



Facultad de Ciencias Exactas, Ingeniería y Agrimensura
UNIVERSIDAD NACIONAL DE ROSARIO

Introducción a la Inteligencia Artificial

Trabajo Práctico Lógica Borrosa

Alumnas:

Cipullo, Inés

Sullivan, Katherine

2022

Descripción del problema

Cantidad de seguidores del perfil

La cantidad de seguidores del perfil

hora en la que se publica el video (H) incidencia de la temática en la actualidad (T) -¿ cantidad de perfiles de prueba – 10k - (C)

duracion de video (L) cantidad de perfiles de prueba (C) tipo de interacción con el video (I) porcentaje de perfiles de prueba que interactuaron de manera positiva con el video (I) -¿ popularidad del video (P) –

determinada por las views del video luego de las primeras 48hs -¿ impopular (menos de 1k) normal (entre 10k - 100k) o popular (entre 100k - 1m)

1 Problema

Es innegable el impacto que tienen las redes sociales hoy en día para definir nuestros valores, comportamientos y creencias. La realidad es que las generaciones actuales no solo nos informamos con el contenido que vemos en ellas sino que también nos formamos a través de él.

Partidarios de todas las ideas y creencias se encuentran en las redes sociales. Son demasiados los mensajes que llegan a miles y millones de personas al mismo tiempo y el potencial de eso podría ser tanto esperanzador como peligroso.

Por eso, resulta más que lógico el estar interesado en cómo es que se generan esos mensajes que nos llegan a todos, es decir, cómo es que se genera el contenido viral.

Se desea, entonces, en este trabajo poder predecir la popularidad videos publicados en la red TikTok publicados por cuentas nuevas, valiéndonos en distintos datos incidentes.

La **hora en la que se publica el video** y la **relevancia de la temática** en la actualidad determinarán la **cantidad de perfiles de prueba** a los que se le mostrará el video en una primera instancia

La **cantidad de perfiles de prueba** del video, la **duración del video** y la **tipo de interacción** con el video influirán enormemente en la **popularidad del video**.

2 Modelado

2.1 Variables lingüísticas

- **Hora en la que se publica el video (H)**: determinada por los conjuntos borrosos hora pico y hora valle. Estos están a su vez determinados por la predicción de la cantidad de personas que estarán haciendo uso de la red social en ese momento del día, que será alta en la hora pico y baja en la hora valle. En Argentina se podría considerar hora pico de 11:00 a 17:00 horas.
- **Relevancia de la temática en la actualidad (T)**: la relevancia de una temática en particular puede ser baja, media o alta y está representada por un índice del 0 al 1.
- **Cantidad de perfiles de prueba (C)**: hace referencia a la cantidad de perfiles a los que se le muestra el video en una primera instancia, y varía de 10 a 10000. La cantidad de perfiles de prueba es baja si es menor de 1500, es media si está entre 4000 y 6000, y es alta si es igual o mayor a 8500.

- **Duración del video (D):** medido en cantidad de minutos, se puede decir que es corta, media o larga y va desde 0 a 10 minutos, siendo corta de 0 a 1 minuto, media de 1 a 3 minutos y larga si es mayor de 3 minutos.
- **Tipo de interacción con el video (I):** medida como el porcentaje de perfiles de prueba que tuvieron una o más interacciones positivas sobre el video. Se considera negativa si es menor al 5%, regular si es alrededor del 10% y positiva si es más del 20%.
- **Popularidad del video (P):** para representar esta variable lingüística se utilizan 4 conjuntos borrosos definidos a partir de un índice del 0 al 1. Estos son: impopular (de 0 a 0.25), semi-popular (de 0.4 a 0.5), popular (de 0.7 a 0.8) y viral (mayor a 0.9).

2.2 Reglas de inferencia

Planteamos las reglas de inferencia utilizando la aritmética lógica adaptada al lenguaje natural, para presentarlas de forma compacta. En la herramienta **fispro**, alguna de estas operaciones lógicas no se soportan y entonces se plantean, en su lugar, un conjunto de reglas por cada una de ellas.

