NOMBRE DEL TRABAJO

4426__1723825472P01.pdf

AUTOR

Deivis Garcias

RECUENTO DE PALABRAS

13183 Words

RECUENTO DE CARACTERES

70529 Characters

RECUENTO DE PÁGINAS

25 Pages

TAMAÑO DEL ARCHIVO

1.0MB

FECHA DE ENTREGA

Aug 16, 2024 11:35 AM GMT-5

FECHA DEL INFORME

Aug 16, 2024 11:35 AM GMT-5

8% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 5% Base de datos de Internet
- 3% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossref
- 8% Base de datos de trabajos entregados

Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 15 palabras)

Indonesian Journal of Electrical Engineering and Computer Science
Vol. 99, No. 1, Month 2099, pp. 1~1x
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v99.i1.pp1-1x

Análisis de sentimientos en Twitter con stacking ensemble para el índice de aprobación política de la presidencia del estado peruano

Deivis R. Garcia[1], Juan M. Villarroel[2], William S. Barzola [3,] Juan J. Soria[4], Nemias Saboya[5]

[1,2,3,4,5]Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Lima, Perú

Article Info

Article history:

Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy

Keywords:

Stacking ensemble
Sentiment analysis
Machine learning
The approval rating
Twitter

ABSTRACT

La creciente complejidad de los problemas políticos de un país destaca la importancia de comprender la percepción pública, para buscar alternativas y estrategias que permitan mejorar el bienestar de la población. Este estudio empleó el método stacking ensemble como una alternativa para mejorar la precisión en las métricas de clasificación y evaluar el índice de aprobación política del estado peruano mediante el análisis de sentimientos en Twitter. Con un enfoque metodológico riguroso que abarcó desde la recolección de datos hasta su presentación, el estudio logró una clasificación eficiente utilizando una muestra de 8724 tuits recolectados de ciudadanos peruanos del año 2023. Se generaron gráficos dinámicos que ilustran las puntuaciones de clasificación de sentimientos, lo que facilita una interpretación intuitiva. La investigación tuvo como objetivo mejorar las métricas de precisión, recall, exactitud y F_1-score en la clasificación del análisis de sentimientos. Los resultados de los modelos individuales mostraron que Support Vector Machine alcanzo un (accuracy=75%, MSE = 0.306). Sin embargo, el método stacking ensemble logró una accuracy del 77% y un MSE de 0.2846, lo cual es más óptimo en comparación con los métodos individuales.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Nemias Saboya

Facultad de Ingeniería y Arquitectura
Universidad Peruana Unión
Ñaña, Lima, Perú
Email: saboya@upeu.edu.pe

1) INTRODUCCION

En los últimos años, los gobiernos y políticos buscan conocer la opinión pública sobre sus acciones y decisiones, esta información se categoriza en términos de aprobación o desaprobación respectivamente [1]. En Perú al igual que en otros estados del mundo, la determinación del índice de aprobación de los líderes políticos se lleva a cabo mediante encuestas mensuales realizadas por empresas especializadas, estos resultados se utilizan para realizar estudios y pronosticar la tendencia del índice de aprobación [2]. Para un país en desarrollo

como Perú, esto puede tener graves consecuencias como fuga de capitales, recesión, inestabilidad económica, aumento del desempleo y desinversión. Hoy en día los líderes políticos buscan una alta tasa de aprobación, como métricas de solides de su gobierno y aceptación en la comunidad nacional e internacional [3]. Este índice de aprobación está fuertemente influenciada por las circunstancias y eventos que afectan a un país. Factores como decisiones políticas, acontecimientos sociales o situaciones económicas, pueden impactar significativamente en el índice de aprobación de la sociedad [4]. De acuerdo con los resultados de la encuesta realizada en el Perú por Ipsos en el año 2023, evidencia que el actual presidente del Perú enfrenta un marcado índice de desaprobación del 83% por parte de la población. La desaprobación según los resultados de Ipsos atribuye a la incapacidad del gobierno para abordar cuestiones contractuales, como corrupción, delincuencia y pobreza. El respaldo del 10% refleja un apoyo limitado, indicando una desconexión entre el gobierno y los ciudadanos. Además, un 7% de la población opta por no expresar opiniones sobre temas políticos [5]. Estas encuestas se basan en la recolección de datos a través de muestras extraídas de la población, buscando capturar la diversidad de opiniones políticas y sociales tanto a nivel nacional como internacional. Sin embargo, estas encuestas no son 100% confiables, ya que ofrecen una imagen desfasada de las percepciones, debido a que no reflejan los cambios diarios de opiniones públicas [6]. En este contexto, el análisis de sentimiento se convierte en una herramienta esencial para perfeccionar las evaluaciones, brindando una comprensión más precisa de las opiniones del público [7]. El análisis de sentimientos es una técnica de procesamiento del lenguaje natural (PLN) que es utilizada para enseñar a las máquinas a detectar y reconocer información emocional y sentimental a partir de textos determinados [8]. Este algoritmo permite interpretar el significado, la orientación y carga de emociones presentes de un texto, por ende, suele emplear datos de entradas provenientes de mensajes de textos, los cuales pueden estar almacenados en hojas de cálculo estructuradas, documentos de textos, páginas web o repositorios no estructurados [9]. Esta técnica no solo abarca evaluación de emociones, sino también mejora la comprensión de opiniones en profundidad [10]. Su característica principal consiste en categorizar la polaridad mediante puntuaciones de satisfacción, clasificando las expresiones digitales de los usuarios en sentimientos positivos, negativos y neutros. Esto posibilita obtener una visión más integral y equilibrada de las opiniones expresadas en línea. [11], [12]. A medida que la cantidad de datos crece, realizar este análisis demanda cada vez más tiempo y esfuerzo. Una solución efectiva para abordar estos desafíos es la aplicación de técnicas de aprendizaje automático. [13]. Desde la perspectiva del aprendizaje automático, el análisis de sentimientos es un problema que requiere aprendizaje supervisado, es decir, clasificación. Por tanto, para entrenar un modelo de aprendizaje automático se necesitan datos con etiquetas de sentimiento [14]. Este método computacional, conocido como aprendizaje automático, se emplea con el propósito de mejorar el rendimiento y generar predicciones precisas [15], [16]. A partir del conjunto de datos proporcionado, la máquina extrae patrones y crea su propio conjunto de reglas [17] [18]. Existen tres categorías distintas de aprendizaje automático: aprendizaje por refuerzo, aprendizaje no supervisado y aprendizaje supervisado [19]. Existen diversas investigaciones enfocadas en el análisis de sentimiento en el ámbito social como, el análisis de críticas de películas [20], la identificación de ciberagresión [21], y la violencia contra las mujeres [22]. Por consiguiente, es crucial no solo en tendencias emocionales, sino que también revela polaridad de las opiniones, ofreciendo información valiosa para mejorar la toma de decisiones [23], [24]. En este contexto Twitter ahora llamado “X”, se destaca como una fuente de datos fundamental para el análisis de sentimiento gracias a su elevada actividad diaria en publicaciones llegando a ser un canal crucial para la difusión de información, opiniones y tendencias, brindando datos valiosos que permiten comprender las dinámicas sociales y políticas de manera inmediata. [25]. A pesar de la abundancia de datos disponibles en twitter, muchos estudios sugieren que los modelos actuales de clasificación para el análisis de sentimiento no proporcionan un análisis preciso [26]. Investigaciones recientes como las de Nadia da Silva demuestran que la precisión en la clasificación de sentimientos de los tweets mejoran significativamente al emplear métodos de ensamblaje [27]. Otros estudios respaldados por los autores como Ammar Hassan, Sam Clark y Joseph Prusa, enfatizan la importancia de combinar clasificadores independientes y diversificados para lograr una mayor precisión, F_1 score y recall en la clasificación de sentimientos [28],[29],[30].

En consecuencia, el propósito de esta investigación es presentar un método de ensamblaje con el objetivo de mejorar la precisión, el recall, la exactitud y el puntaje F_1 en la clasificación del índice de aprobación política de la presidencia del Perú utilizando un conjunto de datos recopilados de la cuenta presidencial de Twitter. Posteriormente, se exploraron diversos modelos de clasificación, como Bernoulli NB, Multinomial NB, Logistic Regression, Random Forest, Decision Tree, Support Vector Machine, y K-Nearest Neighbors. Estos modelos fueron estratégicamente seleccionados para integrarse de manera eficaz en un meta-modelo, el cual desempeña la función de realizar la predicción final. Este enfoque integral y avanzado tiene como objetivo proporcionar una perspectiva mejorada en el análisis de sentimientos en el contexto de los tweets vinculados a la cuenta presidencial de Perú en Twitter.

2) ARQUITECTURA DEL MODELO PROPUESTO

Los métodos ensamblados en el aprendizaje automático involucran el uso de múltiples modelos base para predecir resultados, en lugar de un solo modelo grande, se entrenan varios modelos más pequeños de forma independiente, conocidos como aprendices débiles, estos modelos generan predicciones individuales que se combinan para mejorar la precisión y el rendimiento general del sistema, los tres tipos principales de métodos de conjuntos son Bagging, Boosting y Stacking, que se utilizan comúnmente para realizar predicciones y clasificaciones [31].

2.1 Bagging

Este método entrena múltiples modelos bases simultáneamente con muestras aleatorias del conjunto original, las predicciones de cada modelo base contribuyen con su salida y la predicción final se obtiene mediante votación [32].

2.2 Boosting

Este método entrena modelos bases uno tras otro, cada modelo toma mayor importancia de los datos que fueron clasificados erróneamente por el modelo anterior, lo que permite que los nuevos modelos se centren en áreas específicas y disminuyan los errores de predicción. [33].

2.3 Stacking

El método de stacking, también conocido como apilamiento, entrena varios modelos base simultáneamente, similar al bagging, sin embargo, se diferencia en que no se limita a votaciones para combinar las salidas de los modelos base y generar la predicción final; en cambio, entrena un meta-aprendiz o meta-modelo utilizando las predicciones obtenidas de los modelos base, construyendo una relación entre las salidas de los modelos y la predicción final [34]. Para este estudio, se optó por emplear el método de stacking debido a su mayor flexibilidad y capacidad de adaptación para fusionar múltiples modelos. A diferencia de otras opciones disponibles, el stacking brinda la posibilidad de integrar una diversidad de modelos base a través de un meta-modelo que aprende a combinar sus predicciones. Su habilidad para ajustarse a distintos modelos y optimizar las combinaciones lo posiciona como la elección ideal para nuestro estudio. En la Figura 1 se ilustra de manera gráfica la arquitectura utilizada en el método de stacking ensemble.

Comentado [NS1]: ¿?

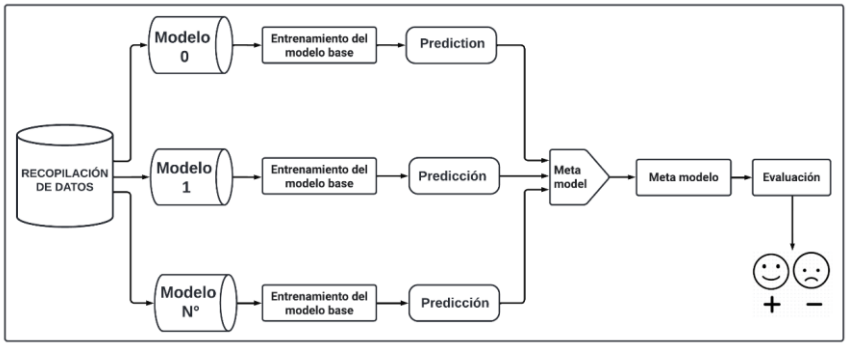


Figura 1. Arquitectura del Método Stacking Ensemble

Comentado [NS2]: Este tititlo pareciera como si asi fuera teoricamente el metodo la pregunta radica justamente en eso? Por lo que lei esto es la propuesta entonces la recomendacion es colocar. Como lo describes "Arquitectura del metodo Stacking Ensemble"

El proceso de stacking ensemble comienza con la recopilación y preparación de datos, que incluye la selección, limpieza y división en conjuntos de entrenamiento y pruebas[35]. Luego, se seleccionan y entrenan modelos base de forma independiente para asegurar su robustez y capacidad de generalización[35]. Las predicciones de estos modelos base se utilizan para crear un meta-modelo, que aprende a combinarlas óptimamente[36]. Este meta-modelo se entrena utilizando las predicciones y los resultados reales para mejorar la precisión y el rendimiento[36]. Finalmente, se evalúa el rendimiento del modelo de apilamiento comparando sus predicciones con los valores reales del conjunto de pruebas, utilizando métricas como precisión, puntaje F1 score, curva ROC, AUC y accuracy[37].

