UNIVERSIDAD PERUANA
CAYETANO HEREDIA

“IDENTIFICACIÓN DE LA CORRELACIÓN ENTRE LOS SENTIMIENTOS IDENTIFICADOS EN TWITTER Y LA MOVILIDAD POBLACIONAL EN EL PERIODO DE CUARENTENA EN EL PERÚ”

TESIS PARA OPTAR EL GRADO DE MAESTRO EN INFORMÁTICA BIOMÉDICA EN SALUD GLOBAL CON MENCIÓN EN INFORMÁTICA EN SALUD

FRANKLIN PAUL BARRIENTOS PORRAS

LIMA – PERÚ

2022

ASESORA:

Dra. Patricia Silvia Mallma Salazar

JURADO DE TESIS

Dr. Giancarlo Ojeda Mercado

PRESIDENTE

Dra. Luz Aurora Carbajal Arroyo

VOCAL

Dr. Walter Humberto Castillo Martell

SECRETARIO (A)

DEDICATORIA

A mi pareja y a mi bebe quienes son los que hacen que mis días sean se llenen de ilusión y alegría, me siento dichoso de tener una familia como la nuestra

A mi mamá, hermanos y padrastro quienes siempre han estado para mí en las buenas y malas extendiéndome sus manos ante cualquier circunstancia, los quiero mucho y admiro

AGRADECIMIENTOS

Quisiera agradecer a CONCYTEC por haberme financiado la Maestría en Informática Biomédica en Salud Global.

Al Doctor César Cárcamo quien siempre estuvo ahí para brindarme consejos no solo como un maestro sino también como persona a quien he llegado a admirar mucho.

A la Facultad de Salud Pública y Administración (FASPA) quien abrió mi mente a nuevos conocimientos y como estas pueden llegar a beneficiar a las personas

A mis compañeros de la maestría (2019-2021) con quienes pasamos momentos muy amenos juntos y aunque el contexto nos distancie yo los tengo presentes muy dentro de mi persona.

A mi amiga Kathy Alva por apoyarme siempre y es gracias a ella que pude continuar desarrollando de manera profesional.

FINANCIAMIENTO

El desarrollo de la tesis fue gracias al financiamiento del Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica CONCYTEC, a través del Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica FONDECYT.

ÍNDICE

RESUMEN

ABSTRACT

I. INTRODUCCIÓN....1

II. PLANTEAMIENTO DE LA INVESTIGACIÓN 3

II.1. Marco Teórico 3

II.1.1 COVID-19....3

II.1.2 Medidas De Contención Para La Prevención Del Contagio 4

II.1.3 Medidas Contra El COVID-19 En El Perú 6

II.1.4 Movilización Poblacional....7

II.1.5. Consecuencias De Las Cuarentenas En La Salud Mental 8

II.1.6 Uso De Redes Sociales Como Medición De Estado De Ánimo....9

II.1.7 Métodos Para Cuantificar Información De Redes Sociales 10

II.2. Planteamiento Del Problema....14

II.3 Pregunta De Investigación 16

II.4 Justificación Del Estudio....16

III. OBJETIVOS 18

III.1. Objetivo General 18

III.2. Objetivos Específicos 18

IV. MATERIAL Y MÉTODOS 18

IV.1. Diseño del estudio 18

IV.2. Población....18

IV.3. Operacionalización De Variables 20

IV.4. Técnicas Y Procedimientos....22

Etapa 1: Determinación De La Opinión Pública 22

Etapa 2: Determinación De La Movilidad En Perú 33

V. PLAN DE ANÁLISIS....34

V.1 Modelos De Categorización De La Variable Sentimiento Inferido A Partir De Los Tweets: 34

V.2 Estadística Descriptiva Del Análisis De Sentimientos: 34

V.3 Análisis De Asociación Entre El Sentimiento Inferido En Los Tweets Y La Movilidad: 35
VI. CONSIDERACIONES ÉTICAS 35
VII. RESULTADOS 37
VII.1 Descripción De La Población De Tweets 37
VII.2 Opinión Pública De Los Tweets 39
VII.2.1 Análisis De Sentimientos 39
VII.3 Descripción De La Movilidad En El Perú 43
VII.4 Análisis De Asociación 45
VIII. DISCUSIÓN 46
IX. CONCLUSIONES 48
X. RECOMENDACIONES 49
XI. DECLARACIÓN DE CONFLICTOS DE INTERÉS 50
XII. REFERENCIAS BIBLIOGRÁFICAS 50
XIII. ANEXOS

ÍNDICE DE TABLAS

Tabla 1. Variables de estudio 20

Tabla 2. Variables de ajuste 21

Tabla 3. Distribución de tweets por departamento..... 37

Tabla 4. Distribución de tweets positivos, negativos y neutros por departamentos y la Provincia Constitucional de Callao 40

Tabla 5. Evaluación del sentimiento del tweet respecto a la movilidad 45

ÍNDICE DE FIGURAS

Figura 1. Metodología de la determinación de la opinión pública..... 23

Figura 2. Número de tweets en la cuarenta..... 38

Figura 3. Nube de palabras de la base de datos de tweets 39

Figura 4. Porcentaje de tweets negativos en el tiempo 42

Figura 5. Porcentaje de tweets positivos en el tiempo 42

Figura 6. Movilidad poblacional del Perú durante la cuarentena 43

Figura 7. Cambios de movilidad poblacional en los departamentos del Perú y la provincia constitucional del Callao. 44

RESUMEN

Introducción: La cuarentena por COVID 19 ha afectado en gran medida en la salud mental de las personas, donde esto podría llevar al incumplimiento de la misma. A su vez, existe evidencia la cual establece que las redes sociales pueden ser usadas para medir el estado emocional de las personas. **Objetivo**: Identificar la correlación entre la opinión pública en Twitter mediante el Análisis de Sentimiento y la movilidad poblacional en el Perú durante el periodo de la cuarentena. **Métodos**: Se realizó un estudio transversal. La muestra está constituida por las publicaciones (tweets) realizadas en Twitter durante el periodo de cuarentena en el Perú y cuya temática esté relacionada a este período de confinamiento. Para la determinación de la opinión pública se utilizó el análisis de sentimientos y los datos de movilidad fueron obtenidos a través de los reportes de movilidad Google. Para identificar la asociación entre estas dos variables hicimos uso de la regresión de Poisson. **Resultados**: Se encontró que la proporción de tweets categorizados como negativos fueron de 84.55% mientras que los tweets positivos fueron un 14.09%. Durante la etapa inicial de la cuarentena se observó una reducción porcentual de hasta un 80% en la movilidad poblacional, pero a medida que transcurrían los días esta se iba incrementando hasta observar una reducción porcentual en la movilidad de solo un 25%. El análisis estadístico no evidencia una asociación estadísticamente significativa entre las variables de interés. **Conclusiones**: No se encontró una asociación entre la opinión pública en Twitter y la movilidad poblacional durante el periodo de la cuarentena en el Perú.

PALABRAS CLAVES: ANÁLISIS DE SENTIMIENTO, MOVILIDAD POBLACIONAL, COVID-19, OPINIÓN PÚBLICA, SALUD MENTAL

ABSTRACT

Background: COVID-19 lockdown has greatly affected people's mental health, where this could lead to non-compliance with it. In turn, there is evidence which establishes that social networks can be used to measure the emotional state of people. **Objective**: Identify the correlation between public opinion on Twitter through Sentiment Analysis and population mobility in Peru in COVID-19 lockdown. **Methods**: A cross-sectional study was carried out. The sample is made up of the publications (tweets) made on Twitter during the lockdown in Peru and whose theme is related to quarantine. To determine public opinion, sentiment analysis was used and mobility data were obtained through Google mobility reports. To identify the association between these two variables we used the Poisson regression. **Results**: It was found that the proportion of tweets categorized as negative were 84.55% meanwhile tweets categorized as positive were 14.09%. During the initial stage of quarantine, a percentage reduction of up to 80% in population mobility was observed, but as the days passed, this increased until a percentage reduction in mobility of only 25% was observed. The statistical analysis does not show a statistically significant association between the variables of interest. **Conclusions**: No association was found between public opinion on Twitter and population mobility during lockdown in Peru.

KEY WORDS: SENTIMENT ANALYSIS, POPULATION MOBILITY, COVID-19, PUBLIC OPINION, MENTAL HEALTH

I. INTRODUCCIÓN

La enfermedad del coronavirus (COVID-19) surge en Wuhan-China en enero del 2020 (1). Debido a la alta tasa de contagios, superando incluso a la epidemia del SARS en el 2002, el 11 de marzo del 2020 la Organización Mundial de la Salud (OMS) declara este brote como una pandemia (2,3). Para este momento, se conoce que el contagio se produce a través de gotículas respiratorias en interacciones persona – persona (4). De ahí un factor importante para la transmisión de la enfermedad es la movilidad de las personas, por tal motivo los países comenzaron a implementar cuarentenas para prevenir los contagios. Las autoridades de Wuhan fueron los primeros en adoptar este tipo de medidas, que mostraron una alta efectividad reflejada en la disminución significativa de la tasa de contagios (5,6).

Sin embargo, los resultados positivos de la implementación de cuarentenas prolongadas en los países comenzaron a opacarse por las consecuencias adversas que conlleva el confinamiento que afectaron la economía y deterioraron la salud física y mental de la población (7–9). La incertidumbre ante el COVID-19, el miedo a la muerte, sentimientos de soledad, la infodemia, la inestabilidad laboral fueron algunos de los factores que contribuyeron al incremento de los niveles de estrés, ansiedad y depresión en las personas que viven principalmente en países emergentes económicamente como el Perú (10–12).

Además, se han encontrado estudios que evidencian que las personas durante la cuarentena escribían y realizaban publicaciones con mayor frecuencia en redes sociales, y a su vez estas publicaciones podrían estar asociadas con la salud mental

de las personas (10–15). En ese sentido, las redes sociales sirvieron como medios a través de las cuales las personas podían expresar sus opiniones, pensamientos, sentimientos, críticas, etc. Además, hay evidencia de que el nivel de salud mental de las personas está relacionado con el cumplimiento de la inmovilización social obligatoria, por esta razón se señala la necesidad de que medidas como la cuarentena vengan acompañadas de acciones, por parte de los gobiernos, que reduzcan la carga emocional que conlleva el aislamiento (12,16–19).

En el Perú, el 6 de marzo del 2020 se reportó oficialmente el primer paciente con coronavirus, estableciéndose el estado de emergencia sanitaria en el país (20). Para poder contener la propagación del COVID-19 se implementó el aislamiento social obligatorio en el país de manera rápida, la cual tuvo una duración de 5 meses desde el 16 de marzo hasta el 31 de julio del 2020, la cual nos referiremos como la primera cuarentena.

Por lo expuesto, el presente estudio busca generar evidencia de una posible asociación entre la opinión pública obtenida en redes sociales y como esta se ve afectada por el aislamiento obligatorio impuesto por el gobierno en el periodo de la primera cuarentena. Esto a través del uso de los informes de movilidad local que proporciona Google, y la determinación de la opinión pública a través del uso de una herramienta como el Análisis de Sentimientos dentro del campo del Machine Learning (21,22).

II. PLANTEAMIENTO DE LA INVESTIGACIÓN

II.1. Marco Teórico

II.1.1 COVID-19

A lo largo de la historia los brotes de coronavirus se debieron a actividades inapropiadas que el hombre ha venido realizando tales como el tráfico ilegal de animales silvestres, la deforestación que altera en enorme medida los ecosistemas naturales entre otras prácticas que conllevan al desarrollo de la zoonosis, siendo estas actividades uno de los principales motivos que dio origen al COVID-19 (23).

El COVID-19 es altamente contagioso manifestándose a través de varios síntomas, entre las más comunes se tiene fiebre, dificultad al respirar, tos seca, pérdida del gusto y/o olfato entre otros pudiendo causar desde infecciones leves en el tracto respiratorio superior hasta provocar síndromes severos que podrían desencadenar en la muerte de los pacientes infectados. Desde que el primer caso del nuevo coronavirus fuese detectado en Wuhan, la capital de la provincia de Hubei en China el 31 de diciembre del 2019, el brote rápidamente se convirtió en una crisis global afectando la vida de muchas personas (3,24,25).

A inicios de marzo del 2020, la alta tasa de transmisión del COVID-19, que llevo al colapso de los establecimientos de salud, y la tasa de mortalidad del virus, aunque baja producía miles de muertes hasta ese momento, motivo a que la Organización Mundial de la Salud (OMS) declaré oficialmente al COVID-19 como pandemia el 11 de marzo del 2020. La OMS lo ha denominado como el Síndrome Respiratorio

Agudo Severo (SARS-CoV-2) convirtiéndose así en la quinta pandemia reportada después de la pandemia de la gripe española de 1918 (24,26,27).

II.1.2 Medidas De Contención Para La Prevención Del Contagio

Las diversas medidas tomadas a nivel mundial estuvieron basadas en las principales formas de transmisión. La transmisión viral del coronavirus es principalmente por gotitas de personas a persona en el momento de hablar, toser o estornudar. Algunas medidas como el uso de mascarillas y escudos faciales, el uso de guantes, entre otros han sido y son parte de investigaciones a fin de mejorar los planes de prevención y expansión del virus (26).

Así mismo, en ausencia de intervenciones farmacéuticas en contra de la COVID-19, la estrategia es reducir el contacto entre personas infectadas y personas susceptibles mediante la determinación temprana de los casos o la reducción del contacto entre las mismas. Por lo tanto, un gran número de países han elegido el distanciamiento social y medidas de cuarentena total o parcial como medidas para disminuir la tasa de contagios (28,29).

Estas medidas estuvieron influenciadas por los reportes brindados por el Imperial College y The Lancet Infectious Diseases publicados en marzo del 2020, las cuales evaluaron los efectos de intervenciones no farmacéuticas sobre la propagación del SARS-CoV-2. Estas intervenciones incluían medidas como el distanciamiento social, cuarentena y más. Sin embargo, el impacto de estas intervenciones fue pequeña comparadas con las medidas impuestas en China en respuesta al COVID-19 las cuales incluían el cierre de escuelas, centros de trabajo, carreteras, cancelación de reuniones públicas, cuarentena obligatoria y vigilancia a gran escala.

Aunque estas acciones fueron elogiadas por la OMS, la posibilidad de imponer medidas similares a otros países generaba nuevos retos (28,30,31).

Las intervenciones combinadas que fueron implementadas llegaron a obtener un resultado propicio en el cual se reducía el número promedio de infecciones en un 99.3%. Aunque, siendo las bases científicas para estas intervenciones robustas, las consideraciones éticas son diversas. Por ende, es importante resaltar que los líderes políticos deben implementar políticas de cuarentena y distanciamiento social de manera eficiente sin afectar a las poblaciones vulnerables ya que estas intervenciones conllevan muchos riesgos afectando de manera desproporcionada a las poblaciones más desfavorecidas (28,29).

II.1.2.1 Inmovilización Social Obligatoria

Existen diferencias entre la cuarentena, el aislamiento y distanciamiento social y el confinamiento, los cuales requieren de definiciones claras para comprender los procedimientos posteriores a su realización. La cuarentena se ha utilizado para el control de brotes de enfermedades transmisibles, hace referencia a la restricción ya sea voluntaria u obligatoria del libre desplazamiento, que es realizada en personas que pudieron estar expuestas a una enfermedad. Puede ser realizado a nivel grupal, comunitario o personal y que implica la permanencia en un determinado lugar para la evaluación y posible asistencia médica (29,30).

El aislamiento es la separación de personas contagiadas y enfermas que al igual que la cuarentena puede ser aplicado a nivel grupal, individual o comunitario y tiene la finalidad de proteger a las personas no infectadas, idealmente un área de aislamiento cumple con características que permitan reducir la transmisión de la enfermedad.

Por otro lado, el aislamiento social reduce las interacciones entre las personas a nivel comunitario, evitando las infecciones a los no infectados y que no están aislados (31).

Por último, el confinamiento es una intervención aún más compleja, debido a que conlleva la contención a nivel comunitario, de ciudad o región, con la finalidad de reducir las interacciones entre personas. Se aplica con la combinación de otras medidas como el distanciamiento social, el uso de mascarillas, la restricción de horarios, cierre de fronteras entre otras medidas (30).

II.1.3 Medidas Contra El COVID-19 En El Perú

Ante la pandemia del COVID el Ministerio de salud del Perú publicó el "Plan Nacional de Preparación y Respuesta frente al riesgo de introducción del Coronavirus 2019-nCoV" el cual establecía los lineamientos generales en torno a los objetivos del plan nacional, el ámbito de aplicación, aspectos técnicos y actividades de articulación estratégica a fin de afrontar el riesgo de ingreso del SARS-CoV-2. Sumado a este plan se estableció el "Protocolo para la Atención de Personas con Sospechas o Infección Confirmada por Coronavirus (2019-nCoN)", ambos documentos fueron contemplados a finales de enero del 2020 y con la disposición ante un posible ingreso del virus en ese entonces (1,27).

El 16 de marzo se declaró emergencia nacional, tomando algunas medidas como el distanciamiento social, el cierre de las escuelas, cierre de fronteras y la cuarentena. Todas estas medidas tuvieron la intención de desacelerar la propagación del virus. Sin embargo, se vio un crecimiento de incidencia lineal, los cuales siguen

acumulando los casos de Covid-19, destacando la necesidad de continuar y mejorar diversas medidas que promuevan la disminución de contagios (28).

A pesar de estas diferencias terminológicas, en el Perú se ha utilizado la terminología cuarentena y aislamiento social para el mismo fin, esto desde el "Decreto Supremo N° 044-2020-PCM que declara Estado de Emergencia Nacional por las graves circunstancias que afectan la vida de la Nación a consecuencia del brote del COVID-19", en la cual se declara por primera vez el estado de emergencia nacional (1) y se determina el aislamiento social obligatorio (cuarentena) para la contención de la enfermedad con el fin de evitar un creciente número de casos, colapso hospitalario y muertes por COVID-19 (20).

II.1.4 Movilización Poblacional

Una forma de verificar si las personas en el mundo se encuentran acatando con la "cuarentena" es a través de su movilidad. Es por ello que Google realiza el lanzamiento de los reportes de movilidad comunitaria COVID-19 los cuales proporcionan información sobre cómo ha cambiado la movilidad en respuesta al trabajo desde el hogar y otras políticas destinadas a contener la pandemia. Los reportes utilizan datos anónimos para estimar el desplazamiento de las personas en el tiempo y en diferentes ubicaciones estratégicas tales como plazas, centros comerciales, lugares de trabajo y más (20).

Los datos muestran cómo cambia el número de visitantes en los lugares categorizados en comparación con días de referencia (Porcentaje de cambio del último día). Los días de referencia representan un valor normal para determinado

día de la semana y dichos valores son la mediana de un periodo de 5 semanas comprendido entre el 3 de enero y el 6 de febrero del 2020 (20).

II.1.5. Consecuencias De Las Cuarentenas En La Salud Mental

La salud mental, se ha visto seriamente afectada debido al cambio brusco en las formas de vida de las personas a raíz de la pandemia del COVID-19. A nivel mundial se han registrado estudios que evidencian el impacto a diversos niveles. En los casos más leves se ha registrado frustración, ira, aburrimiento, baja del estado anímico y en los casos más severos se generaron diversos trastornos mentales como la depresión, angustia, ansiedad y estrés postraumático en casos de cuarentenas prolongadas (32–34).

Por otra parte las personas con enfermedades mentales preexistentes sufrieron de un manejo más complejo de sus padecimientos debido a la cuarentena y a la reducción en el acceso a servicios psiquiátricos (35). Viendo incrementados diversos problemas del estado de ánimo, miedo, estigmatización, baja autoestima y falta de autocontrol que requieren de intervenciones múltiples incluidas medidas que puedan brindar un bienestar psicosocial a las personas más vulnerables (36).

Una propuesta de respuesta ante brotes de riesgo elevado que impidan el servicio de apoyo oportuno estaría compuesta de: 1. Un equipo de intervención psicológica planificada, 2. Sesiones dirigidas por un especialista terapeuta, 3. Acceso a una línea de atención psicológica de forma directa 4. Acceso a un equipo de apoyo con conocimientos y capacitación en primeros auxilios psicológicos. Siendo muy importante el monitoreo de cada uno de los grupos antes mencionados, durante el desarrollo de la pandemia (37).

