

UNIVERSIDAD INTERNACIONAL SEK

FACULTAD DE ARQUITECTURA E INGENIERÍAS

Trabajo de fin de carrera titulado:

**“DESARROLLO DE UN MÓDULO DE ALERTAS TEMPRANAS
ANALIZANDO LOS DATOS PÚBLICOS DE *TWITTER* PARA LA
PREVENCIÓN DE ALTERCADOS EN ESTADIOS DE FÚTBOL CASO
DE ESTUDIO: CAMPEONATO NACIONAL SERIE A ECUADOR -
2017”**

Realizado por:

Ing. Jorge David Herrera Sarango

Director del proyecto:

PhD. Daniel Riofrío

Como requisito para la obtención del título de:

**MÁSTER EN TECNOLOGÍAS DE LA
INFORMACIÓN**

Quito, Junio de 2018

DECLARACIÓN JURAMENTADA

Yo, **JORGE DAVID HERRERA SARANGO**, con cédula de identidad **1104624380**, declaro bajo juramento que el trabajo aquí desarrollado es de mi autoría, que no ha sido previamente presentado para ningún grado o calificación profesional; y, que ha consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración, cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la UNIVERSIDAD INTERNACIONAL SEK, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normativa institucional vigente.

Jorge David Herrera Sarango
CI: 1104624380

DECLARATORIA

El presente trabajo de investigación titulado:

**“DESARROLLO DE UN MÓDULO DE ALERTAS TEMPRANAS ANALIZANDO
LOS DATOS PÚBLICOS DE *TWITTER* PARA LA PREVENCIÓN DE
ALTERCADOS EN ESTADIOS DE FÚTBOL CASO DE ESTUDIO: CAMPEONATO
NACIONAL SERIE A ECUADOR-2017”**

Realizado por:

JORGE DAVID HERRERA SARANGO

Como requisito para la obtención del título de:

**MÁSTER EN TECNOLOGÍAS DE LA
INFORMACIÓN**

Ha sido dirigido por el docente:

PhD. Daniel Riofrío

Quien considera que constituye un trabajo original de su autor

PhD. Daniel Riofrío

DIRECTOR

LOS PROFESORES INFORMANTES

Los profesores informantes:

Diego Fernando Riofrío Luzcando

Verónica Elizabeth Rodríguez Arboleda

Después de revisar el trabajo presentado, lo han calificado como apto para su defensa oral ante el tribunal examinador.

PhD. Diego F. Riofrío

Ing. Verónica E. Rodríguez, MBA

Quito, Junio de 2018

DEDICATORIA

En primer lugar, dedico este logro a Dios por tantas bendiciones recibidas durante toda la vida.

A mis queridos padres Francisco y Rosita, por haberme apoyado en todo momento, con sus consejos y motivación constante, que me ha permitido ser la persona que soy, pero más que nada su amor y comprensión incondicional.

A mis hermanos, tía Patricia, familiares, Adriana y amigos por el gran apoyo que me han brindado a lo largo de mi carrera los quiero mucho.

A mis amigos que me han apoyado continuamente en mi formación profesional y espero que sigamos siendo amigos: Bryan, Diego, Marco y Wilson.

¡Gracias a todos ustedes!

AGRADECIMIENTO

Agradezco infinitamente a mis profesores por todos los conocimientos que han sido impartidos a lo largo de estos dos años, por su constante apoyo en todo lo que respecta a mi formación profesional.

Un agradecimiento especial para la persona que me ha guiado de manera incondicional, Daniel Riofrío, por sus consejos y recomendaciones para que este trabajo de fin de titulación resulte de una manera eficaz y que sea útil, como una fuente de consulta para futuros trabajos.

A mis padres por toda su dedicación y esfuerzo que me han brindado a lo largo de este tiempo.

A los amigos que siempre han sabido estar, y a todas aquellas personas que han sabido apoyarme de una u otra manera.

ÍNDICE GENERAL DE CONTENIDO

RESUMEN.....	XIII
ABSTRACT.....	XIV
CAPÍTULO I INTRODUCCIÓN.....	1
1.1 EL PROBLEMA DE LA INVESTIGACIÓN	1
1.1.1 Planteamiento del Problema.....	1
1.1.2 Formulación del Problema	7
1.1.3 Sistematización del Problema	7
1.1.4 Objetivo General	7
1.1.5 Objetivos Específicos.....	8
1.1.6 Justificación.....	8
1.2 MARCO TEÓRICO.....	13
1.2.1 Marco Conceptual	16
1.2.2 Estado del Arte.....	18
CAPÍTULO II MÉTODO	22
2.1 ANÁLISIS	22
2.1.1 Tendencias de uso de Smartphone y Redes Sociales en Ecuador.....	22
2.1.2 Palabras Clave de los Equipos que conforman el campeonato nacional ecuatoriano de fútbol Serie A 2017.....	24
2.1.3 Datos de Twitter que pueden ser Extraídos.....	24
2.2 DISEÑO DEL SISTEMA PROPUESTO.....	26
2.2.1 Esquema de la Arquitectura	26
2.2.2 Descripción del Funcionamiento del Módulo de Alertas Tempranas.....	27
2.2.3 Capa de Presentación	28
2.2.4 Capa de Negocio	28
2.2.5 Capa de Datos	36
2.3 IMPLEMENTACIÓN DE PROTOTIPO.....	37
2.3.1 Tecnologías Usadas.....	37
2.3.2 Resumen.....	38
CAPÍTULO III RESULTADOS.....	39
3.1 EXPERIMENTACIÓN	39
3.2 EXPERIMENTO UNO: EXTRACCIÓN, ANÁLISIS Y ALMACENAMIENTO DE TWEETS	40
3.3 EXPERIMENTO DOS: PRUEBAS DE FUNCIONAMIENTO DEL ÍNDICE PARA LA GENERACIÓN DE ALERTAS	45
3.4 EXPERIMENTO TRES: ANÁLISIS DE TRES ENCUENTROS DEPORTIVOS DE FÚTBOL DEL CAMPEONATO NACIONAL ECUATORIANO SERIE A.....	49
3.5 ANÁLISIS RESULTADOS	54