3) MATERIALES Y MÉTODOS

En esta sección se describe el método utilizado en la investigación, que consiste en la implementación del método de stacking ensemble diseñado para clasificar el índice de aprobación política del gobierno peruano. Este método emplea la técnica de análisis de sentimiento a través de información obtenida en twiter hoy en dia reconocida como 𝕏, utilizando modelos de clasificación de aprendizaje automatico, que incluyen Bernoulli NB, Multinomial NB, Logist Regression, Random Forest, Decision Tree, Support Vector Machine y K-Nearest Neighbors. Este enfoque destaca su relevancia en el área de investigación al ofrecer mejoras notables en el rendimiento de la clasificación. Cabe resaltar que, en el marco de este estudio se emplearón tweets provenientes de la cuenta presidencial del Perú(@presidenciaperu) correspondientes a junio 2023. Asi mismo se detallan las diversas etapas realizadas que incluyen la recopilación de datos, preprocesamiento, construcción del método ensamble y evaluación del método ensamble, como se muestra en la figura 2.

Comentado [C3]: Areglar la imagen

Comentado [SS4R3]: Listo

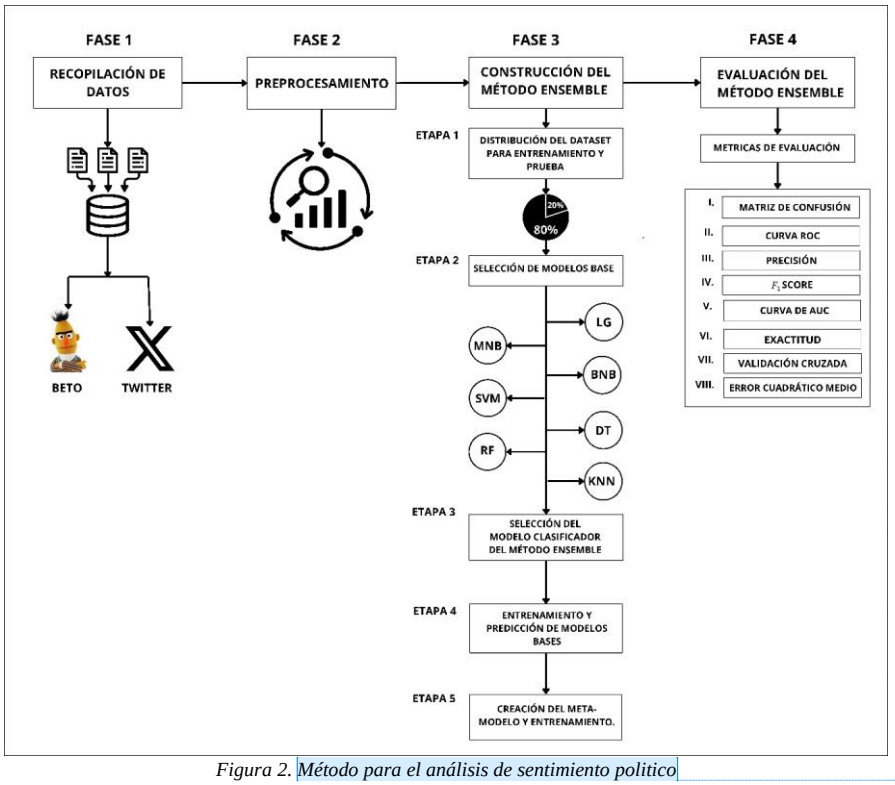


Figura 2. Método para el análisis de sentimiento politico

Comentado [NS5]: Metodo para el analisis de sentimiento politico

A. Fase 1: Recopilación de datos

Hoy en día, enfrentamos una vasta cantidad de datos no estructurados en la web, desde textos hasta multimedia, lo que plantea desafíos en su gestión y análisis. [39]. En base a la información previa, se recolectó los datos de 𝕏 mediante el uso de su API proporcionada por su plataforma, seleccionando los tweets de ciudadanos peruanos que utilizaron el hashtag @presidenciaperu, correspondientes a junio del 2023. La eleccion del año 2023 se argumenta al evidente aumento de corrupción de los políticos del estado peruano. Luego, se aplicó el modelo BERT utilizando su método Beto Sentiment Analysis para predecir la polaridad negativa, neutra y positiva de los textos en español. Al finalizar la predicion, se creó un nuevo conjunto de datos que incorpora la polaridad de los textos. Además, se proporciona el código del algoritmo utilizado en el método stacking ensemble y el conjunto de datos aquí.

B. Fase 2: Preprocesamiento

El preprocesamiento es esencial para lograr un conjunto de datos coherente y de alta calidad, eliminando caracteres y símbolos innecesarios y excluyendo características irrelevantes del texto. [40]. En esta fase, se realizó una meticulosa limpieza del conjunto de datos mediante la eliminación de textos no relacionados con la investigación y la eliminación de elementos no esenciales como símbolos, hashtags, íconos y espacios en blanco. Además, se uniformizó el texto convirtiendo todas las letras mayúsculas a minúsculas. Posteriormente, se efectuó la tokenización de los tuits, fraccionando los textos en unidades más pequeñas denominadas tokens. Seguidamente, se implementó la lematización, que consiste en reducir las palabras a sus formas base o lemas, facilitando la agrupación de términos similares. Este proceso es fundamental para asegurar que los datos estén en un formato adecuado y óptimo para el análisis de sentimientos, lo que permite una interpretación precisa y mejora la identificación de patrones y tendencias en la percepción pública.

Comentado [NS6]: Como hiciete el preprocesamiento y para que te sirvio

C. Fase 3: Construcción del método ensamblado

Para construir el método stacking, se seleccionaron varios modelos base, incluidos Bernoulli NB, Multinomial NB, Regresión Logística, Random Forest, Decision Tree, Support Vector Machine y K-Nearest Neighbors (KNN), debido a sus diferentes capacidades de manejar diversos tipos de datos y mejorar el rendimiento en tareas de clasificación. Posteriormente, se dividió el conjunto de datos en conjuntos de entrenamiento y prueba siguiendo la regla de Pareto 80/20, destinando el 80% del dataset para el entrenamiento de los modelos y el 20% restante para la validación y pruebas. Esto permite evaluar el rendimiento en datos no vistos durante el entrenamiento y garantizar una evaluación más realista de la capacidad de generalización. Dentro del método stacking, se otorgó especial importancia al clasificador de regresión logística como clasificador final, desempeñando un papel crucial en la combinación de predicciones al aprender a partir de las salidas de los modelos base y generar la predicción final. La regresión logística es especialmente adecuada para este rol debido a su capacidad para manejar relaciones lineales y su simplicidad, lo que facilita la interpretación de los resultados. Los pasos para desarrollar el método ensemble y la figura 3 ilustran detalladamente la arquitectura empleada y los siete modelos base utilizados en el método stacking ensemble.

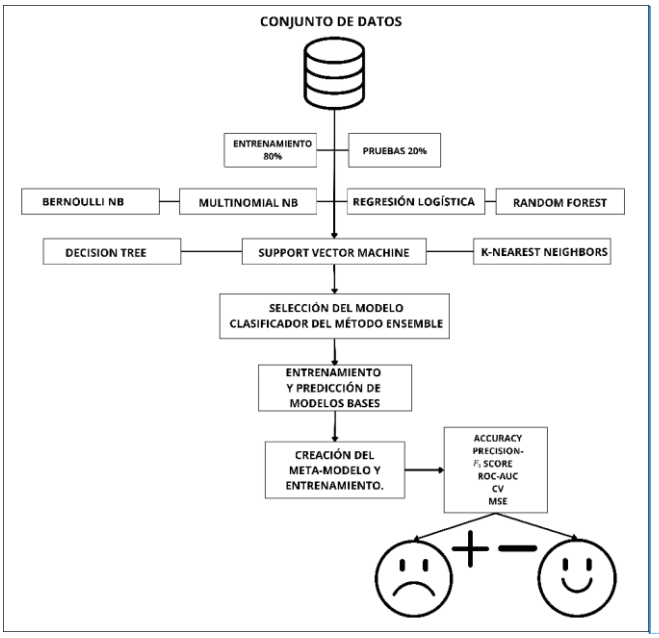


Figura 3. Estructura de la Fase 3 del Proceso de Stacking Ensemble

Comentado [NS7]: Analizar…

i) **Etapa 1: Distribución del Dataset para Entrenamiento y Prueba**
Se realizó la división del conjunto de datos en conjuntos de entrenamiento y prueba siguiendo la regla de Pareto 80/20, asignando el 80% del dataset para entrenar los modelos y el 20% restante para la validación y pruebas.

ii) **Etapa 2: Selección de Modelos Base**

Comentado [NS8]: No coincide con la numeracion de la viñetas

Se seleccionaron los modelos Bernoulli NB y Multinomial NB por su eficiencia en el manejo de datos textuales y su aptitud para dimensiones altas, claves para el análisis de tweets. La Regresión Logística fue elegida debido a su capacidad para ofrecer interpretaciones claras y manejar relaciones lineales, vital para entender las tendencias en las opiniones políticas. Random Forest y Decision Tree se prefirieron por su robustez ante datos imprecisos y su habilidad para identificar patrones no lineales en opiniones diversas. Support Vector Machine destacó por su rendimiento en espacios de alta dimensionalidad y su exactitud en definir hiperplanos para diferenciar sentimientos. K-Nearest Neighbors se incorporó por su método intuitivo para evaluar la cercanía entre datos, facilitando la agrupación de opiniones parecidas. En la Tabla 1 se mencionan los ciete modelos base utilizados en el método de stacking ensemble.

