

Machine Learning: Método supervisado

Discurso discriminatorio hacia la mujer en Twitter

Introducción

En los últimos años han crecido las protestas por los derechos de las mujeres en diferentes partes del mundo. Esto se debe a que las mujeres jóvenes se enfrentan a los límites de los diversos cambios sociales de las últimas décadas; oportunidades laborales, salarios, sexualidad. Para las mujeres, este momento es una oportunidad de solidaridad, pues han callado por mucho tiempo la opresión de un entorno que no se preocupa por ellas.

Campañas como #NiUnaMenos o #MillionWomanRise han mostrado que las redes sociales pueden ayudar a conocer más acerca de un problema que afecta a más de 4 millones 600 mil mujeres tan solo en México.

Luis Bonino es un psicólogo que ha investigado la violencia de género y ha propuesto diferentes clasificaciones de este. Propuso el concepto de micromachismo para referirse a aquellas conductas sutiles y cotidianas que construyen violencia hacia las mujeres. Estas conductas son tan pequeñas que se normalizan en la sociedad.

Objetivos

- Utilizar un método de aprendizaje supervisado para clasificar texto
- Detectar en Twitter el discurso discriminatorio con base en la clasificación de Luis Bonino
- Conocer qué palabras o colocaciones determinan un discurso discriminatorio con mayor certeza

Modelado del problema

En las redes sociales, no es sorpresa que se pueda percibir discriminación en todo tipo de comentarios. La propuesta de este trabajo es detectar el discurso discriminatorio hacia las mujeres en Twitter y clasificarlo según las categorías que propone Luis Bonino.

La propuesta es en 4 categorías de micromachismo: utilitario, encubierto, coercitivo y de crisis. Cada clase está determinada por diferentes acciones que se pueden detectar en texto.

Micromachismos utilitarios: No responsabilizarse sobre lo doméstico No implicación Pseudo implicación Implicación ventajosa Aprovechamiento y abuso de las capacidades “femeninas de servicio” Naturalización y aprovechamiento del rol de cuidadora Delegación del trabajo del cuidado de vínculos y personas Requerimientos abusivos solapados Negación de la reciprocidad Naturalización y aprovechamiento de la “ayuda” al marido Amiguismo paternal
Micromachismos encubiertos: Creación de falta de intimidad Silencio Aislamiento y malhumor manipulativo Puesta de límites Avaricia de reconocimiento y disponibilidad Inclusión invasiva de terceros Seudointimidad y seudocomunicación Comunicación defensiva – ofensiva Engaños y mentiras Seudonegociación Desautorización Descalificación-desvalorización Negación de lo positivo Colisión con terceros Microterrorismo misógino Paternalismo Manipulación emocional Dobles mensajes afectivo/agresivos Enfurruñamiento Abuso de confianza Inocentizaciones Inocentización culpabilizadora Autoindulgencia y autojustificación. Hacerse el tonto (y el bueno) Impericia y olvidos selectivos Comparación ventajosa Minusvaloración de los propios errores Echar balones fuera
Micromachismos coercitivos: Coacciones a la comunicación Control del dinero Uso expansivo – abusivo del espacio y el tiempo para sí Insistencia abusiva Imposición de intimidad Apelación a la “superioridad” de la lógica varonil Toma o abandono repentino del mando
Micromachismos de crisis: Hipercontrol Seudoapoyo Resistencia pasiva y distanciamiento Rehuir la crítica y la negociación Prometer y hacer méritos Victimismo Darse tiempo Dar lástima

Lista de tipos de micromachismos y su categoría

Para realizar el aprendizaje de máquina se propone un etiquetado automático que se encarga de buscar tweets, en una base de datos MongoDB, con base en léxico o colocaciones que identifiquen el discurso discriminatorio.

El primer corpus consta de 2,639,743 registros y pertenecen al tercer cuarto del año 2017. Todos son de México.

Se eligió una muestra básica que tiene la siguiente forma:

Micromachismo utilitario

- perdio la cocina
- mujer multiusos
- vieja multiusos
- mujer sirve para
- vieja sirve para
- puta sirve para

Micromachismo de crisis

- matriarcado opresor
- cosa de hombres
- cosas de hombres
- estar en sus días
- esta en sus días

Micromachismo coercitivo

- la metio
- se la cogio
- neofeminazi
- neofeminista
- neofeminazis
- neofeministas
- por ser mujer

Micromachismo encubiero

- que ser puta
- tenias que ser mujer
- tenias que ser vieja
- parece hombre
- parece mujer
- parece vieja
- pareces hombre
- pareces mujer
- parecer vieja
- andar de puta

El léxico y colocaciones se eligieron con una base no rigurosa: una pequeña encuesta en redes sociales con los tipos de mensajes discriminatorios más comunes.