CAPÍTULO IV DISCUSIÓN	56
4.1 CONCLUSIONES	56
4.2 RECOMENDACIONES Y TRABAJOS FUTUROS	57
REFERENCIAS	58
ANEXOS	62

ÍNDICE DE TABLAS

CAPÍTULO I INTRODUCCIÓN.....	1
Tabla # 1. Número de espectadores del año 2014	3
Tabla # 2. Número de espectadores del año 2015	3
Tabla # 3. Número de espectadores del año 2016	4
Tabla # 4. Equipos de fútbol que participan en la Copa Banco del Pacífico 2017	5
Tabla # 5. Porcentaje de población con celular y redes sociales	9
Tabla # 6. Delitos del COIP.....	12
CAPÍTULO II MÉTODO	22
Tabla # 7. Palabras clave de los equipos de fútbol del campeonato ecuatoriano.....	24
Tabla # 8. Atributos que contiene un tweet	25
Tabla # 9. Etiquetado de los tweets	34
Tabla # 10. Cantidad de datos extraídos de Twitter durante 4 fechas del campeonato de fútbol ecuatoriano	35
Tabla # 11. Resumen de software y librerías.....	38

ÍNDICE DE FIGURAS

CAPÍTULO I INTRODUCCIÓN	1
Figura # 1. Tendencia de incidentes ocurridos desde el año 2013 hasta agosto de 2017.....	6
Figura # 2. Delitos contra la inviolabilidad de la vida.....	10
Figura # 3. Delitos contra el derecho a la propiedad.....	10
Figura # 4. Delitos contra la integridad de la persona.....	11
Figura # 5. Delitos contra el derecho a la igualdad.....	11
Figura # 6. Arquitectura de un Sistema de Procesamiento de lenguaje Natural.....	16
CAPÍTULO II MÉTODO	22
Figura # 7. Tendencia de personas que hacen uso de smartphone y redes sociales en Ecuador.....	23
Figura # 8. Tendencia de asistencia de espectadores en un partido de fútbol Serie A.....	23
Figura # 9. Posibles relaciones de Twitter.....	25
Figura # 10. Arquitectura para el desarrollo del módulo de generación de alertas tempranas.....	26
Figura # 11. Interfaz gráfica del usuario.....	28
Figura # 12. Capa de negocios.....	29
Figura # 13. Proceso para la extracción de tweets.....	29
Figura # 14. Coordenadas del estadio de fútbol Rodrigo Paz Delgado de la ciudad de Quito.....	31
Figura # 15. Procesos para el procesamiento de los tweets.....	32
Figura # 16. Índices de alerta.....	35
Figura # 17. Capa de datos.....	36
Figura # 18. Almacenamiento de los tweets en la base de datos.....	37
CAPÍTULO III RESULTADOS	39
Figura # 19. Tweets analizados y ponderados por varias personas.....	39
Figura # 20. Tweets analizados mediante el modelo PLN.....	40
Figura # 21. Elección del encuentro para hacer analizado Macara VS Universidad católica.....	40
Figura # 22. Mensaje para Twitrear mediante palabra clave.....	41
Figura # 23. Análisis positivo de un tweet mediante palabra clave.....	41
Figura # 24. Almacenamiento de los tweets en la base de datos mediante palabra clave.....	42
Figura # 25. Mensaje para Twitrear mediante geolocalización.....	43
Figura # 26. Análisis positivo de un tweet mediante geolocalización.....	43
Figura # 27. Almacenamiento del tweet en la base de datos mediante geolocalización.....	44
Figura # 28. Presentación del tweet en el módulo de generación de alertas.....	44
Figura # 29. Presentación de índice zona segura.....	45
Figura # 30. Reporte de índice zona segura.....	46
Figura # 31. Gráfica de tweets índice zona segura.....	46
Figura # 32. Presentación de índice zona de alerta.....	47