II.1.6 Uso De Redes Sociales Como Medición De Estado De Ánimo

Las redes sociales contribuyeron durante el periodo de cuarentena como un medio de comunicación masivo, teniendo un impacto tan importante como los medios tradicionales y más especializados (38). Diversos estudios prestan especial atención a las redes sociales debido al gran impacto que tienen sobre la vida de las personas, logrando medir hasta 200 tipos de emociones (39).

Así mismo, se ha considerado a las redes sociales como un método de detección pasiva y discreta que mediante el aprendizaje automático permite inferir la inestabilidad emocional (IE) de las personas. Se han logrado grandes avances en la detección del IE gracias a evaluaciones ecológicas momentáneas móviles las cuales obtuvieron información de los usuarios de la red social Twitter (40).

Además, el acceso al internet por parte de la población se ha incrementado a través de los años, esto se evidencia en el incremento de porcentaje de hogares que tienen servicio de internet con un aumento de 28,2% a 29.8% del 2017 al 2018 a nivel nacional, en tanto que en el departamento de Lima se registró un incremento de 49.8% a 51.8% en los años 2017 y 2018 respectivamente (41).

Así mismo la estadística relacionada al porcentaje de hogares con al menos una persona que tiene celular es de 90.9% a nivel nacional y de 94.8% en el departamento de Lima durante el 2018 (41). Estas estadísticas revelan una aproximación al uso masivo de las redes sociales como medios de entretenimiento, fuentes de información y expresión de diversas opiniones a favor o en contra de un tema determinado.

Las redes sociales en la actualidad son consideradas como plataformas para la formación de opinión y participación pública desencadenando cambios políticos y sociales. Es por ello que se enfatiza la relevancia en la investigación de los datos de las redes sociales en el marco de las opiniones políticas, manejo de crisis, calidad de intervención, eventos geopolíticos o seguridad, estados de salud (42).

II.1.6.1 Twitter

Twitter es un sistema de micro blogs que permite a los usuarios enviar y recibir publicaciones cortas denominadas tweets. Los tweets pueden estar compuestos de hasta 140 caracteres y pueden incluir fotos, vídeos e incluso enlaces a fuentes externas (43).

Debido a la simplicidad de la plataforma en el cual el público puede expresar sus opiniones y realizar comentarios en tiempo real, Twitter es en una de las plataformas proveedoras de datos importantes para los periodistas, políticos y empresarios (44).

II.1.7 Métodos Para Cuantificar Información De Redes Sociales

En la actualidad dentro del campo del Machine Learning existe una rama denominada Procesamiento del Lenguaje Natural (PLN) la cual se centra en el análisis de las comunicaciones humanas y, en concreto, de su lenguaje (45).

Las aplicaciones del PLN tienen diversas ramas, desde detección de similitudes o anomalías en los textos, Chatbots, Clasificación automática de documentos, Análisis de Sentimiento hasta sintetización de textos. Es por ello que las redes

sociales al ser una gran fuente de información de la necesaria para este tipo conocimientos, son usados en la actualidad con mayor frecuencia en el ámbito científico entre las cuales destaca el Análisis de Sentimientos (43,45,46).

II.1.7.1 Análisis De Sentimientos

El análisis de sentimientos la cual es también llamada “opinion mining” es una de las principales ramas de trabajo en el PLN desde los inicios del año 2000. La tarea principal del análisis de sentimientos es poder inferir información proveniente de textos, donde dicha información se encuentra relacionada a las opiniones y/o sentimientos que puedan expresar dichos textos. A su vez esta información podría ser utilizada por un sistema de soporte de decisiones o por un tomador de decisiones (47).

El análisis de sentimientos se enfoca principalmente en las opiniones que expresan o implican sentimientos positivos o negativos, también llamados opiniones positivas o negativas en todos los lenguajes. Este tipo de opiniones es similar al concepto de actitud en la psicología social. Por ejemplo, Eagly y Chaikem definen una actitud como “una tendencia psicológica que es expresada para evaluar una entidad particular con algún grado de acuerdo o desacuerdo”. En la discusión de sentimientos positivos o negativos también se debe considerar expresiones sin algún sentimiento implícito a los cuales se les denomina expresiones neutrales (48).

Las sentencias que expresan opiniones o sentimientos son usualmente sentencias subjetivas lo opuesto a sentencias objetivas, las cuales representan hechos, debido a que tanto opiniones como sentimientos ambos son subjetivos. Sin embargo, las sentencias objetivas pueden implicar sentimientos positivos o negativos de los

autores. Por ejemplo, sabemos que “Compre mi carro ayer y hoy se descompuso” describe hechos indeseables y donde se puede inferir que a partir de la sentencia la cual expresa negatividad acerca de lo ocurrido con el automóvil. El análisis de sentimientos también estudia esos tipos de sentencias objetivas (48).

II. 1.7.2 Análisis De Sentimientos En Twitter

Actualmente redes sociales como Facebook y Twitter son consideradas como las más grandes fuentes de información disponibles. Se han realizado diversos estudios a partir de Twitter enfocados a la salud pública abordando diversas enfermedades como la influenza, obesidad, alergias, insomnio y sobre vacunas, evidenciando así a Twitter como una gran fuente de conocimientos en dicho campo (49–51).

Por otro lado, el análisis de sentimientos hace referencia a la identificación subjetiva de los sentimientos que se pueden expresar a través de un texto (41). Permitiendo de esta forma inferir el estado emocional o mental de un individuo. Actualmente, existen diversas aplicaciones del análisis de sentimientos y metodologías que permiten su uso (42).

Se han descrito una gran variedad de estudios utilizando metodologías del Machine Learning para el análisis de sentimientos orientados a la salud pública, como es el caso del análisis de sentimientos en la atención médica y evaluar la opinión en relación a la atención de la salud (43), así mismo se han realizado investigaciones que prosperaron a un sistema de monitoreo los cuales partieron del análisis de sentimientos para evidenciar la preocupación de los usuarios de Twitter con respecto a un brote de enfermedad (44).

El análisis de sentimientos permite cuantificar la opinión pública de las personas en un determinado momento y sobre un tópico en específico. Es así que, Twitter y el análisis de sentimientos, permitirían en conjunto inferir la opinión pública, a través del sentimiento inferido en los tweets, respecto a la cuarentena durante la pandemia del COVID 19.

II.2. Planteamiento Del Problema

En el Perú la implementación del aislamiento social obligatorio rigió a partir del 15 de marzo del 2020, complementándose con el cierre de las actividades económicas, excepto aquellas relacionadas al área de alimentos. No obstante, las políticas públicas se fueron adaptando en el transcurso del tiempo para optimizar su eficacia (52). Al 10 de diciembre del 2021 no se ha realizado inmovilización social, sin embargo se sigue utilizando de forma obligatoria la doble mascarilla en diversas circunstancias como las vías públicas, ambientes cerrados o en la atención de pacientes con sospecha de infección por COVID 19 (53).

A raíz de las medidas tomadas, tuvieron efectos variados en la población, tanto a nivel del avance del virus como en los ámbitos de salud mental y socioeconómico. Las consecuencias del confinamiento han sido ampliamente descritas en casos como expediciones, prisión, guerras, entre otros, siendo la pandemia del COVID-19 un momento en la historia donde también se recurrió al confinamiento. Algunos efectos en la salud mental estudiados son la ansiedad, depresión, comportamientos adictivos, violencia intradomiciliaria, falta de sueño entre otros (54).

Una forma de percibir estas consecuencias del confinamiento en la salud mental puede verse reflejado a través de las opiniones públicas registradas en las redes sociales. La opinión pública se refiere a las actitudes sociales que experimenta la población ante el cambio y difusión de eventos en un espacio social determinado, así como también su nivel de salud mental. A nivel mundial se ha visto un gran movimiento en las redes sociales, desde compartir información sobre el COVID - 19 hasta expresar opiniones e ideas (47).

La opinión pública influye sobre la prevención y el control de epidemias en diversos aspectos. Un claro ejemplo se refleja en las opiniones hacia las vacunas COVID-19, un estudio realizado en Estados Unidos visibilizó que existía una tendencia general positiva hacia las vacunas que implica confianza y anticipación hacia la misma. Sin embargo, se reveló que también estaban presentes opiniones que expresaron miedo, tristeza e ira (48).

Un estudio realizado en EEUU busco investigar como la pandemia del COVID-19 afectó a los adolescentes en su participación en redes sociales y en su bienestar psicológico. Concluyendo que los adolescentes con problemas de salud mental son más propensos al quebrantamiento de la cuarentena dada su necesidad por interacciones sociales (10). Además, otro estudio realizado en el mismo país introdujo un enfoque basado en redes sociales que cuantifica la proporción pro/anti cuarentena como indicador del riesgo de interacciones humanas. Reflejando a través de sus resultados una asociación entre las opiniones en Twitter y el patrón de movilización de las personas durante la pandemia (55).

En América Latina, se encontró estudios realizados en Colombia y Ecuador que utilizaron las redes sociales para evaluar el nivel de aceptación de la cuarentena por parte de los ciudadanos. En el caso de Colombia se concluyó que temas asociados a la cuarentena generan sentimientos negativos destacando el miedo entre ellas (56), por otro lado, en Ecuador se realizó un análisis de opinión sobre tweets del COVID-19 obteniendo que el 54.37% fueron categorizados como positivos y el 35,18%

como negativos reflejando cierto nivel de aceptación a temas como la cuarentena en dicho país (57).

Por lo anterior se hace necesario determinar si las publicaciones realizadas en redes sociales estuvieron asociadas a la movilidad poblacional en un país de mediano ingreso como el Perú.

Por tanto, se plantea la siguiente pregunta de investigación

II.3 Pregunta De Investigación

¿La opinión pública en redes sociales está relacionada con la movilidad durante la cuarentena en el Perú?

II.4 Justificación Del Estudio

A pesar que el período de cuarentena, no fue totalmente exitoso en nuestro país por factores tales como lo económico, sociocultural, laboral, entre otras (58). Se hace necesario evaluar su cumplimiento a través de la movilidad poblacional y asociarla con la opinión pública a través de las redes sociales.

Algunos estudios realizados en Israel, Egipto y UK han demostrado que las publicaciones en redes sociales tienden a reflejar el comportamiento de las personas, el estado de ánimo, entre otros y que podrían verse plasmados en el cumplimiento de la cuarentena (12,55,58,59). Otros estudios realizados en América Latina determinaron el nivel de aceptación de la cuarentena a través de las redes sociales, mostrando sentimientos negativos hacia el contexto en el que se vivía (60)

Se ha encontrado poca evidencia de la relación entre publicaciones en redes sociales y el cumplimiento de la cuarentena. Sin embargo, un estudio reportó que las personas que realizan publicaciones en contra de la cuarentena tuvieron una mayor tendencia a movilizarse (55). Evaluar la asociación entre las variables de opinión pública y la movilización social, podrían brindar información a los tomadores de decisión locales o de otros países parecidos al nuestro, con el fin de fortalecer medidas como la cuarentena u otros para evitar más contagios debido a esta pandemia.

Además, el uso de metodología novedosa como el Machine Learning que a través del análisis de sentimientos permite cuantificar la opinión pública en redes sociales de forma rápida. Por ende, es uno de los métodos más utilizados en la determinación de la percepción pública en redes sociales orientados a tópicos específicos. Y el uso de información tales como la movilidad en de Google, hace de este estudio uno de los primeros en desarrollarse en el mundo hasta la fecha de publicación de este estudio.

Por tanto, el presente estudio busca determinar la asociación entre la opinión pública obtenida en Twitter y la movilidad de las personas durante el periodo de la primera cuarentena en el Perú, a través de métodos novedosos como el Machine Learning.

III. OBJETIVOS

III.1. Objetivo General

- Identificar la correlación entre la opinión pública en Twitter mediante el Análisis de Sentimientos y la movilidad poblacional en el periodo de cuarentena en el Perú.

III.2. Objetivos Específicos

- Realizar análisis de sentimientos para determinar la opinión pública en Twitter respecto a la cuarentena en el Perú.
- Describir la movilidad poblacional por regiones en el Perú.

IV. MATERIAL Y MÉTODOS

IV.1. Diseño del estudio

Estudio ecológico, donde se realizará un análisis de datos secundario de acceso público, para determinar la opinión pública en redes sociales en la cual la unidad de análisis será las publicaciones realizadas en la plataforma de Twitter (tweets) durante el periodo de la cuarentena.

IV.2. Población

La población son todos los tweets que cumplan los criterios de inclusión y exclusión establecidos en el periodo de tiempo correspondiente a la cuarentena en el Perú.

Criterios de inclusión:

1. Tweets de usuarios pertenecientes a la región de Perú.
2. Tweets de usuarios que permiten visualizar su locación.
3. Tweets escritos en el idioma español.
4. Tweets que hacen referencia a la cuarentena, pandemia, COVID-19 y medidas tomadas por el gobierno.
5. Tweets que pertenezcan al periodo de tiempo donde se implementaron las medidas de cuarentena en el Perú.

Criterios de exclusión:

1. No serán incluidos en el estudio los retweets de otros usuarios.
2. No serán incluidos tweets correspondientes a medios de comunicación tanto públicos como privados.
3. No serán incluidos tweets los cuales contienen URLs.

IV.3. Operacionalización De Variables

Tabla 1. Variables de estudio

VARIABLE	TIPO DE VARIABLE	ESCALA DE MEDICIÓN	DEFINICIÓN CONCEPTUAL	FORMA DE REGISTRO
Sentimiento del tweet	Cualitativo	Nominal	Opinión de los usuarios determinado con respecto al modelo de clasificación con respecto a el tópico de interés del proyecto.	- Neutral - Positivo - Negativo
Movilidad	Cuantitativo	Razón	Los valores muestran la variación porcentual en la movilidad poblacional con respecto a un año de referencia.	Porcentaje de cambio con respecto a un valor de referencia

Tabla 2. Variables de ajuste

VARIABLE	TIPO DE VARIABLE	ESCALA DE MEDICIÓN	DEFINICIÓN CONCEPTUAL	FORMA DE REGISTRO
Movilidad poblacional uno días antes de publicación de tweet	Cuantitativo	Razón	Los valores muestran la variación porcentual en la movilidad poblacional con respecto a un año de referencia un día antes de la publicación del tweet.	Porcentaje de cambio con respecto a un valor de referencia
Movilidad poblacional dos días antes de publicación de tweet	Cuantitativo	Razón	Los valores muestran la variación porcentual en la movilidad poblacional con respecto a un año de referencia dos días antes de la publicación del tweet.	Porcentaje de cambio con respecto a un valor de referencia
Dia de la semana	Categórico	Ordinal	Variable que incluye un factor de temporalidad con respecto al día de la semana	0, 1, 2, 3, 4, 5, 6 y 7
Dia del año	Cuantitativo	Razón	Variable que añade el factor de temporalidad en la información obtenida	Numérico
$(Dia\ del\ año)^2$	Cuantitativo	Razón	Variable que añade el factor de temporalidad	Numérico

			cuadrática en la información obtenida	
$(Dia\ del\ año)^3$	Cuantitativo	Razón	Variable que añade el factor de temporalidad cúbica en la información obtenida	Numérico
Origen del tweet	Categórico	Nominal	Variable que muestra si un tweet fue publicado en el departamento de Lima o no	- Lima - No Lima

IV.4. Técnicas Y Procedimientos

Para el cumplimiento de los objetivos del presente estudio es necesario determinar la opinión pública acerca de la cuarentena y la movilidad local en el Perú. Es por ello que la metodología consta de dos etapas previas al análisis estadístico las cuales son:

Etapa 1: Determinación De La Opinión Pública

Para inferir la opinión pública extraída de Twitter en las categorías: positivo, negativo y neutro, se siguieron los pasos mostrados en la Figura 1.

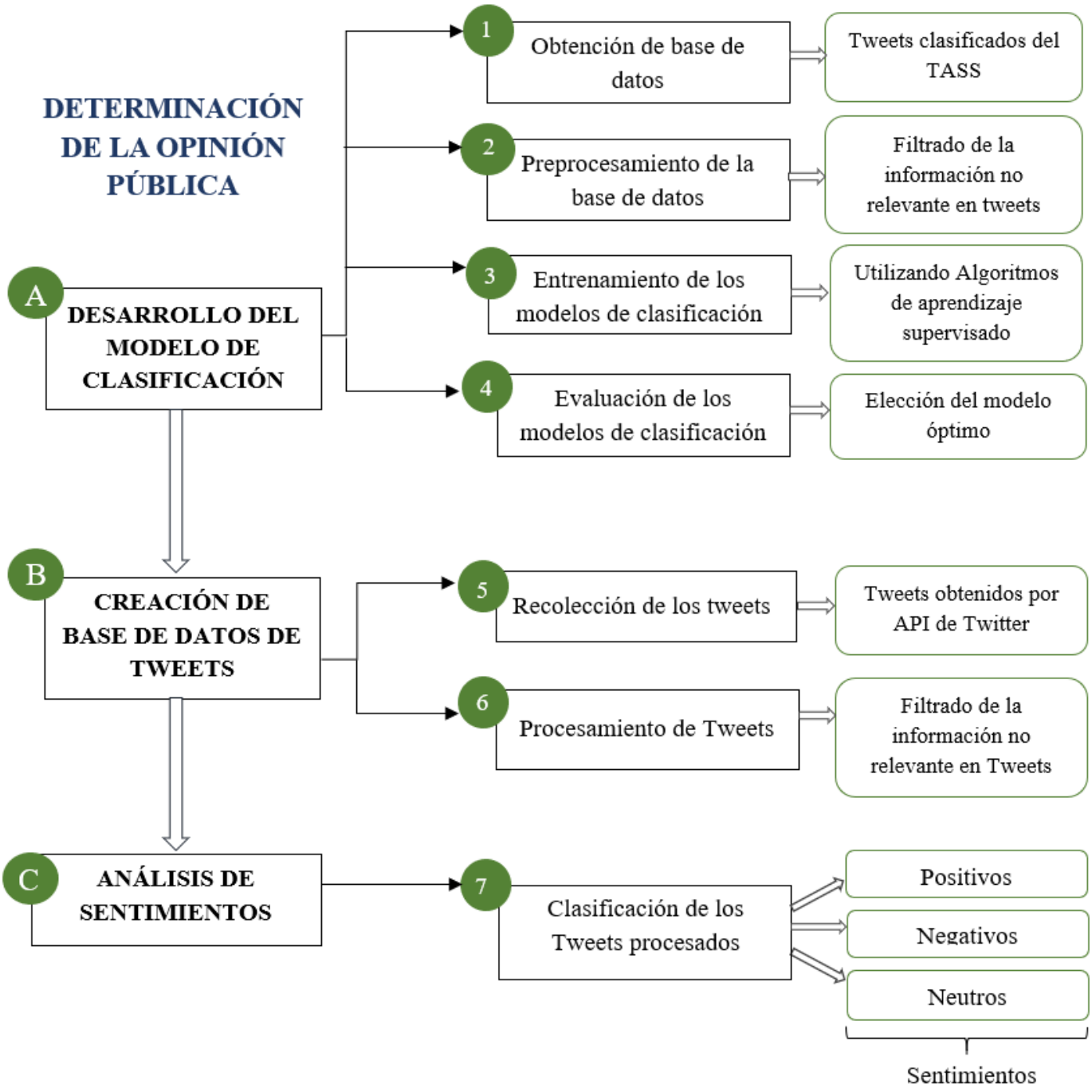

Figura 1. Metodología de la determinación de la opinión pública

A. Desarrollo Del Modelo De Clasificación

El modelo de clasificación es un sistema entrenado a partir de un conjunto de datos y cuya función es determinar la opinión, sentimiento inferido, que este envuelto en un tweet.

A.1 Obtención De Base De Datos

Los datos de Twitter para el desarrollo del sistema de clasificación (será positivo, negativo, neutro) de tweets fueron obtenidos del Taller de Análisis de Sentimientos en la SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural) conocido por la comunidad investigadora como TASS, el cual desde el año 2012 vienen fomentando la investigación en el campo del análisis de sentimientos en la red social Twitter, centrándose específicamente en el idioma español y donde hasta la actualidad vienen incrementando su base de datos (corpus) las cuales están disponibles para investigación.

La base de datos de tweets obtenida de la TASS está preparada para el desarrollo de modelos que realicen tareas como el análisis de sentimientos. Además, los tweets se encuentran clasificados entre las siguientes categorías positivo (P), negativo (N) y neutral (NEU). La base de datos se encuentra en formato XML es por ello que haciendo uso de algoritmos en Python se transformó dicha información a un formato tabla para su uso en el desarrollo del modelo de clasificación (Anexo1).