- Reglas relativas a la primera etapa
 - R1** H hora pico y T alta o media \rightarrow C alta
 - R2** H hora pico o valle y T baja \rightarrow C baja
 - R3** H hora valle y T alta o media \rightarrow C media
- Reglas relativas a la segunda etapa
 - S1** C alta, D corta e I positiva \rightarrow P viral
 - S2** C media o alta, D corta o media e I positiva \rightarrow P popular
 - S3** C media, D larga e I positiva \rightarrow P semi-popular

S4 C baja o media, D media e I regular \rightarrow P semi-popular

S5 I semi-regular o negativa \rightarrow P impopular

3 Pruebas y ajustes de parámetros

4 Resultados

5 Conclusiones

6 Bibliografía

Resultados de la encuesta - reglas del modelo

Las encuestas realizadas por la cooperativa arrojaron la siguiente información:

1. En relación a la longitud de la nota, la velocidad de escritura y el tiempo de redacción
 - (a) Si la longitud de la nota es corta y la velocidad de escritura de quien escribe es rápida entonces el tiempo que toma redactar definitivamente la nota es corto
 - (b) Si la longitud de la nota es media y la velocidad de escritura de quien escribe es media entonces el tiempo que toma redactar definitivamente la nota es medio
 - (c) Si la longitud de la nota es larga y la velocidad de escritura de quien escribe es lenta entonces el tiempo que toma redactar definitivamente la nota es largo
2. En relación al tiempo de redacción, la reputación del periodista, la popularidad de la temática y la cantidad de lectores
 - (a) Si el tiempo de redacción de una nota es corto, la reputación del periodista es buena y la popularidad de la temática es media o alta entonces la cantidad de lectores es alta
 - (b) Si el tiempo de redacción de una nota es medio, la reputación del periodista es buena y la popularidad de la temática es media o alta entonces la cantidad de lectores es medio-alta
 - (c) Si el tiempo de redacción de una nota es medio, la reputación del periodista es mala y la popularidad de la temática es alta entonces la cantidad de lectores es medio-baja
 - (d) Si el tiempo de redacción de una nota es largo, la reputación del periodista es mala y la popularidad de la temática es alta entonces la cantidad de lectores es medio-baja

- (e) Si el tiempo de redacción de una nota es largo, la reputación del periodista es mala y la popularidad de la temática es baja entonces la cantidad de lectores es baja

Enunciado

Sobre la información provista en la anterior sección se solicita:

1. Determinar variables lingüísticas de entrada y salida con sus conjuntos borrosos asociados.
2. Analizar si se puede reescribir el modelo en un número de reglas menor admitiendo el uso de más operadores lógicos.
3. Determinar el tiempo de redacción para una nota de longitud de 470 palabras y una velocidad de quien la escribe de 40 palabras por minuto. Indicar qué reglas se dispararon con qué grado de veracidad.
4. Repetir el apartado anterior para una nota de longitud de 240 palabras y una velocidad de redacción de quien la escribe de 63 palabras por minuto.
5. Según el valor crisp obtenido en el apartado 3 y considerando que la reputación de quien escribe la nota es 0 (la mínima) y P 1 (la máxima) estimar la cantidad de lectores de dicha nota. Indicar qué reglas se dispararon con qué grado de veracidad.
6. Según el valor crisp obtenido en el apartado 4 y considerando que la reputación de quien escribe la nota es 0.6 y la temática de la que se escribe es economía y finanzas, estimar la cantidad de lectores de dicha nota. Indicar qué reglas se dispararon con qué grado de veracidad.

Resolución del ejercicio

Ejercicio 1 - Variables lingüísticas

El problema se puede dividir en dos etapas. En la primera intervienen las variables lingüísticas de entrada longitud de una nota y velocidad de escritura de quien redacta la nota. Mientras que la variable de salida de la primer etapa es el tiempo que toma redactar la nota.

En la segunda etapa las variables de entrada son: el tiempo que toma redactar la nota, la reputación de quien escribe la nota y la popularidad de la temática de la que se escribe la nota. La variable de salida de la segunda etapa es la cantidad de lectores de dicha nota.

Los conjuntos borrosos asociados a cada una de las variables se listan a continuación

- longitud de una nota: corta, media y larga
- velocidad de escritura de quien redacta la nota: lenta, media y rápida
- tiempo que toma redactar la nota: corto, medio y largo
- reputación de quien escribe la nota: mala y buena
- popularidad de la temática de la que se escribe la nota: baja, media o alta
- cantidad de lectores de dicha nota: baja, medio-baja, medio-alta y alta

En las figuras (??), (??) (??) y (??) se pueden visualizar las variables de entrada y salida de cada una de las etapas.