Figura # 33. Reporte de índice zona de alerta.....	47
Figura # 34. Gráfica de tweets índice zona de alerta	48
Figura # 35. Presentación de índice zona de peligro	48
Figura # 36. Reporte de índice zona de peligro	49
Figura # 37. Gráfica de tweets índice zona de peligro.....	49
Figura # 38. Análisis del encuentro Independiente del Valle VS Liga de Quito	50
Figura # 39. Reporte del índice del encuentro Independiente del Valle VS Liga de Quito	50
Figura # 40. Gráfica estadística de los tweets analizados en encuentro Independiente del Valle VS Liga de Quito.....	51
Figura # 41. Análisis del encuentro Liga de Quito VS Emelec	51
Figura # 42. Reporte del índice del encuentro Liga de Quito VS Emelec	52
Figura # 43. Gráfica estadística de los tweets analizados en encuentro Liga de Quito VS Emelec.....	52
Figura # 44. Análisis del encuentro Emelec VS Delfín	53
Figura # 45. Reporte del índice del encuentro Emelec VS Delfín	53
Figura # 46. Gráfica estadística de los tweets analizados en encuentro Emelec VS Delfín	54
ANEXOS.....	62
Figura # A1. Calendario de juego primera etapa-primera vuelta serie A, Copa Banco del Pacífico 2017 ...	63
Figura # A2. Calendario de juego primera etapa-segunda vuelta serie A, Copa Banco del Pacífico 2017 ..	64
Figura # A3. Calendario de juego segunda etapa serie A, Copa Banco del Pacífico 2017	65

ÍNDICE DE ECUACIONES

CAPÍTULO I INTRODUCCIÓN	1
Ecuación # 1. Fórmula de un hiperplano	14
Ecuación # 2. Fórmula de un hiperplano sujeta a un conjunto	14
CAPÍTULO II MÉTODO	22
Ecuación # 3. Fórmula de Haversine	32
Ecuación # 4. Fórmula para normalización de datos	35
Ecuación # 5. Relación de tweets geolocalizados y key Words	35

RESUMEN

En los últimos años, en los eventos deportivos del campeonato nacional ecuatoriano de fútbol Serie A, se han producido varios incidentes, previo, durante y al finalizar partidos de fútbol, en dichos incidentes se han visto involucradas las personas aficionadas a este deporte y aquellas que no tienen relación con el mismo.

Estos disturbios no solo han sido de agresiones verbales o físicas, sino que han llegado a casos extremos como la pérdida de vidas. Es por ello, que se ha visto necesario el realizar el diseño y desarrollo de un módulo de generación de alertas tempranas, el cual permita la extracción, análisis y etiquetado de *tweets* referentes a partidos de fútbol, para generar alertas tempranas en caso que exista algún incidente o altercado dentro o fuera de un estadio de fútbol y así poder precautelar la seguridad de los aficionados.

En este trabajo se detalla el problema a investigar, así como los efectos que tendrá este en el futuro, los incidentes que han ocurridos en eventos de fútbol del campeonato ecuatoriano Serie A, la evolución del uso *smartphone* y redes sociales en el Ecuador, así como las leyes, reglamentos y normativas legales sobre la violencia en escenarios y eventos deportivos. Además, se realiza una investigación sobre estudios previos para la extracción, análisis de datos de *Twitter*, algoritmos de predicción y generación de alertas tempranas.

Adicionalmente, se describe la arquitectura del módulo de generación de alertas temprana el cual se la lleva a cabo utilizando un modelo de tres capas utilizando Python como lenguaje central. Además, se presentan tres experimentos los cuales sirvieron para verificar el funcionamiento del módulo de generación de alertas tempranas.

Finalmente, se presentan las conclusiones, recomendaciones y trabajos futuros que se desprenden de este trabajo de fin de titulación.

Palabras clave: *Twitter*, PLN, alerta temprana, extracción de datos, Python

ABSTRACT

In the past years, the national championship soccer Serie A in Ecuador has presented violent incidences. These incidents occur prior, during and at the end of soccer matches. Generally, the incidents involve soccer followers and people who do not have any connection with this sport.

These disturbances have been verbally, physically and have reached extreme situations as the dead of people. In order to prevent these kind of violent situations, this work proposes the design and development of an early warning generation system. This system module allows the extraction, analysis and labelling of tweets regarding soccer matches. With this information, it is possible to generate early warnings about any incident inside or outside of a soccer stadium, to safeguard the security of the fans.

The present work details the research problem. Furthermore, we analyze the incidents that have occurred in soccer events of the Ecuadorian championship series A, the evolution of the smartphone and use social networks in Ecuador, as well as the law, regulations and legal regulations about violence in sport events and sport scenarios. We also review previous studies about analysis of Twitter data, prediction algorithms and generation of early warnings.

In addition, we describe the architecture of an early warning generation system. This module uses three layers and Python as its main programming language. We present three experiments to verify the operation of our system and show our results.

Finally, we present conclusions, recommendations and future work that will continue this project.

Keywords: Twitter, NLP, early warning systems, data mining, Python.