Indonesian Journal of Electrical Engineering and Computer Science
Vol. 99, No. 1, Month 2099, pp. 1~1x
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v99.i1.pp1-1x

Tabla 1 Modelos Bases Seleccionados

Nombre	Formula		Algoritmo
Logistic Regression [41], [42].	$decisión(x) = \begin{cases} 1\,; if\ \ P(y=1\|x) > 0.5 \\ 0\,; \quad otherwise \end{cases}$	(1)	***TRAINLOGISTICREGRESSION*** 1. For i←1 to k 2. For each training data instance di: 3. Set the target value for the regression to $Z_I \leftarrow \frac{y_j - P(1\|d_j)}{\left[P(1\|d_j).\left(1-P(d_j)\right)\right]}$ 4. Initialize the weight of instance d_j to $P(1\mid d_j).\left(1-P(1\mid d_j)\right)$ 5. Finalize a f(j) to the data with class value (zj) & weights (wf) Classification Label Decision 6. Assign (class label: 1) if $P(1\mid d_j) > 0.5$, otherwise (class label: 2)
Multinomial NB [44]	$P(d\|q) \propto P(d)\prod_{t\in q}(1-\lambda)P(t\|M_c) + \lambda P(t\|M_d)$ $P(d\|q) \propto P(d)\prod_{t\in q} P(t\|M_d)$	(2) (3)	***TRAINMULTINOMIALNB(C, D)*** 1. V ← EXTRACTVOCABULARY(D) 2. N ← COUNTDOCS(D) 3. for each c ∈ C 4. do Nc ← COUNTDOCSINCLASS(D, c) 5. $prior[c] \leftarrow \frac{N_c}{N}$ 6. $text_c$← CONCATENATETEXTOFALLDOCSINCLASS(D, c) 7. for each t ∈ V 8. do T_{ct}← COUNTTOKENSOFTERM($text_c$, t) 9. for each t ∈ V

Bernoulli NB [44], [45]	$P(x = \text{éxito})$	(4)
	$P(x = \text{fracaso}) = 1 - P$	(5)

10. do condprob[t][c] ← $\frac{T_{et}+1}{\sum_{t'}(T_{tc'}1+)}$
11. return V, prior, condprob

APPLYMULTINOMIALNB(C, V, prior, condprob, d)
1. w ← EXTRACTTOKENSFROMDOC(V, d)
2. for each c ∈ C
3. do score[c] ← log prior[c]
4. for each t ∈ W
5. do score[c] += log condprob[t][c]
6. return arg $max_{c \in C}$ score[c]

TRAINBERNOULLINB(C, D)
1. V ← EXTRACTVOCABULARY(D)
2. N ← COUNTDOCS(D)
3. for each c ∈ C
4. do N_c ← COUNTDOCSINCLASS(D, c)
5. prior[c] ← $\frac{N_c}{N}$
6. for each t ∈ V
7. do Nct ← COUNTDOCSINCLASSCONTAININGTERM(D, c, t)
8. condprob[t][c] ← $\frac{N_{ct}+1}{N_c+2}$
9. return V, prior, condprob
10. APPLYBERNOULLINB(C, V, prior, condprob, d)
11. V_d ← EXTRACTTERMSFROMDOC(V, d)
12. for each c ∈ C
13. do score[c] ← log prior[c]
14. for each t ∈ V
15. do if t ∈ V_d
16. then score[c] += log condprob[t][c]

17. else score[c] += log(1 - condprob[t][c])
18. return arg $max_{c \in C}$ score[c]

Support Vector Machine [46], [47]

$$\begin{cases} w^T x_i + b \geq 1 \ , \ si\ y_i = 1 \\ w^T x_i + b \leq -1\ , si\ y_i = -1 \end{cases} ;\ \forall (x_i; y_i), i\ 1,2,3,..,n \quad (6)$$

7 **Input**: α := 0 or α := partially trained SVM, X and y

1. C := Approximate upper bound constant value
2. Loop: $\forall \{x_i\ ,\ y_i\}\ and\ \{x_j\ , y_j\}$
7 3. do
4. Optimize αi and α_j
5. End loop
6. Until no changes in α (or some resource constraint encounters)

Output: Hold merely support vectors (SV) $(\alpha_i > 0)$

Decision tree [43], [48], [49], [50]

$$Entropy(S) = \sum_{i=1}^{n} - Pi\ \times \log_2(Pi) \quad (7)$$

$$Gain(S, A) = Entropy(S)\ \sum_{i=1}^{n} \frac{|Si|}{|S|}\ \times Entropy(S) \quad (8)$$

2 **Input:** Data

1. Set the value of k
2. Loop: 1 to N // To get predicted class 2.1. Calculate the distance Di (Euclidean/Cosine/Chebyshev) between Data instance in training data and test data
3. Increasingly arrange the computed distances (Di)
4. Populate the upper k results from the arranged list
5. Pick up the most frequent class from the list

Output: resultant class

Modelo Random Forest [21], [51], [52]

Tabel 1. Funcionamiento del Random Forest

Modelo: Random Forest for clasificación
1. Sea N el número de casos de entrenamiento y M el número de variables.
2. Seleccione el valor de m, para determinar la decisión en un nodo dado, que debe ser menor que el valor de M.
3. Varíe y elige N para el árbol (completando todos los casos de entrenamiento) y utiliza el resto de los casos de prueba para estimar el error.
4. Para cada nodo del árbol, elija aleatoriamente m. Obtenga el voto de cada árbol y clasifique según el voto de la mayoría.

Modelo K-nearest neighbors (KNN) [53], [54], [57],

$$\vec{\mu}(c) = \frac{1}{|D_c|}\sum_{d \in D_c} \vec{v}(d) \quad (9)$$

Input: training set S with F features

1. Set the value of k
2. Loop: 1 to N // To get predicted class
   2.1. Calculate the distance Di (Euclidean/Cosine/Chebyshev) between Data instance in training data and test data
3. Increasingly arrange the computed distances (Di)
4. Populate the upper k results from the arranged list
5. Pick up the most frequent class from the list

Output: resultant class

TRAIN-kNN (C, D)

1. D′←PREPROCESS(D)
2. k←SELECT-k(C,D′)
3. Return D′,k APPLY-kNN (C,D′, k, d)
   1. $S_k \leftarrow$ COMPUTENEARESTNEIGHBORS(D′,k,d)
   2. for each $c_j \in C$
      1. do $p_j \leftarrow \left|\frac{|S_k \cap c_k|}{k}\right|$
4. return arg $max_j\ P_j$

iii) Etapa 3: Selección del modelo clasificador del método stacking

Se selecciono el modelo de Regresión Logística como clasificador para el método de stacking ensemble, debido a su simplicidad y la claridad con que permite interpretar cómo las características de los tweets impactan las predicciones finales. Su eficiente uso de recursos y su limitada propensión al sobreajuste contribuyen a una mejor generalización del modelo. Esta elección también potencia significativamente el rendimiento de clasificación del ensemble, asegurando predicciones precisas y robustas en análisis políticos.

iv) Etapa 4: Entrenamiento y Predicción de Modelos Bases

Los modelos base seleccionados se entrenaron de manera independiente, empleando el 80% del conjunto de datos disponible. Esta estrategia permitió maximizar las fortalezas particulares de cada modelo, asegurando un entrenamiento sólido y adaptado a cada algoritmo. Posteriormente, se usaron los modelos ya entrenados para hacer predicciones sobre el 20% restante del conjunto de datos, aplicando así de manera práctica los conocimientos obtenidos durante la fase de entrenamiento

v) Etapa 5: Creación del Meta-Modelo y Entrenamiento.

Las predicciones generadas por los modelos bases fueron empleados para desarrollar el meta-modelo del método de stacking ensemble. Este meta-modelo permitió aprender eficazmente cómo fusionar estas predicciones, con el objetivo de maximizar la sinergia entre los modelos y capitalizar sus fortalezas individuales. De esta manera, se logró superar las capacidades de los modelos por separado, aumentando la precisión y la robustez del metodo. Además, se realizó el entrenamiento utilizando tanto las predicciones combinadas como los datos reales, lo que permitió un ajuste óptimo que mejoró la precisión y el rendimiento general del método. Este enfoque integrado mejoró significativamente la evaluación política, optimizando la integración de las predicciones y la calibración del modelo para obtener resultados superiores.

D. Fase 4. Evaluación del Método Stacking

Comentado [NS9]: Tiene qe ser igual al metodo staking

La evaluación del rendimiento del método stacking se realizó comparando las predicciones del modelo de apilamiento con los valores reales del conjunto de prueba. Para medir la eficacia del método, se utilizó la validación cruzada y se aplicaron diversas métricas de la matriz de confusión, como la precisión, F_1 score, recall, la curva ROC, el AUC y el accuracy. Estas métricas proporcionaron una visión integral del rendimiento del método, permitiendo una evaluación completa de su capacidad predictiva y su adaptación a datos no vistos. Además, se ofreció una explicación detallada de cada métrica utilizada en el análisis y se realizó una comparación entre el método stacking ensemble y los modelos individuales seleccionados.

i) Matriz de Confusión

La matriz de confusión es una herramienta útil para evaluar el rendimiento de un modelo de clasificación del aprendizaje automático, es utilizada para calcular la precisión, la sensibilidad, la especificidad y el F_1 score [55]. Asi mismo se describe las 4 categorias de esta [56], [57].

Verdaderos Positivos (VP), Representa los casos en donde el modelo predijo correctamente clase positiva. En la ecuación (10) describe, la formula utilizada para calcular los VP.

15

Falsos Positivos (FP), Representa los casos donde el modelo predijo erróneamente la clase positiva, la clase verdadera era la negativa. En la ecuación (11) describe, la formula utilizada para los FP.

Verdaderos Negativos (VN), Representa los casos que el modelo predijo correctamente la clase negativa. En la ecuación (12) describe, la formula utilizada para los VN.

Falsos Negativos (FN), Representa los casos donde el modelo predijo erróneamente la clase negativa, la clase verdadera era positiva. En la ecuación (13) describe, la formula utilizada.

$VP = N$ (Número de instancias positivas correctamente clasificadas). (10)
$FP = M$ (Número de instancias negativas incorrectamente clasificadas como positivas). (11)
$VN = O$ (Número de instancias negativas correctamente calificadas). (12)
$FN = P$ (Número de instancias positivas incorrectamente clasificadas como negativas). (13)

Ademas se define la ecuación 14, donde se describe la formula para calcular la tasa de falsos positivos.

$$TFP = \frac{M}{M+O} \quad (14)$$

ii) Sensibilidad y Especificidad

La sensibilidad se define como la fracción de las verdades positivas, representando la fracción de acierto en la identificación de casos positivos por el modelo, mientras que la especificidad representa la proporción de verdaderos negativos, indicando la fracción de aciertos al identificar casos negativos por el modelo [58]. En la ecuación (15) y (16) se describen, las formulas utilizada para calcular la sensibilidad y especificidad respectivamente.

$$Sensibilidad = \frac{N}{N+P} \quad (15)$$

$$Especificidad = \frac{O}{0+M} \quad (16)$$

iii) Curva ROC

La curva ROC es una herramienta visual importante para evaluar modelos de clasificación binaria, ya que muestra la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos a medida que se ajusta el umbral de clasificación, podemos analizar el equilibrio de un modelo entre sensibilidad y especificidad utilizando la curva ROC y elegir el umbral ideal para nuestras necesidades [56].

iv) Precisión

Es la distribución de los valores totales obtenidos a partir de mediciones repetidas de una determinada magnitud, una mayor precisión corresponde a una menor dispersión y puede expresarse como la proporción de verdaderos positivos dividida por el número total de resultados positivos [58]. La fórmula utilizada para evaluar esta métrica se expresa en la ecuación (17).

$$Precisión = \frac{VP}{VP+FP} \quad (17)$$

v) F_1 Score

El F_1 score, es una métrica consolidada al combinar sensibilidad y precisión en una única medida, ofreciendo una visión más completa del rendimiento del modelo [59]. Este enfoque es especialmente útil al equilibrar la capacidad del modelo para detectar positivos verdaderos y minimizar falsos positivos, proporcionando una evaluación más holística y equitativa de su eficacia en diversas situaciones. La fórmula utilizada para evaluar esta métrica se expresa en la ecuación (18).

$$F_1 = 2 * \frac{Precisión * Sensibilidad}{Precisión + Sensibilidad} \quad (18)$$

vi) Accuracy

Es la proximidad de los resultados de una medición al valor real, en estadística, está relacionada con el sesgo de una estimación y la proporción de resultados verdaderos (tanto verdaderos positivos (VP) como verdaderos negativos (VN)) dividida entre el número total de casos examinados (verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos) , cuyo resultado es la cantidad de predicciones positivas que fueron correctas [60]. La fórmula utilizada para evaluar esta métrica se expresa en la ecuación (19).

$$\text{Accuracy} = \frac{(VP+VN)}{(VP+FP+FN+VN)} \quad (19)$$

vii) Curva de AUC

La Curva AUC es una representación gráfica del rendimiento de un modelo de clasificación variando el umbral de decisión, esta métrica proporciona una medida numérica del rendimiento global el auc más alto indica una mejor capacidad de distinguir entre clases positivas y negativas, sin depender de un umbral de clasificación específica [61].

viii) Validacion Cruzada

La validación cruzada es una técnica para evaluar el rendimiento de un modelo dividiendo los datos en subconjuntos, entrenando el modelo en algunos y evaluándolo en otros, cuyo objetivo es estimar la capacidad de generalización del modelo y evitar el sobreajuste al proporcionar una evaluación más realista del rendimiento en datos no vistos [62].

ix) Error Cuadrático Medio

El Error Cuadrático Medio (MSE) es una métrica empleada para evaluar la calidad de un modelo de regresión. Se obtiene al calcular la media de los cuadrados de las diferencias entre los valores predichos por el modelo y los valores reales. Un MSE más bajo indica una mejor capacidad del modelo para predecir los valores de la variable objetivo. [63].