(a) Longitud de la nota

(b) Velocidad de escritura

Figure 1: Funciones de pertenencia de los conjuntos borrosos asociados a las variables de entrada de la primera etapa

Figure 2: Tiempo que toma redactar una nota - Variable de salida de la primera etapa

	(b) Reputación del peri-	(c) Popularidad de la
(a) Tiempo de redacción	odista	temática

Figure 3: Funciones de pertenencia de los conjuntos borrosos asociados a las variables de entrada de la segunda etapa

Figure 4: Cantidad de lectores por nota - Variable de salida de la segunda etapa

Ejercicio 2 - Reglas nuevas basadas en el modelo

El listado de reglas escritas de forma más compacta haciendo uso de operadores lógicos se puede ver a continuación

- Reglas relativas a la primera etapa
 - * $L \text{ corta y } V \text{ rápida} \rightarrow T \text{ corto}$
 - * $L \text{ media y } V \text{ media} \rightarrow T \text{ medio}$
 - * $L \text{ larga y } V \text{ lenta} \rightarrow T \text{ largo}$
- Reglas relativas a la segunda etapa
 - * $T \text{ corto, } R \text{ buena y } P \text{ media o alta} \rightarrow C \text{ alta}$
 - * $T \text{ medio, } R \text{ buena y } P \text{ media o alta} \rightarrow C \text{ medio-alta}$
 - * $T \text{ medio o largo, } R \text{ mala y } P \text{ alta} \rightarrow C \text{ medio-baja}$
 - * $T \text{ largo, } R \text{ mala y } P \text{ baja} \rightarrow C \text{ baja}$

Como la herramienta fispro no permite agregar reglas que incluyen el operador de disyunción, reformulamos las reglas para satisfacer estas necesidades.

- Reglas relativas a la primera etapa
 - R1** $L \text{ corta y } V \text{ rápida} \rightarrow T \text{ corto}$
 - R2** $L \text{ media y } V \text{ media} \rightarrow T \text{ medio}$
 - R3** $L \text{ larga y } V \text{ lenta} \rightarrow T \text{ largo}$
- Reglas relativas a la segunda etapa
 - S1** $T \text{ corto, } R \text{ buena y } P \text{ media} \rightarrow C \text{ alta}$
 - S2** $T \text{ corto, } R \text{ buena y } P \text{ alta} \rightarrow C \text{ alta}$
 - S3** $T \text{ medio, } R \text{ buena y } P \text{ media} \rightarrow C \text{ medio-alta}$
 - S4** $T \text{ medio, } R \text{ buena y } P \text{ alta} \rightarrow C \text{ medio-alta}$
 - S5** $T \text{ medio, } R \text{ mala y } P \text{ alta} \rightarrow C \text{ medio-baja}$
 - S6** $T \text{ largo, } R \text{ mala y } P \text{ alta} \rightarrow C \text{ medio-baja}$

S7 T largo, R mala y P baja \rightarrow C baja

En la figura (??) se pueden ver las reglas tal cual las enumeramos previamente.

Figure 5: Reglas cargadas en la herramienta fispro

Ejercicio 3 - Inferir el tiempo de redacción

Dada una nota de 470 palabras de largo (L media, L larga) y considerando la velocidad de la persona que la redacta de 40 palabras por minuto (V lenta, V media) pudimos inferir con la herramienta fispro que la redacción de dicha nota tomará 4.85 días, es decir, 4 días y 20 horas (T medio, T largo) en completarse.

Figure 6: Inferencia del caso de estudio en la herramienta fispro

Nota Para inferir utilizamos de *borrosificador* la Norma mínimo y para defusificar utilizamos el método mean max.