A.2 Preprocesamiento De La Base De Datos

Los tweets obtenidos de la TASS para el desarrollo del modelo de clasificación fueron sometidos a un preprocesamiento de textos con el objetivo de transformar la

información desestructurada de los textos a un formato legible para la computadora. Es por ello que se aplicó una serie de técnicas y así remover información irrelevante para el análisis tales como: nombre de usuarios, URLs, caracteres que no estén en formato ASCII, retweets, letras singulares, palabras de interrupción (por ejemplo, do, la, tu, etc.), espacios adicionales, palabras en otros lenguajes. Además, se realizó la tokenización de los textos la cual involucra aislar las palabras presentes en un tweet. Para finalizar la etapa procesamiento se extrajo las raíces de cada una de las palabras, es decir la mínima porción de la palabra que conserva su significado, siendo estas los datos de entrada en nuestro modelo a entrenar (Anexo2).

A.3 Entrenamiento De Los Modelos De Clasificación

Cada modelo de clasificación se obtiene a partir del uso de un "Algoritmos de Aprendizaje Supervisado". Los cuales fueron entrenados y posteriormente se evaluó el rendimiento haciendo uso de los datos de entrenamiento y evaluación respectivamente. Los algoritmos de aprendizaje supervisado utilizados en el estudio se muestran a continuación:

Clasificador de Naïve Bayes

El método de Naïve Bayes es un clasificador que hace uso del teorema de Bayes con la asunción "naive" la cual hace referencia que la presencia o ausencia de alguna variable no está relacionada a la presencia o ausencia de alguna otra por lo tanto son independientes. Esta suposición de independencia entre las variables hace que este algoritmo sea extremadamente rápido en comparación con algoritmos más sofisticados.

Actualmente existen variantes de este algoritmo las cuales fueron usadas en el presente estudio, dichas variantes son:

1. ***Multinomial Naive Bayes***: Esta variante implementa el algoritmo de bayes para datos distribuidos de forma multinomial y es una de las dos variantes que se utilizan en la clasificación de textos (donde los datos se representan normalmente como un vector de frecuencia de palabras)
2. ***Complement Naive Bayes***: Este algoritmo es una variante del algoritmo Multinomial Naive Bayes que es adecuado para trabajar con datos desbalanceados.
3. ***Bernoulli Naive Bayes***: Este algoritmo implementa el algoritmo de Naive Bayes para datos que están distribuidos acorde a una distribución de Bernoulli multivariada

Regresión Logística

La regresión logística es un método estadístico similar a la regresión lineal, ya que la regresión logística encuentra una ecuación que predice un resultado para una variable binaria a partir de una o más variables. La regresión logística asume independencia entre las variables predictoras lo que no siempre se cumple en los conjuntos de datos.

K-Nearest Neighbors (KNN)

Es un método para clasificar datos basándose en su parecido a otros datos, donde cada dato es un punto en un hiperplano. El principio detrás del método KNN es encontrar un conjunto de puntos lo más cercanos en distancia a un punto nuevo y determinar la categoría de estas. El número de conjuntos puede ser un valor fijo

establecido por el usuario o variar basado en la densidad de puntos. Además, cabe mencionar que este algoritmo posee una variante para tareas de clasificación denominado KNN Classifier.

Árboles de Decisión

Los árboles de decisión son algoritmos de aprendizaje supervisado los cuales son utilizados en tareas como clasificación o predicción. El objetivo es crear un conjunto de reglas las cuales puedan dividir nuestro conjunto de datos entre sus diferentes categorías.

Random Forest

Random Forest es un conjunto de árboles combinados los cuales son entrenados con un conjunto distinto de datos. De esta forma, al combinar los resultados, unos errores se compensan con otros y tenemos así una mejor predicción.

Clasificación SGD

Este estimador implementa modelos lineales regularizados con un aprendizaje mediante gradiente descendente estocástica (SGD). La gradiente de la función de costo se estima en cada muestra a la vez y el modelo es actualizado a lo largo del camino hasta llegar a un mínimo global.

Support Vector Machine

El modelo de Support Vector Machine (SVM) es un algoritmo de aprendizaje supervisado. Dicho algoritmo tiene como objetivo es encontrar un hiperplano que separe de la mejor manera dos conjuntos diferentes de puntos. Este algoritmo

presenta variaciones las cuales fueron implementadas en el estudio, estas variantes son:

1. ***SVC***: Es una variante del algoritmo de Support Vector Machine (SVM) cuyo desarrollo se encuentra orientado a clasificaciones múltiples haciendo uso del enfoque "uno-versus-uno" en la cual dividimos el conjunto de datos total con N clases en N*(N-1) /2 subconjuntos de datos de clasificación binaria para cada par de clases.
2. ***Linear SVC***: Es otra implementación (más rápida) de SVM haciendo uso de kernels lineales en el caso de la clasificación de datos con múltiples categorías implementando la estrategia de "uno-versus-el-resto" en la cual para para un conjunto de datos con N clases, se generará un total de N modelos de clasificación binaria.

MultiLayer Perceptron Neural Network

Considerando un problema de aprendizaje supervisado donde se tiene acceso a los datos de entrenamiento con sus respectivas etiquetas. Las redes neuronales brindan una forma de definir una hipótesis compleja y no lineal, con un listado de parámetros que se ajustaran a nuestros datos.

Clasificador AdaBoost

AdaBoost es un método de aprendizaje ensamblado el cual inicialmente es creado para incrementar la eficiencia de clasificadores binarios. AdaBoost usa un enfoque iterativo para aprender de los errores de los clasificadores "débiles" y así obtener un clasificador más robusto.

A.4 Evaluación De Los Modelos De Clasificación

En esta etapa se evaluó el rendimiento de los algoritmos de clasificación presentados anteriormente. Posteriormente se eligió aquel modelo que presentó un mayor rendimiento con respecto a los demás, siendo este el modelo a utilizar en el análisis de sentimientos de los tweets.

Para la realización de esta tarea se utilizó el siguiente método:

Validación Cruzada De K Iteraciones

La validación cruzada es una estrategia utilizada para la evaluación de modelos de clasificación que asegura una independencia entre el conjunto de entrenamiento y prueba. Dicha independencia es asegurada debido a la manera en la que se construyen los conjuntos. Para determinar la precisión del modelo de clasificación se usará el promedio de las precisiones obtenidas del modelo aplicado sobre cada uno de los conjuntos distintos.

Este método consiste en dividir de manera aleatoria el conjunto de datos en k grupos del mismo tamaño, de los cuales k-1 grupos fueron utilizados para entrenar el modelo y el grupo restante será usado como conjunto de evaluación. Este conjunto de pasos se repite k veces seleccionando un grupo diferente como conjunto de evaluación. Las métricas del modelo de clasificación se obtendrán como el promedio de las métricas de cada uno de los modelos obtenidos en cada una de las diferentes iteraciones.

El modelo de clasificación que se utilizó en el análisis de sentimientos es aquel que presenta la mayor precisión. La implementación del método de validación cruzada se realizó en Python haciendo uso de las librerías de Scikit-Learn (Anexo3).

B. Creación De Base De Datos De Tweets

B.1 Recolección De Tweets

La fuente de información elegida en el estudio fue la red social Twitter, debido a que es considerado el microblog más popular en el mundo, con más de 200 millones de publicaciones (tweets) diarias. Las publicaciones obtenidas de los usuarios de Twitter se focalizaron en relación a la cuarentena del año 2020 en el Perú.

Los tweets para la realización de este estudio fueron extraídos a través de la Interfaz de Programación de Aplicaciones (API) de Twitter a través del servicio PREMIUM, la cual nos permite tener acceso a la transmisión de datos con antigüedad mayor a los 7 días. Se definieron términos de búsqueda que abarcan palabras cómo: “quedateencasa”, “peruentusmanos”, “COVID19”, “covidenperú”, “AislamientoSocial”, “Cuarentenaperu”, “covid19peru”, “coronavirusenperu”, “Reactiva Perur”, “cuarentenatotal”, “UnidosEnCasa”, “YoApoyoAVizcarra”, “Comando COVID-19”, “CuarentenaNacional”, “CuarentenaExtendida”, etc las cuales fueron elegidas debido a que fueron hashtags tendencia (tópicos populares en Twitter) que hacen referencia tanto a la pandemia como a la cuarentena establecida por el gobierno del Perú (39,40).

Los tweets que contienen al menos una de las palabras claves establecidas fueron recolectadas mediante el uso de algoritmos desarrollados en Python usando la

librería SQLAlchemy la cual está conectada a un gestor de base de datos llamado PostgreSQL (Anexo 4).

La información luego del proceso de recolección de datos a través de la API de Twitter se encuentra en formato JSON y está a través de un algoritmo en Python junto con la librería “Pandas” reestructuramos esta información a un formato tabla. Además, se hace mención que los tweets obtenidos vienen con sus respectivos metadatos en las cuales se tiene información que abarca geolocalización, fecha de publicación, id de usuario, nickname, número de “me gustas”, etc (Anexo 5).

PREMIUM API entre sus funciones permite recolectar los tweets a lo largo del periodo de duración del estado de emergencia además que el número de tweets por unidad de tiempo es mayor pudiendo de esta manera conseguir un volumen mayor de datos.

B.2. Procesamiento De Tweets

Sobre los tweets obtenidos se realizó un preprocesamiento con el objetivo de transformar dichos textos no estructurados a un formato usable el cual es requerido para su futuro análisis. Esta práctica implica el uso de técnicas como (40–44):

1. Eliminación de caracteres Unicode: En esta fase eliminamos signos de puntuación, URLs y emojis las cuales aparecen frecuentemente en los tweets. Esto se realizó mediante el uso de la librería “re” de Python la cual trabaja con expresiones regulares en textos.
2. Tokenización: Etapa en la que procedimos a dividir los tweets en una lista de palabras. Al igual que en el apartado anterior se hizo uso de la librería “re” de

Python la cual nos permite dividir textos siguiendo un patrón (por ejemplo, el espacio entre palabras).

3. Normalización del texto: El siguiente paso realizado en nuestro flujo de trabajo consistió en reconocer el formato de las palabras, minúsculas y mayúsculas, y transformarlas a minúsculas, se removió los nombres de los textos, se filtró los números que se encuentren presentes en el texto, se eliminó los caracteres repetidos dentro de la misma palabra. La realización de los pasos anteriores mencionados se realizó con la librería "re" de Python que está orientada al procesamiento de textos.
4. Filtrado de palabras de interrupción: En este paso se procedió a filtrar palabras que no contribuyen al significado más profundo de la frase. Para la realización de esta tarea se hizo uso de la librería NLTK de Python la cual proporciona una lista de palabras de interrupción para una variedad de idiomas.
5. Lematización/Radicalización: Esta tarea implicó la reducción de una palabra a su forma de raíz, con el objetivo de reducir la variación de la misma palabra dentro del mismo texto. Esto con el propósito de reducir el conjunto de palabras a incluir en nuestro modelo.

6. Geolocalización de los tweets: Luego del preprocesamiento de los tweets, se procedió con la geolocalización de los mismos, esto fue posible mediante el uso de los metadatos **"Location_user"** y **"Place_tweet"** siendo la primera de ambas la ubicación que el usuario, dueño del tweet, comparte en su perfil personal de su cuenta de Twitter mientras que la siguiente corresponde a la ubicación en la cual se realizó la publicación del tweet. Los formatos en cuales se presenta la información de los campos anteriormente mencionados no son estándares por ende se tuvo que

realizar un preprocesamiento de dicha información y clasificar las ubicaciones disponibles por provincias.

Todo el procesamiento realizado para la geolocalización de los tweets se realizó a través de algoritmos desarrollados en Python en conjunto con las librerías “Pandas” y “Numpy” (Anexo 2).

C. Análisis De Sentimientos

C.1 Clasificación De Los Tweets Procesados

Luego de obtenido el modelo con mayor rendimiento haciendo uso del método de validación cruzada, se procedió con el Análisis de Sentimientos sobre nuestra base de datos de tweets obtenidos en el periodo de la cuarentena del 2020.

El análisis de sentimientos es la tarea de extraer información subjetiva a partir de los textos, es decir ofrecer un juicio positivo o negativo a un comentario, opinión, frase o documento. Por lo tanto, en esta etapa se asignan las etiquetas positivo (P), negativo (N) y neutral (NEU) a los tweets que conforman nuestra base de datos.

Etapa 2: Determinación De La Movilidad En Perú

A. Obtención De Datos

La obtención de los datos de movilidad de Google se obtuvo a través de la plataforma de acceso público que tiene como nombre INFORMES DE MOVILIDAD LOCAL SOBRE EL COVID-19 el cual sigue vigente hasta la fecha.

B. Procesamiento De Datos

En esta etapa se realizó un procesamiento sobre los datos que hacen referencia a la movilidad filtrando la información que corresponde exclusivamente a Perú.

Además, los datos contienen información de cómo varía la movilidad en lugares específicos a lo largo de las diferentes provincias del Perú. Es por ello que se tomó el promedio de las variaciones por provincia como dato único que representa los cambios de movilidad en una provincia (Anexo 6).

V. PLAN DE ANÁLISIS

V.1 Modelos De Categorización De La Variable Sentimiento Inferido A Partir De Los Tweets:

Para poder inferir la variable sentimiento asociado al tweet, se utilizó aquel modelo de categorización o clasificación (Entre todos los modelos descritos en la sección *A.3 Entrenamiento de los modelos de clasificación* tales como: Clasificador de Naive Bayes, Regresión Logística, Arboles de decisión, etc.) que presento una mayor precisión al momento de inferir los sentimientos tanto positivos como negativos asociados a un tweet.

Cabe resaltar que, el modelo de categorización con mayor precisión para poder inferir el sentimiento asociado a un tweet, fue el modelo de regresión logística.

V.2 Estadística Descriptiva Del Análisis De Sentimientos:

Se hizo una descripción de los resultados obtenidos a partir del análisis de sentimientos tales como la evolución temporal del número de tweets negativos y

positivos a lo largo de la duración de la primera cuarentena. Además, se determinó la proporción de tweets clasificados como positivos y negativos dentro de nuestra base de datos.

V.3 Análisis De Asociación Entre El Sentimiento Inferido En Los Tweets Y La Movilidad:

El análisis estadístico fue realizado a través de la regresión de Poisson y de esta manera se determinó la asociación entre la variable asociada a la opinión pública denominada sentimiento inferido del tweet y la variable asociada a la movilidad poblacional, con un nivel de significancia del 5%. Además, se evaluó como esta asociación se ven afectada de manera gradual por las variables de ajuste, tales como: la movilidad poblacional uno y dos previos a la publicación del tweet, el día de semana la cual añade un factor de periodicidad en el sentimiento inferido en un tweet, día del año en la cual se realizó la publicación del tweet así mismo se añadió componentes que reflejen una posible dependencia no lineal entre el día del año y el sentimiento inferido del tweet para finalizar ser añadió un factor de ajuste que representa si la publicación se realizó en el departamento de lima o no y así reflejar si el sentimiento inferido del tweet se ve afectada por el origen de la publicación.

Todos los análisis descritos en el párrafo anterior se realizaron utilizando el lenguaje de programación R el cual es un entorno de software libre.

VI. CONSIDERACIONES ÉTICAS

Los datos que se obtuvieron conforman la base de datos de Twitter International Company, el cual presenta las consideraciones éticas de la privacidad de los datos

personales que son aceptadas al momento de descargar Twitter. Los datos publicados por cada usuario autorizan a transferir, almacenar y usar la información por cada país en el que operen. La información pública incluye además de los mensajes los metadatos, los cuales contienen información tales como cuándo se ha Twitteado, el idioma, el país, ubicación, entre otros. Así mismo Twitter en su política de privacidad hace de conocimiento que los datos colectados son distribuidos a motores de búsqueda, desarrolladores, editores, clientes, organizaciones, agencias de salud pública, universidades y empresas de investigación a fin de analizar la información en búsqueda de tendencias y conocimiento.

En el presente estudio los datos colectados estuvieron alineados a la política de privacidad previamente aceptada por el usuario. Se utilizaron los Tweets redactados por cada usuario, el análisis fue en base a estas redacciones cortas que incluyen información de opiniones. La ubicación está considerada como parte del presente estudio solo para los usuarios que permitieron la visibilización del mismo de manera abierta, esto acorde con la política de privacidad que indica que el usuario al hacer de conocimiento su ubicación a través del GPS permite el tratamiento de estos datos para hacerlos parte de la base de datos de Twitter.

Finalmente, el protocolo de estudio fue enviado al comité de ética de la Universidad Peruana Cayetano Heredia el cual fue aprobado con número de SIDISI 203561.

VII. RESULTADOS

VII.1 Descripción De La Población De Tweets

Se recolectaron 118,451 tweets correspondientes a un total de 22,238 usuarios quienes en promedio realizaron 4 publicaciones en Twitter durante el periodo de la primera cuarentena en el Perú la cual inició el 16 de marzo y finalizó el 31 de julio del 2020. De acuerdo a los criterios de inclusión y exclusión se analizaron 89,928 tweets. La Tabla N°3 nos muestra que el departamento de Lima contiene un mayor volumen de tweets con respecto a los demás departamentos.

Tabla 3. Distribución de tweets por departamento

DEPARTAMENTO	**TWEETS (%)**
Lima	64.99
Cusco	4.21
Arequipa	2.45
La Libertad	2.40
Callao	1.88
Otros	24.07

El número de publicaciones realizadas en el Perú enfocadas a la pandemia durante la cuarentena del 2020 fue incrementándose a lo largo del tiempo. Además, existe un salto diferencial en el número de tweets finalizando el mes de mayo e iniciando el mes de junio tal y cual se puede observar en la figura N° 2.

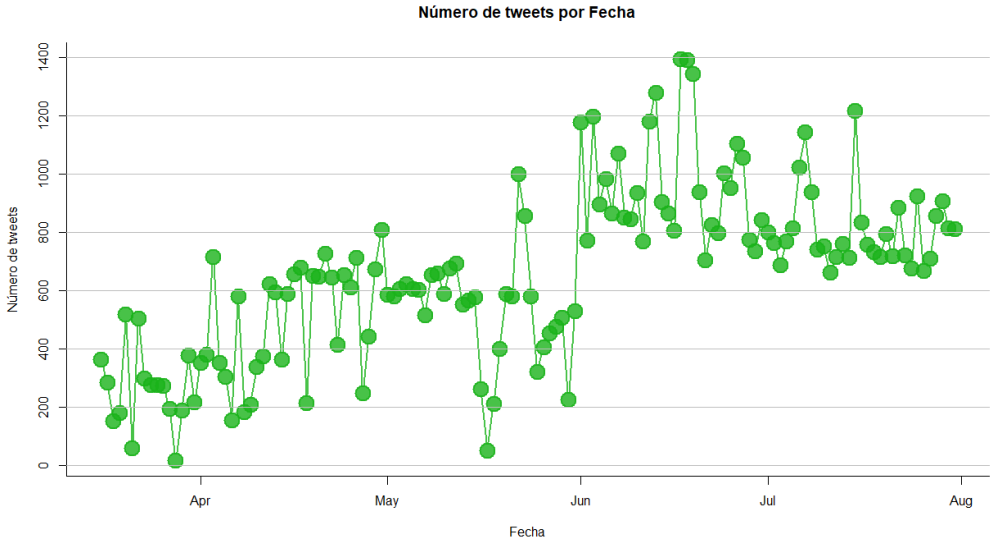

Figura 2. Número de tweets en la cuarenta

En adición, se extrajo las palabras que poseen una alta frecuencia dentro la base de datos de tweets y estas se encuentran plasmadas en la figura N° 3 a través de una nube de palabras. Se puede visualizar que palabras como “Vizcarra”, “covid”, “cuarentena” entre otras reflejan que el tema principal que abordan los tweets es la cuarentena validando así el método usado para la recolección de tweets.

Figura 3. Nube de palabras de la base de datos de tweets

VII.2 Opinión Pública De Los Tweets

El Análisis de Sentimientos se realizó utilizando el modelo de clasificación obtenido a través del algoritmo de regresión logística la cual presento un desempeño mayor desempeño con respecto a los demás modelos entrenados esto se puede evidenciar en el Anexo 7.