CAPÍTULO I

INTRODUCCIÓN

1.1 El Problema de la Investigación

1.1.1 Planteamiento del Problema

En los espectáculos deportivos del Ecuador, concretamente en las exhibiciones de partidos de fútbol de la Serie A, se han presentado enfrentamientos o incidentes previo, durante y al finalizar los espectáculos, como son: “Los 38 supuestos hinchas de Barcelona que fueron procesados por los incidentes son investigados por el delito de rebelión, hechos ocurridos en los exteriores del Estadio de Rumiñahui” (Zapata, 2013) y según Puruncajas (2017) en su artículo en diario El Comercio: “Édison Lucero, hincha de El Nacional, fue apuñalado luego de una pelea ocurrida cerca del estadio Casa Blanca, de Liga de Quito, 13 personas fueron detenidas tras el incidente, dos supuestos hinchas del nacional fueron detenidos por generar disturbios en el ingreso al escenario el Batán”. Estos eventos representan un riesgo tanto para las personas que acuden a estos espectáculos públicos como para las personas que se encuentran a sus alrededores debido a que desconocen en donde se están desarrollando un disturbio o incidente, por lo que quedan expuestos a los mismos.

1.1.1.1 Diagnóstico del Problema

La violencia en los partidos de fútbol ecuatoriano se ha convertido en una gran problemática, lo que obliga a buscar medios eficaces para evitar los incidentes ocurridos.

A continuación, se exponen los acontecimientos más significativos en orden cronológico desde el año 2013 al 2016, recolectados desde los principales medios de comunicación escritos del país:

24 de agosto del 2013

El policía Segundo Lema salió herido del estadio Atahualpa tras ser empujado dentro de la fosa del escenario por supuestos hinchas de Liga de Quito, de la barra Muerte Blanca. Algunos de ellos también se enfrentaron con los hinchas del Deportivo Quito, tras el encuentro. (Puruncajas, 2017)

28 de septiembre de 2013

“Los 38 supuestos hinchas de Barcelona que fueron procesados por los incidentes son investigados por el delito de rebelión por los hechos ocurridos en los exteriores del Estadio de Rumiñahui” (Zapata, 2013).

7 de marzo del 2014

César Mayorga, hincha de Emelec, fue asesinado por supuestos seguidores de otro equipo mientras se transportaba en un bus de la cooperativa Panorama, en Guayaquil. Recibió impactos de piedra lanzados desde el exterior, que lo dejaron inconsciente. Posteriormente continuó siendo atacado, según la Policía. (Puruncajas, 2017)

20 de septiembre del 2015

“Édison Lucero, hincha de El Nacional, fue apuñalado luego de una pelea ocurrida cerca del estadio Casa Blanca, de Liga de Quito. 13 personas fueron detenidas tras el incidente” (Puruncajas, 2017).

6 de diciembre del 2015

“Aficionado de Liga de Quito fue herido con un cuchillo al final del primer tiempo entre el club albo y Universidad Católica, en el Olímpico Atahualpa. Según versiones de los allegados a la víctima, aficionado se negó a entregar dinero” (Puruncajas, 2017).

21 de septiembre del 2016

“Hinchas de Liga de Quito patearon y arrancaron sillas en el sector de palco, del estadio Monumental, de Guayaquil. Los aficionados destrozaron aproximadamente 100 sillas metálicas, del sector donde se los ubicó. Ocurrió en el cotejo ante Barcelona” (Puruncajas, 2017).

4 de diciembre de 2016

“Dos supuestos hinchas del nacional fueron detenidos por generar disturbios en el ingreso al escenario El Batán” (Puruncajas, 2017).

21 de enero del 2017

“En el estadio cantonal de Quero, Tungurahua, futbolistas e hinchas del Deportivo Quero agredieron a los árbitros. Ocurrió luego de su derrota 2-0 ante Liga juvenil, que se proclamó campeón” (Puruncajas, 2017).

5 de febrero del 2017

41 personas resultaron heridas luego de un entre hinchas, en el sector de la barra Sur Oscura, de Barcelona, en el estadio Monumental de Guayaquil. Por esto, a los 13 minutos, se suspendió el encuentro del ‘Ídolo’ ante Macará por falta de garantías. La directiva del cuadro torero anunció que no permitirá más el ingreso de la Sur Oscura a su escenario. (Puruncajas, 2017)

El campeonato nacional es uno de los eventos deportivos con mayor afluencia de personas las cuales apoyan a su equipo de preferencia y de igual manera disfrutan de un momento de

distracción. A continuación, se expone el número de espectadores que se han dado cita a un evento deportivo de fútbol desde el año 2014 al 2016 (ver tablas # 1, 2 y 3):

Tabla # 1. Número de espectadores del año 2014

N°	Equipos	Partidos Jugados	Espectadores
1	Barcelona	23	375.295
2	Emelec	22	317.981
3	El Nacional	21	193.916
4	Liga de Quito	21	152.304
5	Des. Cuenca	21	143.449
6	del. Quito	22	105.831
7	Liga de Loja	22	89.805
8	Olmedo	22	77.188
9	Univ. Católica	19	58.021
10	Mushuc Runa	22	58.197
11	Indo. José Terán	20	39.300
12	Manta	22	40.040
Total		257	1'651.327

Fuente: (Federación Ecuatoriana de Fútbol; Mifutbolecuador, 2014)