Nota: Esta parte es muy importante porque determina el aprendizaje de máquina. Es el corpus de entrenamiento. Un lingüista puede generar un diccionario para generar mejores resultados.

El corpus de entrenamiento es un archivo XML con la siguiente forma:

```
<tweets>
  <tweet discriminacion="Encubierta" favoritos="0" id="890979118174621696" retweets="0">
    <texto>Pena ser pendeja? Pena andar de puta y ofrecida!!! Jajajaja</texto>
    <hashtags/>
    <menciones/>
  </tweet>
</tweets>
```

discriminación: determina el tipo de micromachismo que es.

favoritos: el número de favoritos recibidos. (No se utiliza)

id: el identificador del tweet (No se utiliza)

retweets: el número de retweets del tweets (No se utiliza)

texto: el mensaje del tweet

hashtags: los hashtags asociados al mensaje (No se utiliza)

menciones: los usuarios mencionados en el mensaje (No se utiliza)

Nota: Los atributos y etiquetas que no se utilizan pueden servir para determinar la relevancia de un mensaje y asociarlo a un usuario en específico.

Para la representación vectorial se pre procesó el corpus etiquetado eliminando caracteres especiales y convirtiendo todo a minúsculas.

Se determinó una métrica de TF-IDF para encontrar la relevancia de palabras clave en cada tweet de entrenamiento. Esta representación le ayuda al árbol de decisión a clasificar los tweets de forma que lo más relevante defina la derivación del árbol.

El formato del JSON con la representación vectorial es el siguiente:

```
{"training": {"target_names": ["Utilitaria", "Encubierta", "Coercitiva", "Crisis"], "data": [[0.0, 0.0, 0.3860729711040596, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0], ...], "target": [1], "feature_names": ["Hogar", "Rol de cuidadora", "Servicio", "Descalificacion", "Forma de verse", "Insinuacion", "Violacion normalizada", "Represion", "Inferioridad", "Control", "Exclusion", "Victimismo"]}}
```

training: raíz del JSON

target_names: Clases de micromachismos

data: matriz con los valores de TF-IDF

target: identificador de la clase asociada a la matriz

feature_names: subclasificaciones por elemento del léxico o colocación

Nota: los 12 campos de la matriz “data” corresponden al léxico o colocaciones que hicieron “match” en un tweet. Por ejemplo, el tercer subelemento (0.3860729711040596) corresponde

a un tweet donde encontró la subclasificación “Servicio” (mujer sirve para|vieja sirve para|puta sirve para) en un tweets.

Se muestran las subclasificaciones como aparece en el código:

```
# Hogar
regex_u_1 = re.compile('perdio la cocina')
# Rol de cuidadora
regex_u_2 = re.compile('mujer multiusos|vieja multiusos')
# Servicio
regex_u_3 = re.compile('mujer sirve para|vieja sirve para|puta sirve para')

# Discriminacion encubierta Definicion de las Features
# Descalificacion
regex_e_1 = re.compile('tenias que ser puta|tenias que ser mujer|tenias que ser vieja')
# Forma de verse
regex_e_2 = re.compile('parece hombre|parece mujer|parece vieja|pareces hombre|pareces
mujer|parecer vieja')
# Insinuacion
regex_e_3 = re.compile('andar de puta|pareces puta|parece puta')

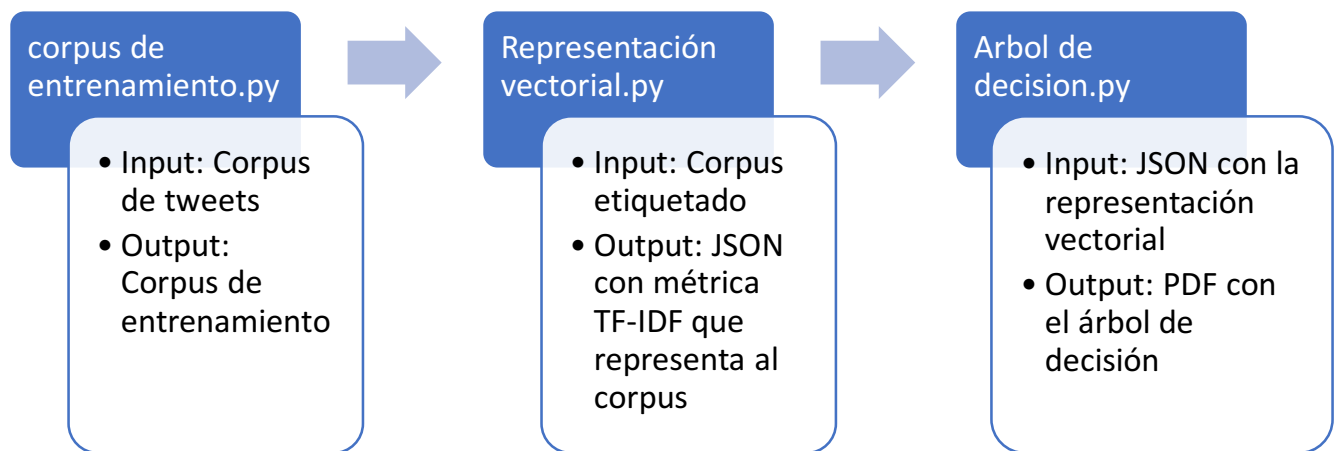
# Discriminacion coercitiva Definicion de las Features
# Violacion normalizada
regex_co_1 = re.compile('se la metio|se la cogio')
# Represion
regex_co_2 = re.compile('neofeminazi|neofeminista|neofeminazis|neofeministas')
# Inferioridad
regex_co_3 = re.compile('por ser mujer|por ser vieja')

# Discriminacion de crisis Definicion de las Features
# Control
regex_cr_1 = re.compile('matriarcado opresor')
# Exclusion
regex_cr_2 = re.compile('cosa de hombres|cosas de hombres')
# Victimismo
regex_cr_3 = re.compile('estar en sus dias|esta en sus dias')
```