Las reglas que se activaron con este caso de estudio son las reglas **R2** y **R3** con valores de veracidad de

Regla	Longitud de la nota	Velocidad de redacción	Veracidad según el mínimo
R2	Media - ≈ 0.5	Media - ≈ 0.7	Medio - 0.5
R3	Larga - ≈ 0.7	Lenta - 1	Largo - 0.7

El tiempo de redacción obtenido, 4 días 20 horas, pertenece al conjunto borroso TiempoRedacción Largo con un grado de pertenencia de 1 y al conjunto borroso TiempoRedacción medio con un grado de pertenencia ≈ 0.1 . Es decir, el grado de pertenencia a TiempoRedacción largo es mayor al grado de pertenencia TiempoRedacción medio. Este resultado además, se

condice con que el grado de veracidad de la regla **R3** es mayor al grado de veracidad de **R2**. Teniendo todo esto en cuenta, el valor obtenido como resultado fue el esperado.

Ejercicio 4 - Inferir el tiempo de redacción

Dada una nota de 240 palabras de largo (L corta, L media) y considerando la velocidad de la persona que la redacta de 63 palabras por minuto (V media, V rápida) pudimos inferir con la herramienta fispro que la redacción de dicha nota tomará 1.2 días, es decir, 1 día y 5 horas (T corto, T medio) en completarse.

Figure 7: Inferencia del caso de estudio en la herramienta fispro

Nota Para inferir utilizamos de *borrosificador* la Norma mínimo y para defusificar utilizamos el método mean max.

Las reglas que se activaron con este caso de estudio son las reglas **R1** y **R2** con valores de veracidad de

Regla	Longitud de la nota	Velocidad de redacción	Veracidad según el mínimo
R1	Corta - ≈ 0.6	Rápida - 1	Corto - ≈ 0.6
R2	Media - ≈ 0.6	Media - ≈ 0.45	Medio - ≈ 0.45

El tiempo de redacción obtenido, 1 día y 5 horas, pertenece al conjunto borroso TiempoRedacción Corto con un grado de pertenencia de 1 y al conjunto borroso TiempoRedacción Medio con un grado de pertenencia ≈ 0.2 . Es decir, el grado de pertenencia a TiempoRedacción medio es mayor al grado de pertenencia TiempoRedacción corto. Este resultado además, se condice con que el grado de veracidad de la regla **R1** es mayor al grado de veracidad de **R2**. Teniendo todo esto en cuenta, el valor obtenido como resultado fue el esperado.

Ejercicio 5 - Inferir la cantidad de lectores

Dado el tiempo de redacción de 4 días y 20 horas (4.85 días; T medio, T largo), 0 de reputación de quien escribe la nota (R mala) y considerando 1 de popularidad de la temática (P alta), pudimos inferir con la herramienta fispro que la cantidad de lectores de dicha nota será de 300 personas (C media-baja)

Figure 8: Inferencia del caso de estudio en la herramienta fispro

Nota Para inferir utilizamos de *borrosificador* la Norma Lukasciewicz y para defusificar utilizamos el método mean max.

Las reglas que se activaron con este caso de estudio son las reglas **S5** y **S6** con valores de veracidad de

Regla	Tiempo de redacción	Reputación de le periodista	Popularidad de la temática	Veracidad según
S5	Medio - ≈ 0.1	Mala - 1	Alta - 1	Medio-baja
S6	Largo - 1	Mala - 1	Alta - 1	Medio-ba

La cantidad de lectores obtenida, 300 lectores, pertenece al conjunto borroso CantidadLectores Medio-baja con un grado de pertenencia de 1. Este resultado además, se condice con que el grado de veracidad de la regla **R5** y **R6** de 1. Teniendo todo esto en cuenta, el valor obtenido como resultado fue el esperado.

¹ $\max\{0, \max\{0, 0.1 + 1 - 1\} + 1 - 1\}$

² $\max\{0, \max\{0, 1 + 1 - 1\} + 1 - 1\}$

Ejercicio 6 - Inferir la cantidad de lectores

Dado el tiempo de redacción de 1 día 17 horas (1.7 días; T corto, T medio), 0.6 de reputación de quien escribe la nota (R mala, R buena) y considerando 0.85 de popularidad de la temática (P media, P alta), pudimos inferir con la herramienta fispro que la cantidad de lectores de dicha nota será de 913 personas (C media-alta, C alta)

Figure 9: Inferencia del caso de estudio en la herramienta fispro

Nota Para inferir utilizamos de *borrosificador* la Norma Lukasciewicz y para defusificar utilizamos el método mean max.

