

SentiMovie



Apellido: Suárez

Habilitación Profesional 2023

Índice

1 Introducción	3
1.1 Planteo del problema	3
1.2 Propósito	3
1.3 Alcance	3
1.4 Definiciones, acrónimos y abreviaturas	3
2 Descripción	4
2.2 Requisitos funcionales	4
2.1.2 Interfaces de software	4
2.3 Actores	4
3 Detalles funcionales	4
3.1 Generar reporte de reseñas	4
3.2 Buscar película	5
4 Proceso de negocio	5

1 Introducción

1.1 Planteo del problema

En la era digital actual, las reseñas de películas han adquirido un papel fundamental en la toma de decisiones de los cinéfilos. Sin embargo, la abrumadora cantidad de opiniones disponibles en internet puede resultar desafiante para aquellos que buscan tomar decisiones informadas.

Es en este contexto que surge este sistema para evaluar y determinar la polaridad de las reseñas de películas. Este sistema tiene como objetivo principal ayudar a los amantes del cine a identificar de manera rápida y precisa si una reseña es positiva o negativa, brindando una valiosa herramienta para la toma de decisiones informadas.

1.2 Propósito

El propósito del sistema de información es desarrollar un modelo de análisis de sentimiento de las reseñas de películas, aplicando técnicas de procesamiento de lenguaje natural y aprendizaje automático, con el objetivo de determinar la polaridad (positiva o negativa) de dichas reseñas presentando dichos resultados en gráficos.

1.3 Alcance

El sistema está destinado a:

- consumidores habituales de cine,
- aficionados a las películas,
- críticos de cine,
- empresas de distribución y producción de películas,
- plataformas de streaming.

El sistema brinda a estos usuarios un reporte del análisis de los sentimientos de las reseñas con el fin de tomar decisiones informadas en lo que se refiere a la selección, promoción y comercialización de películas.

El sistema no recomendará el consumo de una película.

1.4 Definiciones, acrónimos y abreviaturas

Lenguaje natural: Es el lenguaje que usamos los seres humanos para comunicarnos entre nosotros de forma cotidiana, ya sea hablado o escrito.

Aprendizaje automático (Machine Learning): Se refiere a un enfoque de la inteligencia artificial que permite a las máquinas aprender y mejorar automáticamente a partir de los datos sin ser programadas explícitamente para cada tarea específica.

Polaridad: En el contexto del análisis de sentimientos, se refiere a la orientación o el tono de una determinada crítica. Indica la dirección del sentimiento expresado, ya sea positivo o negativo.

Extracción de características (Feature Extraction): Se refiere al proceso de seleccionar y transformar datos no estructurados, como el texto de las reseñas, en un formato que un algoritmo de aprendizaje automático puede utilizar. Esto implica convertir el texto en características vectoriales, lo que permite que el modelo de análisis de sentimientos lo interprete.

Clasificación de crítica: La clasificación de crítica es el proceso de asignar una etiqueta o categoría a una crítica. En este caso, se trata de asignar la polaridad (positiva o negativa) a las críticas de películas. El objetivo es entrenar un modelo de aprendizaje automático que pueda clasificar automáticamente nuevas críticas en una de estas categorías de acuerdo con el tono o sentimiento expresado en el texto.

2 Descripción

2.2 Requisitos funcionales

- RF01 Mostrar el resultado del análisis de críticas presentando un indicador, resumen o gráfico para que el usuario pueda tomar decisiones acerca del consumo o recepción sobre una película.
- RF02 Utilizar herramientas gráficas para mostrar a los usuarios los resultados de los análisis realizados.
- RF03 Permitir ingresar cualquier película para que luego el sistema realice el análisis correspondiente.

2.1.2 Interfaces de software

- Se utiliza la API de themoviedb.org para obtener las reseñas de la película deseada.

