

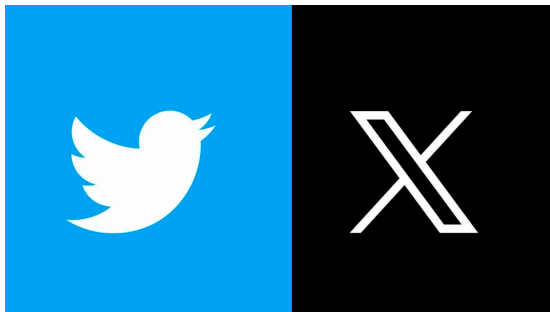
Análisis Reputacional de Aerolíneas via Análisis de Sentimiento en Twitter

Rodolfo Jesús Ramírez Lucario

Centro de Investigación en Matemáticas A.C.

Motivación

- Twitter es una fuente rica de opiniones sobre servicios, en particular sobre aerolíneas estadounidenses.
- La reputación y la satisfacción del cliente impactan directamente en la preferencia de los usuarios.
- Objetivo general:
 - Analizar el sentimiento de tweets sobre aerolíneas de EE. UU. (positivo, negativo, neutral).
 - Comparar el desempeño de modelos clásicos de *Machine Learning* vs. modelos de *Deep Learning*.

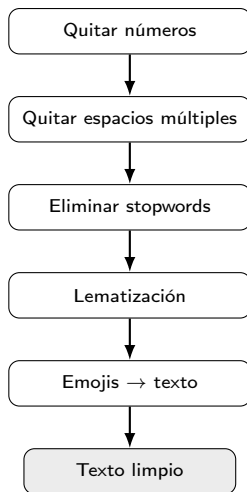
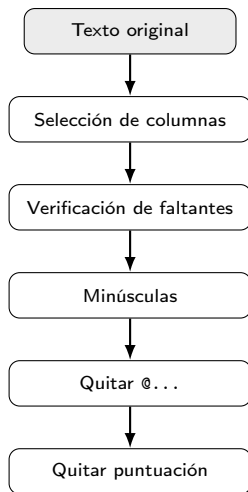


Contamos con 14640 tweets, sin datos faltantes.

| Variable | Descripción |
|---------------------------|--|
| tweet_id | Identificador único del tweet. |
| airline_sentiment | Sentimiento asignado al tweet (positivo, negativo, neutral). |
| airline_sentiment | Nivel de confianza del sentimiento asignado. |
| negativereason | Razón específica del sentimiento negativo (si aplica). |
| negativereason_confidence | Nivel de confianza de la razón negativa. |
| airline | Aerolínea mencionada en el tweet. |
| airline_sentiment_gold | Sentimiento corregido manualmente (si existe). |
| name | Nombre del usuario que publicó el tweet. |
| negativereason_gold | Razón negativa corregida manualmente (si existe). |
| retweet_count | Número de veces que el tweet fue retuiteado. |
| text | Texto completo del tweet. |
| tweet_coord | Coordenadas geográficas del tweet (si existen). |
| tweet_created | Fecha y hora en que se creó el tweet. |
| tweet_location | Ubicación declarada del usuario. |
| user_timezone | Zona horaria del usuario. |

Cuadro: Descripción de variables del dataset de sentimiento de aerolíneas

Tratamiento de Texto



Ejemplo de Tratamiento

Cómo podemos ver en la siguiente imagen, los emogis no fueron alterados por la demás transformaciones. La palabra del Emoji también será tomada en cuenta dentro del corpus de nuestros modelos.

```
'@SouthwestAir 🥵 you won't let me change my reservation online so now I'm just wasting my time. http://t.co/mHA3xXaeD5'
```

```
'emoji_grinning_face_with_sweat let change reservation online wasting time'
```

Figura: Limpieza de texto.

Análisis Exploratorio pt. 1

En la siguiente imagen podemos ver que hay un claro des balance de clases (negativo, neutral y positivo).