4) RESULTADOS Y DISCUCIÓN

La investigación recopiló 8274 de tuits vinculados con la cuenta presidencial del gobierno peruano, los cuales fueron recolectados de perfiles de tuits pertenecientes a ciudadanos peruanos durante el mes de junio del 2023. La elección de estos tuits se realizó mediante la identificación de palabras claves como “asesina”, “asesinos”, “delincuentes”, “cárcel”, así como palabras comunes relacionadas con el desacuerdo o desaprobación política hacia la presidencia del Perú. Estos tuits fueron utilizados para generar el conjunto de datos, el cual fue utilizado en el análisis de sentimiento mediante el método stacking ensamble con la finalidad de evaluar la polaridad de los perfiles de ciudadanos peruanos en Twitter con respecto a la aprobación política del gobierno peruano.

4.1 Resultado descriptivos

Durante la investigación se realizó un análisis descriptivo de los tweets pre procesados, en el cual se identificó la polaridad de las palabras positivas, negativas y neutras. Como resultado de este análisis, se presentó histogramas, N-gramas y tablas para visualizar y comprender mejor la distribución de las palabras en los tweets y su polaridad asociada.

4.1.1 Caracteristicas de usuarios de 𝕏

El estudio identificó características de los perfiles de ciudadanos peruanos de Twitter, las cuales están detalladas en la tabla 2. Estas características ofrecen un panorama completo de las cuentas de usuarios y sus ubicaciones, lo que permitió realizar un análisis detallado de las regiones del Perú donde la desaprobación política hacia la presidencia es más profunda. Además, se resaltan las cuentas gubernamentales con mayor mención, lo que proporciona una visión más relevante de las tendencias predominantes en las publicaciones sobre el estado político peruano. Al mismo tiempo, se mostraron los hashtags más utilizados, lo que refleja las preocupaciones más relevantes de los ciudadanos sobre la aprobación política de la presidencia del estado peruano.

Tabla 2.Caracteristicas de usuarios de Twitter

Acount Users	Location	mentioned accounts	Hashtags
https://twitter.com/ronaldgamarra	Lima, Perù	@FiscaliaPeru	IzquierdaMiserable
https://twitter.com/radiouno_pe	Puno, Perù	@Poder_Judicial	DinaRemuncia
https://twitter.com/ovidartea	Lima, Perù	@TC_Peru	DinaAsesina
https://twitter.com/RaulHAsensio	Lima, Perù	@presidenciaperu	EleccionesGeneralesYa
https://twitter.com/pattyromero11	Lima, Perù	@pcmperu	Dinaasesinarenunciaya
https://twitter.com/jackymc10	Lima, Perù	@CancilleriaPeru	DinaBuluarte
https://twitter.com/leonbravo42	*EE.UU	@MinjusDH_Peru	Presidenciaperu
https://twitter.com/YehudeSimonM	Lima, Perù	@AlbertoOtarolaP	Despiertaperudespierta
https://twitter.com/jfowks	Lima, Perù	@CCFFAA_PERU	PeruEnDictadura

* Estos individuos son ciudadanos peruanos que residen en los Estados Unidos.

4.1.2 Resultados de Frases Coincidentes con Palabras Clave de Búsqueda

El estudio llevó a cabo coincidencias de palabras claves de búsquedas, las cuales se presentan en la tabla 3. Estas frases sirvieron para identificar las opiniones relevantes para el análisis de sentimiento, ya que establecen una correlación entre los términos de búsqueda y el contenido textual. Este análisis es fundamental para comprender cómo las opiniones de interés se reflejan en el conjunto de datos recopilados, proporcionando información valiosa para la evaluación del índice de aprobación política de la presidencia del estado peruano.

Tabla 3. Sentimiento en Tweets Peruanos: Frases Coincidentes con Palabras Claves

Sentiment	Tweets de la Investigaciòn
Negative	Dina asesina infeliz.
Negative	Dina murderer, you will fall.
Negative	Dina la asesina sabrá que otarola si puede huir del país en cualquier momento y ella no el premier no necesita ninguna autorización de viaje la presidenta por ahora no puede salir del país pues no tiene reemplazo otra vez en manos de una futura presidiaria.
Negative	El pueblo peruano no sola la asesina de dina y sus delincuentes q están con ella.
Negative	No te queda de otra asesina de lo contrario terminarias presa antes de lo previsto.
Negative	Es la jefa, pero no saben lo que ellos hacen la jefa no tiene la capacidad de detener los asesinatos cometidos eres un títere e inútil la justicia te va alcanzar ahora toda poderosa te sientes protegida por un congreso igual de corrupto dina asesina.
Negative	Solo espero las horas que esa asesina sea encerrada en una cárcel los crímenes son imprescriptibles muchos años de cárcel a diferencia de castillo que tal vez cumpla su pena no por asesinato pero saldrá y volverá a candidatear.
Negative	Aunque te esfuerces diciendo que respetas los derechos humanos irán a carcel por lesa humanidad la sangre derramada no quedará impune.
Positive	Señores yo no soy simpatizante de la señora dina boluarte sin embargo debo admitir que va haciendo una labor regular a buena.
Positive	Gracias estimada presidenta fue un día histórico lleno de tradiciones viejas y nuevos miramos hacia el futuro con confianza y esperanza celebrando nuestro nuevo rey charles iii.
Neutral	Bien señores congresistas buen trabajo, pero ya se han dado cuenta de que en el sur hay un problema que requiere atención seria que este problema debe ser atendido por la y no lo esta haciendo control político pues señores.
Neutral	No es que no quiera a ella le corresponde estar en la silla presidencial y todos lo tenemos que respetar, así como tuvimos que aguantar al asno y que haya cometido su error en el golpe de estado fue una bendición para el país.

4.1.3 Análisis de Longitud de Textos por Sentimiento

El estudio elaboró un diagrama de caja que representa la longitud de los textos según el sentimiento, el cual se muestra en la figura 4. Este análisis representa la variabilidad en la longitud de los textos para cada categoría de sentimiento, y los puntos que se encuentran fuera de la caja indican valores atípicos en la longitud de los textos. Esto brinda una comprensión más profunda de la naturaleza y la intensidad de las opiniones expresadas por los ciudadanos peruanos en relación con la aprobación política de la presidencia del estado peruano. Identificar estas variaciones permite resaltar tendencias y patrones importantes en los datos que de otra manera podrían pasar desapercibidos, lo que resulta crucial para entender la percepción pública y tomar decisiones acertadas en el ámbito político.

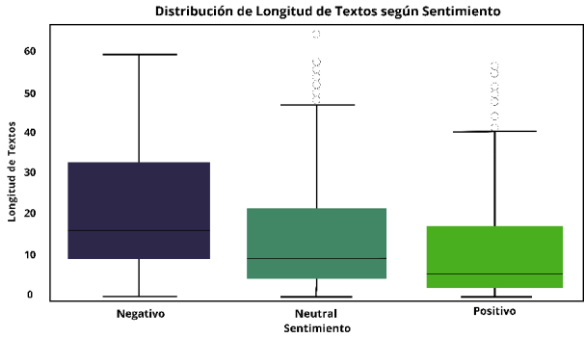


Figura 4. Distribución de Longitud de textos segùn sentimiento

4.1.4 Exploración de las palabras más comunes en el análisis de sentimiento

En el estudio, se examinaron las palabras más frecuentes, como se muestra en la figura 5. Estas palabras fueron procesadas y categorizadas utilizando el modelo Beto Sentiment Analysis, el cual predijo la polaridad de cada palabra como negativa, neutra o positiva. Los resultados revelaron que las palabras más frecuentes asociadas con la polaridad negativa (mostradas en color rojo) incluyen "asesina" con 529 menciones, "asesinos" con 318 menciones, "cárcel" con 217 menciones, 'delincuentes' con 200 menciones y "masacre" con 187 menciones. Por otro lado, las palabras neutras (mostradas en color gris) fueron "ley" con 313 menciones, "izquierda" con 180 menciones, "firma" con 148 menciones, 'modifica' con 126 menciones y "perú" con 122 menciones mientras que las palabras positivas (mostradas en color verde) incluyeron "generando" con 350 menciones, "mejor" con 330 menciones, "excelente" con 250 menciones y "felicitaciones" con 200 menciones. Este análisis permitió identificar las opiniones predominantes en la desaprobación sobre la presidencia del estado peruano.

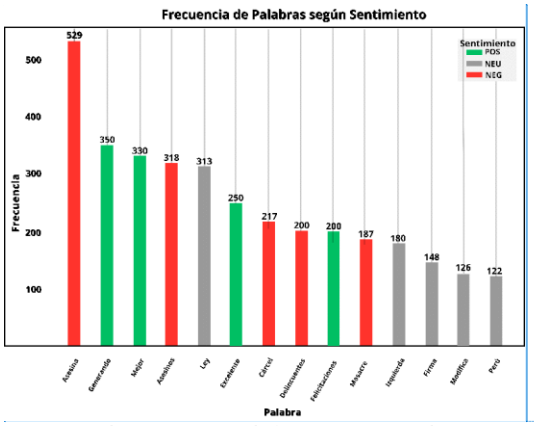


Figura 5. Distribución de Palabras Más frecuentes en el Análisis de Sentimiento

Comentado [NS10]: Tienen que estar en major resolucion

4.2 Resultado inferenciales

En este estudio, se llevó a cabo un análisis de sentimiento para evaluar la aprobación política de la presidencia del estado peruano. Se implementó el método de stacking ensamble, el cual fusiona siete modelos de clasificación de machine learning. Cada modelo fue desarrollado individualmente y sometido a validación cruzada y al cálculo del error cuadrático medio (MSE) para mejorar la precisión en la clasificación.

La validación cruzada, la matriz de confusión y el MSE se utilizaron para evaluar el rendimiento tanto de los modelos individuales como del método de stacking ensamble. La matriz de confusión evaluó el rendimiento del modelo por cada polaridad, mientras que la validación cruzada y el MSE evaluaron el rendimiento general del modelo. Al finalizar esta sección, se presentó una comparación detallada entre el rendimiento del método de stacking ensamble y los siete modelos base desarrollados individualmente. Este análisis permitió evaluar la eficacia de cada enfoque y demostrar cómo el método de stacking ensamble mejoró la precisión, el F_1-score, el recall, la exactitud y la estabilidad en la clasificación de sentimientos en comparación con los modelos individuales.