VII.2.1 Análisis De Sentimientos

Una vez establecido el modelo de clasificación, se procedió con la aplicación de dicho modelo sobre los tweets recolectados en el periodo de la primera cuarentena. La aplicación de dicho modelo nos dio como resultado que el 84.55% de los tweets (n = 63835) fueron categorizados como negativos, mientras que el 14.09% (n = 10636) fueron clasificados como positivos. Esto significa que los usuarios de Twitter tienen una perspectiva negativa a temas relacionados a la cuarentena. Cabe mencionar que los tweets categorizados como neutros no entran al análisis debido a que no reflejan una posición especifica con respecto a la cuarentena.

El resultado luego de realizado la tarea de clasificación está plasmado en la siguiente tabla:

Tabla 4. Distribución de tweets positivos, negativos y neutros por departamentos y la Provincia Constitucional de Callao

PROVINCIA	NÚMERO DE USUARIOS	NÚMERO DE TWEETS	NÚMERO DE TWEETS POSITIVOS	NÚMERO DE TWEETS NEGATIVOS	NÚMERO DE TWEETS NEUTROS
Lima	13828	584345	8168	49491	803
La Libertad	824	2159	382	1755	25
Cusco	796	3794	522	3211	61
Arequipa	558	2206	260	1929	28
Callao	506	1696	258	1422	16
Lambayeque	369	1002	174	816	13
Piura	328	1212	145	1046	22
Ica	235	692	110	575	8
Ancash	197	533	84	447	2
Loreto	197	531	101	424	6
Cajamarca	156	494	70	421	3
San Martin	149	348	65	280	3
Junín	146	431	64	362	5
Tacna	134	487	53	427	8

Huánuco	65	348	45	299	4
Ayacucho	63	269	29	238	2
Ucayali	51	191	28	162	1
Puno	45	127	10	115	2
Moquegua	40	146	19	125	2
Tumbes	37	69	8	59	3
Amazonas	36	107	21	86	0
Apurímac	20	76	7	69	0

La figura N° 4, muestra la evolución temporal del porcentaje de tweets negativos. Se puede observar que en los primeros 50 días de la cuarentena existe un crecimiento en el número de tweets negativos la cual viene acompañada de picos de descenso en los cuales el porcentaje de tweets negativos disminuye hasta en un 12%. Sin embargo, posterior a los 50 días de cuarentena tenemos que el porcentaje de tweets negativos presenta un crecimiento sostenido sin picos descendientes similares a los primeros 50 días.

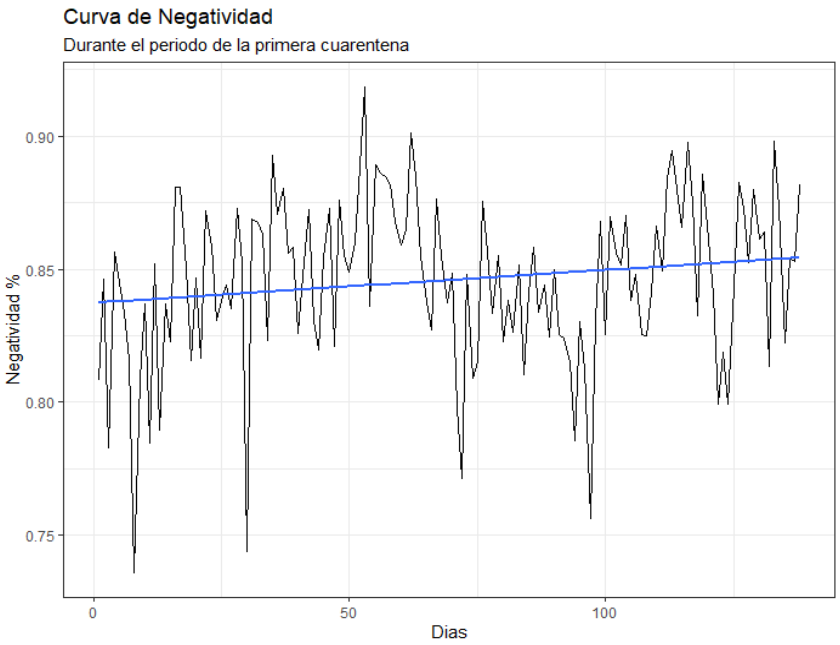

Figura 4. Porcentaje de tweets negativos en el tiempo

Como contraparte, la figura N° 5 representa la evolución temporal del porcentaje de tweets positivos a lo largo del periodo de duración de la cuarentena. Debido a que el total de tweets en el análisis lo conforman los tweets positivos y negativos, los patrones que presenta con complementarias a la figura N° 4.

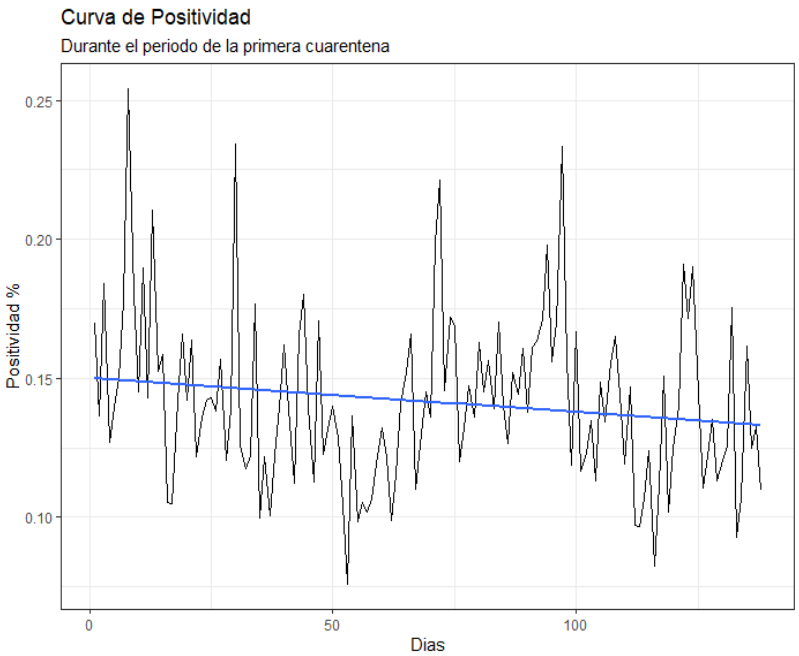

Figura 5. Porcentaje de tweets positivos en el tiempo

VII.3 Descripción De La Movilidad En El Perú

Dado que los reportes de movilidad Google muestran cómo cambiaron las visitas a diferentes lugares en comparación a un valor de referencia. En la Gráfica N° 6 se puede observar que, durante los primeros días de cuarentena, la movilidad en el Perú, tuvo una disminución de hasta un 80% con respecto a su valor referencial. Por otro lado, se aprecia una tendencia al incremento a lo largo de los días, donde los picos inferiores presentes en la gráfica representan los feriados y domingos en los cuales el gobierno del Perú estableció la inmovilidad total para todos los sectores

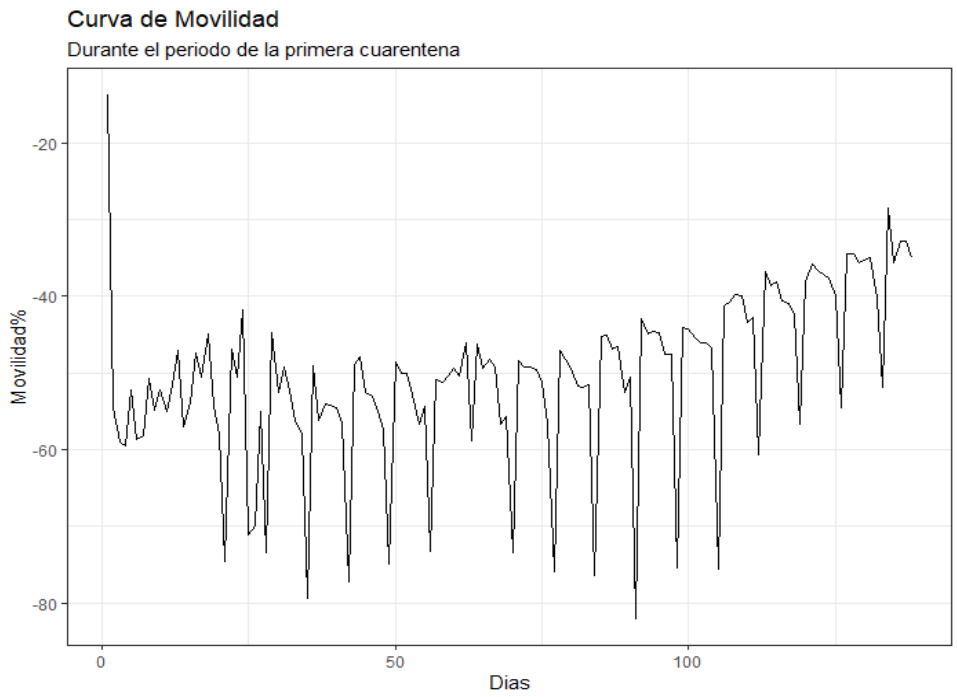

Figura 6. Movilidad poblacional del Perú durante la cuarentena

Se puede observar en la figura N° 7 que los departamentos presentan patrones semejantes a lo largo de la duración de la cuarentena. Sin embargo, departamentos

como Loreto, Huancavelica y Ucayali destacan en la gráfica debido a que sus patrones de movilidad poblacional prácticamente retomaron valores normales. Caso contrario, el departamento de Ica desde implementada la cuarentena ha mantenido su movilidad poblacional casi constante.

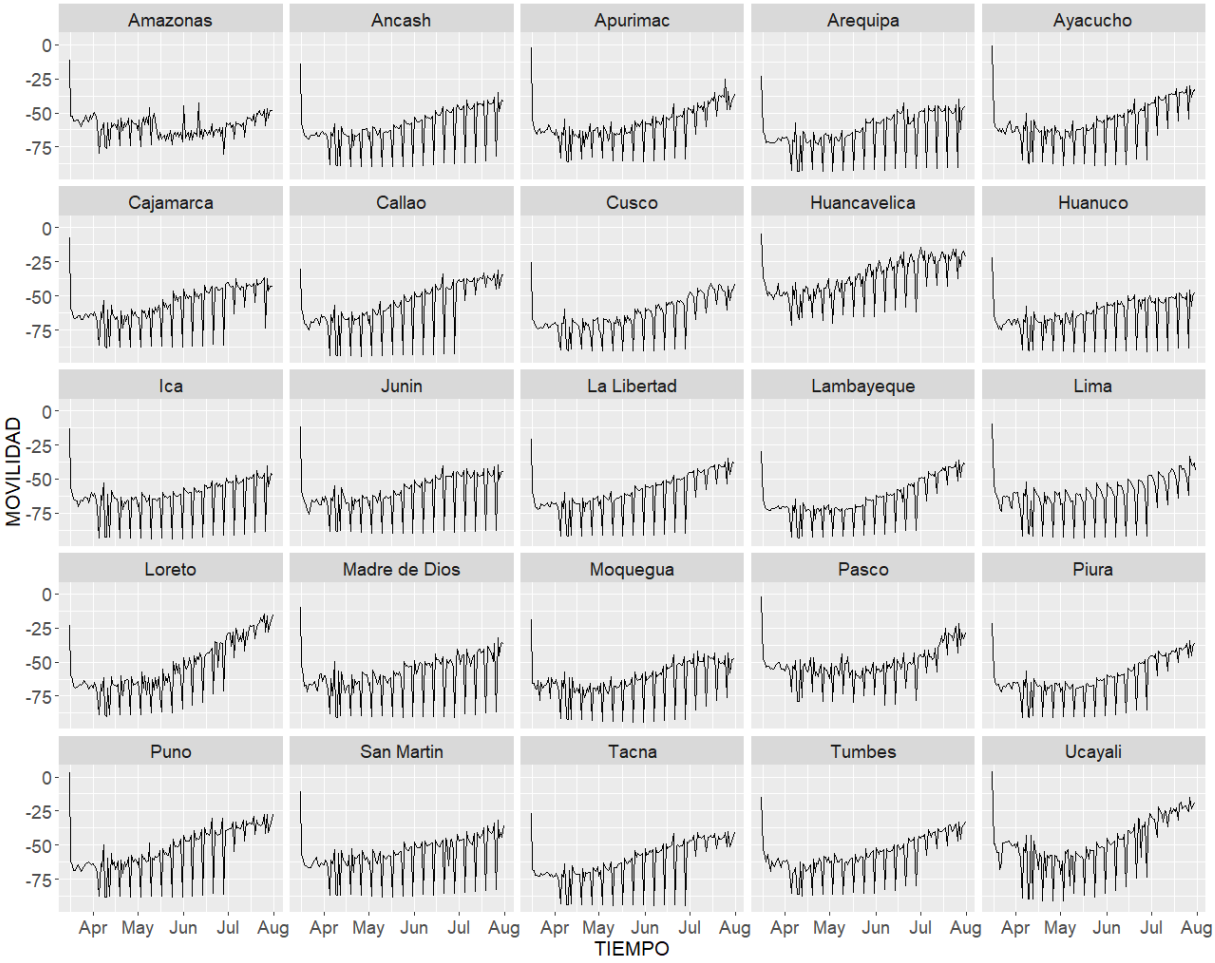

Figura 7. Cambios de movilidad poblacional en los departamentos del Perú y la provincia constitucional del Callao.

VII.4 Análisis De Asociación

Las razones de prevalencias cruda y ajustadas obtenidas a partir de las regresiones de Poisson no mostraron evidencia de algún tipo de asociación significativa entre el sentimiento del tweet y la movilidad durante la cuarentena (tabla 5).

Tabla 5. Evaluación del sentimiento del tweet respecto a la movilidad

Modelo	**Razón de Prevalencia**	**Intervalo de Confianza 95%**		**Valor de P**
Crudo	1.000	0.999	1.001	0.714
Ajustado *	0.999	0.999	1.001	0.946
Ajustado **	0.996	0.993	0.999	0.0512
Ajustado ***	0.998	0.994	1.002	0.4314
Ajustado ****	0.998	0.994	1.002	0.4029

* Modelo crudo ajustado por la movilidad 1 y 2 días antes de la publicación del tweet

** Modelo previo ajustado por día de semana

*** Modelo previo ajustado por día del año, $(día\ del\ año)^2$ y $(día\ del\ año)^3$

****Modelo previo ajustado por origen del tweet (Lima/No Lima)

VIII. DISCUSIÓN

Nuestro estudio no encontró asociación estadísticamente significativa entre la opinión pública obtenida a partir de los tweets enfocados al contexto de la cuarentena y la movilidad local en el Perú. Cabe mencionar que este estudio a nuestro parecer es el primero que usó el análisis de sentimientos para relacionarlo con la movilidad en el Perú.

Un estudio similar realizado en EEUU reveló una conexión entre la opinión expresada en Twitter y los patrones de movilización de las personas (55). Dichos patrones fueron obtenidos a través del índice de distanciamiento social de la Plataforma de Análisis de Impacto COVID-19 publicado por el Instituto de Transporte de Maryland (61). A diferencia de nuestro estudio, el desarrollo de su modelo para la determinación de la opinión en los tweets fue realizado con datos que ellos prepararon y validaron. Además, cabe resaltar que en dicho estudio el índice de movilidad fue obtenido a partir de datos de GPS de teléfonos, vehículos, buses, sistemas de aerolíneas, entre otros, reflejando mejor la realidad que a diferencia de nuestro estudio que utilizo como movilidad los reportes de movilidad de Google.

Nuestro estudio evidencio una percepción negativa en la opinión pública, dado que el 84% de los tweets fueron categorizados como negativo. Otro estudio similar realizado en Colombia busco determinar los sentimientos en Twitter relacionados al aislamiento social encontrando que los sentimientos negativos están presentes en un 51% de los tweets (56). Resultados obtenidos en Perú y Colombia pueden ser

debidos a las implicancias que conlleva la implementación de cuarentenas en países emergentes económicamente ya que según las evidencias la cuarenta trae consigo desempleo, temor, depresión y ansiedad sobre la población (8,60,62–65). Así mismo dos investigaciones realizadas en Ecuador, encontraron diferentes resultados en relación a la percepción del COVID-19, una de las cuales se realizó en una en la población en general (57) y otra en la población de estudiantes universitarios (66). En la primera se encontró un 54.37% de tweets positivos mientras que un 35.18% fueron negativos. A diferencia del estudio realizado en universitarios obtuvo 64% de tweets negativos y 25% positivos evidenciando un comportamiento similar a lo obtenido en el presente estudio y Colombia.

El análisis de movilidad poblacional es de fundamental importancia para la evaluación de la movilización de agentes infecciosos en especial los virus (67). El presente estudio consideró de muy relevante la evaluación de la misma, debido al alto grado de relación ente la movilidad y el aumento de casos de COVID-19. Obteniendo una reducción de la movilidad al inicio de la cuarentena de un 80% y al final de un 20%. De manera similar, un estudio realizado en India, detectó una disminución en la movilidad poblacional del 77% a inicios de la cuarentena, siendo muy similar a nuestro estudio (68).

Una de las limitaciones del estudio fue la falta de poder computacional, ya que esto impedía el uso de técnicas más sofisticadas en el área del Análisis de Sentimientos tales como BERT. Sin embargo, se hicieron uso de 12 algoritmos de aprendizaje supervisado que actualmente se utilizan en un gran número de estudios en el campo

del análisis de sentimientos debido a su alto poder de segmentación de los sentimientos que pueden ser expresados en un texto.

Otra limitación que tuvimos en el estudio, fue la carencia de librerías especializadas para el procesamiento de textos en español, ya que la mayoría de los recursos se encontraban en el idioma inglés. Esto se solucionó mediante el uso recursos disponibles de estudios similares que abordaron la misma problemática.

Una fortaleza de este estudio es que al utilizar el Machine Learning, permitió cerrar brechas presentadas para el desarrollo de investigaciones en épocas de pandemia. Evitando el riesgo de contagio en los pacientes e investigadores. Además, del uso de datos de acceso libre (redes sociales) y a nivel nacional permitieron analizar grandes volúmenes de datos. Por todo lo mencionado, el uso de redes sociales como fuentes de información en conjunto de técnicas para su computacionales para su análisis permiten grandes avances en materia de salud pública.

Así mismo, cabe mencionar que la movilidad de la población durante la cuarentena no solo se encuentra asociada a un ámbito de salud mental sino también se ve afectada por aspecto económicos y culturales los cuales no se ha incluido en el presente estudio siendo también una limitación dentro de la misma (62).

IX. CONCLUSIONES

1. No se encontró asociación entre los sentimientos percibidos en Twitter y la movilidad poblacional durante el periodo de cuarentena en el Perú.

2. El análisis de sentimientos revelo que existió una percepción negativa a temas relacionados a la cuarentena, dado que el 84.55% de los tweets fueron categorizados como negativos y 14.09% como positivos en la etapa del análisis de sentimientos.
3. La movilidad poblacional en todos los departamentos del Perú presenta patrones semejantes. Además, departamentos como Madre de Dios, Puno, Ucayali y Lambayeque muestran un retorno a valores normales de movilidad mucho más rápido que el resto. Sin embargo, departamentos como Amazonas, Arequipa y Tacna muestran los valores más bajos de movilidad haciendo atribución al cumplimiento de las restricciones de la cuarentena por parte de estos departamentos.

X. RECOMENDACIONES

1. El presente estudio plantea el uso de tecnologías novedosas como el Machine Learning en redes sociales para futuras investigaciones que involucren análisis de opinión, sobre todo el Análisis de Sentimientos. Debido a la gran importancia que tienen las opiniones reflejando el comportamiento sociocultural de las personas.
2. El uso de métodos más actuales como BERT para la determinación de la opinión pública podría reflejar con mayor precisión la opinión obtenida a partir de los textos.
3. La salud pública requiere de la multidisciplinariedad y el involucramiento de diversas profesiones asistidas por métodos de aprendizaje automático, para un mejor entendimiento de los posibles factores del quebrantamiento de una medida que contribuye con el bienestar poblacional.

4. Para un análisis más eficaz entre la opinión pública y la movilidad se recomienda la inclusión de factores demográficos, económicos y sanitarios.

XI. DECLARACIÓN DE CONFLICTOS DE INTERÉS

El investigador declara no tener conflictos de interés.

XII. REFERENCIAS BIBLIOGRÁFICAS

1. Coronavirus Disease (COVID-19) - events as they happen [Internet]. [cited 2021 Dec 6]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

2. Kumar A, Singh R, Kaur J, Pandey S, Sharma V, Thakur L, et al. Wuhan to World: The COVID-19 Pandemic. Front Cell Infect Microbiol. 2021;11:242.

