificación	Todos los integrantes
Resultados	Todos los integrantes

c. Coevaluación

Los 100 puntos para repartir serían repartidos de tal forma que los 3

estudiantes cuenten con la misma cantidad de puntos. Esto debido a que la cantidad de trabajo y la calidad de este fue relativamente similar, mostrando que todos trabajamos por igual y presentamos un buen trabajo de forma consistente. Esto diciendo que cada uno tendría alrededor de 33.3 puntos.

Puntos por mejorar:

- Comunicación: Es necesario para la siguiente etapa tener una comunicación mas efectiva para así evitar confusiones y que nuestra calidad de trabajo sea mucho mejor
- Establecer objetivos: Es crucial mejorar el establecimiento de objetivos para trabajar de forma constante y que pueda haber revisiones más seguidas.
- Tiempo de trabajo: Es importante iniciar a trabajar con mayor anterioridad para así evitar días de trabajo pocos días antes de la entrega.
- Realizar reuniones grupales: Esto con el fin de hacer que todos los integrantes del grupo estén al tanto de las decisiones tomadas en el proyecto. Por otro lado, ayudaría a presentar objetivos grupales para mayor coordinación.

Referencias

- <https://angular.io/docs>
- <https://fastapi.tiangolo.com>
- https://scikit-learn.org/stable/user_guide.html