4.2.1 Clasificación mediante los Modelos Base Seleccionados

Los resultados del análisis de sentimiento utilizando diferentes modelos exhiben notables variaciones en las tasas de clasificación correcta para cada polaridad. Por ejemplo, el modelo de regresión logística logró una precisión del 85.34% para los casos negativos, 57.26% para los neutrales y 27.87% para los positivos. En el caso del modelo Multinomial NB, se alcanzaron tasas de precisión del 84.85% para los casos negativos, 58.37% para los neutrales y 44.23% para los positivos. Del mismo modo, el modelo Bernoulli NB presentó tasas de clasificación correctas del 64.46% para los casos negativos, 77.43% para los neutrales y 15.38% para los positivos. Asimismo, el modelo SVM obtuvo una tasa de precisión del 86.41% para los casos negativos, 60.12% para los neutrales y 36.54% para los positivos. Por otro lado, el modelo de árbol de decisión demostró una tasa de clasificación correcta del 73.46% para la polaridad negativa, 57.00% para la polaridad neutral y 36.54% para las instancias positivas. Igualmente, el modelo de Random Forest presentó tasas de precisión del 86.23% para los casos negativos, 59.73% para los neutrales y 28.85% para los positivos. Finalmente, el modelo

de KNN mostró resultados menos favorables, con una precisión del 14.51% para los casos negativos, 97.67% para los neutrales y 17.31% para los positivos. Sin embargo, se observaron errores importantes en la clasificación de los modelos, como se detalla en la tabla 4.

Table 4. Errores de clasificación basados en la matriz de confusión de los modelos bases

Error de clasificación	Error % Regresión Logistica	Error % Multinomial NB	Error % Bernoulli NB	Error % SVM	Error % Desicion Tree	Error % Random Forest	Error % KNN
Negativos clasificados incorrectamente como neutrales	13.31	13.50	34.16	13.13	23.14	13.67	85.49
Negativos clasificados incorrectamente como positivos	1.35	1.65	1.38	0.46	3.40	0.09	0.00
Neutrales clasificados incorrectamente como negativos	38.17	36.19	21.60	36.38	37.35	39.30	1.75
Neutrales clasificados incorrectamente como positivos	4.56	5.45	0.97	3.50	5.64	0.97	0.58
Positivos clasificados incorrectamente como negativos	32.79	21.15	7.69	23.08	32.69	34.62	1.92
Positivos clasificados incorrectamente como neutrales	39.34	34.62	76.92	36.54	30.77	28.85	80.77

Además, se llevó a cabo una comparación del rendimiento de varios modelos de clasificación utilizando métricas como precisión, F_1-score, recall y exactitud, tanto mediante la matriz de confusión como mediante la validación cruzada.

En la tabla 5 y 6 se muestran las comparaciones entre la matriz de confusión y la validación cruzada para los diferentes modelos de clasificación utilizados en el análisis de sentimientos. Para el modelo de regresión logística, los valores en la matriz de confusión son los siguientes: para la polaridad negativa, una precisión de 0.82, un F_1-score de 0.84 y un recall de 0.85; para la neutra, una precisión de 0.62, un F_1-score de 0.59 y un recall de 0.57; y para la positiva, una precisión de 0.31, un F_1-score de 0.30 y un recall de 0.28. La exactitud total del modelo fue de 0.75. Mediante la validación cruzada, se obtuvo una precisión de 0.76, un F_1-score de 0.72, un recall de 0.75 y una exactitud de 0.75. Para el modelo Multinomial NB, los valores en la matriz de confusión fueron: para la polaridad negativa, una precisión de 0.82, un F_1-score de 0.84 y un recall de 0.85; para la neutra, una precisión de 0.65, un F_1-score de 0.61 y un recall de 0.58; y para la positiva, una precisión de 0.33, un F_1-score de 0.38 y un recall de 0.44. La exactitud total fue de 0.75. A través de la validación cruzada, se obtuvo una precisión de 0.69, un F_1-score de 0.60, un recall de 0.69 y una exactitud de 0.69. Para el modelo BernoulliNB, se encontraron los siguientes resultados en la matriz de confusión: para la polaridad negativa, una precisión de 0.86, un F_1-score de 0.74 y un recall de 0.64; para la neutra, una precisión de 0.49, un F_1-score de 0.60 y un recall de 0.77; y para la positiva, una precisión de 0.29, un F_1-score de 0.20 y un recall de 0.15. La exactitud total fue de 0.67. Mediante la validación cruzada, se determinó una precisión de 0.73, un F_1-score de 0.71, un recall de 0.73 y una exactitud de 0.73. Para el modelo SVM, los valores en la matriz de confusión fueron: para la polaridad negativa, una precisión de 0.83, un F_1-score de 0.84 y un recall de 0.86; para la neutra, una precisión de 0.65, un F_1-score de 0.63 y un recall de 0.60; y para la positiva, una precisión de 0.45, un F_1-score de 0.40 y un recall de 0.37. La exactitud total fue de 0.77. Mediante la validación cruzada, se revelaron valores de precisión de 0.75, F_1-score de 0.73, recall de 0.75 y exactitud de 0.75. Para el modelo de árbol de decisión, los resultados en la matriz de confusión fueron: para la polaridad negativa, una precisión de 0.78, un F_1-score de 0.78 y un recall de 0.79; para la neutra, una precisión de 0.55, un F_1-score de 0.55 y un recall de 0.55; y para la positiva, una precisión de 0.33, un F_1-score de 0.30 y un recall de 0.27. La exactitud total fue de 0.70. Mediante la validación cruzada, se observaron valores de precisión de 0.68, F_1-score de 0.67, recall de 0.68 y exactitud de 0.68. Además, para el modelo Random Forest, se obtuvieron los siguientes resultados en la matriz de confusión: para la polaridad negativa, una precisión de 0.78, un F_1-score de 0.85 y un recall de 0.93; para la neutra, una precisión de 0.75, un F_1-score de 0.59 y un recall de 0.49; y para la positiva, una precisión de 0.80, un F_1-score de 0.36 y un recall de 0.23. La exactitud total fue de 0.77. Mediante la validación cruzada, se evidenciaron valores de precisión de 0.74, F_1-score de 0.72, recall de 0.74 y exactitud de 0.74. Finalmente, para el modelo KNN, se observaron los siguientes resultados en la matriz de confusión: para la polaridad negativa, una precisión de 0.60, un F_1-score de 0.25 y un recall de 0.15; para la neutra, una precisión de 0.34, un F_1-score de 0.50 y un recall de 0.98; y para la positiva, una precisión de 0.75, un F_1-score de 0.28 y un recall de 0.17. La exactitud total fue del 0.40. Mediante la validación cruzada, se revelaron valores de precisión del 0.38, F_1-score del 0.30, recall del 0.38 y exactitud del 0.38.

Tabla 5. Métricas de Matriz de confusion y MSE.

Modelo Base	Metrica	Negativo	Neutro	Positivo	Accuracy	MSE
Logistica regresion	Precisión	0.82	0.62	0.31	0.75	0.3129
Logistica regresion	F_1-score	0.84	0.59	0.30	0.75	0.3129
Logistica regresion	Recall	0.85	0.57	0.28	0.75	0.3129
Multinomial NB	Precisión	0.82	0.65	0.33	0.75	0.382
Multinomial NB	F_1-score	0.84	0.61	0.38	0.75	0.382
Multinomial NB	Recall	0.85	0.57	0.44	0.75	0.382
Bernoulli NB	Precisión	0.86	0.60	0.29	0.67	0.297
Bernoulli NB	F_1-score	0.74	0.60	0.20	0.67	0.297
Bernoulli NB	Recall	0.64	0.77	0.15	0.67	0.297
SVM	Precisión	0.83	0.65	0.45	0.77	0.306
SVM	F_1-score	0.84	0.63	0.40	0.77	0.306
SVM	Recall	0.86	0.60	0.37	0.77	0.306
Decicion tree	Precisión	0.78	0.55	0.33	0.70	0.425
Decicion tree	F_1-score	0.78	0.55	0.30	0.70	0.425
Decicion tree	Recall	0.79	0.55	0.27	0.70	0.425
Random Forest	Precisión	0.78	0.75	0.80	0.77	0.274
Random Forest	F_1-score	0.85	0.59	0.36	0.77	0.274
Random Forest	Recall	0.93	0.49	0.23	0.77	0.274
KNN	Precisión	0.60	0.34	0.75	0.40	0.597
KNN	F_1-score	0.25	0.50	0.28	0.40	0.597
KNN	Recall	0.15	0.98	0.17	0.40	0.597

Tabla 6. Métricas de validación cruzada.

Metrica	Logistra Regresion	Multinomial NB	Bernoulli NB	SVM	Decision Tree	Random Forest	KNN
Precisión	0.76	0.69	0.73	0.75	0.69	0.74	0.38
F_1-score	0.72	0.60	0.71	0.73	0.67	0.72	0.30
Recall	0.75	0.69	0.73	0.75	0.68	0.74	0.38
Accuracy	0.75	0.69	0.73	0.75	0.68	0.74	0.38

4.2.2 Clasificación mediante el Método Stacking Ensemble

En esta sección, se resumen los resultados obtenidos del análisis de sentimiento al aplicar el método de stacking ensemble para evaluar la aprobación política de la presidencia del estado peruano. Stacking ensemble es una técnica de aprendizaje automático que combina las predicciones de varios modelos base para mejorar la capacidad de clasificación. En este estudio, utilizamos siete modelos base, previamente detallados, para construir un modelo integrado que fusiona las predicciones individuales y produce la clasificación final. Ademas, se analizo los resultados del análisis de sentimiento y se detalla las métricas de evaluación que se emplearon para medir el rendimiento del método stacking ensemble, así como una comparación de estos resultados con las métricas obtenidas de los modelos base individuales. Asimismo, se exploro las fortalezas del método de stacking en el contexto del análisis de sentimientos y se analizo la consistencia del rendimiento obtenido. A través de este análisis, nuestro objetivo es ofrecer una evaluación detallada de la efectividad del método de stacking para mejorar el rendimiento de la clasificación en comparación con los modelos base individuales.

4.2.3 Resultados del análisis de sentimiento

Los resultados obtenidos del conjunto de datos revelaron una predominancia abrumadora de negatividad, con un 66.1% de las opiniones de los ciudadanos peruanos publicadas en Twitter mostrando sentimientos negativos hacia la presidencia del estado peruano. Por otro lado, solo una pequeña fracción, aproximadamente un 4.1%, exhibía sentimientos positivos, mientras que el 29.8% restante mostraba un sentimiento neutral. Estos hallazgos se confirmaron mediante el análisis de sentimientos de las publicaciones en 𝕏, que respaldaron la percepción mayoritariamente negativa de los ciudadanos hacia la presidencia del estado peruano, como se ilustra en la figura.

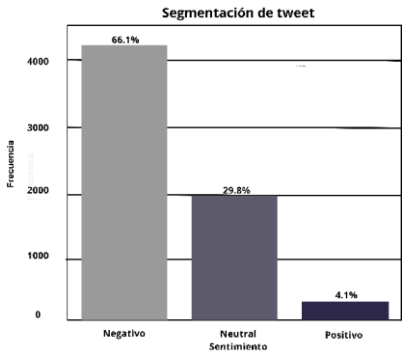


Figura 6. Segmentación de Tweets de Ciudadanos Peruanos en 𝕏

De igual forma se incluyó el uso de nubes de palabras para representar visualmente las opiniones públicas de los ciudadanos peruanos sobre la aprobación política de la presidencia del estado peruano, categorizadas en polaridades negativas, neutrales y positivas. En la figura 7 y 8 se presenta la nube de palabras para la polaridad negativa, que refleja una fuerte desaprobación hacia la actual presidencia del estado peruano. Por otro lado, en la figura 9 se muestra la nube de palabras para la polaridad positiva, la cual destaca las opiniones favorables hacia la presidencia. Finalmente, en la figura 10 se exhibe la nube de palabras para la polaridad neutral, indicando la falta de relevancia que algunos ciudadanos otorgan a la política estatal peruana. Es relevante señalar que la nube de palabras negativas resalta más, ya que recibe una mayor atención por parte de los ciudadanos peruanos en 𝕏. Esto sugiere que las críticas y desaprobaciones hacia la presidencia del estado peruano son temas de considerable interés y debate en la plataforma social.