Tabla # 2. Número de espectadores del año 2015

N°	Equipos	Partidos Jugados	Espectadores
1	Liga de Quito	23	269.601
2	Barcelona	22	248.533
3	Emelec	23	222.332
4	El Nacional	21	179.342
5	del. Cuenca	22	170.988
6	Aucas	22	169.243
7	Dep. Quito	20	143.000
8	Liga de Loja	20	80.379
9	Univ. Católica	21	72.852
10	River Ecuador	22	65.618
11	Mushuc Runa	22	41.304
12	Indep. del Valle	22	37.786
Total		260	1'700.978

Fuente: (Federación Ecuatoriana de Fútbol; Mifutbolecuador, 2016)

Tabla # 3. Número de espectadores del año 2016

N°	Equipos	Partidos Jugados	Espectadores
1	Barcelona	22	437.733
2	Emelec	22	324.920
3	El Nacional	21	151.863
4	Dep. Cuenca	22	145.756
5	Liga de Quito	22	115.645
6	Aucas	22	115.279
7	Fuerza Amarilla	22	92.064
8	Delfín S.C	19	50.267
9	Mushuc Runa	21	51.387
10	Univ. Católica	22	53.307
11	River Plate	18	38.888
12	Indep. del Valle	22	25.319
	Total	255	1'602.428

Fuente: (Federación Ecuatoriana de Fútbol; Mifutbolecuador, 2016b)

Como se pudo observar en las tablas descritas anteriormente en los últimos 3 años ha existido un promedio de 1'651.577 personas, entre mujeres, niños y adultos que asisten a un encuentro de fútbol durante el desarrollo del campeonato nacional.

Según Cheing (2017) expresa que el campeonato Ecuatoriano de Fútbol 2017 está en la quincuagésima novena (59ª) edición cuyo nombre comercial es Copa Banco del Pacífico 2017. Este campeonato consta de tres etapas, la primera y segunda tienen un sistema de juego de todos contra todos mientras que en la tercera etapa se disputa el título de campeón entre los equipos ganadores de la primera y segunda etapa, en el caso de que el ganador de la primera y segunda etapa sea el mismo se proclamará «campeón nacional» directamente. El Campeonato Nacional de Fútbol Serie A 2017 consta de 12 equipos, los cuales se detallan en la tabla # 4, en esta tabla se ha visto necesario el detallar no sólo el nombre del equipo sino también ciudad, nombre del estadio y coordenadas del mismo ya que esta información es de gran relevancia para el desarrollo de este trabajo.

La primera etapa del campeonato inició el 29 de enero y culminó el 9 de julio, los encuentros que se disputaron en esta etapa se detallan en la figura # A1 y A2 la cual se encuentra en el anexo 1, la segunda etapa fue del 16 de julio al 10 de diciembre y los encuentros que se disputaron en esta etapa se detallan en la figura # A3 que se encuentra en el anexo 1.

Tabla # 4. Equipos de fútbol que participan en la Copa Banco del Pacífico 2017

Nombre del Club	Ciudad	Estadio	Coordenadas
Barcelona	Guayaquil	Monumental Banco Pichincha	-2.185856 -79.924876
Clan Juvenil	Sangolquí	General Rumiñahui	-0.329395 -78.445080
Delfín	Manta	Jocay	-0.964150 -80.702772
Deportivo Cuenca	Cuenca	Alejandro Serrano Aguilar Banco del Austro	-2.906551 -79.005951
El Nacional	Quito	Olímpico Atahualpa	-0.177486 -78.476623
Emelec	Guayaquil	Arena Banco del Pacífico	-2.206802 -79.893799
Fuerza Amarilla	Machala	9 de Mayo	-3.254750 -79.962302
Guayaquil City	Guayaquil	Christian Benítez Betancourt	-2.102915 -79.903766
Independiente del Valle	Sangolquí	General Rumiñahui	-0.328702 -78.444895
Liga de Quito	Quito	Rodrigo Paz Delgado	-0.107669 -78.489114
Macará	Ambato	Bellavista	-1.244608 -78.623240
Universidad Católica	Quito	Olímpico Atahualpa	-0.177486 -78.476623

Fuente:(Federación Ecuatoriana de Fútbol)

1.1.1.2 Pronóstico



Figura # 1. Tendencia de incidentes ocurridos desde el año 2013 hasta agosto de 2017

Fuente: (El Universo, 2017; Ministerio del Interior, 2012; Puruncajas, 2017; Zapata, 2013)

Tomando en consideración la figura # 1 se puede apreciar que durante los últimos 4 años ha existido un incremento de incidentes en eventos deportivos, lo que ya marca una tendencia crítica fluctuante como pronóstico de lo que sucederá si no se actúa sobre el problema. En el año 2017 las cifras son alarmantes, debido a que se presentan en conjunto tres tipos de delitos que involucran la inviolabilidad de la vida, la integridad personal y la vulneración del derecho a la propiedad e igualdad, que traen como resultado:

La pelea entre espectadores, lo que conlleva a la falta de asistencia a los estadios ya que ningún espectador va a un evento deportivo con la predisposición de ser agredido físicamente o que exista daños a su propiedad. De igual manera, se verán afectados los estadios de fútbol ya que si los incidentes son ocasionados en el escenario deportivo recibirán sanciones y suspensiones.