2.3 Actores

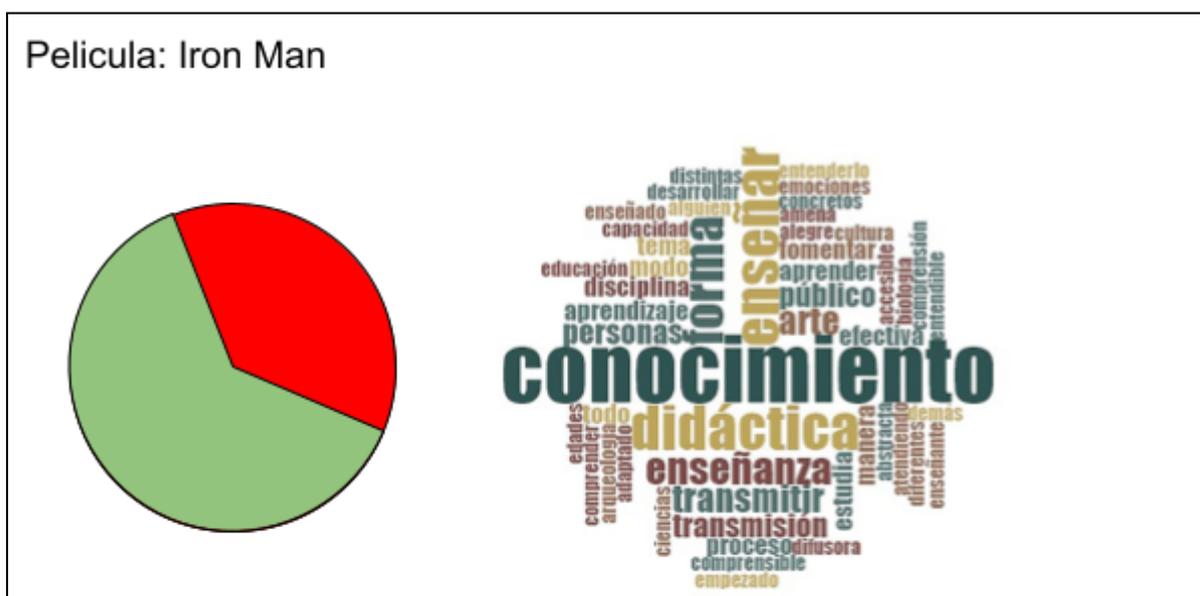
- **Cinéfilo:** Será quien solicite el reporte gráfico acerca de las reseñas introduciendo la película deseada.
- **Crawler:** Quien se encarga de consumir los datos para realizar el análisis de los sentimientos.

3 Detalles funcionales

3.1 Generar reporte de reseñas

Descripción Breve

El sistema va a generar un reporte que cuente con información resumida acerca de las reseñas de una película. Este reporte tendrá información tal como un resumen de los sentimientos en las reseñas, junto a las palabras más utilizadas en las reseñas para describir la película.



Funcionamiento

- El usuario ingresa el título de una película.
- El sistema realizará el análisis correspondiente a dicha película

- El sistema generará un reporte indicando la información mencionada previamente
- El sistema muestra en pantalla el reporte generado

3.2 Buscar película

Descripción Breve

El sistema debe permitir al usuario ingresar el título de una película para su correspondiente análisis.

Funcionamiento

- El usuario ingresa el título de la película en el cuadro correspondiente.
- El sistema lista películas relacionadas con el título colocado por el usuario.
- El usuario selecciona una película del listado.
- El sistema busca la información relacionada con el título ingresado.
- Una vez obtenida dicha información, el sistema procederá a realizar el análisis correspondiente.

Buscador de Películas

Ingresa el nombre de la película:

Resultados:

[Spider-Man: Homecoming \(2017-07-05\) - Popularidad: 97.332](#)
[A Fan's Guide to Spider-Man: Homecoming \(2017-06-03\) - Popularidad: 2.229](#)

4 Proceso de negocio

- 1- Búsqueda del texto ingresado:
 - Con el texto escrito se deben buscar películas relacionadas en la API <https://www.themoviedb.org/> y se debe asociar el ID de cada una.
 - Ordenarlas por fecha de estreno (más nuevas primero).
 - Se listan las posibles concordancias con el texto ingresado.
 - El usuario selecciona un ítem válido que representa a una película.
- 2- Buscar las críticas de esa película:
 - El sistema debe consumir de la API las críticas basándose en el ID relacionado.
 - El sistema recibe las críticas y las almacena en memoria.

- Preprocesamiento: Se coloca todo el texto de la crítica en minúscula, se elimina signos de puntuación junto con artículos definidos e indefinidos, preposiciones, pronombres, etc.

Antes	Después
Una de las mejores versiones de Batman;mas reales y mas apasionadas.Reparto de lujo y muy bien elegido para encarnar los personajes	Una mejores versiones Batman mas reales mas apasionadas.Reparto lujo bien elegido encarnar personajes

3- Realizar el análisis semántico:

- Se aplica el algoritmo de Naves Bayes, el cual básicamente es un clasificador probabilístico. En este contexto del análisis de sentimientos de críticas de películas, se utiliza para determinar si una crítica es positiva o negativa.

El enfoque de este modelo se basa en el teorema de Bayes, que calcula la probabilidad de una etiqueta (por ejemplo, "positiva" o "negativa") dadas ciertas características o palabras presentes en una crítica. El algoritmo asume independencia condicional de las características, es decir, que cada característica contribuye de manera independiente a la probabilidad de la etiqueta.

- El resultado es una tabla con las principales palabras y la frecuencia de la misma, que se utilizará para la posterior creación de gráficos.