Figura 7. Nube de Palabras: Polaridad Negativa

Figura 8. Nube de Palabras: Polaridad Negativa

Figura 9. Nube de Palabras: Polaridad Positiva

Figura 10. Nube de Palabras: Polaridad Neutra

4.2.4 Métricas de Evaluación del Método Stacking Enseemble

En la evaluación del método stackin ensemble, se emplearon diversas métricas de la matriz de confusión para evaluar su desempeño en el análisis de sentimiento sobre la aprobación política de la presidencia del estado peruano. Estas métricas, que abarcan precisión, recall, exactitud, F1-score, AUC-ROC y validación cruzada, proporcionan una evaluación completa del rendimiento del método en la clasificación de sentimientos.

Matriz de Confución

Como resultado de la matriz de confusión realizada al método de ensemble stacking, se obtuvieron las siguientes tasas de precisión para las tres polaridades estudiadas, para la polaridad negativa, se registró una

tasa del 91.46%, para la polaridad neutra, se registró una tasa del 57.20%, para la polaridad positiva, se registró una tasa del 26.92%. Sin embargo, el método tambien realizo clasificaciones incorrectas, que se detallan de la siguiente manera, para la polaridad negativa, el 8.54% fue clasificado como neutra cuando la clase real era negativa, para la polaridad neutra, el 42.41% fue clasificado como negativa cuando la clase real era neutra, y el 0.39% fue clasificado como positiva cuando la clase real era neutra, para la polaridad positiva, el 30.77% fue clasificado como negativa cuando la clase real era positiva, y el 42.31% fue clasificado como neutra cuando la clase real era positiva. Estos resultados destacan tanto las clasificaciones correctas como las incorrectas realizadas por el método de ensemble stacking, como se muestra en la figura 11.

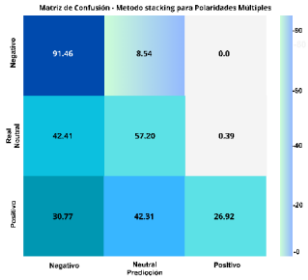


Figure 11. Matriz de confusion del método stacking ensemble

Tras examinar las tasas de precisión y los errores de clasificación de las polaridades de sentimiento analizadas, es esencial mejorar nuestra evaluación utilizando métricas adicionales como la curva ROC-AUC, recall, F_1 score y accuracy.

Curva ROC - AUC

La curva ROC proporciona una perspectiva más detallada sobre la capacidad de clasificación del método de stacking ensemble, especialmente en la distinción entre las polaridades negativas, neutrales y positivas. Mientras que las tasas de precisión nos dan una idea del porcentaje de clasificaciones correctas, la curva ROC muestra cómo varía la tasa de verdaderos positivos frente a los falsos positivos en diferentes umbrales de clasificación. Esto es fundamental para evaluar la capacidad del modelo para discernir entre las clases positivas neutras y negativas en el análisis de sentimientos políticos. Al integrar la curva ROC con las tasas de precisión, obtenemos una evaluación más detallada y precisa del rendimiento del método de ensemble stacking en la clasificación de polaridades, lo que nos permite tomar decisiones más fundamentadas en el análisis de sentimientos políticos. En la figura 12, se muestra la curva ROC correspondiente al método stacking ensemble. Los resultados revelaron que las curvas de las tres polaridades se ubicaron cerca de la esquina superior izquierda del gráfico, indicando así una alta sensibilidad (verdaderos positivos) y una baja tasa de falsos positivos para todas las clases. Además, se determinó que el área bajo la curva (AUC) para la clase negativa fue de 0.87, para la clase neutral fue de 0.85 y para la clase positiva fue de 0.89. Como promedio macro final, la curva ROC mostró un AUC de 0.87 para el método stacking, lo que sugiere un rendimiento aceptable para la clasificación en el análisis de sentimientos. Estos hallazgos destacan la capacidad del método stacking ensemble para distinguir entre las diferentes polaridades.

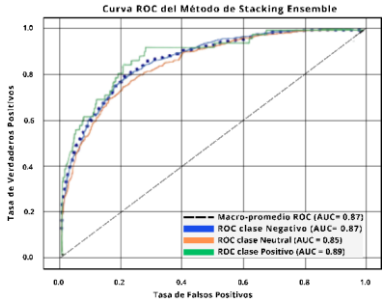


Figura 12. Curva de ROC: Método Stacking Ensemble

Recall

El recall es una métrica valiosa para revelar cómo el método de stacking ensemble identifica las instancias por cada polaridad. En las Figura 13, se presento gráficamente los valores de recall para las polaridades negativa, neutra y positiva respectivamente. Para la polaridad negativa, se observo un recall de 0.92, lo que indica que el 92% de los tweets con polaridad negativa fueron clasificados correctamente como negativos. En contraste, el recall para la polaridad neutral es de 0.55, indicando que solo el 55% de los tweets neutrales se clasifican correctamente como neutrales. Finalmente, para la polaridad positiva, se observo un recall de 0.31, lo que sugiere que solo el 31% de los tweets positivos se clasifican correctamente como positivos. Este análisis revelo que la polaridad negativa tiene la mayor capacidad para identificar correctamente las instancias negativas del conjunto de datos, mientras que la polaridad positiva tiene la menor capacidad para identificar correctamente las instancias positivas. La polaridad neutra muestra una capacidad moderadamente baja para identificar instancias neutrales. Es importante tener en cuenta que estos porcentajes pueden variar según el equilibrio de clases en el conjunto de datos.

F_1 Score

El F_1-score, es una métrica que fusiona precisión y recall, resulta esencial en problemas de clasificación. En este estudio, se aplicó para evaluar el equilibrio que ofrece el método stacking ensemble en términos de precisión y recall al realizar predicciones en el análisis de sentimientos. En la Figura 14, se presentan gráficamente los porcentajes obtenidos. Para la polaridad negativa, se obtuvo un F_1-score de 0.86, indicando un equilibrio óptimo en la identificación correcta de las clases negativas. Para la polaridad neutral, se logró un F_1-score de 0.62, lo que muestra un buen equilibrio en la identificación precisa de las clases neutrales. Por último, para la polaridad positiva se obtuvo un F_1-score de 0.39, reflejando un bajo equilibrio al identificar las clases positivas. Es importante considerar que estos porcentajes pueden variar dependiendo del equilibrio de clases en el conjunto de datos.

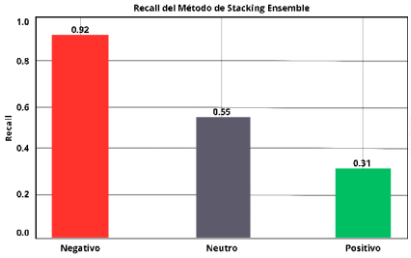


Figura 13. Recall: Método Stacking Ensemble

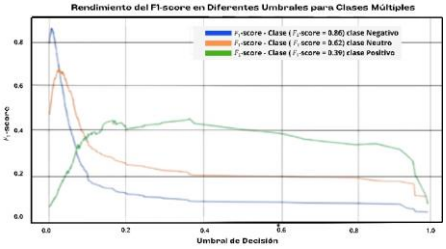


Figura 14. F_1 Score: Método stacking ensemble

Exactitud

En la evaluación del método de stacking ensemble, se observó que el modelo logró una exactitud de 0.78 en la clasificación de sentimientos, lo que demuestra un rendimiento sobresaliente. Esta alta exactitud indica que el modelo ha realizado un número significativo de predicciones correctas en relación con el total de predicciones realizadas, subrayando su eficacia en la clasificación de los sentimientos expresados en los tweets.

Validación Cruzada

Como evaluación final del método de stacking ensemble, se utilizó la técnica de validación cruzada, que proporciona una evaluación detallada del desempeño del modelo. Los resultados, que se muestran en la Tabla 7, revelan que el método logró una precisión total del 0.77. Además, el F1-score, el recall y el accuracy alcanzaron valores de 0.76, 0.76 y 0.77, respectivamente. El Error Cuadrático Medio (MSE) se situó en 0.2846, indicando un error cuadrático menor en comparación con los modelos base. Estos resultados subrayan que el método de stacking ensemble no solo proporciona una alta calidad en la predicción de sentimientos, sino que también exhibe una excelente capacidad de generalización en diferentes contextos.

Tabla 7. Métricas de la Validación Cruzada y MSE.

Metricas	Porcentaje
Precisión	0.78
F_1-score	0.76

Recall	0.76
Accuracy	0.77
MSE	0.2846

4.2.5 Comparación del Método Stacking Ensemble y Modelos Individuales

Este estudio comparó el desempeño del método stacking ensemble con los modelos base individuales en el análisis de sentimientos para evaluar el índice de aprobación política de la presidencia del estado peruano. Se encontró que el método stacking demostró una mayor precisión, recall, F_1-score y accuracy en comparación con los modelos base individuales, con una precisión del 91% para polaridad negativa, 57.20% para polaridades neutras y 26.92% para polaridades positivas. Además, el método stacking exhibió el menor error cuadrático medio en comparación con los modelos bases individuales. Estos resultados respaldan investigaciones previas, como la de Dou y Peng, que implementaron un método de ensamble para mejorar la clasificación de imágenes. Utilizando tres conjuntos de datos relacionados con imágenes multiespectrales, hiperespectrales y de series temporales, lograron aumentar la precisión en un 3.73%, 7.66% y 9.16%, respectivamente [64]. Del mismo modo, la investigación de Nouri y Zahra introdujo un nuevo método de ensamble llamado RUEO, que alcanzó una precisión media de 0.69 en la clasificación, en comparación con el promedio de 0.66 de los modelos bases [65]. Además, Zeyad Ghaleb Al-Mekhlafi desarrolló un método ensamblado para mejorar la precisión de detección de sitios web de phishing, logrando un aumento del 97.16% en la precisión de detección [66]. Por último, en la investigación de Padmavathi Radhakrishnan, se presentó un método de ensamblado para clasificar diferentes perturbaciones de la calidad de la energía (PQDs) en la red eléctrica integrada fotovoltaica (PV). Este método de ensamble logró una precisión mayor en las clasificaciones que los clasificadores base en condiciones de prueba estándar (92.22%) y condiciones ambientales dinámicas de la energía solar fotovoltaica (91%), así como en la presencia de ruido en el clasificador (89.33%) [67]. Estos resultados, junto con los mencionados en este artículo, evidencian que el método stacking logra una mejor precisión que los modelos bases individuales.

5) CONCLUSION

El estudio tuvo como objetivo implementar el método stacking ensemble para analizar el índice de aprobación política de la presidencia del estado peruano, utilizando datos de Twitter. Se utilizó Beto Sentiment Analysis para evaluar los sentimientos expresados en los tweets. Los resultados mostraron que el método stacking logró una precisión del 91% en la clasificación de polaridad negativa, del 57.20% en la polaridad neutra y del 26.92% en la polaridad positiva. Además, las curvas ROC indicaron un rendimiento del 87% para polaridad negativa, 85% para polaridad neutra y 87% para polaridad positiva. El F1-score fue del 86% para la polaridad negativa, 62% para la polaridad neutra y 39% para la polaridad positiva. La precisión general del análisis de sentimientos alcanzó el 78%. La validación cruzada arrojó una precisión del 78%, un recall del 76%, un F1-score del 76% y un accuracy del 77%. Además, el error cuadrático medio del método stacking fue del 0.28, mostrando el menor error en comparación con los modelos base individuales. Estos resultados destacan que el método stacking ensemble supera significativamente a los modelos individuales en términos de precisión y consistencia en las clasificaciones, lo cual es fundamental para evaluar el índice de aprobación política. En conclusion el método de stacking ensemble obtuvo una optima precisión en el análisis de sentimientos político en comparación con los modelos base individuales, esto permite concluir que los ciudanos peruanos al alfrontar la crisis política gubernamental difiere de las leyes dadas por el gobierno peruano y esto hace que el mensaje transmitido por la red social twitter sea el reclamo para tener un estado solido y democratico.