Para las personas que se encuentran siendo parte de los incidentes serán privadas de su libertad de acuerdo al tipo de delito que hayan cometido.

1.1.1.3 Control del Pronóstico

El modelo de seguridad que existe actualmente en los estadios de fútbol del Ecuador es: seguridad privada, resguardo policial y circuitos cerrados de televisión, a pesar de esto es

insuficiente ya que se han producido incidentes en los estadios. (Federación Ecuatoriana de Fútbol, 2009).

Esta última afirmación en conjunto con la falta de canales de comunicación que permitan notificar eficaz y eficientemente a los ciudadanos acerca de los peligros a los cuales pueden exponerse agravan aún más la situación.

Por lo que, considerando el incremento de incidentes en eventos deportivos Serie A, en los últimos años como se puede apreciar en la figura # 1, se plantea la creación de una aplicación innovadora que utilice tecnologías nuevas para el análisis de texto del contenido de redes sociales, lo cual promete ser un camino para mitigar el número de víctimas y daños. La aplicación busca aprovechar la gran influencia en la sociedad de las redes sociales como medio de información para proveer alertas tempranas de incidentes que ocurran dentro un escenario deportivo.

1.1.2 Formulación del Problema

La falta de conocimiento por parte de las personas, de la localización de un disturbio o incidente relacionado con un espectáculo de fútbol, ha generado que estas se vean involucradas en los mismos, provocándose daños materiales y contra su integridad.

1.1.3 Sistematización del Problema

- ¿Cómo se obtendrá la información relevante a disturbios en el escenario deportivo de la red social *Twitter*?
- ¿Cómo catalogar la información recolectada de *Twitter*?
- ¿De qué manera se puede analizar la información obtenida de *Twitter* acerca de incidentes de fútbol?
- ¿Cómo se evidenciará la información procesada a las personas sobre la existencia de incidente o disturbio en el escenario deportivo y sus alrededores?

1.1.4 Objetivo General

Desarrollar un módulo de alertas tempranas, basado en los datos obtenidos de la red social *Twitter*, con el objeto de evitar y disminuir actos de violencia generados en estadios de fútbol y sus alrededores.

1.1.5 Objetivos Específicos

- Extraer los datos de *Twitter* en tiempo real, para almacenarlos en una base de datos, mediante la utilización de un módulo de la misma red social.
- Clasificar los *tweets* por medio de un filtrado y contrastar su prioridad por geolocalización o palabras clave.
- Establecer un método que permita analizar *tweets* referentes a incidentes en eventos deportivos de fútbol, en base a un algoritmo para en análisis de *tweets* de eventos deportivos.
- Implementar un módulo de notificaciones, que les permita visualizar alertas sobre incidentes de una forma oportuna.

1.1.6 Justificación

Las personas en la actualidad utilizan las redes sociales para comunicarse con amigos, familiares y compartir lo que les sucede, de igual manera expresar sus inquietudes y sentimientos. Las redes sociales se han convertido en medios de información online en donde las personas buscan encontrar personas que tengan sus mismos intereses, lo que permite generar contactos para fines sociales y comerciales. (Tocci, 2016)

Las redes sociales constituyen uno de los medios de comunicación más utilizados, las cuales con frecuencia son instaladas en un teléfono inteligente o *Smartphone* debido a la facilidad de acceso y comodidad, por lo que es fundamental conocer el número de personas que tienen acceso a estas tecnologías.

En el Ecuador 4,2 millones de personas usaron redes sociales a través de su teléfono inteligente en 2016 (Acosta, 2017) y tomando en cuenta los datos desde el 2013 al 2016 ha existido un incremento anual del 7% es decir que cada año 1.170,045 de personas forman parte de una red social (Instituto Nacional de Estadística y Censos, 2017a).

Tabla # 5. Porcentaje de población con celular y redes sociales

Año	Población	Personas con celular activo	Personas que utilizan un Smartphone	Personas que utilizan redes Sociales
2013	15.872,755	7.453.781	1.261.944	1.081.620 (6,81%)
2014	16.148.648	7.820.597	1.928.108	1.722.159 (10,66%)
2015	16.404.531	8.174.520	3.084.886	2.807.282 (17,11%)
2016	16.714.929	8.482.236	4.484.087	4.224.984 (25,28%)

Fuente: (Instituto Nacional de Estadística y Censos, 2017)

Como podemos observar en la tabla # 5 el incremento de personas que utilizan redes sociales es de orden ascendente teniendo un crecimiento aproximado del 7% por año, es decir 1.170.045 personas.