Palabra	Frecuencia
spiderman	21
good	12
nice	6
character	6
marvel	6
great	5
make	5
superhero	5
tom	5
better	4

4- Mostrar los resultados:

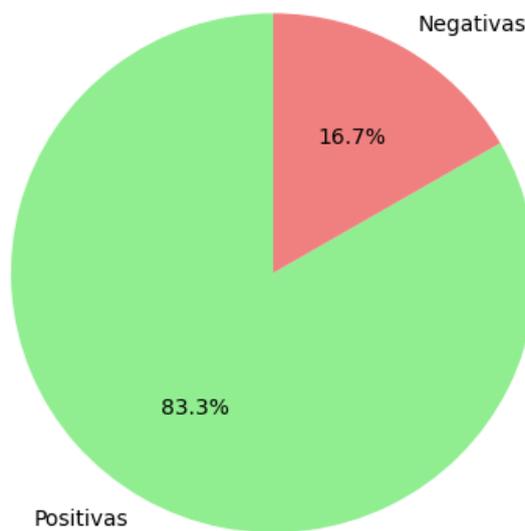
- Se realiza una nube de palabras con las palabras resultantes del punto 3 y la ocurrencia de cada una.



- Se muestra un gráfico de torta con porciones que identifican las categorías positiva y negativa.

Análisis de Sentimiento

Porcentaje de Criticas Positivas y Negativas





SentiMovie

¿Qué dicen las estadísticas sobre las reseñas en línea?

90% de los consumidores lee reseñas en línea antes de tomar decisiones de compra.

El 88% de los clientes confía tanto en las reseñas online como en las recomendaciones personales.

Análisis de sentimiento

- Es el proceso de determinar si el lenguaje refleja sentimientos positivos o negativos.
- También se la conoce como minería de opinión.
- Permite obtener información sobre las opiniones de clientes acerca de un tema.



¿Cómo se aplica en la industria cinematográfica?

- Evaluación de la recepción de películas.
- Predicción de taquilla y éxito de la película.
- Optimización de estrategias de marketing.
- Desarrollo de contenido.
- Gestión de crisis y reputación.
- Análisis comparativo.



SentiMovie



Principales funcionalidades

Buscar película

Permite realizar búsquedas de películas para su análisis

Generar reporte de reseñas

Permite generar reportes para la toma de decisiones

Buscador de Películas

Ingresa el nombre de la película:

Buscar

Resultados:

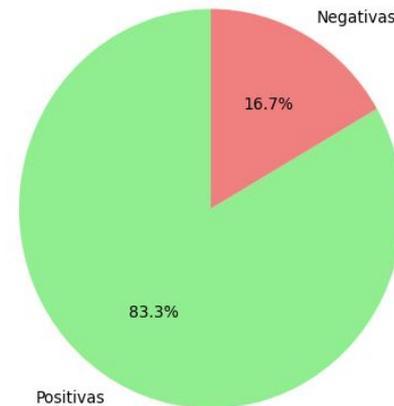
[Spider-Man: Homecoming \(2017-07-05\) - Popularidad: 97.332](#)

[A Fan's Guide to Spider-Man: Homecoming \(2017-06-03\) - Popularidad: 2.229](#)

make better
spiderman
character
marvel bad good tom
great superhero

Análisis de Sentimiento

Porcentaje de Criticas Positivas y Negativas



¿Cómo funciona?

1. Tokenización de las reseñas.
2. Eliminación de palabras irrelevantes.
3. Normalización de palabras.
4. Vectorización de texto.
5. Entrenamiento del modelo.
6. Análisis de sentimiento.



1. Tokenización de las reseñas

- Se divide la reseña en tokens
- Los tokens pueden ser palabras o frases más pequeñas
- Permite trabajar con cada palabra de manera individual

Natural Language Processing

['Natural', 'Language', 'Processing']

2. Eliminación de palabras irrelevantes

- Se eliminan palabras que no aportan al análisis
- Principalmente son palabras comunes con poco significado
- Ejemplos: “a”, “ante”, “bajo”, “con”, “contra” o en ingles “the”, ”and”, “is”

Sample text with Stop Words	Without Stop Words
GeeksforGeeks – A Computer Science Portal for Geeks	GeeksforGeeks , Computer Science, Portal ,Geeks
Can listening be exhausting?	Listening, Exhausting
I like reading, so I read	Like, Reading, read

3. Normalización de palabras

- Se eliminan signos de puntuación
- Se convierte a minúscula todo el texto
- Se eliminan caracteres especiales

4. Vectorización de texto

- Se genera una representación numérica de las palabras
- Se cuenta la frecuencia de cada palabra en cada reseña
- Se construye una matriz donde se representan los datos