ACKNOWLEDGEMENTS

Los autores expresan su agradecimiento a la Universidad Peruana Union por el apoyo brindado en la realización de la investigacion.

REFERENCES

[1] D. Kydros, M. Argyropoulou, and V. Vrana, "A Content and Sentiment Analysis of Greek Tweets during the Pandemic," *Sustain.*, vol. 13, no. 11, 2021, doi: 10.3390/su13116150.

[2] J. Yauri, L. Solis, E. Porras, M. Lagos, and E. Tinoco, "Approval rating of peruvian politicians and policies using sentiment analysis on twitter," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 1–8, 2022, [Online]. Available: www.ijacsa.thesai.org

[3] D. Grgić, M. Karaula, M. Bagić Babac, and V. Podobnik, "Predicting dependency of approval rating change from twitter activity and sentiment analysis," *Smart Innov. Syst. Technol.*, vol. 186, pp. 103–112, Jan. 2020, doi: 10.1007/978-981-15-5764-4_10.

[4] I. Savin and N. Teplyakov, "Topics of the nationwide phone-ins with Vladimir Putin and their role for public support and Russian economy," *Inf. Process. Manag.*, vol. 59, no. 5, p. 103043, 2022, doi: 10.1016/j.ipm.2022.103043.

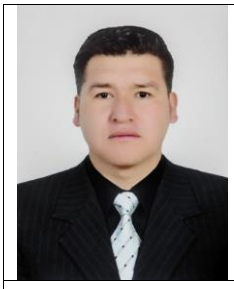
[5] ipsos, "Encuesta Ipsos," 2023. https://www.ipsos.com/sites/default/files/ct/news/documents/2023-12/Informe de Opinión_Perú

21 Ipsos_Noviembre 2023.pdf
[6] C. G. Rodríguez and L. G. Tule, "Honduras 2019: Persistent economic and social instability and institutional weakness[Honduras 2019: Persistente inestabilidad económica y social y debilidad institucional]," *Rev. Cienc. Polit.*, vol. 40, no. 2, pp. 379–400, Jun. 2020.
[7] P. Khurana Batra, A. Saxena, Shruti, and C. Goel, "Election result prediction using twitter sentiments analysis," *PDGC 2020 - 2020 6th Int. Conf. Parallel, Distrib. Grid Comput.*, pp. 182–185, 2020, doi: 10.1109/PDGC50313.2020.9315789.
[8] O. B. Deho, W. A. Agangiba, F. L. Aryeh, and J. A. Ansah, "Sentiment analysis with word embedding," *IEEE Int. Conf. Adapt. Sci. Technol. ICAST*, vol. 2018-Augus, pp. 1–4, 2018, doi: 10.1109/ICASTECH.2018.8506717.
[9] P. Balakesava Reddy, S. Ramasubbareddy, G. Viswanath, and K. Govinda, "Sentiment Analysis of Tweets Related to SUS Before and During COVID-19 Pandemic," *Lect. Notes Networks Syst.*, vol. 385, no. 1, pp. 385–393, 2022, doi: 10.1007/978-981-16-8987-1_41.
[10] D. G. Mandhasiya, H. Murfi, and A. Bustamam, "The hybrid of BERT and deep learning models for Indonesian sentiment analysis," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 1, pp. 591–602, 2024, doi: 10.11591/ijeecs.v33.i1.pp591-602.
[11] E. C. Soujanya Poria, Amir Hussain, *Multimodal Sentiment Analysis*, Cham, Spri. Springer Cham, 2018. doi: https://doi.org/10.1007/978-3-319-95020-4.
[12] S. Y. Lin, Y. C. Kung, and F. Y. Leu, "Predictive intelligence in harmful news identification by BERT-based ensemble learning model with text sentiment analysis," *Inf. Process. Manag.*, vol. 59, no. 2, p. 102872, 2022, doi: 10.1016/j.ipm.2022.102872.
[13] A. Nandi and P. Sharma, "Comparative Study of Sentiment Analysis Techniques," *Interdiscip. Res. Technol. Manag.*, no. June, pp. 456–460, 2021, doi: 10.1201/9781003202240-72.
[14] T. Gowandi, H. Murfi, and S. Nurrohmah, "Performance Analysis of Hybrid Architectures of Deep Learning for Indonesian Sentiment Analysis," S. A. M. A. M. U. T. M. S. A. S. M. B. W. Y. U. T. M. S. A. M. J. M. Z. U. of T. K. T. U. M. W. B. Editors and Affiliations Universiti Teknologi MARA, Ed., Springer, Singapore, 2021, pp. 18–27. doi: 10.1007/978-981-16-7334-4_2.
[15] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 4487–4496, 2020, doi: 10.18653/v1/p19-1441.
[16] D. Kannan *et al.*, "Virtual analysis of machine learning models for diseases prediction in muskmelon," vol. 33, no. 3, pp. 1748–1759, 2024, doi: 10.11591/ijeecs.v33.i3.pp1748-1759.
[17] A. Chaudhary, K. S. Chouhan, J. Gajrani, and B. Sharma, *Deep learning with PyTorch*. 2020. doi: 10.4018/978-1-7998-3095-5.ch003.
[18] A. Chunawale and M. Bedekar, "Electroencephalogram based human emotion classification for valence and arousal using machine learning approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 2, pp. 920–931, 2024, doi: 10.11591/ijeecs.v33.i2.pp920-931.
[19] K. P. Murphy, "Machine Learning A Probabilistic Perspective."
[20] M. Umer, I. Ashraf, A. Mehmood, S. Kumari, S. Ullah, and G. Sang Choi, "Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model," *Comput. Intell.*, vol. 37, no. 1, pp. 409–434, Feb. 2021, doi: 10.1111/coin.12415.
[21] G. O. Gutiérrez-Esparza, M. Vallejo-Allende, and J. Hernández-Torruco, "Classification of cyber-aggression cases applying machine learning," *Appl. Sci.*, vol. 9, no. 9, 2019, doi: 10.3390/app9091828.
[22] M. Alzyout, E. Al Bashabsheh, H. Najadat, and A. Alaiad, "Sentiment Analysis of Arabic Tweets about Violence against Women using Machine Learning," *2021 12th Int. Conf. Inf. Commun. Syst. ICICS 2021*, no. May, pp. 171–176, 2021, doi: 10.1109/ICICS52457.2021.9464600.
[23] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," *Comput. Linguist.*, vol. 35, no. 3, pp. 399–433, 2009, doi: 10.1162/coli.08-012-R1-06-90.
[24] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *Processing*, vol., pp. 1–6, 2009.
[25] W. Aslam, W. H. Butt, and M. W. Anwar, "A Systematic Review on Social Network Analysis," in *Proceedings of the 2018 International Conference on Computing and Big Data*, New York, NY, USA: ACM, Sep. 2018, pp. 92–97. doi: 10.1145/3277104.3277112.
[26] J. Prusa, T. M. Khoshgoftaar, and D. J. DIttman, "Using Ensemble Learners to Improve Classifier Performance on Tweet Sentiment Data," *Proc. - 2015 IEEE 16th Int. Conf. Inf. Reuse Integr. IRI 2015*, pp. 252–257, 2015, doi: 10.1109/IRI.2015.49.
[27] N. F. F. Da Silva, E. R. Hruschka, and E. R. Hruschka, "Tweet sentiment analysis with classifier ensembles," *Decis. Support Syst.*, vol. 66, pp. 170–179, 2014, doi: 10.1016/j.dss.2014.07.003.
[28] A. Hassan, A. Abbasi, and D. Zeng, "Twitter sentiment analysis: A bootstrap ensemble framework," *Proc. - Soc. 2013*, pp. 357–364, 2013, doi: 10.1109/SocialCom.2013.56.
[29] S. Clark and R. Wicentowski, "SwatCS: Combining simple classifiers with estimated accuracy," **SEM 2013 - 2nd Jt. Conf. Lex. Comput. Semant.*, vol. 2, no. SemEval, pp. 425–429, 2013.
[30] K. T. and J. GHOSHt, "ANALYSIS OF DECISION BOUNDARIES IN LINEARLY COMBINED NEURAL CLASSIFIERS KAGAN," vol. 29, no. 2, pp. 341–348, 1996, doi: https://doi.org/10.1016/0031-3203(95)00085-2.
[31] C. Li, L. Wang, J. Li, and Y. Chen, "Application of multi-algorithm ensemble methods in high-dimensional and small-sample data of geotechnical engineering: A case study of swelling pressure of expansive soils," *J. Rock Mech. Geotech. Eng.*, Feb. 2024, doi: 10.1016/J.JRMGE.2023.10.015.
[32] P. S. Maciąg, R. Bembenik, A. Piekarzewicz, J. Del Ser, J. L. Lobo, and N. K. Kasabov, "Effective air pollution prediction by combining time series decomposition with stacking and bagging ensembles of evolving spiking neural networks," *Environ. Model. Softw.*, vol. 170, p. 105851, Dec. 2023, doi: 10.1016/J.ENVSOFT.2023.105851.
[33] M. Menor-Flores and M. A. Vega-Rodríguez, "Boosting-based ensemble of global network aligners for PPI network alignment," *Expert Syst. Appl.*, vol. 230, p. 120671, Nov. 2023, doi: 10.1016/J.ESWA.2023.120671.
[34] T. Peng *et al.*, "Research and application of a novel selective stacking ensemble model based on error compensation and parameter optimization for AQI prediction," *Environ. Res.*, vol. 247, p. 118176, Apr. 2024, doi: 10.1016/J.ENVRES.2024.118176.
[35] M. S. Reza, R. Amin, R. Yasmin, W. Kulsum, and S. Ruhi, "Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data," *Heliyon*, vol. 10, no. 2, p. e24536, Jan. 2024, doi: 10.1016/J.HELIYON.2024.E24536.
[36] R. Dey and R. Mathur, "Ensemble Learning Method Using Stacking with Base Learner, A Comparison," *Lect. Notes Networks*