Las reglas que se activaron con este caso de estudio son las reglas **S1**, **S2** y **S4** con valores de veracidad de

Regla	Tiempo de redacción	Reputación de le periodista	Popularidad de la temática	Veracidad según
S1	Corto - 1	Buena - ≈ 0.8	Media - ≈ 0.25	Alta - \approx
S2	Corto - 1	Buena - ≈ 0.8	Alta - 1	Alta - \approx
S4	Medio - ≈ 0.5	Buena - ≈ 0.8	Alta - ≈ 0.75	Medio-alto

La cantidad de lectores obtenida, 913 personas, pertenece al conjunto borroso CantidadLectores Medio-alto con un grado de pertenencia de ≈ 0.3 y al conjunto borroso CantidadLectores Alto con un grado de pertenencia de 1. Es decir, el grado de pertenencia a CantidadLectores Alto es mayor al grado de pertenencia CantidadLectores Medio. Este resultado además, se condice con que el grado de veracidad de la regla **S2** es mayor al grado de veracidad de **S1** y **S4**. Teniendo todo esto en cuenta, el valor obtenido como resultado fue el esperado.

¹ $\max\{0, \max\{0, 1 + 0.8 - 1\} + 0.25 - 1\}$

² $\max\{0, \max\{0, 1 + 0.8 - 1\} + 0.75 - 1\}$

³ $\max\{0, \max\{0, 0.5 + 0.8 - 1\} + 0.25 - 1\}$

Ejercicio 7 - Inferir la cantidad de lectores

Dado el tiempo de redacción de 4 días (T medio, T largo), 0.6 de reputación de quien escribe la nota (R mala, R buena) y considerando 0.5 de popularidad de la temática (P media), pudimos inferir con la herramienta fispro que la cantidad de lectores de dicha nota será de 687 personas (C media-alta, C alta).

Figure 10: Inferencia del caso de estudio en la herramienta fispro

Nota Para inferir utilizamos de *borrosificador* la Norma Lukasciewicz y para defusificar utilizamos el método mean max.

La regla que se activó con este caso de estudio es **S3** con valor de veracidad de

Regla	Tiempo de redacción	Reputación de le periodista	Popularidad de la temática	Veracidad según
S3	Medio - ≈ 0.75	Buena - ≈ 0.75	Media - 1	Medio-alta

La cantidad de lectores obtenida, 687 lectores, pertenece al conjunto borroso CantidadLectores Medio-alta con un grado de pertenencia de 1, lo cual se condice con un alto grado de veracidad de la regla **S3** (considerando lo ajustado de la norma de Lukasciewicz).

⁶ $\max\{0, \max\{0, 0.75 + 0.75 - 1\} + 1 - 1\}$

Conclusiones

El modelo resultó completo y práctico en términos de los resultados que pudimos obtener a partir de él. Apartado por apartado los resultados fueron verificando nuestra intuición. Además, éstos se correspondieron con las medidas (borrosas) que utilizamos para ganar confianza sobre nuestro modelo.

Para los apartados 3 y 4 que pedían inferir el tiempo de redacción de una nota, usamos la norma mínimo dado que es la operación menos restrictiva. Creemos que elegir una operación que no subestime el tiempo de una nota se alinea con los ideales que podría tener una cooperativa: genera un entorno de trabajo amigable en el que los trabajadores no se sienten presionados a cumplir plazos muy cortos de redacción.

Para los apartados, 5, 6 y 7, en contraste, decidimos operar con la norma de Lukasiewicz, norma más restrictiva que el mínimo, para no sobreestimar la cantidad de lectores que comprará el ejemplar con la nota. Pensamos que tiene sentido que una cooperativa no pueda imprimir más de lo que le permite un presupuesto acotado, por lo que ajustar la estimación de cantidad de lectores va a actuar en su favor.

Como único comentario extra, en la teoría la norma de Lukasiewicz estaba solamente para dos variables de entrada mientras que en nuestro trabajo, tuvimos que aplicar reglas con 3 variables de entrada. Lo que decidimos en este caso es extender la norma para 3 variables de entrada de la siguiente manera:

$$L(L(V1, V2), V3)$$

donde L es la norma de Lukasiewicz.

Los resultados que obtuvimos a partir de la extensión de esta regla van de la mano con los valores que esperábamos.