3. Mahalmani VM, Mahendru D, Semwal A, Kaur S, Kaur H, Sarma P, et al. COVID-19 pandemic: A review based on current evidence. Indian J Pharmacol. 2020;52(2):117–29.

4. Lotfi M, Hamblin MR, Rezaei N. COVID-19: Transmission, prevention, and potential therapeutic opportunities. Clin Chim Acta. 2020 Sep 1;508:254–66.

5. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med. 2020 May 18;27(3):taaa037.

6. Oliver N, Letouzé E, Sterly H, Delataille S, De Nadai M, Lepri B, et al. Mobile phone data and COVID-19: Missing an opportunity? 2020 Mar 27 [cited 2021 Aug 8]; Available from: https://arxiv.org/abs/2003.12347v1

7. Sanabria-Mazo JP, Useche-Aldana B, Ochoa PP, Rojas-Gualdrón DF, Mateo-Canedo C, Carmona-Cervelló M, et al. Social Inequities in the Impact of COVID-19 Lockdown Measures on the Mental Health of a Large Sample of the Colombian Population (PSY-COVID Study). J Clin Med. 2021 Jan;10(22):5297.

8. Effects on Mental Health After the COVID-19 Lockdown Period: Results From a Population Survey Study in Lima, Peru - Hever Krüger-Malpartida, Bruno Pedraz-Petrozzi, Martin Arevalo-Flores, Frine Samalvides-Cuba, Victor Anculle-Arauco, Mauricio Dancuart-Mendoza, 2020 [Internet]. [cited 2021 Dec 11]. Available from: https://journals.sagepub.com/doi/full/10.1177/1179557320980423

9. Cueva MAL, Cortez ADC. Repercusión del aislamiento social por COVID-19 en la salud mental en la población de Perú: Síntomas en el discurso del ciberespacio. Discurso Soc. 2021;15(1):215–43.

10. Zhang S, Liu M, Li Y, Chung JE. Teens' Social Media Engagement during the COVID-19 Pandemic: A Time Series Examination of Posting and Emotion on Reddit. Int J Environ Res Public Health. 2021 Jan;18(19):10079.

11. Yang Y, Liu K, Li S, Shu M. Social Media Activities, Emotion Regulation Strategies, and Their Interactions on People's Mental Health in COVID-19 Pandemic. Int J Environ Res Public Health. 2020 Jan;17(23):8931.

12. Kaim A, Siman-Tov M, Jaffe E, Adini B. Factors that enhance or impede compliance of the public with governmental regulation of lockdown during COVID-19 in Israel. Int J Disaster Risk Reduct. 2021 Dec 1;66:102596.

13. Pancani L, Marinucci M, Aureli N, Riva P. Forced Social Isolation and Mental Health: A Study on 1,006 Italians Under COVID-19 Lockdown. Front Psychol. 2021;12:1540.

14. Al-Dwaikat TN, Aldalaykeh M, Ta'an W, Rababa M. The relationship between social networking sites usage and psychological distress among undergraduate students during COVID-19 lockdown. Heliyon. 2020 Dec 1;6(12):e05695.

15. Elmer T, Mepham K, Stadtfeld C. Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. PLOS ONE. 2020 Jul 23;15(7):e0236337.

16. Bellato A. Psychological factors underlying adherence to COVID-19 regulations: A commentary on how to promote compliance through mass media and limit the risk of a second wave. Soc Sci Humanit Open. 2020 Jan 1;2(1):100062.

17. Smith LE, Amlôt R, Lambert H, Oliver I, Robin C, Yardley L, et al. Factors associated with adherence to self-isolation and lockdown measures in the UK: a cross-sectional survey. Public Health. 2020 Oct 1;187:41–52.

18. Gilan D, Müssig M, Hahad O, Kunzler AM, Samstag S, Röthke N, et al. Protective and Risk Factors for Mental Distress and Its Impact on Health-Protective Behaviors during the SARS-CoV-2 Pandemic between March 2020 and March 2021 in Germany. Int J Environ Res Public Health. 2021 Jan;18(17):9167.

19. Hills S, Eraso Y. Factors associated with non-adherence to social distancing rules during the COVID-19 pandemic: a logistic regression analysis. BMC Public Health. 2021 Feb 13;21(1):352.

20. Decreto Supremo que declara Estado de Emergencia Nacional por las graves circunstancias que afectan la vida de la Nación a consecuencia del brote del COVID-19-DECRETO SUPREMO-N° 044-2020-PCM [Internet]. [cited 2020 Aug 21]. Available from: http://busquedas.elperuano.pe/normaslegales/decreto-supremo-que-

declara-estado-de-emergencia-nacional-po-decreto-supremo-n-044-2020-pcm-1864948-2/?fbclid=IwAR3F6AYj9jITWsthivOyq-qTUF4v_-wwplzqzUcWHlS6Puagk1g8weyvNnE

21. Bagheri H, Islam MJ. Sentiment analysis of twitter data. ArXiv171110377 Cs [Internet]. 2017 Dec 15 [cited 2020 Aug 21]; Available from: http://arxiv.org/abs/1711.10377

22. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health. 2016 Nov 17;107(1):e1–8.

23. Contini C, Nuzzo MD, Barp N, Bonazza A, Giorgio RD, Tognon M, et al. The novel zoonotic COVID-19 pandemic: An expected global health concern. J Infect Dev Ctries. 2020 Mar 31;14(03):254–64.

24. Liu Y-C, Kuo R-L, Shih S-R. COVID-19: The first documented coronavirus pandemic in history. Biomed J [Internet]. 2020 May 5 [cited 2020 Aug 21]; Available from: http://www.sciencedirect.com/science/article/pii/S2319417020300445

25. Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, Farahmandian N, Miresmaeili S-M, Bahreini E. A comprehensive review of COVID-19 characteristics. Biol Proced Online. 2020 Aug 4;22(1):19.

26. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020 Apr;5(4):536–44.

27. Muralidar S, Ambi SV, Sekaran S, Krishnan UM. The emergence of COVID-19 as a global pandemic: Understanding the epidemiology, immune response and potential therapeutic targets of SARS-CoV-2. Biochimie. 2020 Dec 1;179:85–100.

28. Lewnard JA, Lo NC. Scientific and ethical basis for social-distancing interventions against COVID-19. Lancet Infect Dis. 2020 Jun 1;20(6):631–3.

29. Siedner MJ, Harling G, Reynolds Z, Gilbert RF, Haneuse S, Venkataramani AS, et al. Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest–posttest comparison group study. PLOS Med. 2020 ago;17(8):e1003244.

30. De Rosis S, Lopreite M, Puliga M, Vainieri M. The early weeks of the Italian Covid-19 outbreak: sentiment insights from a Twitter analysis. Health Policy. 2021 Aug 1;125(8):987–94.

31. Dai Z, Locasale JW. Cooperative virus propagation in COVID-19 transmission. medRxiv. 2020 Sep 18;2020.05.05.20092361.

32. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis. 2004 Jul;10(7):1206–12.

33. Jeong H, Yim HW, Song Y-J, Ki M, Min J-A, Cho J, et al. Mental health status of people isolated due to Middle East Respiratory Syndrome. Epidemiol Health. 2016;38:e2016048.

34. Luo Y, Chua CR, Xiong Z, Ho RC, Ho CSH. A Systematic Review of the Impact of Viral Respiratory Epidemics on Mental Health: An Implication on the Coronavirus Disease 2019 Pandemic. Front Psychiatry. 2020;11:1247.

35. Venkatesh A, Edirappuli S. Social distancing in covid-19: what are the mental health implications? BMJ. 2020 Apr 6;369:m1379.

36. Hossain MM, Sultana A, Purohit N. Mental health outcomes of quarantine and isolation for infection prevention: a systematic umbrella review of the global evidence. Epidemiol Health. 2020 Jun 2;42:e2020038.

37. The potential impact of COVID-19 on mental health outcomes and the implications for service solutions [Internet]. ARC West. [cited 2021 Dec 20]. Available from: https://arc-w.nihr.ac.uk/covid-response/rapid-reports/potential-impact-of-covid-19-on-mental-health-outcomes-and-the-implications-for-service-solutions/

38. Pérez-Escoda A, Jiménez-Narros C, Perlado-Lamo-de-Espinosa M, Pedrero-Esteban LM. Social Networks' Engagement During the COVID-19 Pandemic in Spain: Health Media vs. Healthcare Professionals. Int J Environ Res Public Health. 2020 Jan;17(14):5261.

39. Choudhury MD, Counts S, Gamon M. Not All Moods Are Created Equal! Exploring Human Emotional States in Social Media. Proc Int AAAI Conf Web Soc Media. 2012;6(1):66–73.

40. Saha K, Chan L, De Barbaro K, Abowd GD, De Choudhury M. Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2017 Sep 11;1(3):95:1-95:27.

41. Twitter Sentiment Analysis | Implement Twitter Sentiment Analysis Model [Internet]. Analytics Vidhya. 2021 [cited 2021 Dec 9]. Available from: https://www.analyticsvidhya.com/blog/2021/06/twitter-sentiment-analysis-a-nlp-use-case-for-beginners/

42. Nandwani P, Verma R. A review on sentiment analysis and emotion detection from text. Soc Netw Anal Min. 2021;11(1):81.

43. Gohil S, Vuik S, Darzi A. Sentiment Analysis of Health Care Tweets: Review of the Methods Used. JMIR Public Health Surveill. 2018 Apr 23;4(2):e5789.

44. Ji X, Chun S, Geller J. Monitoring Public Health Concerns Using Twitter Sentiment Classifications. Proceedings - 2013 IEEE International Conference on Healthcare Informatics, ICHI 2013. 2013. 335 p.

45. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011 Sep 1;18(5):544–51.

46. Khachidze M, Tsintsadze M, Archuadze M. Natural Language Processing Based Instrument for Classification of Free Text Medical Records [Internet]. BioMed Research International. 2016 [cited 2020 Aug 21]. Available from: https://www.hindawi.com/journals/bmri/2016/8313454/

47. Chen L, Liu Y, Chang Y, Wang X, Luo X. Public opinion analysis of novel coronavirus from online data. J Saf Sci Resil. 2020 Dec 1;1(2):120–7.

48. Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, et al. Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. J Med Internet Res. 2021 Sep 10;23(9):e30854.

49. Paul M, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. Proc Int AAAI Conf Web Soc Media. 2011;5(1):265–72.

50. Bornmann L, Haunschild R, Patel VM. Are papers addressing certain diseases perceived where these diseases are prevalent? The proposal to use Twitter data as social-spatial sensors. PLoS ONE. 2020 Nov 20;15(11):e0242550.

51. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. J Med Internet Res. 2021 Apr 5;23(4):e26627.

52. Jose L. Segovia Juarez. Estado de la epidemia causada por el novel coronavirus SARS-CoV-2 y sus posibles implicancias en el Perú. Sugerencias de medidas urgentes [Internet]. 2020 [cited 2020 Aug 21]. Available from: http://167.249.11.60/anc_j28.1/index.php?option=com_content&view=article&id=461:estado-de-la-epidemia-causada-por-el-novel-coronavirus-sars-cov-2-y-sus-posibles-implicancias-en-el-peru-sugerencias-de-medidas-urgentes&catid=61&Itemid=28

53. Coronavirus: Recomendaciones para el uso de mascarillas [Internet]. [cited 2021 Dec 7]. Available from: https://www.gob.pe/8804

54. Pellecchia U, Crestani R, Decroo T, Bergh RV den, Al-Kourdi Y. Social Consequences of Ebola Containment Measures in Liberia. PLOS ONE. 2015 dic;10(12):e0143036.

55. Li L, Ma Z, Lee H, Lee S. Can social media data be used to evaluate the risk of human interactions during the COVID-19 pandemic? Int J Disaster Risk Reduct. 2021 Apr 1;56:102142.

56. Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning [Internet]. [cited 2021 Dec 11]. Available from: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2215-910X2021000100001

57. A JAT. Análisis de opinión sobre tuits del COVID-19 generados por usuarios ecuatorianos. CEDAMAZ. 2021 Jul 15;11(1):70–7.

58. Vázquez-Rowe I, Gandolfi A. Peruvian efforts to contain COVID-19 fail to protect vulnerable population groups. Public Health Pract. 2020 Nov 1;1:100020.

59. Padidar S, Liao S, Magagula S, Mahlaba TAM, Nhlabatsi NM, Lukas S. Assessment of early COVID-19 compliance to and challenges with public health and social prevention measures in the Kingdom of Eswatini, using an online survey. PLOS ONE. 2021 Jun 29;16(6):e0253954.

60. Cabezas J, Moctezuma D, Fernández-Isabel A, Martín de Diego I. Detecting Emotional Evolution on Twitter during the COVID-19 Pandemic Using Text Analysis. Int J Environ Res Public Health. 2021 Jan;18(13):6981.

61. University of Maryland COVID-19 Impact Analysis Platform [Internet]. [cited 2021 Dec 22]. Available from: https://data.covid.umd.edu/

62. Diseases TLI. The intersection of COVID-19 and mental health. Lancet Infect Dis. 2020 Nov 1;20(11):1217.

63. Lust J. A Class Analysis of the Expansion of COVID-19 in Peru: The Case of Metropolitan Lima. Crit Sociol. 2021 Jul 1;47(4–5):657–70.

64. Johnson MC, Saletti-Cuesta L, Tumas N. Emociones, preocupaciones y reflexiones frente a la pandemia del COVID-19 en Argentina. Ciênc Saúde Coletiva. 2020 Jun 5;25:2447–56.

65. Ramírez ML, Martinez SM, Bessone C del V, Allemandi DA, Quinteros DA. COVID-19: Epidemiological Situation of Argentina and its Neighbor Countries after Three Months of Pandemic. Disaster Med Public Health Prep. 2021 Mar 25;1–7.

66. Pazmiño R, Badillo F, González MC, García-Peñalvo FJ. Ecuadorian Higher Education in COVID-19: A Sentiment Analysis. In: Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality [Internet]. New York, NY, USA: Association for Computing Machinery; 2020 [cited 2021 Dec 22]. p. 758–64. (TEEM'20). Available from: https://doi.org/10.1145/3434780.3436679

67. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci. 2009 Dec 22;106(51):21484–9.

68. Kishore K, Jaswal V, Verma M, Koushal V. Exploring the Utility of Google Mobility Data During the COVID-19 Pandemic in India: Digital Epidemiological Analysis. JMIR Public Health Surveill. 2021 Aug 30;7(8):e29957.

69. GESTIÓN N. Estas son las redes sociales en las que más interactúan los peruanos | TENDENCIAS [Internet]. Gestión. NOTICIAS GESTIÓN; 2021 [cited 2021 Dec 11]. Available from: https://gestion.pe/tendencias/estas-son-las-redes-sociales-en-las-que-mas-interactuan-los-peruanos-noticia/

XIII. ANEXOS

Anexo 1. Código para transformar la base de datos del TASS a formato tabla

```
# # CONVIRTIENDO LA BASE DE DATOS DE LA TASS A FORMATO TABLA

#

# Esta sección contiene la serie de pasos para la
transformación de la base de datos de tweets de la TASS a un
formato tabla para su posterior uso

from lxml import etree

import pandas as pd

import numpy  as np



get_ipython().system('DIR TASS*')



xmlFileTrain1 = "TASS2019_country_CR_train.xml"

xmlFileTrain2 = "TASS2019_country_ES_train.xml"

xmlFileTrain3 = "TASS2019_country_MX_train.xml"

xmlFileTrain4 = "TASS2019_country_PE_train.xml"

xmlFileTrain5 = "TASS2019_country_UY_train.xml"



xmlFileTest1 = "TASS2019_country_CR_dev.xml"

xmlFileTest2 = "TASS2019_country_ES_dev.xml"

xmlFileTest3 = "TASS2019_country_MX_dev.xml"

xmlFileTest4 = "TASS2019_country_PE_dev.xml"

xmlFileTest5 = "TASS2019_country_UY_dev.xml"



def parseXML(xmlFile):
```

```
    with open(xmlFile, 'rb') as fobj:
        xml = fobj.read()
    root = etree.fromstring(xml)

    num   = 1
    data = dict()

    for appt in root.getchildren():
        rows_val = []
        tweet    = 'tweet_' + str(num)
        for elem in appt.getchildren():
            if not elem.text:
                text = "None"
            elif elem.tag == 'sentiment':
                polar= elem.getchildren()[0].getchildren()[0]
                text = polar.text
            else:
                text = elem.text

            rows_val.append(text)
        data[tweet] = rows_val
        num = num + 1

    return data

def dict_df(dictionary):
```

```
    columns = ['tweetid', 'user', 'content', 'date', 'lang',
'sentiment']



    return pd.DataFrame.from_dict(dictionary, orient='index',
columns= columns)



dbtrain1 = dict_df(parseXML(xmlFileTrain1))

dbtrain2 = dict_df(parseXML(xmlFileTrain2))

dbtrain3 = dict_df(parseXML(xmlFileTrain3))

dbtrain4 = dict_df(parseXML(xmlFileTrain4))

dbtrain5 = dict_df(parseXML(xmlFileTrain5))



dbtest1 = dict_df(parseXML(xmlFileTest1))

dbtest2 = dict_df(parseXML(xmlFileTest2))

dbtest3 = dict_df(parseXML(xmlFileTest3))

dbtest4 = dict_df(parseXML(xmlFileTest4))

dbtest5 = dict_df(parseXML(xmlFileTest5))




dbtrain1['country'] = 'CR'

dbtrain2['country'] = 'ES'

dbtrain3['country'] = 'MX'

dbtrain4['country'] = 'PE'

dbtrain5['country'] = 'UY'



dbtest1['country'] = 'CR'

dbtest2['country'] = 'ES'

dbtest3['country'] = 'MX'
```

```
dbtest4['country'] = 'PE'

dbtest5['country'] = 'UY'



db_train = pd.concat([dbtrain1, dbtrain2, dbtrain3, dbtrain4,
dbtrain5,

                      dbtest1, dbtest2, dbtest3, dbtest4,
dbtest5])



db_train.drop_duplicates(inplace=True)

db_train = db_train.reset_index()



db_train.drop('index', axis = 1, inplace=True)

db_train.to_csv('training_set.csv')
```

Anexo 2. Código para el procesamiento de textos

```
# # PROCESAMIENTO DE TEXTOS

#

# El siguiente documento contiene los pasos a seguir para
filtrar la información irrelevante que contienen los tweets



import pandas as pd

import numpy as np

import re

import string

import stanza



## method and stopwords text processing

from nltk.corpus   import stopwords
```

```
from nltk.tokenize import word_tokenize

from nltk.stem     import SnowballStemmer

from nltk.stem     import WordNetLemmatizer

from sklearn.feature_extraction.text import CountVectorizer

from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.feature_extraction.text import TfidfTransformer

from sklearn.model_selection         import train_test_split



import warnings

warnings.filterwarnings("ignore")

stanza.download('es', package='ancora',
processors='tokenize,mwt,pos,lemma', verbose=True)




# ## Spanish Stopwords

## Createing a stopwords set

import nltk

nltk.download('punkt')

nltk.download('wordnet')

nltk.download('stopwords')

stop_words = set(stopwords.words('spanish'))




# ## Preprocessing the Tweet Text

#  1. Casing

#  2. Noise Removal

#  3. Tokenization
```

```
#  4. Stopword Removal

#  5. Text Normalization(Stemming and Lemmatization)



DIACRITICAL_VOWELS = [('á', 'a'), ('é', 'e'), ('í', 'i'), ('ó',
'o'), ('ú', 'u'), ('ü', 'u')]

SLANG              = [('d', 'de'), ('[qk]', 'que'), ('xo',
'pero'), ('xa', 'para'), ('[xp]q', 'porque'), ('es[qk]', 'es
que'),

                      ('fvr', 'favor'),
('(xfa|xf|pf|plis|pls|pofa)', 'por favor'), ('dnd', 'donde'),
('tb', 'tambien'),

                      ('(tq|tk)', 'te quiero'), ('(tqm|tkm)',
'te quiero mucho'), ('x', 'por'), ('\+', 'mas')]




# ### Lemmatizer



lemmatizer  = stanza.Pipeline(lang       = 'es',

                              processors =
'tokenize,mwt,pos,lemma',

                              use_gpu    = True)



def preprocess_tweet_text(tweet):

    """

    Runs a set of transformational steps to

    preprocess the text of the tweet.

    """



    # remove user @ references and '#' from tweet

    tweet = re.sub(r'\@\w+|\#', "", tweet)
```

```
    # remove numbers from tweet
    tweet = re.sub(r'[0-9]', "", tweet)

    # remove signals from tweet
    #tweet = re.sub(r'\?|\¿', "", tweet)

    # remove jaja expressions from tweet
    tweet = re.sub(r'ja\w+|ha\w+|he\w+|ji\w+|je\w+', "", tweet)

    # remove any urls
    tweet = re.sub(r"http\S+|www\S+|https\S+", "", tweet, flags
= re.MULTILINE)

    # remove vowels with diacritical marks
    for s, t in DIACRITICAL_VOWELS:
        tweet = re.sub(r'{0}'.format(s), t, tweet)

    # translate slang
    for s, t in SLANG:
        tweet = re.sub(r'\b{0}\b'.format(s), t, tweet)

    # Substituting multiple spaces with single space
    tweet = re.sub(r'\s+', ' ', tweet, flags=re.I)

    # remove retweet
    tweet  = re.sub(r'RT[\s]+', '', tweet)
```

```
    # remove repeated characters

    tweet  = re.sub(r'(.)\1{2,}', r'\1', tweet)



    # convert all text lowercase

    tweet = tweet.lower()



    # remove punctuations

    tweet = tweet.translate(str.maketrans("", "",
string.punctuation))



    # remove stopwords

    tweet_tokens   = word_tokenize(tweet)

    filtered_words = [word for word in tweet_tokens if word not in
stop_words]



    # stemming

    #ps             = SnowballStemmer('spanish')

    #stemmed_words = [ps.stem(w) for w in filtered_words]



    # lemmatizing

    lemma_words = [lemmatizer(w).get("lemma")[0] for w in
filtered_words]



    return " ".join(lemma_words)



#dataset = pd.read_csv("training_set.csv")

#columns = ['tweetid', 'user', 'content', 'date', 'lang',
'sentiment', 'country']

#dataset = dataset[columns]
```

```
dataset =pd.read_csv('BaseDeDatosPeru.csv')

dataset.drop('Unnamed: 0', inplace =True, axis=1)



dataset2         =dataset.copy()

dataset2.text    =
dataset2['text'].apply(preprocess_tweet_text)

dataset2.to_csv('BaseDeDatosPeru_preproc_Lemma.csv')
```