El entrenamiento de máquina se realizó con la herramienta “scikit-learn” y el método de árbol de decisión.

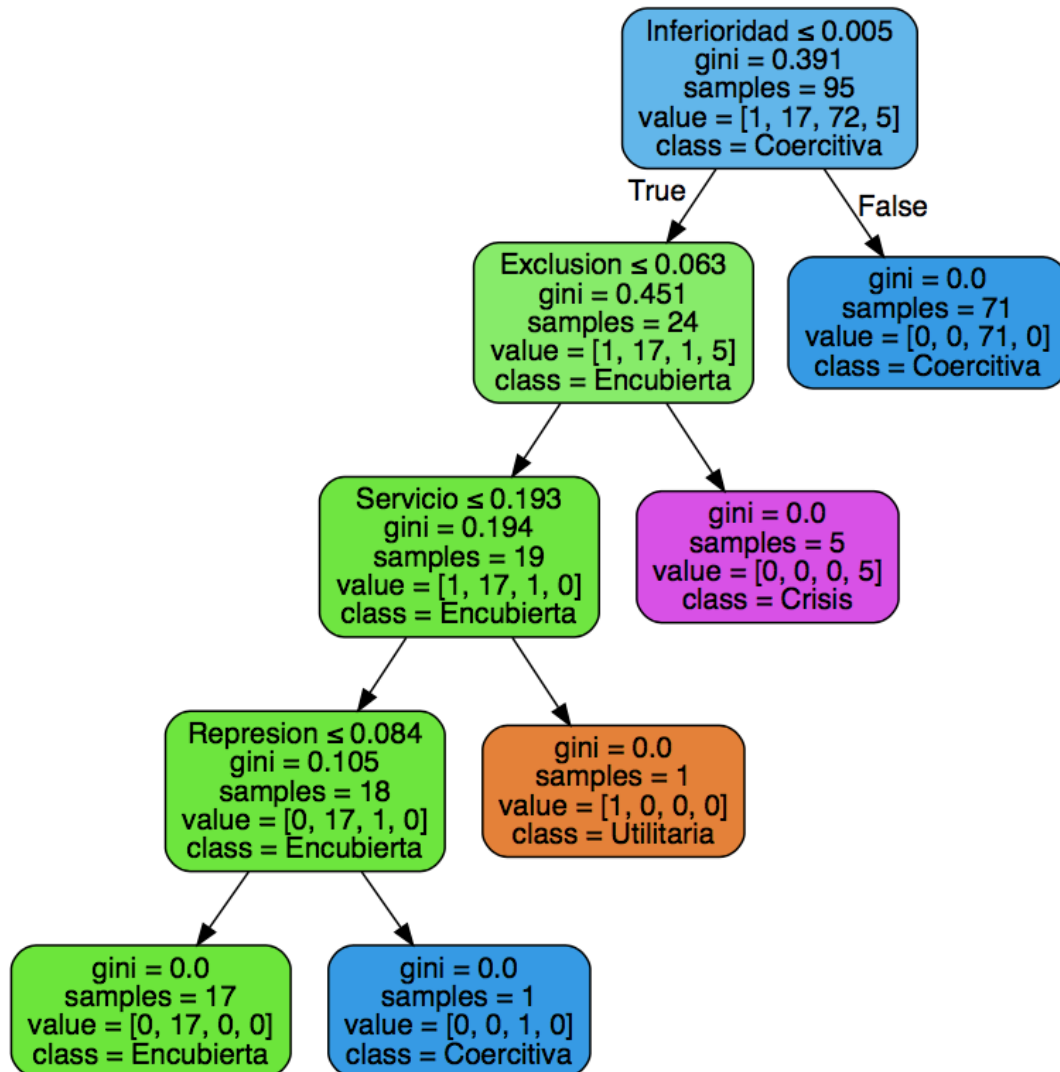
Se utiliza este modelo porque es relativamente fácil interpretar los resultados para un humano. Como es un problema social el que se plantea en este proyecto, mientras más personas que puedan entenderlo es mejor. Además, se requiere identificar qué palabras o colocaciones determinan con mayor certeza qué tweets tienen contenido micromachista de cierta clasificación.