	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

Word Vector (Passage Vector) →

Document Vector ↗

5. Entrenamiento del modelo

- Se utiliza un algoritmo de aprendizaje automático supervisado
- Se utilizan datos preprocesados y etiquetados
- El modelo reconocerá patrones
- Se realizan predicciones



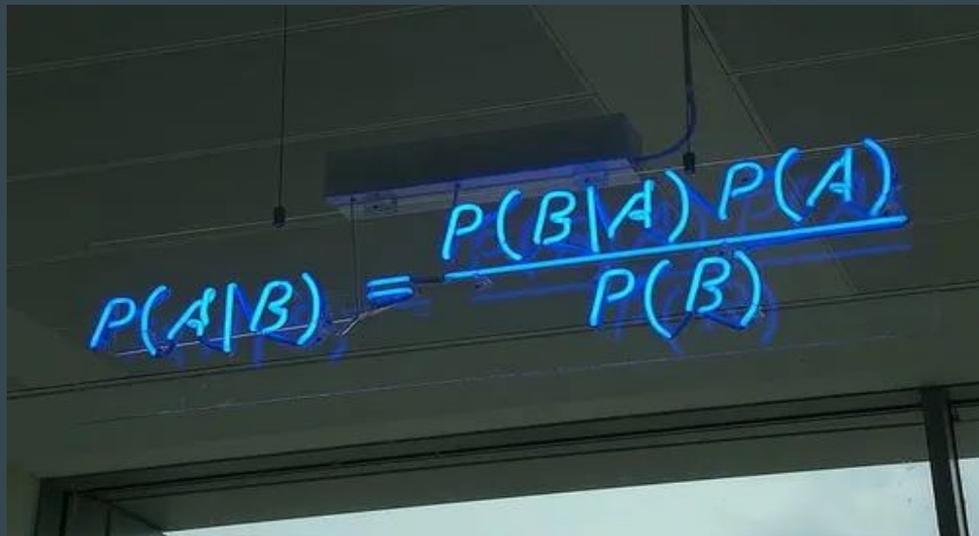
6. Análisis de sentimiento

- Se basa en el aprendizaje previo
- Se clasifican las nuevas reseñas en positivas o negativas
- Permite obtener una evaluación automática de la polaridad de las reseñas



¿Cómo clasifica?

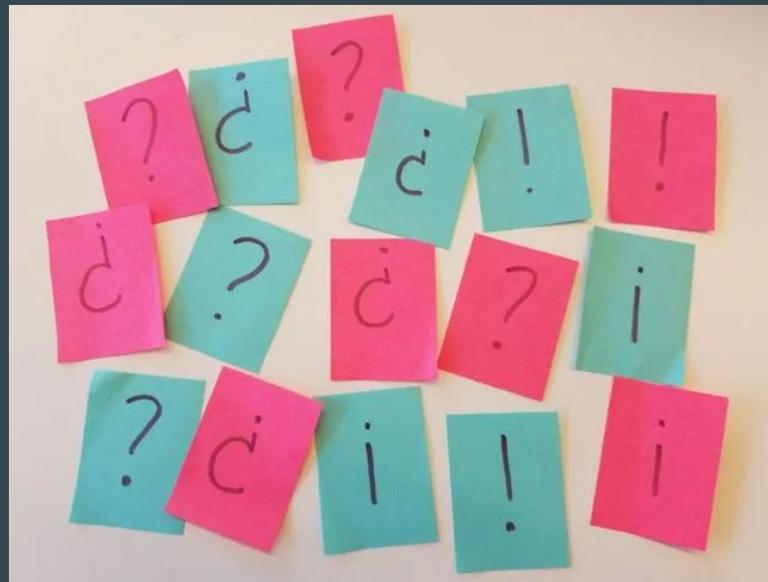
- Modelo de Naive Bayes basado en el teorema de Bayes:
 - Se calcula la probabilidad condicional de que una palabra aparezca en una reseña dado un sentimiento específico.
- Clasificación de los sentimientos en base a las palabras.



$$P(\text{Sentimiento} | \text{Palabras}) = \frac{P(\text{Palabras} | \text{Sentimiento}) * P(\text{Sentimiento})}{P(\text{Palabras})}$$

¿Por qué Naive Bayes?

- Eficiencia y rapidez.
- Resultados sólidos.
- Manejo de alta dimensionalidad.
- Requiere pocos datos de entrenamiento.



Conclusión

¿Por qué usar SentiMovie?

- Facilita la toma de decisiones para consumidores y productores de películas.
- Permite un análisis de sentimientos automatizado en tiempo real.
- Brinda una visualización clara de los resultados y permite su fácil interpretación.
- La tecnología que utiliza SentiMovie presenta posibilidades de mejoras y expansión hacia infinitudes de temáticas.



SentiMovie