Anexo 3. Evaluación de modelos de clasificación

```
# # EVALUACIÓN DE LOS MODELOS DE CLASIFICACIÓN

#

# Para la evaluación del rendimiento en la clasificacion de
modelos las cuales hacen uso de tecnicas de machine learning me
guie de la siguiente pagina: <a
href="https://www.learndatasci.com/tutorials/predicting-reddit-
news-sentiment-naive-bayes-text-classifiers/"
target="_top">Reference Page</a>



## importing libraries

## data manipulation

import pandas as pd

import numpy as np

import re

import string

import nltk



## method and stopwords text processing

from nltk.corpus   import stopwords

from nltk.tokenize import word_tokenize
```

```
from nltk.stem      import SnowballStemmer

from nltk.stem      import WordNetLemmatizer

from nltk.classify  import SklearnClassifier

from functools      import reduce

from sklearn.feature_extraction.text  import CountVectorizer

from sklearn.feature_extraction.text  import TfidfVectorizer

from sklearn.feature_extraction.text  import TfidfTransformer

from sklearn.model_selection          import train_test_split

from imblearn.over_sampling           import SMOTE



## Machine learning libraries

from sklearn.model_selection        import ShuffleSplit

from sklearn.metrics                import classification_report,
accuracy_score, f1_score, confusion_matrix

from sklearn.naive_bayes            import BernoulliNB,
ComplementNB, MultinomialNB

from sklearn.linear_model           import LogisticRegression,
SGDClassifier

from sklearn.svm                    import SVC, LinearSVC

from sklearn.neighbors              import KNeighborsClassifier

from sklearn.tree                   import DecisionTreeClassifier

from sklearn.ensemble               import
RandomForestClassifier, AdaBoostClassifier

from sklearn.neural_network         import MLPClassifier

from sklearn.discriminant_analysis  import
QuadraticDiscriminantAnalysis




import tensorflow as tf
```

```
from tensorflow.keras.models     import Sequential

from tensorflow.keras.layers     import Dense, Activation,
Dropout

from tensorflow.keras.callbacks import EarlyStopping



## Plotting libraries

import matplotlib.pyplot as plt

import seaborn            as sns



sns.set_style(style='white')

sns.set_context(context='notebook', font_scale=1.3,
rc={'figure.figsize': (16,9)})



import warnings

warnings.filterwarnings("ignore")




# ## Loading Dataset



# Load dataset

dataset  = pd.read_csv("DatasetTASS_preproc_lem.csv")

dataset  = dataset[['content', 'sentiment']]

dataset  = dataset[dataset['sentiment'] != 'NONE']



nan_value= dataset[dataset['content'].isnull()].index.item()

dataset.drop(index=nan_value, inplace = True)

dataset
```

```
dataset.sentiment.value_counts()



#sm = SMOTE()


#X_train, y_train = sm.fit_resample(X_train, y_train)



#unique, counts = np.unique(y_train, return_counts=True)
#print(list(zip(unique, counts)))



from sklearn.linear_model import SGDClassifier
from sklearn.svm           import LinearSVC

X  = dataset.content
y  = dataset.sentiment

cv = ShuffleSplit(n_splits=20, test_size=0.2)

models = [
    MultinomialNB(),
    BernoulliNB(),
    ComplementNB(),
    LogisticRegression(),
    KNeighborsClassifier(),
    DecisionTreeClassifier(),
```

```
    RandomForestClassifier(),
    SGDClassifier(),
    SVC(),
    LinearSVC(),
    MLPClassifier(max_iter=1000),
    AdaBoostClassifier()
]


#sm = SMOTE()


# Init a dictionary for storing results of each run for each
model
results = {
    model.__class__.__name__: {
        'accuracy'     : [],
        'f1_score'     : [],
        'confusion_matrix'  : [],
        'classification_report': [],
        'model'      : []
    } for model in models
}


for train_index, test_index in cv.split(X):
    X_train, X_test  = X.iloc[train_index], X.iloc[test_index]
    y_train, y_test  = y.iloc[train_index], y.iloc[test_index]


    vector       = TfidfVectorizer(sublinear_tf=True)
```

```
    X_train_vect =
vector.fit_transform(np.array(X_train).ravel())

    X_test_vect  = vector.transform(np.array(X_test).ravel())



    #X_train_res, y_train_res = sm.fit_sample(X_train_vect,
y_train)



    for model in models:

        model.fit(X_train_vect, y_train)

        y_pred = model.predict(X_test_vect)



        acc = accuracy_score(y_test, y_pred)

        f1 = f1_score(y_test, y_pred, average='micro')

        cm = confusion_matrix(y_test, y_pred)

        cr = classification_report(y_test,y_pred)



        results[model.__class__.__name__]['accuracy'].append(ac
c)

        results[model.__class__.__name__]['f1_score'].append(f1
)

        results[model.__class__.__name__]['confusion_matrix'].a
ppend(cm)

        results[model.__class__.__name__]['classification_repor
t'].append(cr)

        results[model.__class__.__name__]['model'].append(model
)




for model, d in results.items():

    avg_acc = sum(d['accuracy']) / len(d['accuracy']) * 100
```

```
    avg_f1 = sum(d['f1_score']) / len(d['f1_score']) * 100

    avg_cm = sum(d['confusion_matrix']) /
len(d['confusion_matrix'])



    slashes = '-' * 30



    s = f"""{model}\n{slashes}

        Avg. Accuracy: {avg_acc:.2f}%

        Avg. F1 Score: {avg_f1:.2f}

        Avg. Confusion Matrix:

        \n{avg_cm}

        """

    print(s)




'''fig, (ax1, ax2) = plt.subplots(2, 1, sharex=True,
figsize=(16,9))



acc_scores = [round(a * 100, 1) for a in accs]

f1_scores  = [round(f * 100, 2) for f in f1s]



x1 = np.arange(len(acc_scores))

x2 = np.arange(len(f1_scores))



ax1.bar(x1, acc_scores)

ax2.bar(x2, f1_scores, color='#559ebf')



# Place values on top of bars
```

```
for i, v in enumerate(list(zip(acc_scores, f1_scores))):

    ax1.text(i - 0.25, v[0] + 2, str(v[0]) + '%')

    ax2.text(i - 0.25, v[1] + 2, str(v[1]))



ax1.set_ylabel('Accuracy (%)')

ax1.set_title('Naive Bayes')

ax1.set_ylim([0, 100])



ax2.set_ylabel('F1 Score')

ax2.set_xlabel('Runs')

ax2.set_ylim([0, 100])



sns.despine(bottom=True, left=True)  # Remove the ticks on axes
for cleaner presentation

plt.show()'''
```