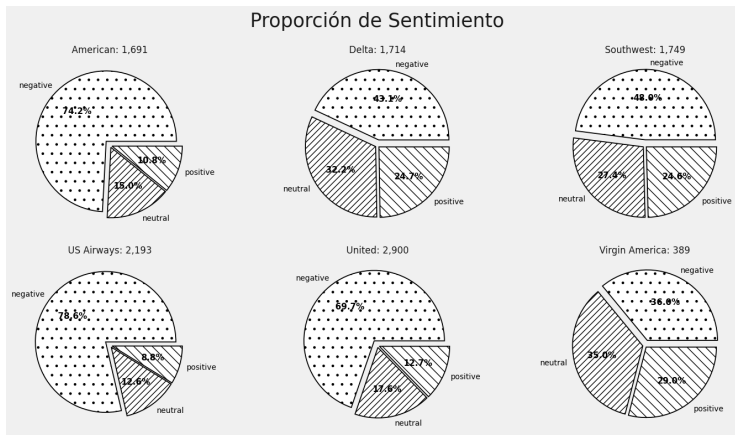


Figura: Proporción de sentimiento de tweets por Aerolínea

Análisis Exploratorio pt. 2

Cómo se puede notar, hay palabras que se repiten en las tres gráficas, de aquí podemos darnos una idea de que el sentimiento reflejado en un tweet no proviene de las palabras aisladas, si no del contexto en el que se usan. Requerimos de un modelo que tome esto en consideración.



Figura: Palabras más frecuentes en los tweets por sentimiento.

Modelación

Dado que queremos evitar sesgos positivos o negativos, es decir, evitar que el modelo tienda a predecir más tweets como positivos que como negativos, tomamos la decisión de balancear las muestras para el entrenamiento y evaluar en un conjunto que refleje la proporción verdadera del sentimiento de los tweets. Para balancear las clases minoritarias, usamos el método de oversampling.

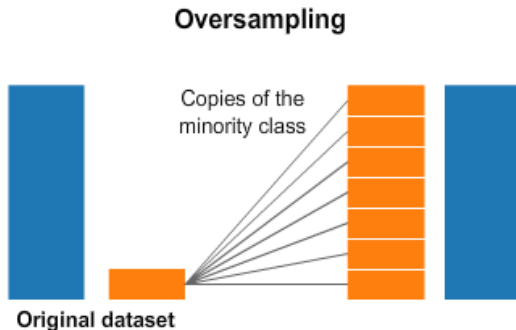
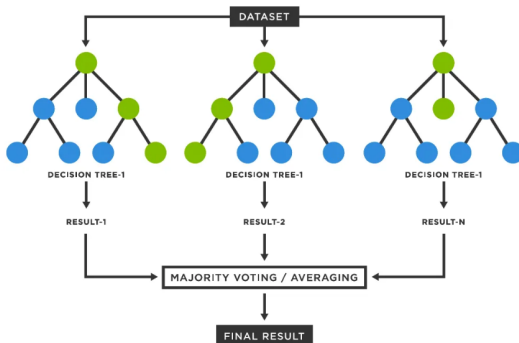


Figura: Oversampling

Modelación Random Forest

Es importante mencionar que para ingresar el texto a este modelo se le realizó una transformación usando el módulo de python llamado **TfidfVectorizer**. El **TfidfVectorizer** transforma el conjunto de tweets en una matriz numérica basada en la ponderación *TF-IDF* (*Term Frequency-Inverse Document Frequency*), que mide la relevancia de cada término dentro de un tweet y en el corpus completo. Se ajustó un Random Forest y se utilizó Cross Validation para escoger el número de árboles.



Modelación LSTM

Como segundo modelo, proponemos una red recurrente integrada con una LSTM (Long Short-Term Memory) Bidireccional. Antes de ingresar el texto a la Red, se utilizó un módulo de python llamado Word2Vec para inicializar los embeddings, estos son una representación vectorial (de una longitud a elegir) para cada una de las palabras de nuestro corpus.

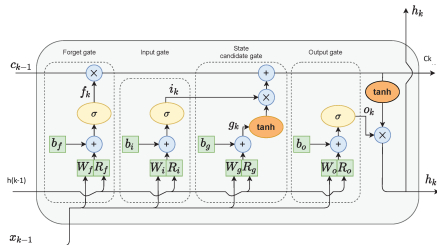


Figura: Gráfico de LSTM

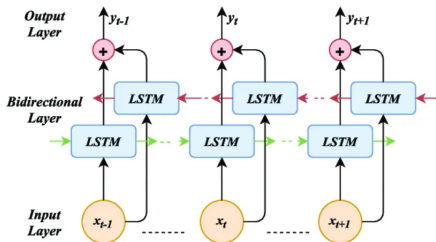


Figura: LSTM Bidireccional

Resultados Random Forest pt 1.

Se escogió el número de árboles usando CV. Podemos ver que el Accuracy promedio va subiendo hasta que llega a los 200 árboles, a partir de ahí baja, por lo que decidimos quedarnos con el modelo que tiene 200 árboles. El Top 10 de variables más importantes para el Random Forest las podemos ver en la Figura de la derecha.