De forma gráfica el sistema es el siguiente:



Resultados

Como resultado se tuvo el siguiente árbol



La raíz y los nodos intermedios tienen 5 características:

- Criterio de separación con base en la medida TF-IDF
- El criterio “gini” para medir la calidad de una división de rama
- Número de muestras
- Valores encontrados en cada clasificación
- Clase determinada

Desde la raíz se puede observar que el corpus de entrenamiento no está balanceado. Tiene un tweet de muestra para machismo utilitario, 17 para encubierto, 72 para coercitivo y 5 para crisis.

“Inferioridad” se refiere a la subclasificación dada a “por ser mujer | por ser vieja” que se encuentra en la clase coercitiva. El valor mayor de 0.005 se refiere a los valores de referencia para clasificar como verdadero y falso.

El valor “gini” muestra la pureza del corte en la rama, por esto las hojas tienen valor 0. A menor valor mejor clasificación.

Conclusiones

Los resultados obtenidos no fueron del todo satisfactorios. Se tienen las siguientes intuiciones para mejorar el árbol:

- Mejor representatividad de datos. Con ayuda de un lingüista, identificar las muestras representativas de cada clase para poder generar un corpus.
- Tener un corpus balanceado.

Lo que dice el árbol generado es que un tweet con un micromachismo coercitivo de tipo “Inferioridad” puede determinar la clase directamente. Si revisamos el corpus etiquetado podemos buscar las colocaciones “por ser mujer” o “por ser vieja” y encontramos que 71 de 96 tweets obtenidos contienen esta cadena.

Se agregó un tweet al final del corpus etiquetado:

```
<tweet discriminacion="Utilitaria" favoritos="0" id="919777058246594562" retweets="0">  
  <texto>@m_zamarripa @RDBonillaA @betocabcam @Ale_BarralesM La mujer sirve para  
  cuidar a los hijos</texto>  
  <hashtags>@m_zamarripa@RDBonillaA@betocabcam@Ale_BarralesM</hashtags>  
  <menciones/>  
</tweet>
```

Esto fue porque, del corpus etiquetado que representa cerca del 0.00003% del total de tweets, no se encontró ni una muestra de micromachismo utilitario. Y el árbol quedaba peor.

El siguiente nodo corresponde a la subclasificación “Exclusión” que determina si la clase es Encubierta o va directamente a Crisis. Es decir, del corpus etiquetado se sabe que un tweet es de clase “Crisis” si en su contenido aparece “cosas de hombres”.

Finalmente, las hojas del último nivel muestran que, si todos los tweets revisados no cumplen con alguna característica, se clasifica como “Coercitivo”. Esto por la gran cantidad de tweets que representan a esta clasificación.

Bibliografía

Ferrer Pérez, Victoria A.; Bosch Fiol, Esperanza; Navarro Guzmán, Capilla; Ramis Palmer, M. Carmen; García Buades, M. Esther *Los micromachismos o microviolencias en la relación de pareja: Una aproximación empírica Anales de Psicología*, vol. 24, núm. 2, diciembre, 2008, pp. 341-352 Universidad de Murcia Murcia, España

Instalación y consultas de para Machine Learning de <http://scikit-learn.org/stable/>