Syst., vol. 727 LNNS, pp. 159–169, 2023, doi: 10.1007/978-981-99-3878-0_14/COVER.
[37] J. Martinez-Gil, "A comprehensive review of stacking methods for semantic similarity measurement," *Mach. Learn. with Appl.*, vol. 10, no. September, p. 100423, 2022, doi: 10.1016/j.mlwa.2022.100423.
[38] Fathoni, Erwin, and Abdiansah, "Extraction of Event Sentence Information in the Covid-19 Distribution Location Detection System based on the Indonesian Language Corpus," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2022-October, pp. 383–388, 2022, doi: 10.23919/EECSI56542.2022.9946530.
[39] Di. Baviskar, S. Ahirrao, and K. Kotecha, "Multi-Layout Unstructured Invoice Documents Dataset: A Dataset for Template-Free Invoice Processing and Its Evaluation Using AI Approaches," *IEEE Access*, vol. 9, pp. 101494–101512, 2021, doi: 10.1109/ACCESS.2021.3096739.
[40] Fathoni, Erwin, and Abdiansah, "Multilabel sentiment analysis for classification of the spread of COVID-19 in Indonesia using machine learning," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 31, no. 2, pp. 968–978, 2023, doi: 10.11591/ijeecs.v31.i2.pp968-978.
[41] J. Bhattacharjee, *Practical machine learning with Rust: Creating intelligent applications in Rust*. 2019. doi: 10.1007/9781484251218.
[42] J. Daniel and J. H. Martin, "Chapter 5 - Speech and Language Processing," *Speech Lang. Process.*, 2023.
[43] V. Kumar, "Evaluation of computationally intelligent techniques for breast cancer diagnosis," *Neural Comput. Appl.*, vol. 33, no. 8, pp. 3195–3208, 2021, doi: 10.1007/s00521-020-05204-y.
[44] M. Christopher, "An Introduction to Information Retrieval," *Libr. Rev.*, vol. 38, no. 9, pp. 462–463, 2009, doi: 10.1210/endo-38-3-156.
[45] C. Dewi, R.-C. Chen, H. J. Christanto, and F. Cauteruccio, "Multinomial Naïve Bayes Classifier for Sentiment Analysis of Internet Movie Database," *Vietnam J. Comput. Sci.*, vol. 10, no. 04, pp. 485–498, 2023, doi: 10.1142/s2196888823500100.
[46] R. Pedersen and M. Schoeberl, "An Embedded Support Vector Machine," no. May, pp. 1–11, 2007, doi: 10.1109/wises.2006.329117.
[47] C. C. Chang and C. J. Lin, "LIBSVM," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, May 2011, doi: 10.1145/1961189.1961199.
[48] R. Bemthuis, W. Wang, M. E. Iacob, and P. Havinga, "Business rule extraction using decision tree machine learning techniques: A case study into smart returnable transport items," *Procedia Comput. Sci.*, vol. 220, pp. 446–455, 2023, doi: 10.1016/j.procs.2023.03.057.
[49] S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, pp. 660–674, 1991, doi: 10.1109/21.97458.
[50] A. M. Segre, *Proceedings of the Sixth International Workshop on Machine Learning, Cornell University, Ithaca, New York, June 26-27, 1989*. M. Kaufmann Publishers, 1989. Accessed: Mar. 12, 2024. [Online]. Available: http://www.sciencedirect.com:5070/book/9781558600362/proceedings-of-the-sixth-international-workshop-on-machine-learning
[51] LEO BREIMAN, "Random Forests," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, p. 28, 2001, doi: 10.1023/A:1010933404324.
[52] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[53] J. Maillo, S. Ramírez, I. Triguero, and F. Herrera, "kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data," *Knowledge-Based Syst.*, vol. 117, pp. 3–15, 2017, doi: 10.1016/j.knosys.2016.06.012.
[54] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Syst. Appl.*, vol. 223, no. March, 2023, doi: 10.1016/j.eswa.2023.119862.
[55] F. J. Ariza-López, J. Rodríguez-Avi, and V. Alba-Fernández, "CONTROL ESTRICTO DE MATRICES DE CONFUSIÓN POR MEDIO DE DISTRIBUCIONES MULTINOMIALES," *GeoFocus Rev. Int. Cienc. y Tecnol. la Inf. Geográfica*, pp. 215–226, Jul. 2018, doi: 10.21138/gf.591.
[56] Q. Zhu, W. Ding, M. Xiang, M. Hu, and N. Zhang, "Loan Default Prediction Based on Convolutional Neural Network and LightGBM," *Int. J. Data Warehous. Min.*, vol. 19, no. 1, pp. 1–16, 2022, doi: 10.4018/IJDWM.315823.
[57] S. Nayak, Savita, and Y. K. Sharma, "A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis," *Decis. Anal. J.*, vol. 8, no. June, p. 100289, 2023, doi: 10.1016/j.dajour.2023.100289.
[58] D. Valero-Carreras, J. Alcaraz, and M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, no. December 2022, p. 106131, 2023, doi: 10.1016/j.cor.2022.106131.
[59] Juan Ignacio Barrios Arce, "La matriz de confusión y sus métricas," *BIG DATA, Ciencia de datos, Informatica Médica, Inteligencia Artificial, machinelearning*, 2019. https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/
[60] N. Mekhaznia and R. Khenfer, "Diagnosis of PV module based on neural network using performance indices," *Int. J. Power Electron. Drive Syst.*, vol. 14, no. 4, pp. 2347–2353, 2023, doi: 10.11591/ijpeds.v14.i4.pp2347-2353.
[61] N. Anđelić, S. Baressi Šegota, and Z. Car, "Improvement of Malicious Software Detection Accuracy through Genetic Programming Symbolic Classifier with Application of Dataset Oversampling Techniques," *Comput. 2023, Vol. 12, Page 242*, vol. 12, no. 12, p. 242, Nov. 2023, doi: 10.3390/COMPUTERS12120242.
[62] Y. Liang, X. R. Yin, Y. Sen Zhang, Y. Guo, and Y. L. Wang, "Predicting lncRNA–protein interactions through deep learning framework employing multiple features and random forest algorithm," *BMC Bioinformatics*, vol. 25, no. 1, p. 108, Dec. 2024, doi: 10.1186/s12859-024-05727-4.
[63] W. Zhou, Z. Yan, and L. Zhang, "A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction," *Sci. Rep.*, vol. 14, no. 1, p. 5905, Dec. 2024, doi: 10.1038/s41598-024-55243-x.
[64] P. Dou, C. Huang, W. Han, J. Hou, Y. Zhang, and J. Gu, "Remote sensing image classification using an ensemble framework without multiple classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 208, pp. 190–209, Feb. 2024, doi: 10.1016/j.isprsjprs.2023.12.012.
[65] Z. Nouri, V. Kiani, and H. Fadishei, "Rarity updated ensemble with oversampling: An ensemble approach to classification of imbalanced data streams," *Stat. Anal. Data Min.*, vol. 17, no. 1, p. e11662, Feb. 2024, doi: 10.1002/sam.11662.
[66] Z. G. Al-Mekhlafi *et al.*, "Phishing websites detection by using optimized stacking ensemble model," *Comput. Syst. Sci. Eng.*, vol. 41, no. 1, pp. 109–125, 2022, doi: 10.32604/csse.2022.020414.
[67] P. Radhakrishnan, K. Ramaiyan, A. Vinayagam, and V. Veerasamy, "A stacking ensemble classification model for detection and classification of power quality disturbances in PV integrated power network," *Measurement*, vol. 175, p. 109025, Apr. 2021, doi: 10.1016/J.MEASUREMENT.2021.109025.

BIOGRAPHIES OF AUTHORS (10 PT)

The recommended number of authors is at least 2. One of them as a corresponding author.
Please attach clear photo (3x4 cm) and vita. Example of biographies of authors:

	**Garcia Cercado Deivis Ronald** Systems Engineer with more than 3 years of experience specialized in software development and with basic skills in the field of data engineering. Expert in teamwork, effective communication and with solid technical skills for all stages of the software project cycle and data management. My quick learning and decision making skills make me the ideal candidate for companies looking for a young, innovative and demanding team. I seek to adapt to the needs of the company and take the opportunity to grow as a professional in your team. Committed to excellence and ready to contribute to the success of the organization. My proactive approach and passion for technology, both in software development and data management, are valuable assets to any team. He can be contacted at email: deyvisgarcia049@gmail.com
	**Villarroel Lopez, Juan Manuel** Systems Engineer graduated from UPeU Unión Peruana, Peru. Professional with over 6 years of experience, specialized in the field of data engineering. He has excelled in various roles at leading industry companies, such as Bluetab, an IBM subsidiary, where he served as a Data Developer. Here, he had the opportunity to apply his technical skills to develop innovative solutions and contribute to the success of large-scale projects. Additionally, he has had the privilege of working on significant projects, such as the implementation of institutional repositories at UPeU Unión Peruana, providing consultancy on repository management and process analysis in the Directorate of Evaluation and Knowledge Management. Previously, he held the position of IT Analyst at Ogreen, where he strengthened his system analysis and problem-solving skills. He is a professional passionate about technology and innovation, committed to the success of the projects he participates in and always ready to face new challenges and opportunities for growth. He can be contacted at email: villarroellopezjm@gmail.com
	**Barzola Torres, William Salvadory** Systems Engineer graduated from the Universidad Peruana Union (UpeU), Peru, Professional with over 4 years of experience, stands out in fields of information technology IT, has excelled in various leading companies in the Peruvian sector, such as, FCISA, ATENTO, SERPROVISASAC, served in the area technology and information as a developer in frond end, data mining, traditional infrastructure, Implementation of ERP Starsoft and all ADS in marketing with advertising campaigns, thanks to this I had the opportunity to apply his skills for development providing innovative and timely solutions, in addition, he has had the privilege of working on projects such as the implementation of a complete ERP in STARSOFT system in its different modules and areas, providing advice to the management of the modules and process analysis in the use of the system and knowledge management in addition to using his skills in the company as an expert in social media marketing in creating approaches and campaigns. He can be contacted at email: ingsalvatoretorres@gmail.com
	**Prof. Dr. Soria Quijaite Juan Jesús** PhD in Systems Engineering, Master in University Teaching and Educational Management, master's in applied mathematics and specialist in Statistics for Research. Statistical consultant at the Ministry of Production Cite Agroindustrial, research professor at the Peruvian Union University, member of the global network of researchers AUTHOR AID, member of the Network of Teachers of Latin America and the Caribbean, member of the National System of Science, Technology and Technological Innovation RENACYT (CONCYTEC) at Level VI, thesis advisor with the approaches of the scientific method, good practices of PMBOK version 7 and systems thinking in different specialties of undergraduate and graduate universities. He can be contacted at email: jesussoria@upeu.edu.pe.

Prof. MSc Nemias Saboya (Member, IEEE) received the B.S. degree in systems engineering from Universidad Peruana Unión (UPeU), Peru, and the M.S. degree in computer and systems engineering, with a specialization in information technology management, USMP, Peru. He is currently pursuing the Ph.D. degree in systems engineering with UPeU. His main research interests include data governance, statistical machine learning, IT, and data science, member of the National System of Science, Technology and Technological Innovation RENACYT (CONCYTEC). He can be contacted at email: saboya@upeu.edu.pe. (9 pt)

8% de similitud general

Principales fuentes encontradas en las siguientes bases de datos:

- 5% Base de datos de Internet
- 3% Base de datos de publicaciones
- Base de datos de Crossref
- Base de datos de contenido publicado de Crossref
- 8% Base de datos de trabajos entregados

FUENTES PRINCIPALES

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	cse.iitd.ac.in Internet	1%
2	coursehero.com Internet	1%
3	Universiti Teknikal Malaysia Melaka on 2023-11-22 Submitted works	1%
4	repository.ung.ac.id Internet	1%
5	sifisheriessciences.com Internet	<1%
6	de Sá, Rafael Castilho Corrêa. "Movie's Box Office Performance Predict... Publication	<1%
7	srmap on 2024-03-02 Submitted works	<1%
8	Universiti Malaysia Perlis on 2023-10-26 Submitted works	<1%

9	Universidad Tecnica De Ambato- Direccion de Investigacion y Desarrol... Submitted works	<1%
10	Universidad Internacional de la Rioja on 2024-01-06 Submitted works	<1%
11	Pontificia Universidad Catolica del Ecuador - PUCE on 2023-08-01 Submitted works	<1%
12	Corporación Universitaria Minuto de Dios,UNIMINUTO on 2024-08-12 Submitted works	<1%
13	Grupo IOE on 2023-11-06 Submitted works	<1%
14	hdl.handle.net Internet	<1%
15	Universidad Loyola Andalucia on 2024-01-16 Submitted works	<1%
16	Universiti Malaysia Perlis on 2023-08-25 Submitted works	<1%
17	Universidad Carlos III de Madrid - EUR on 2024-07-08 Submitted works	<1%