Twitter actualmente permite escribir mensajes de 280 caracteres, en los cuales se puede incluir texto o archivos multimedia a estos mensajes se los llama *tweets* y podrán ser leídos por las personas que tengan acceso a esta plataforma. Esta forma de comunicación permite una interacción en tiempo real entre sus usuarios, lo cual proporciona hechos o sucesos del diario vivir en el menor tiempo posible, a diferencia de los medios de comunicación tradicionales como: Televisión, radio y prensa impresa. (Alvarado, 2016)

Según estadísticas recientes *Twitter* cuenta con más de 300 millones de usuarios mensuales activos, que publican cada día más de 600 millones de tweets (mensajes) a nivel mundial (Amirkhanyan & Meinel, 2016). “En el Ecuador se promedia que 2’000.000 de usuarios utilizan *Twitter*, de los cuales el 53 % en promedio utiliza la red social desde dispositivos móviles.” (Alcazar, 2015)

En cuanto al cuerpo legal que regula los actos violentos, en el Ecuador, “El Código Integral Penal es el cuerpo legal que limita o prohíbe comportamientos que pueden considerarse contrarios al orden jurídico, aquellos que atenten con la violación de derechos individuales o colectivos de la sociedad.” (Defensoría del Pueblo Ecuador, 2014)

Su aplicación rige a partir del 10 de Agosto del 2014, en la cual se reflejan todos los principios que emanan de la Constitución y de los instrumentos internacionales de derechos humanos, incorporando de manera más activa procedimientos como la conciliación.

Su finalidad es normar el poder punitivo del Estado, tipificar las infracciones penales, establecer el procedimiento para el juzgamiento de las personas con estricta observancia del debido proceso, promover la rehabilitación social de las personas sentenciadas y la reparación integral de las víctimas.

En su Capítulo Segundo, Sección Primera tipifica: DELITOS CONTRA LA INVIOLABILIDAD DE LA VIDA:

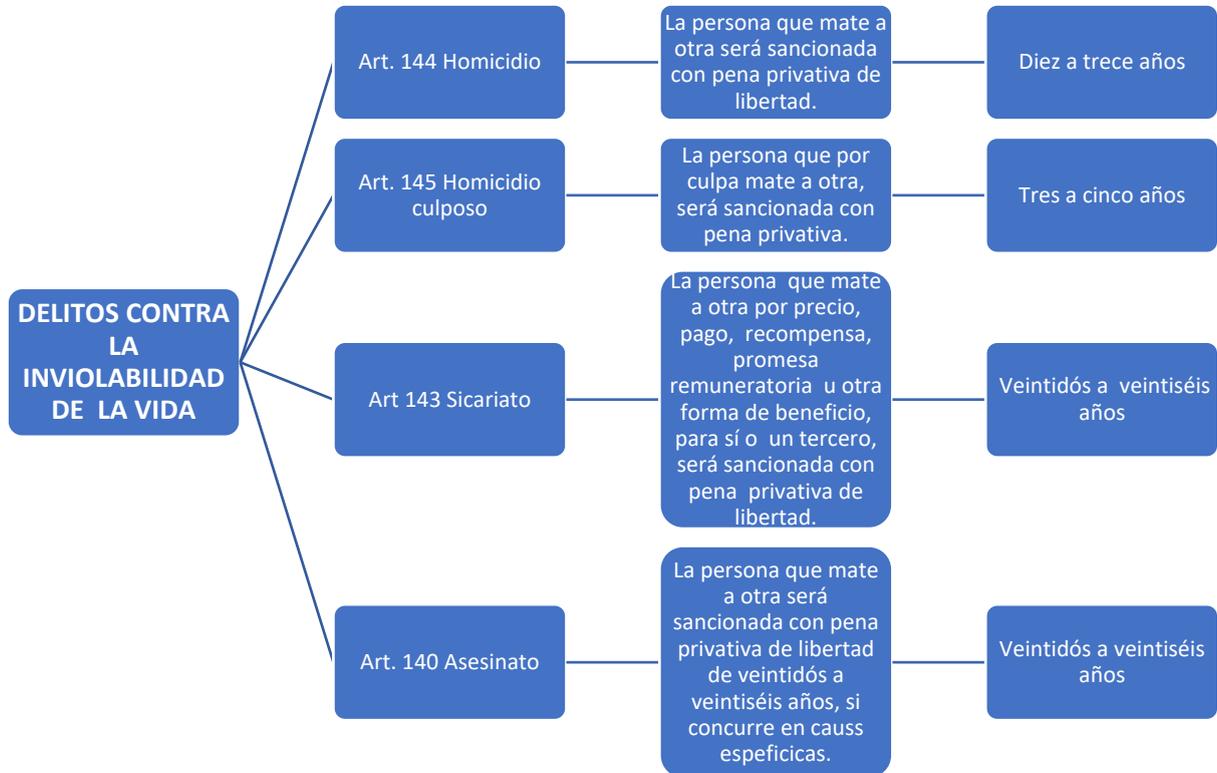


Figura # 2. Delitos contra la inviolabilidad de la vida

Fuente: (Ministerio de Justicia Derechos Humanos y Cultos, 2014)

En su Capítulo Segundo, Sección Sexta tipifica: DELITOS CONTRA EL DERECHO A LA PROPIEDAD:

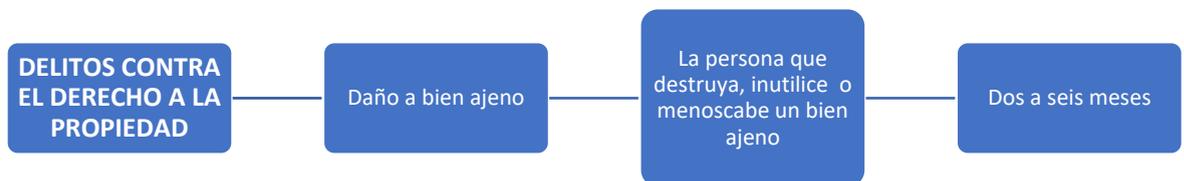


Figura # 3. Delitos contra el derecho a la propiedad

Fuente: (Ministerio de Justicia Derechos Humanos y Cultos, 2014)