Anexo 4. Código Python para la recolección de tweets

```
# # CREACIÓN DE BASE DE DATOS

#

# Este código fue usado para la recolección de tweets a traves de la
API de Twitter

from sqlalchemy import create_engine, Column, Integer, String, DateTime

from sqlalchemy.orm import sessionmaker

from sqlalchemy.ext.declarative import declarative_base



engine =
create_engine('postgresql+psycopg2://postgres:velocidad@localhost:5432
/twitterdb', echo= False)
```

```
session = sessionmaker(bind=engine)

session = session()

Base = declarative_base()



class TweetsTab(Base):

    __tablename__ = 'TweetsTab'



    id     = Column(String, primary_key=True)

    id_user      = Column(String)

    text         = Column(String)

    created_at   = Column(DateTime)

    Location_user= Column(String)

    Place_tweet  = Column(String)



#Base.metadata.create_all(engine)

# dictionary = {date: [query, from_date, to_date]}

Marzo = {'Marzo_16' : ['((Día 1) OR EstadosDeEmergencia OR covidenperú
OR perujuntoscontraelcovid_19 OR AislamientoSocial OR CuarentenaPeru
OR QuedateEnTuCasaCarajo OR CuarentenaCoronavirus OR
QuedateEnLaCasa)', '202003162359','202003160000'],

        'Marzo_17' : ['((Día 2) OR EstadoDeEmergencia OR
yomequedoencasa OR QuedateEnTuCasaCarajo OR presidente OR covidenperú
OR PerúEstáEnNuestrasManos OR CuarentenaNacional OR FFAA OR
cuidemonosentretodos)', '202003172359','202003170000'],

        'Marzo_18' : ['((Día 3) OR yomequedoencasa OR QuedateEnLaCasa
OR QuedateEnTuCasaCarajo OR CuarentenaCoronavirus OR coronavirusenperu
OR PerúEstaEnNuestrasManos OR LoBuenoDeLaCuarentena OR AFPs OR FFAA)',
'202003182359','202003180000'],

        'Marzo_19' : ['((Día 4) OR coronavirusenperu OR
yomequedoencasa OR (Toque de Queda) OR ContigoPerú OR COVID2019 OR
cuarentenatotal OR AFPs OR toquedequedaperu OR
SueltaMiPlataMalditaAFP)', '202003192359','202003190000'],

        'Marzo_20' : ['((Día 5) OR cuarentenatotal OR covid19peru OR
coronavirusenperu OR yomequedoencasa OR toquedequedaperu OR
CuarentenaNacional OR QuedateEnCasa OR COVID2019ec OR AFPs OR
PerúEstáEnNuestrasManos)', '202003202359','202003200000'],
```

```
        'Marzo_21' : ['((Día 6) OR covid19peru OR cuarentenatotal OR
Essalud OR Minsa OR Dia5deCuarentena OR coronacrisis OR
coronavirusenperu OR AislamientiObligatorio OR YoApoyoAVizcarraGratis
OR CuarentenaPorLaVida)', '202003212359','202003210000'],

        'Marzo_22' : ['(Dia6 OR (Ministro de Salud) OR (Víctor Zamora)
OR CuarentenaPorLaVida OR QuedateEnCasaCTM OR COVID_19 OR (Día 7) OR
InfelizDiaIncapaz OR Abuso OR (Sr. Presidente) OR
CuarentenaPorLaVida)', '202003222359','202003220000'],

        'Marzo_23' : ['((Día 8) OR Dia8 OR (Sr. Presidente) OR
QuedateEnCasaSubnormal OR CuarentenaEnFamilia OR
YoApoyoAVizcarraGratis OR QuedateEnCasaCTM OR vizcarralover OR
Dia7decuarentena OR YoMeSumo OR Coronavid19)',
'202003232359','202003230000'],

        'Marzo_24' : ['((Día 9) OR Dia8 OR YoMeSumo OR Policía OR
ApoyoALasFuerzasArmadas OR Coronavid19 OR cuarentenaobligatoriaya OR
Dia9 OR 24Mar OR FFAA OR Lloronavirus)',
'202003242359','202003240000'],

        'Marzo_25' : ['((Día 10) OR Día 9 OR Coronavid19 OR
cuarentenaobligatoriaya OR YoMeSumo OR Lloronavirus OR Dia10 OR 25MAr
OR COVID2019)', '202003252359','202003250000'],

        'Marzo_26' : ['((Día 11) OR quedateEnTuCasa OR CuarentenaPeru
OR SaliendoDeCuarentena OR (ministerio público) OR municípios OR
Municipalidades OR AFPs OR Día 10 OR AFPLiberacionDeMiPlataYa OR
LiberenFondosAFP)', '202003262359','202003260000'],

        'Marzo_27' : ['((Día 12) OR Dia12 OR SaliendoDeCuarentena OR
cuarentenatotal OR UnidosEnCasa OR (UCI y 19) OR 27Mar OR
13DíasMásSin)', '202003272359','202003270000'],

        'Marzo_28' : ['((Día 13) OR COVID2019 OR 27marzo OR (Día
12))', '202003282359','202003280000'],

        'Marzo_29' : ['(SiVizcarraMeDice OR CODVID19 OR
CuandoAcabeLaCuarentena OR tepongoencuarentenasi OR UnidosVenceremos
OR (Día 14) OR (25% de la AFP) OR 29Mar OR (Ley de Protección
Policial) OR Loreto)', '202003292359','202003290000'],

        'Marzo_30' : ['((Las AFP) OR Encuarentena OR CODVID19 OR
QuédateEnCasaYaOR AFPs OR SueltaMiPlataMalditaAFP OR
EnEstaCuarentenaSupeQue OR AFPDevuelvemeMiDinero OR
ConMiPlataNoTeMetas)', '202003302359','202003300000'],

        'Marzo_31' : ['((Día 16) OR AFPs OR SueltaMiPlataMalditaAFP OR
(Plaza de Acho) OR (Miró Quesada) OR AFPDevuelvemeMiDinero OR
QuédateEnCasaYa OR CODVID19 OR Encuarentena OR (Piura y Loreto) OR
JuntosNosCuidamos)', '202003312359','202003310000']}



Abril ={'Abril_1' : ['((Día 17) OR AFPs OR (Plaza de Acho) OR
QuédateEnCasaYa OR AFPDevuelvemeMiDinero OR JuntosNosCuidamos OR
PlazaDeAcho OR SueltaMiPlataMalditaAFP OR PlazaDeAcho OR
DevuelveMiPlataMalditaAfp)', '202004012359','202004010000'],
```

```
        'Abril_2' : ['((Día 18) OR AFPs OR YaPasaron17DíasY OR (Plaza
de Acho) OR QuédateEnCasaYa OR AFPDevuelvemeMiDinero OR
DevuelvemeMiPlataMalditaAfp OR PlazaDeAcho OR quedateencasa OR
CuarentenaOblgatoria)', '202004022359','202004020000'],

        'Abril_3' : ['(quedateencasa OR Dia18 OR (Día 18) OR Vizcarra
OR SoloSalgoSi OR AFPs OR (NADIE SALE) OR parasobrevivirlacuarentenayo
OR dia19 OR (Reactiva Perú) OR (Día 19) OR QuedOTEnCasa3A OR
2DeAbril)', '202004032359','202004030000'],

        'Abril_4' : ['((Reactiva Perú) OR (Perú En Tus Manos) OR (Ciro
Maguiña) OR NOmasAFP OR quedateencasa OR CuandoEstoTermineYo OR
peruentusmanos OR CuarentenaSinTransfobia OR (hospital de la policia)
OR AFPs)', '202004042359','202004040000'],

        'Abril_5' : ['((Día 21) OR Dia20 OR (Día 19) OR (Perú En Tus
Manos) OR CuarentenaSinTransfobia OR dia19 OR COVID2019 OR 5DeAbril OR
Dia21 OR (HOY 5) OR (el 5) OR cuarentenatotal)',
'202004052359','202004050000'],

        'Abril_6' : ['((Día 21) OR Dia21 OR cuarentenatotal OR
(Reactiva Perú) OR 6Abr OR 5Abril)', '202004062359','202004060000'],

        'Abril_7' : ['((Día 22) OR (Reactiva Perú) OR
TeCuidoComoVizcarra OR COVID_19 OR coronavirusenperu OR Dia23 OR 7Abr
OR Día22)', '202004072359','202004070000'],

        'Abril_8' : ['(Dia23 OR (Día 23) OR (AFP Prima) OR
JuroQueCuandoEstoTermine OR (Colegio Médico del Perú) OR 8Abr OR
CuarentenaSinViolencia OR (Prima AFP) OR BonoIndependiente OR
CuarentenaNacional)', '202004082359','202004080000'],

        'Abril_9' : ['(LaCuarentenaTerminaraCuando OR Dia24 OR (Dia
24) OR (Prima AFP) OR CuarentenaNacional OR CuarentenaSinViolencia OR
(Dia 25) OR 9Abr OR BonoIndependiente)',
'202004092359','202004090000'],

        'Abril_10' : ['(Día25 OR (Dia 25) OR (Pilar Mazzetti) OR
(Comando COVID-19) OR (Ministra de Trabajo) OR ValoroCuando OR
CuandoPorFinPodamosSalir OR LaCuarentenaTerminaraCuando OR Día26 OR
SOSAmazonia OR 10Abr)', '202004102359','202004100000'],

        'Abril_11' : ['(Día26 OR (Dia 26) OR (Pilar Mazzetti) OR Dia27
OR (día 27) OR Mazzetti OR mazzetti OR CuarentenaExtendida OR 11Abr OR
Día26)', '202004112359','202004110000'],

        'Abril_12' : ['(Dia27 OR (día 27) OR UnidosDesdeCasa OR
CuarentenaExtendida OR EstoTerminaraY OR COVID_19 OR Dia28 OR (Día 28)
OR EstoTerminaraY OR Dia27)', '202004122359','202004120000'],

        'Abril_13' : ['((Día 28) OR COVID_19 OR Dia28 OR
CuarentenaExtendida OR COVID2019 OR (CTS y AFP) OR (Día 29) OR
(Ministra de Trabajo) OR 12Abr)', '202004132359','202004130000'],

        'Abril_14' : ['(COVID2019 OR Dia29 OR (Día 29) OR (Reactiva
Perú) OR MINTRA OR reactivaperu OR cuandoestésaqui OR Día30 OR (Día
30) OR (La Ministra de Trabajo) OR Vizcarrita OR (CTS y AFP) OR mypes
OR COVID2019)', '202004142359','202004140000'],
```

```
        'Abril_15' : ['(Perú OR Vizcarra OR Día30 OR (Día 30) OR
COVID2019  OR Dia31 OR (Dia 31) OR 15Abr OR Vizcarra)',
'202004152359','202004150000'],

        'Abril_16' : ['(Perú OR Vizcarra OR Dia31 OR (Dia 31) OR
BonoUniversal OR Dia32 OR (Martín Vizcarra))',
'202004162359','202004160000'],

        'Abril_17' : ['(Perú OR Vizcarra OR Dia32 OR (Día 32) OR
SeparadosPeroJuntos OR BonoUniversal OR Dia33 OR (Día 33) OR 17Abr OR
Perú OR Lima)', '202004172359','202004170000'],

        'Abril_18' : ['(Perú OR Dia33 OR TogetherAtHome OR Dia34 OR
MartínVizcarra OR (Villa Panamericana) OR Perú)',
'202004182359','202004180000'],

        'Abril_19' : ['(TogetherAtHome OR Perú OR Dia34 OR Vizcarra OR
Dia35 OR (Día 35) OR 19Abr OR Perú OR Lima)',
'202004192359','202004190000'],

        'Abril_20' : ['(TogetherAtHome OR Dia35 OR (Día 35) OR
Vizcarra OR (Ciro Maguiña) OR quedateencasa OR CuarentenaSinAbusos OR
BonoUniversal OR Dia36 OR Perú OR Zamora)',
'202004202359','202004200000'],

        'Abril_21' : ['(Dia36 OR (Día 36) OR Vizcarra OR 20Abr OR
Dia37 OR (Día 37) OR Perú OR Lima OR (Colegio Médico) OR
YoYaNoRecuerdo)', '202004212359','202004210000'],

        'Abril_22' : ['(Dia37 OR (Día 37) OR Perú OR EstoAcabaraCuando
OR 21Abril OR Dia38 OR (Colegio Médico del Perú) OR Vizcarra OR Dia37
OR CuandoPaseLa40TenaYo OR AlBordeDelColapso OR EstoAcabaraCuando)',
'202004222359','202004220000'],

        'Abril_23' : ['((Día 39) OR Dia38 OR Perú OR 22Abril OR
CuarentenaExtendida OR Dia39 OR BonoFamiliarUniversal OR Dia38)',
'202004232359','202004230000'],

        'Abril_24' : ['(LaCuarentenaSeExtiendeHasta OR Dia40 OR
Vizcarra OR (Día 39) OR Perú OR Dia39 OR BonoUniversalparaResistir OR
23Abr)', '202004242359','202004240000'],

        'Abril_25' : ['(Dia40 OR (Día 40) OR SinPlataYo OR Mininter OR
CuarentenaExtendida OR 24Abr OR COVID_19 OR BonoUniversalparaResistir
OR dia41 OR distanciamiento OR Loreto OR 25Abr)',
'202004252359','202004250000'],

        'Abril_26' : ['(dia41 OR (Día 41) OR distanciamiento OR 25Abr
OR dia42 OR (Ministra de Economía) OR Vizcarra OR Loreto OR 26Abr)',
'202004262359','202004260000'],

        'Abril_27' : ['((Con Calma) OR dia42 OR (DÍA 42) OR
Irresponsables OR DomingoDeCuarentena OR Dia43 OR 27Abr OR (Con Calma)
OR (Colegio Médico del Perú) OR Mininter OR (Ministerio del
Interior))', '202004272359','202004270000'],

        'Abril_28' : ['(Dia43 OR Ministra OR (DIA 43) OR (Con Calma)
OR Dia44 OR Ministra OR (Día 44) OR Dia43)',
'202004282359','202004280000'],
```

```
        'Abril_29' : ['((Pilar Mazzetti) OR (Día 44) OR Zamora OR
Ministra OR Dia44 OR COVID_19 OR 28Abr OR (25% de afp) OR
AFPDevuelvemeMiDinero OR (Ministro de Justicia) OR
SueltaMiPlataMalditaAFP OR FueraZamora)',
'202004292359','202004290000'],

        'Abril_30' : ['(Vizcarra OR COVID_19 OR (Pilar Mazzetti) OR
presidente OR Zamora OR (25% de afp) OR SueltaMiPlataMalditaAFP OR
VizcarraNoSeasFresco OR AFPs OR (Ministra de Justicia) OR COVID_19)',
'202004302359','202004300000']}



Mayo ={'Mayo_1' : ['(AFPs OR VizcarraNoSeasFresco OR Dia46 OR
NoSeanFrescos OR (Día 46) OR (AFP Habitat) OR (25% de afp) OR COVID_19
OR 30Abril OR DevuelvemeMiDinero OR 1deMayo OR Dia47 OR AFPs OR (1ro
de Mayo))', '202005012359','202005010000'],

        'Mayo_2' : ['((Día 47) OR AFPs OR Dia47 OR COVID_19 OR
VizcarraNoSeasFresco OR CuandoVuelvaASalir OR Dia48 OR 1deMayo)',
'202005022359','202005020000'],

        'Mayo_3' : ['((Bien Vizcarra) OR UnidosDesdeCasa OR Dia48 OR
(Día 48) OR COVID_19 OR cuarentenatotal OR Dia49 OR Iquitos OR (Fase
1) OR 3Mayo OR (Día 49))', '202005032359','202005030000'],

        'Mayo_4' : ['((Día 49) OR (Bien Vizacarra) OR (Fase 1) OR
Dia49 OR Dia50 OR Iquitos OR Loreto OR (Día 50) OR (Fase 1) OR COVID-
19)', '202005042359','202005040000'],

        'Mayo_5' : ['(Dia50 OR (Día 50) OR COVID-19 OR dia51 OR
(Presidente Martín Vizcarra) OR (Día 51) OR Iquitos OR Loreto)',
'202005052359','202005050000'],

        'Mayo_6' : ['((Día 51) OR (Ministerio del Interior) OR COVID-
19 OR Dia52 OR (Día 51) OR Loreto OR Iquitos)',
'202005062359','202005060000'],

        'Mayo_7' : ['(Dia52 OR (Día 52) OR (Ministro de Agricultura)
OR Dia53 OR 7Mayo OR (Ciro Maguiña) OR (Ministro de Defensa) OR (Día
53) OR (Maternidad de Lima) OR BCRP)', '202005072359','202005070000'],

        'Mayo_8' : ['((Día 53) OR Dia53 OR COVID_19 OR Dia54 OR
CuarentenaPeru OR ToqueDeQueda OR (Presidente Martín Vizcarra) OR
(Estado de Emergencia) OR FFAA OR (Ministro de Defensa))',
'202005082359','202005080000'],

        'Mayo_9' : ['(Dia54 OR Vizcarra OR CuarentenaPeru OR (Día 54)
OR (Estado de Emergencia) OR BonoUniversal OR cuarentenatotal OR dia55
OR (Día 55) OR CuarentenaExtendida)', '202005092359','202005090000'],

        'Mayo_10' : ['(FuerzaZamora OR dia55 OR (Día 55) OR
CuarentenaPeru OR COVID_19 OR Iquitos OR FueraZamora)',
'202005102359','202005100000'],

        'Mayo_11' : ['((Día 56) OR FuerzaZamora OR Dia56 OR Zamora OR
Maguiña OR Dia57 OR (Ministro de Salud))',
'202005112359','202005110000'],
```

```
        'Mayo_12' : ['(Zamora OR Maguiña OR Dia57 OR (Día 57) OR dia58
OR (Colegio Médico) OR (Ministro de Salud) OR Iquitos)',
'202005122359','202005120000'],

        'Mayo_13' : ['(Zamora OR dia58 OR (Día 58) OR (Colegio Médico)
OR Dia59 OR (Martín Vizcarra) OR (Ministro de Salud) OR 13MAy OR
13DeMayo)', '202005132359','202005130000'],

        'Mayo_14' : ['(Dia59 OR (Día 59) OR (Martín Vizcarra) OR
COVID-19 OR Dia60 OR TeCuidoPerú)', '202005142359','202005140000'],

        'Mayo_15' : ['(Dia60 OR (Día 60) OR COVID-19 OR Dia61 OR (Día
61))', '202005152359','202005150000'],

        'Mayo_16' : ['((Día 61) OR Dia61 OR Día62)',
'202005162359','202005160000'],

        'Mayo_17' : ['((Día 62) OR idiotasencuarentena OR Día62 OR
(María Antonieta Alva) OR (25% de afp) OR 17May)',
'202005172359','202005170000'],

        'Mayo_18' : ['((Día 63) OR (Pilar Mazzetti) OR Dia63 OR Dia64
OR 18May)', '202005182359','202005180000'],

        'Mayo_19' : ['((Día 64) OR Dia64 OR LaPoliciaNoMeCuida OR
Dia65 OR (Presidente Martín Vizcarra) OR (Pilar Mazzetti) OR (Colegio
Médico del Perú) OR MIDIS OR 19May)', '202005192359','202005190000'],

        'Mayo_20' : ['(Dia65 OR (Día 65) OR LaPoliciaNoMeCuida OR
COVID_19 OR VIzcarraEsELProblema OR Dia66 OR BonoUniversalFamiliar OR
(Día 66) OR 20May)', '202005202359','202005200000'],

        'Mayo_21' : ['(VizcarraEsElProblema OR BonoFamiliarUniversal
OR SinBonoYo OR (Día 66) OR Dia66 OR COVID_19 OR Dia67 OR (Día 67) OR
SinBonoYo OR 21May)', '202005212359','202005210000'],

        'Mayo_22' : ['(VizcarraEsElProblema OR (Día 67) OR Dia67 OR
BonoFamiliarUniversal OR BonoUniversal OR (Estado de Emergencia) OR
Cuarentena OR Dia68 OR ToqueDeQueda OR 30dejunio OR COVID-19)',
'202005222359','202005220000'],

        'Mayo_23' : ['(UnMesMásSin OR FloreoComoVizcarra OR (Estado de
Emergencia) OR Cuarentena OR (Día 68) OR (CUA REN TE NA) OR Dia68 OR
COVID-19 OR ToqueDeQueda OR 23May OR Vizcarra OR UnMesMásSin OR COVID-
19)', '202005232359','202005230000'],

        'Mayo_24' : ['(sabadodecuarentena OR Vizcarra OR (Estado de
Emergencia) OR Dia70 OR (Día 70) OR 24Mayo OR sabadocuarentena OR
dia69 OR ToqueDeQueda OR 23May)', '202005242359','202005240000'],

        'Mayo_25' : ['(Dia70 OR (Día 70) OR Perú OR Dia71 OR (Día 71)
OR Perú)', '202005252359','202005250000'],

        'Mayo_26' : ['(Dia71 OR (Día 71) OR Dia72 OR (día 72))',
'202005262359','202005260000'],

        'Mayo_27' : ['(Dia72 OR (día 72) OR (Pilar Mazzetti)  OR BCRP
OR Dia73)', '202005272359','202005270000'],
```

```
        'Mayo_28' : ['((Día 73) OR (Pilar Mazzetti) OR Dia73 OR
NuevaConvivencia OR (Día 74) OR Dia74 OR CuarentenaConDSDR)',
'202005282359','202005280000'],

        'Mayo_29' : ['((Día 74) OR Dia74 OR SoloPorCuarentena OR Dia75
OR NuevaConvivencia OR SoloPorCuarentena)',
'202005292359','202005290000'],

        'Mayo_30' : ['((Día 75) OR Dia75 OR NuevaConvivencia OR dia76
OR (Día 76) OR 30May OR FFAA)', '202005302359','202005300000'],

        'Mayo_31' : ['(sabadodecuarentena OR 30May OR Dia77 OR (Día
77) OR COVID_19 OR 31May)', '202005312359','202005310000']}



Junio ={'Junio_1' : ['(Dia77 OR Dia78)',
'202006012359','202006010000'],

        'Junio_2' : ['((dia 78) OR Dia78 OR COVID_19 OR 1Jun OR Dia79
OR (Día 79) OR 2Jun)', '202006022359','202006020000'],

        'Junio_3' : ['(Pandemio OR Dia79 OR 2Jun OR Dia80)',
'202006032359','202006030000'],

        'Junio_4' : ['(dia80 OR Día80 OR Dia81 OR SuSalud OR (Fase 2)
OR (Ministra de Economía) OR (Día 81) OR (Emergencia Sanitaria) OR
4Jun OR COVID-19)', '202006042359','202006040000'],

        'Junio_5' : ['((Fase 2) OR (Día 81) OR Susalud OR Dia81 OR
COVD-19 OR Fase2 OR PeruAbreLosOjos OR (Día 81) OR (Emergencia
Sanitaria) OR Dia82)', '202006052359','202006050000'],

        'Junio_6' : ['((Día 82) OR (Fase 2) OR PeruAbreLosOjos OR
Dia82 OR 5Junio OR Dia83 OR (Día 83) OR Iquitos OR PeruAbreLosOjos)',
'202006062359','202006060000'],

        'Junio_7' : ['(UnidosDesdeCasa OR (Día 83) OR Dia83 OR
sabadodecuarentena OR (Día 84) OR (Pilar Mazzetti) OR Día84 OR
Iquitos)', '202006072359','202006070000'],

        'Junio_8' : ['(Perú OR (Día 84) OR 7dejunio OR Día84 OR
Dia85)', '202006082359','202006080000'],

        'Junio_9' : ['(LevantenCuarentena OR (Día 85) OR 8junio OR
día86)', '202006092359','202006090000'],

        'Junio_10' : ['((Martín Vizcarra) OR (Día 86) OR día86 OR
LevantenCuaarentena OR Dia87)', '202006102359','202006100000'],

        'Junio_11' : ['((Día 87) OR Dia87 OR COVID_19 OR (Día 88) OR
Dia88)', '202006112359','202006110000'],

        'Junio_12' : ['((día 88) OR Dia88 OR PrimeroMiSalud OR
Dia89)', '202006122359','202006120000'],
```

```
        'Junio_13' : ['(Dia89 OR PrimeroMiSalud OR
VizcarraEsElProblema OR Dia90 OR (Día 90))',
'202006132359','202006130000'],

        'Junio_14' : ['(VizcarraEsELProblema OR (Día 90) OR DIA90 OR
PrimeroMiSalud OR Dia91 OR Reactiva OR 14Junio OR (Día 91))',
'202006142359','202006140000'],

        'Junio_15' : ['((Reactiva Perú) OR (Pilar Mazzetti) OR Día 91
OR reactivaperu OR DevuelvanLos5200 OR VizcarraEsELProblema OR
(Arranca Perú) OR NoMasCuarentenaPeru OR (Ministra de Economía) OR
PrimeroMiSalud)', '202006152359','202006150000'],

        'Junio_16' : ['(NoMasCuarentenaPeru OR (Reactiva Perú) OR Día
92 OR (Pilar Mazzetti) OR ArrancaPeru OR Día92 R AisladosParaVivir OR
PrimeroMiSalud OR Dia93 OR Día 93 OR mypes OR 16Jun OR Día92)',
'202006162359','202006160000'],

        'Junio_17' : ['((Reactiva Perú) OR (Día 93) OR reactivaperu OR
Dia93 OR NoMasCuarentenaPeru OR ArrancaPeru OR PrimeroMiSalud OR Dia94
OR (Día 93) OR mypes)', '202006172359','202006170000'],

        'Junio_18' : ['(VizcarraMentiroso OR (Reactiva Perú) OR (Día
94) OR Dia94 OR reactivaperu OR Dia95 OR (Ciro Maguiña) OR
VizcarraMentiroso OR PrimeroMiSalud)', '202006182359','202006180000'],

        'Junio_19' : ['(Día 95 OR VizcarraMentiroso OR
ClasicoFloroBarato OR Dia95 OR PrimeroMiSalud OR Dia96)',
'202006192359','202006190000'],

        'Junio_20' : ['((Día 96) OR PorUnPerúSinHambre OR Dia96 OR
Dia97)', '202006202359','202006200000'],

        'Junio_21' : ['((Día 97) OR Dia97 OR Día98)',
'202006212359','202006210000'],

        'Junio_22' : ['(Perú OR VizcarraEsELProblema OR COVID_19 OR
dia99 OR (el estado) OR (Día 99) OR 22Jun)',
'202006222359','202006220000'],

        'Junio_23' : ['(dia99 OR COVID_19 OR (Día 99) OR (gino costa)
OR Día 100 OR (Consejo de Ministros) OR dia99 OR 100DíasSin)',
'202006232359','202006230000'],

        'Junio_24' : ['(Día100 OR (Día 100) OR RRHH OR COVID_19 OR
100diasdecuarentena OR IngresoBásicoUniversal OR Vizcarra OR 24dejunio
OR 100Dias OR (Después de 100))', '202006242359','202006240000'],

        'Junio_25' : ['(Vizcarra OR (Día 101) OR Dia101 OR COVID_19 OR
24Jun OR VizcarraDictador OR (María Antonieta Alva) OR Essalud OR
(Martín Vizcarra) OR (Día 101) OR Minsa OR Dia102)',
'202006252359','202006250000'],

        'Junio_26' : ['((Día 102) OR (María Antonieta Alva) OR Essalud
OR VizcarraDictador OR Dia102 OR Vizcarra OR dia103)',
'202006262359','202006260000'],

        'Junio_27' : ['((Estado de Emergencia) OR (Día 103) OR Dia103
OR EsUnaEmergencia)', '202006272359','202006270000'],
```

```
        'Junio_28' : ['((Día 104) OR Dia104 OR Ministerio OR (Día
105))', '202006282359','202006280000'],

        'Junio_29' : ['(Día106 OR 29Jun OR SinCuarentenaYo)',
'202006292359','202006290000'],

        'Junio_30' : ['((Día 106) OR Día106 OR Dia107 OR COVID-19)',
'202006302359','202006300000']}



Julio ={'Julio_1' : ['(Dia107 OR (Día 107) OR 30Junio OR COVID-19 OR
NuevaNormalidad OR Dia1)', '202007012359','202007010000'],

        'Julio_2' : ['(MiCuarentenaFue OR COVID-19 OR
AisladosParaVivir)', '202007022359','202007020000'],

        'Julio_3' : ['(COVID-19 OR MedioAñoSin OR 3Jul OR
NuevaMortalidad OR AisladosParaVivir)',
'202007032359','202007030000'],

        'Julio_4' : ['(COVID)', '202007042359','202007040000'],

        'Julio_5' : ['((Podemos Perú) OR Vizcarra OR (mensaje a la
nación) OR (El Presidente) OR 5Jul)', '202007052359','202007050000'],

        'Julio_6' : ['((Presidente Vizcarra) OR (presidente de la
república) OR vizcarra OR CongresoVerguenzaNacional OR (Estado de
Derecho))', '202007062359','202007060000'],

        'Julio_7' : ['(Vizcarra)', '202007072359','202007070000'],

        'Julio_8' : ['(Presidente Martín Vizcarra)',
'202007082359','202007080000'],

        'Julio_9' : ['(COVID)', '202007092359','202007090000'],

        'Julio_10' : ['(LasMentirasDeVizcarra OR COVID_19)',
'202007102359','202007100000'],

        'Julio_11' : ['(LasMentirasDeVizcarra OR COVID_19 OR
IRRESPONSABLE)', '202007112359','202007110000'],

        'Julio_12' : ['(LasMentirasDeVizcarra OR COVID_19)',
'202007122359','202007120000'],

        'Julio_13' : ['(COVID_19)', '202007132359','202007130000'],

        'Julio_14' : ['(COVID_19)', '202007142359','202007140000'],

        'Julio_15' : ['((Ministro de Trabajo) OR mazzetti OR (Palacio
de Gobierno))', '202007152359','202007150000'],

        'Julio_16' : ['((Ministro de Trabajo) OR Premier OR Zamora OR
Ministros OR MINTRA OR (Primer Ministro))',
'202007162359','202007160000'],
```

```
        'Julio_17' : ['((Ministro de Trabajo) OR Premier OR Zamora OR
(Pilar Mazzetti) OR COVID_19)', '202007172359','202007170000'],

        'Julio_18' : ['(COVID)', '202007182359','202007180000'],

        'Julio_19' : ['(COVID-19)', '202007192359','202007190000'],

        'Julio_20' : ['(VizcarraBastaDeMentiras OR (NO MIENTAS) OR
COVID-19 OR Iquitos)', '202007202359','202007200000'],

        'Julio_21' : ['(VizcarraBastaDeMentiras OR COVID-19  OR
(Comisión de Salud) OR Iquitos OR (Ministro de Defensa))',
'202007212359','202007210000'],

        'Julio_22' : ['((Pilar Mazzetti) OR (QUÉ LE DIRÍAS))',
'202007222359','202007220000'],

        'Julio_23' : ['((Vamos Perú) OR (Arriba Perú))',
'202007232359','202007230000'],

        'Julio_24' : ['(SinElVirusYo OR DeLaNormalidadExtraño OR
24Jul)', '202007242359','202007240000'],

        'Julio_25' : ['(UnidosDesdeCasa OR AhoraTeTocaATi OR
SinElVirusYo OR DeLaNormalidadExtraño)',
'202007252359','202007250000'],

        'Julio_26' : ['(NOTICIAS EL COMERCIO PERÚ)',
'202007262359','202007260000'],

        'Julio_27' : ['(ReactivaMisDerechos OR Municpalidad)',
'202007272359','202007270000'],

        'Julio_28' : ['(Vizcarra OR (mensaje a la nación) OR
MensajeALaNación OR mensajepresidencial OR Perú OR Vizcarra OR Pdte OR
República)', '202007282359','202007280000'],

        'Julio_29' : ['(Vizcarra OR Essalud OR (NO VALE MENTIR) OR
MensajeALaNación OR COVID_19 Or VizcarraBastaDeMentiras OR (banco de
la nación) OR (mensaje a la nación) Or DestinoPeruano OR
VizcarraNoEscuchaAlPueblo)', '202007292359','202007290000'],

        'Julio_30' : ['(Ministra OR COVID_19 Or Muñoz OR (NO VALE
MENTIR) OR Essalud OR 28deJulio)', '202007302359','202007300000'],

        'Julio_31' : ['((Estado de Emergencia Nacional) OR Ministra)',
'202007312359','202007310000']}



#strong_key = 'Vizcarra OR Peru OR Perú OR Minsa OR Zamora OR
QuedateEnCasa OR Mazzetti OR (Antonieta Alva)'

strong_key ='Vizcarra OR Minsa OR Zamora OR (Antonieta Alva) OR
Mazzetti OR covid OR COVID OR cuarentena OR Essalud'

tweetdate  = Julio['Julio_31']
```

```
query      = '(' + tweetdate[0] + ' OR ' + ' (' + strong_key + ')' + ')' + '
lang:es' + ' (profile_country:PE OR profile_region:Lima OR
profile_region:Arequipa OR profile_region:Cusco OR
profile_region:Ayacucho OR profile_region:Iquitos OR
profile_region:Pasco OR profile_region:Puno OR profile_region:Piura OR
profile_region:Ancash OR profile_region:Loreto OR profile_region:Junin
OR profile_region:Callao OR profile_region:Trujillo OR
profile_region:Ica OR profile_region:Tacna OR profile_region:Ucayali
OR profile_region:Madre de Dios OR profile_region:La Libertad OR
profile_region:Lambayeque OR profile_region:Moquegua OR
profile_region:Amazonas OR profile_region:Huanuco OR
profile_region:Apurimac OR profile_region:San Martin OR place:Peru)' +
' -is:retweet'

from_date  = tweetdate[2]

to_date    = tweetdate[1]



data = '"query":"{}", "maxResults": "500", "fromDate":"{}",
"toDate":"{}"'.format(query, from_date, to_date)

data = '{' + data + '}'

## Making query to search the next page in the results obtained by the
last query

query      = '(' + tweetdate[0] + ' OR ' + ' (' + strong_key + ')' + ')' + '
lang:es' + ' place:Peru' + ' -is:retweet'

from_date  = tweetdate[2]

to_date    = tweetdate[1]



data = '"query":"{}", "maxResults": "500", "fromDate":"{}",
"toDate":"{}", "next":"{}"'.format(query, from_date, to_date,
r2json['next'])

data = '{' + data + '}'

import json

import requests



endpoint =
'https://api.twitter.com/1.1/tweets/search/fullarchive/Dev.json'

headers  = {'authorization': 'Bearer XXXXXXXXXXXXXXXXXXXXXXXXXXXX',

            'content-type' : 'application/json'}
```

```
response = requests.post(endpoint, headers= headers, data= data)


if response.status_code != 200:
    raise Exception(response.status_code, response.text)


r2json = response.json()


for tweet in r2json['results']:
    try:


        new_data = TweetsTab(id    = tweet['id'],
                             id_user      = tweet['user']['id_str'],
                             text         = tweet['text'],
                             created_at   = tweet['created_at'],
                             Location_user= tweet['user']['location'],
                             Place_tweet  = tweet['place']['full_name']
if type(tweet['place']) == dict else tweet['place'])
        session.add(new_data)
        session.commit()
    except:
        session.rollback()
        raise
session.close()
prueba = session.query(TweetsTab)


for data in prueba:
    print(data)
```

```
#next_vect = []

for data in r2json['results']:

    print(data['text'])

#next_vect.append(r2json['next'])
```