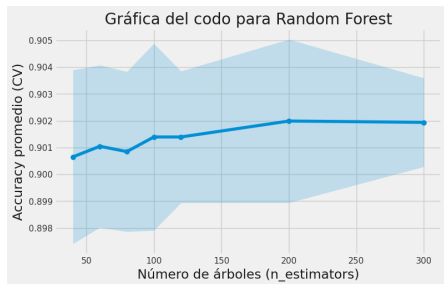


Figura: Gráfico del codo Random Forest.

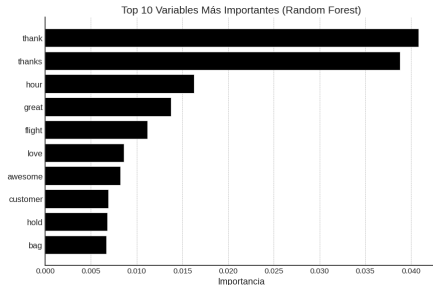


Figura: Variables más importantes.

Resultados Random Forest pt 2.

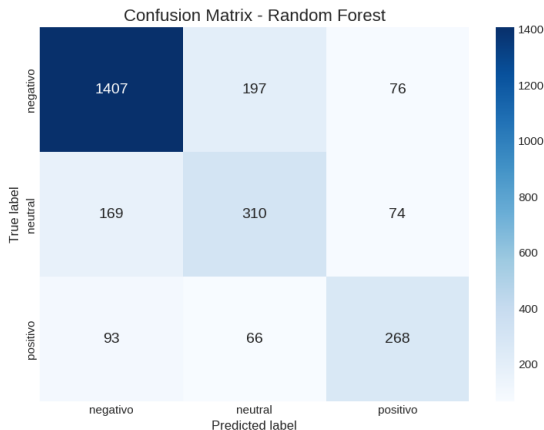
El modelo Random Forest alcanzó un *accuracy* global del 75 % en validación, aunque su desempeño varía entre clases. La categoría *negative* es la mejor identificada, con precisión, *recall* y F1-score de 0.84, lo que indica patrones claros y fáciles de detectar. En contraste, la clase *neutral* resulta la más difícil, con métricas alrededor de 0.55, evidenciando frecuentes confusiones. Las medias macro (0.68) y ponderada (0.75) confirman que, aunque el rendimiento general es aceptable, existe una notable disparidad entre categorías: el modelo es sólido para negativos, adecuado para positivos y limitado en la detección de textos neutrales.

| Clase | Precisión | Recall | F1-score | Soporte |
|---------------------|-----------|--------|----------|---------|
| negative | 0.84 | 0.84 | 0.84 | 1680 |
| neutral | 0.54 | 0.56 | 0.55 | 553 |
| positive | 0.64 | 0.63 | 0.63 | 427 |
| accuracy | | | 0.75 | 2660 |
| macro avg | 0.68 | 0.68 | 0.68 | 2660 |
| weighted avg | 0.75 | 0.75 | 0.75 | 2660 |

Cuadro: Reporte de clasificación para el modelo Random Forest.

Resultados Random Forest pt. 3

La matriz de confusión puede verse en el Figura de abajo. El color azul fuerte prevalece en la diagonal de la matriz; sin embargo pareciera haber algo de confusión entre los tweets neutrales y negativos, pues en ocasiones los negativos se predicen como neutrales y viceversa.



Resultados LSTM pt. 1

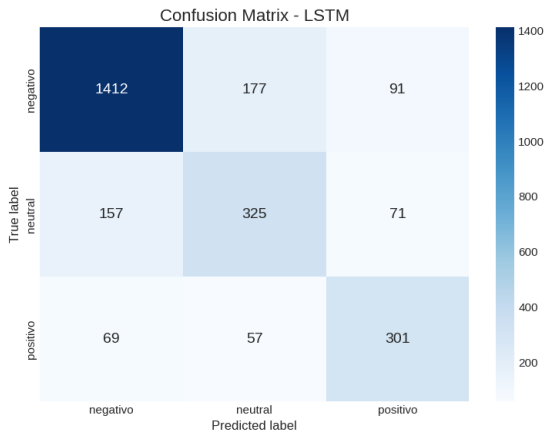
El modelo LSTM alcanzó un *accuracy* del 77 % en validación, superando ligeramente al Random Forest y mostrando mayor equilibrio entre clases. La categoría *negative* obtuvo las mejores métricas , superando claramente al modelo clásico. La clase *neutral* mejora de forma moderada (0.58–0.59), aunque sigue siendo la más difícil de identificar. La clase *positive* logra métricas intermedias, con un F1-score de 0.68 y *recall* de 0.70. En general, las medias macro (0.70) y ponderada (0.77) reflejan que la LSTM ofrece un desempeño más sólido y balanceado.