Anexo 5. Geolocalización de los Tweets

```
# # GEOLOCALIZACIÓN DE LOS TWEETS

#

# Este código esta orientado a determinar cual es el departamento
al cual pertenece cada tweet de la base de datos



import pandas as pd

import numpy  as np

import operator



dataset                                                          =
pd.read_csv('BaseDeDatos_Tweets_Sent_Analy_PERU_RegresionLogist
ica.csv')

dataset.drop('Unnamed: 0', inplace=True, axis = 1)



dataset.sentiment.value_counts()



dataset['created_at']  = dataset['created_at'].apply(lambda  x:
x[:-9])



dataset['Location_user'] =dataset['Location_user'].apply(lambda
x: str(x).lower())

dataset['Place_tweet']    = dataset['Place_tweet'].apply(lambda
x: str(x).lower())
```

```
# Provincia Diccionario

prov_peru  = {'amazonas'  :  ['amazonas',  'chachapoyas',  'bagua
grande', 'san nicolás'],

 'ancash'  : ['ancash', 'chimbote', 'nuevo chimbote', 'coishco',
'huaraz', 'caraz', 'santo toribio',

                'pomabamba',  'huarmey',  'carhuaz',  'casma',
'chacas', 'nepeña', 'cáceres del perú',

                'aija',     'alfonso     ugarte',     'tauca',
'huaripampa'],

 'apurimac' : ['apurimac', 'abancay', 'talavera', 'andahuaylas',
'san marcos'],

 'arequipa'  :  ['arequipa',  'jose  luis  bustamante  y  rivero',
'cayma', 'uchumayo', 'jacobo hunter',

                'santa catalina', 'cerro colorado', 'arequipa',
'yanahuara', 'mollendo', 'sachaca',

                'acari',  'paucarpata',  'socabaya',  'mariano
melgar', 'alto selva alegre', 'vitor',

                'atico',   'yauca',   'caylloma',   'cayarani',
'tiabaya', 'orcopampa', 'camaná', 'acari'],

 'ayacucho'   :   ['ayacucho',   'llochegua',   'sarhua',   'jesus
nazareno', 'puquio', 'carmen alto',

                'huancaraylla', 'huanta'],

 'cajamarca'  :  ['cajamarca',  'cajamara','cutervo',  'jaen',
'jaén','san ignacio', 'los baños del inca', 'cajabamba',

                'chilete', 'chirinos', 'chota'],

 'callao'  : ['callao', 'la perla', 'carmen de la legua reynoso',
'la punta', 'bellavista',

                'hospital nacional alberto sabogal sologuren'],

 'cusco'  : ['cusco', 'wanchaq', 'santa ana', 'san sebastian',
'santiago', 'ollantaytambo',

                'san jerónimo', 'lamay', 'maras', 'chinchero',
'kimbiri', 'espinar', 'sicuani',

                'velille',      'paucartambo',      'oropesa',
'chamaca', 'santa teresa', 'machupicchu',

                'cuzco', 'urubamba', 'cuzco'],
```

```
 'huancavelica':['huancavelica',              'castrovirreyna',
'surcubamba', 'pampas'],

 'huanuco'  : ['huanuco', 'huánuco', 'amarilis', 'huacrachuco',
'rupa-rupa', 'la unión', 'puerto inca',

                'tingo maría', 'huacar', 'tingo maria', 'tingo
maría'],

 'ica'   : ['ica', 'marcona', 'san juan bautista', 'la tinguiña',
'los aquijes', 'sunampe',

                'pisco', 'subtanjalla', 'nazca', 'rio grande',
'palpa', 'chincha alta', 'paracas',

                'parcona',  'san  andres',  'chincha  baja',
'chincha'],

 'junin'  : ['junin', 'junín', 'el tambo', 'tarma', 'huancayo',
'pilcomayo', 'huancan', 'jauja',

                'santa  rosa  de  sacco',  'perene',  'pangoa',
'matahuasi', 'chanchamayo', 'mazamari',

                'san agustin', 'apata', 'san ramón', 'satipo',
'yauli', 'concepción', 'morococha',

                'san jerónimo de tunan', 'la oroya', 'huanca'],

 'la  libertad':  ['la  libertad',  'trujillo',  'cachicadan',
'jequetepeque', 'pacasmayo', 'la esperanza', 'simbal',

                'victor  larco  herrera',  'huanchaco',  'pueblo
nuevo', 'casa grande', 'el porvenir',

                'chicama', 'moche', 'laredo', 'pacanga', 'san
pedro de lloc', 'guadalupe', 'chepen',

                'chepén', 'paijan', 'huamachuco', 'florencia de
mora', 'trujiillo', 'sayapullo'],

 'lambayeque': ['lambayeque', 'ferreñafe', 'olmos', 'chiclayo',
'pimentel', 'jayanca', 'eten puerto',

                'monsefu',  'jose  leonardo  ortiz',  'eten',
'salas', 'tucume', 'motupe', 'chilayo',

                'íllimo', 'reque', 'pomalca'],

 'lima'   : ['lima', 'san juan de lurigancho', 'miraflores', 'san
isidro', 'mercado caqueta',

             'villa  maria  del  triunfo',  'san  juan  de
miraflores', 'cerro azul', 'santa maria del mar',
```

```
            'magdalena del mar', 'lurin', 'santiago de surco',
'nuevo imperial', 'santa eulalia',

            'san martin de porres', 'santa anita', 'breña',
'los olivos', 'av. caminos del inca',

            'comas', 'magdalena vieja', 'la victoria', 'la
molina', 'surquillo', 'pativilca',

            'lurigancho', 'san borja', 'ate', 'santa rosa',
'chorrillos', 'av. carlos izaguirre',

            'el agustino', 'barranco', 'villa el salvador',
'san miguel', 'yauyos', 'pueblo libre',

            'asia', 'lince', 'san bartolo', 'jesús maria',
'puente huachipa', 'punta hermosa',

            'pachacamac', 'colegio médico del perú', 'huacho',
'independencia', 'surco', 'huaura',

            'carabayllo', 'rimac', 'puente piedra', 'san
mateo', 'ventanilla', 'chancay', 'canta',

            'ancón', 'san luis', 'mala', 'hualmay', 'villa
militar de chorrillos', 'san bartolome',

            'chaclacayo', 'estación uni - metropolitano',
'huaral', 'ancon', 'embajada de suiza',

            'open plaza angamos', 'punta negra', 'san vicente
de cañete', 'santa cruz de flores',

            'paramonga', 'vegueta', 'supe', 'san antonio',
'chilca', 'cieneguilla', 'barranca',

            'palacio de gobierno', 'calango', 'san juan de
iris', 'hospital octavio mongrut essalud',

            'escuela de equitación del ejercito', 'matucana',
'lurín', 'l i m a', 'lim@', 'magdalena',

            'quilmana', 'lim', 'pamplona', 'las palmas', 'jesús
maría', '***Lima***', 'lims', 'salamanca',

            'cañete', 'l!m@', 's. isidro', 'jesus maria',
'monterrico', 'san iisidro', 'canete',

            'san martín de porres', 'gamarra', 'huarochiri'],

 'loreto'  : ['loreto', 'iquitos', 'yurimaguas', 'punchana',
'san lorenzo', 'putumayo', 'lagunas',
```

```
                'ramón castilla', 'torres causana', 'napo',
'manseriche'],

 'madre de dios': ['madre de dios', 'tambopata', 'las piedras',
'iberia', 'pto maldonado',

                  'puerto maldonado'],

 'moquegua' : ['moquegua', 'ilo', 'pacocha'],

 'pasco'   : ['pasco', 'chaupimarca', 'chontabamba', 'simón
bolívar', 'villa rica', 'huariaca',

                'oxapampa'],

 'piura'   : ['piura', 'cristo nos valga', 'marcavelica',
'ayabaca', 'sullana', 'chulucanas',

                'castilla', 'suyo', 'catacaos', 'pariñas',
'ignacio escudero', 'la brea', 'paita',

                'los organos', 'mancora', 'sechura', 'colan',
'la huaca', 'talara'],

 'puno'   : ['puno', 'anapia', 'desaguadero', 'juliaca', 'san
gaban'],

 'san martin' : ['san martin', 'tarapoto', 'moyobamba',
'tocache', 'cacatachi', 'tres unidos',

                'tabalosos', 'rioja', 'morales', 'lamas',
'polvora', 'soritor', 'saposoa',

                'la banda de shilcayo', 'picota', 'juanjui',
'san martín'],

 'tacna'   : ['tacna', 'pocollay', 'sama', 'ciudad nueva',
'ilabaya', 'calana', 'inclan',

                'alto de la alianza', 'coronel gregorio
albarracin lanchipa'],

 'tumbes'  : ['tumbes', 'zarumilla', 'aguas verdes', 'corrales',
'zorritos', 'tumbez'],

 'ucayali' : ['ucayali', 'calleria', 'yarinacocha', 'manantay',
'campoverde', 'pucallpa']}




def set_place(Place):
```

```
    prov_list = []
    end       = False
    for option in Place:
        reco = 0
        for prov in prov_peru:
            for elem in prov_peru[prov]:
                if elem in option:
                    prov_list.append(prov)
                    end = True
                    break
            if end == True:
                end = False
                break
            reco += 1
        if reco == 25:
            prov_list.append('NOPROVPERU')
    return prov_list



dataset['prov_place_tweet']                                    =
set_place(dataset['Place_tweet'])

dataset['prov_location_user']                                  =
set_place(dataset['Location_user'])



dataset['Province'] = dataset['prov_place_tweet']

vec_noprov          = dataset['prov_place_tweet']=='NOPROVPERU'
```

```
dataset.loc[vec_noprov, 'Province'] = dataset.loc[vec_noprov, 'prov_location_user']



dataset[dataset['Province']=='NOPROVPERU']



dataset['Country'] = dataset['Province']



dataset.loc[dataset['Country']!='NOPROVPERU', 'Country'] = 'PERU'



vec_indx = dataset['Country']=='NOPROVPERU'



vector = list(((dataset['Place_tweet'][vec_indx] == 'peru')|(dataset['Location_user'][vec_indx] == 'peru'))|((dataset['Place_tweet'][vec_indx] == 'perú')|(dataset['Location_user'][vec_indx] == 'perú')))

final = [i*'PERU' for i in vector]



dataset.loc[vec_indx, 'Country'] = final



dict_month = {'03':'marzo', '04':'abril', '05':'mayo', '06':'junio', '07':'julio'}



dataset['month'] = dataset['created_at'].apply(lambda x: dict_month[x[5:7]])



dataset['month'].value_counts().plot.bar()




db_mob = pd.read_csv('Mobility_Report_Peru_withAVGValues.csv')

db_mob.drop('Unnamed: 0', inplace=True, axis = 1)
```

```
db_mob['sub_region_1']  = db_mob['sub_region_1'].map(lambda x:
x.lower() if isinstance(x,str) else x)



vect = []

for   row_date,    row_loc    in   zip(dataset['created_at'],
dataset['Province']):

    #print(row_date + '---' + row_loc)

    if row_loc == 'NOPROVPERU':

        vect.append(np.nan)

    else:

        #print(db_mob[(db_mob['date']==row_date)              &
(db_mob['sub_region_1']==row_loc)].values)

        vect.append(db_mob['AVG_Mobility'][(db_mob['date']==row
_date) & (db_mob['sub_region_1']==row_loc)].values)




vec_mob = [x.item() if not np.isnan(x) else np.nan for x in vect]



dataset['AVG_Mobility'] = vec_mob



base = dataset[['id', 'created_at', 'text','month', 'Province',
'sentiment', 'AVG_Mobility']].copy()



base['Province'].value_counts()



base.to_csv('BaseDeDatos_Analisis_Thesis_2.csv')
```

Anexo 6. Procesamiento de los Reportes de Movilidad de GOOGLE

```
# # PROCESAMIENTO DE LOS REPORTES DE MOVILIDAD DE GOOGLE
```

```
#

# Para el uso de los datos de movilidad se realizó el siguiente
código



import pandas as pd



path =
r'Region_Mobility_Report_CSVs\2020_PE_Region_Mobility_Report.cs
v'

data = pd.read_csv(path)

data = data[data['sub_region_2'].isna()]



data2 = data[['date',

              'sub_region_1',

              'retail_and_recreation_percent_change_from_baseli
ne',

              'grocery_and_pharmacy_percent_change_from_baselin
e',

              'parks_percent_change_from_baseline',

              'transit_stations_percent_change_from_baseline',

              'workplaces_percent_change_from_baseline']].copy(
)



time = ['2020-03-16', '2020-03-17', '2020-03-18', '2020-03-19',

        '2020-03-20', '2020-03-21', '2020-03-22', '2020-03-23',

        '2020-03-24', '2020-03-25', '2020-03-26', '2020-03-27',

        '2020-03-28', '2020-03-29', '2020-03-30', '2020-03-31',

        '2020-04-01', '2020-04-02', '2020-04-03', '2020-04-04',

        '2020-04-05', '2020-04-06', '2020-04-07', '2020-04-08',

        '2020-04-09', '2020-04-10', '2020-04-11', '2020-04-12',
```

```
        '2020-04-13', '2020-04-14', '2020-04-15', '2020-04-16',
        '2020-04-17', '2020-04-18', '2020-04-19', '2020-04-20',
        '2020-04-21', '2020-04-22', '2020-04-23', '2020-04-24',
        '2020-04-25', '2020-04-26', '2020-04-27', '2020-04-28',
        '2020-04-29', '2020-04-30', '2020-05-01', '2020-05-02',
        '2020-05-03', '2020-05-04', '2020-05-05', '2020-05-06',
        '2020-05-07', '2020-05-08', '2020-05-09', '2020-05-10',
        '2020-05-11', '2020-05-12', '2020-05-13', '2020-05-14',
        '2020-05-15', '2020-05-16', '2020-05-17', '2020-05-18',
        '2020-05-19', '2020-05-20', '2020-05-21', '2020-05-22',
        '2020-05-23', '2020-05-24', '2020-05-25', '2020-05-26',
        '2020-05-27', '2020-05-28', '2020-05-29', '2020-05-30',
        '2020-05-31', '2020-06-01', '2020-06-02', '2020-06-03',
        '2020-06-04', '2020-06-05', '2020-06-06', '2020-06-07',
        '2020-06-08', '2020-06-09', '2020-06-10', '2020-06-11',
        '2020-06-12', '2020-06-13', '2020-06-14', '2020-06-15',
        '2020-06-16', '2020-06-17', '2020-06-18', '2020-06-19',
        '2020-06-20', '2020-06-21', '2020-06-22', '2020-06-23',
        '2020-06-24', '2020-06-25', '2020-06-26', '2020-06-27',
        '2020-06-28', '2020-06-29', '2020-06-30', '2020-07-01',
        '2020-07-02', '2020-07-03', '2020-07-04', '2020-07-05',
        '2020-07-06', '2020-07-07', '2020-07-08', '2020-07-09',
        '2020-07-10', '2020-07-11', '2020-07-12', '2020-07-13',
        '2020-07-14', '2020-07-15', '2020-07-16', '2020-07-17',
        '2020-07-18', '2020-07-19', '2020-07-20', '2020-07-21',
        '2020-07-22', '2020-07-23', '2020-07-24', '2020-07-25',
        '2020-07-26', '2020-07-27', '2020-07-28', '2020-07-29',
```

```
        '2020-07-30', '2020-07-31']



data2 = data2[[value in time  for value in data2['date']]]

data2.sub_region_1.unique()



data2 = data2.loc[data2['sub_region_1']!='Metropolitan
Municipality of Lima',:]



vec1 = data2['sub_region_1']=='Lima Region'

vec2 = data2['sub_region_1']=='Callao Region'

data2.loc[vec1, 'sub_region_1'] = 'Lima'

data2.loc[vec2, 'sub_region_1'] = 'Callao'



data2['AVG_Mobility'] = data2.iloc[:,2:5].mean(axis=1,
skipna=True)



data2.to_csv('Mobility_Report_Peru_withAVGValues.csv')
```

Anexo 7. Tabla de rendimiento de los modelos entrenados

MODELOS	**PRECISIÓN**	**F1**
MultinomialNB	63.54%	63.54
BernoulliNB	63.73%	63.73
ComplementNB	61.91%	61.91

LogisticRegression	**64.39%**	**64.39**
KNeighborsClassifier	53.33%	53.33
DecisionTreeClassifier	52.20%	52.20
RandomForestClassifier	61.39%	61.39
SGDClassifier	61.17%	61.17
SVC	64.29%	64.29
LinearSVC	60.94%	60.94
MLPClassifier	53.55%	53.55
AdaBoostClassifier	60.88%	60.88

Para la elección del modelo clasificación se compararon métricas como la Precisión y valoración F1 (Para datos desbalanceados) promedios, obtenidos a partir del proceso de validación cruzada con K = 5 iteraciones. Los resultados del proceso de validación se pueden apreciar en la siguiente tabla.

Los modelos que presentan mayor rendimiento según lo presentado en la tabla X son los modelos de Regresión Logística y una variante de SVM con precisiones del 64.39% y 64.29% respectivamente. En consecuencia, el modelo a utilizar para la clasificación de polaridad de los tweets fue el modelo de Regresión Logística.