| Clase | Precisión | Recall | F1-score | Soporte |
|---------------------|-----------|--------|----------|---------|
| negativo | 0.86 | 0.84 | 0.85 | 1680 |
| neutral | 0.58 | 0.59 | 0.58 | 553 |
| positivo | 0.65 | 0.70 | 0.68 | 427 |
| accuracy | | | 0.77 | 2660 |
| macro avg | 0.70 | 0.71 | 0.70 | 2660 |
| weighted avg | 0.77 | 0.77 | 0.77 | 2660 |

Cuadro: Reporte de clasificación para el modelo LSTM.

Resultados LSTM pt. 2

En la Figura de abajo, podemos apreciar la matriz de confusión del modelo, vemos que el color azul fuerte prevalece en la diagonal de la matriz; sin embargo pareciera haber algo de confusión entre los tweets neutrales y negativos, pues en ocasiones los negativos se predicen como neutrales y viceversa.



Elección del mejor modelo y Limitantes

En conjunto, el modelo LSTM demuestra ser ligeramente más competente que el Random Forest para la clasificación tanto de textos positivos como neutrales, manteniendo un aceptable desempeño en la detección de textos negativos. El modelo tiene algunas limitantes

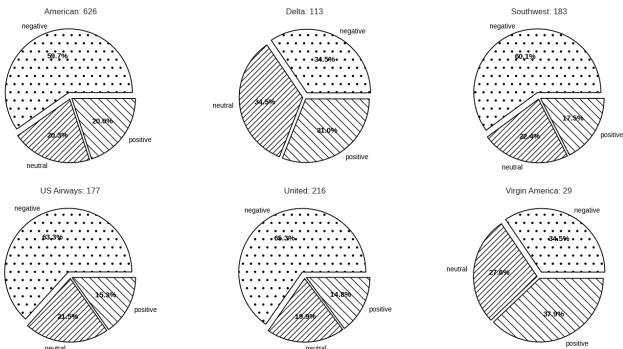
- Confunde muchos tweets entre neutrales y negativos
- Después de la lematización nos quedamos con casi 6000 palabras, considero que se requiere un corpus más grande.

| text | true_sentimen | pred_sentiment |
|--|---------------|----------------|
| @VirginAmerica What @dhepburn said. | 1 | 1 |
| @VirginAmerica plus you've added commercials to the experience... tacky. | 2 | 1 |
| @VirginAmerica I didn't today... Must mean I need to take another trip! | 1 | 1 |
| @VirginAmerica it's really aggressive to blast obnoxious "entertainment" in your guests' faces & they have little recourse | 0 | 0 |
| @VirginAmerica and it's a really big bad thing about it | 0 | 2 |
| ... | ... | ... |
| @AmericanAir WORST SERVICE EVER!! Delayed flights for more than 5 hours plus you missed my bag! And your employees are rude | 0 | 0 |
| @AmericanAir thanks for the generic computer generated response. How about you accommodate your travelers instead of just saying sorry | 0 | 0 |
| @AmericanAir they are giving cots to the people that did not get hotel rooms... that is terrible.. | 0 | 0 |
| @AmericanAir I did twice got a letter back saying that your company "doesn't issue refunds for phone bills" helpline will be shut off | 0 | 0 |
| @AmericanAir oh, yeah. I guess those are two different things. 3 am does weird things to my brain. Thanks again! xox | 1 | 2 |

Implementación del Modelo

A partir de este gráfico las Aerolíneas pueden saber cómo está su reputación para cualquier periodo de tiempo dado. Esto puede llegar a ser muy útil después de implementar políticas nuevas, anuncios sobre el servicio del cliente y estrategias comerciales. En la práctica, de poseer información diaria, o semanal podríamos ir calculando KPI's sobre la reputación de la Aerolínea en función de las predicciones y visualizar series de tiempo para analizar su evolución.

Proporción de predicciones de sentimiento por aerolínea



- Rivera, M. *Introducción a Redes Neuronales Recurrentes (RNN)*. 2018.
- Zhang, A. et al. *Dive into Deep Learning*, capítulo Word2Vec.
- Lim, S. *Twitter Sentiment Analysis — EDA and ML/DL*. Kaggle.
- Material de cursos de Deep Learning e Introducción a la Ciencia de Datos del CIMAT.

¡Gracias!