En su Capítulo Segundo, Sección Segunda tipifica: DELITOS CONTRA LA INTEGRIDAD PERSONAL:

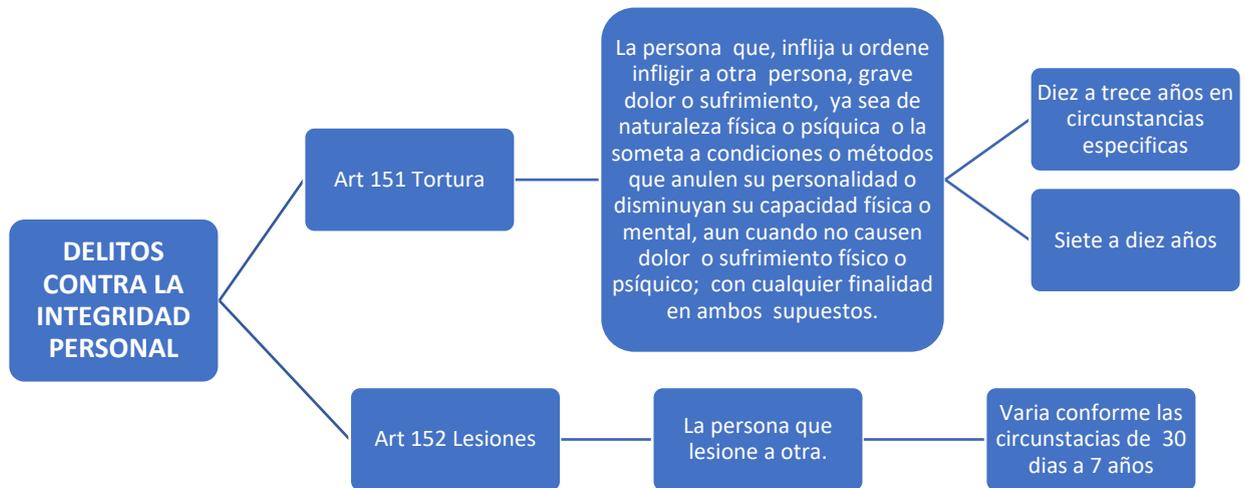


Figura # 4. Delitos contra la integridad de la persona

Fuente: (Ministerio de Justicia Derechos Humanos y Cultos, 2014)

En su Capítulo Segundo, Sección Quinta tipifica: DELITOS CONTRA EL DERECHO A LA IGUALDAD:

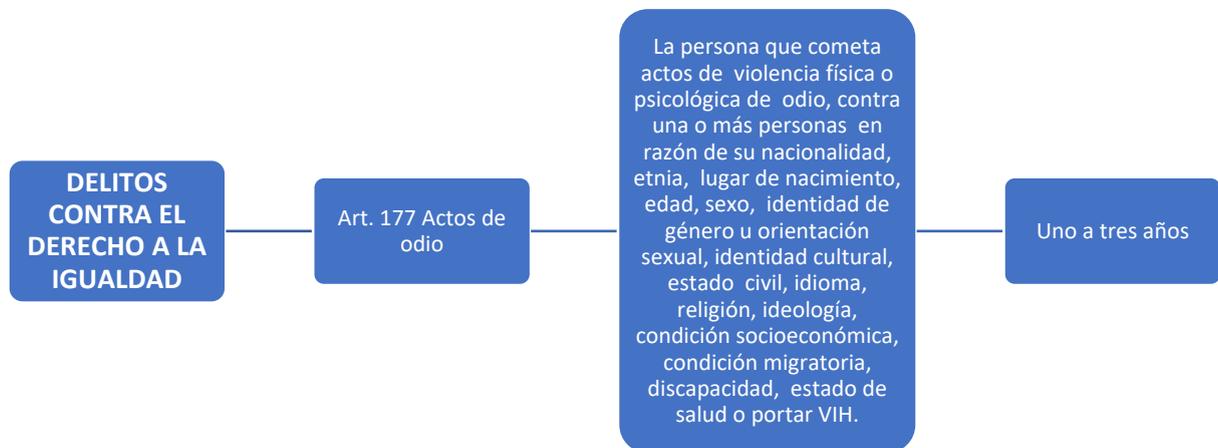


Figura # 5. Delitos contra el derecho a la igualdad

Fuente: (Ministerio de Justicia Derechos Humanos y Cultos, 2014)

A continuación, se presenta un resumen de los posibles delitos que se pueden cometer en un evento deportivo y que se encuentran tipificados y sancionados en nuestro código orgánico integral penal. (Ver tabla # 6)

Tabla # 6. Delitos del COIP

Clasificación de delitos COIP	Tipo de Delito	Categoría	Artículos
Contra la inviolabilidad de la vida	Asesinato	Todo lo que	Art. 140, 145.
Contra la integridad personal	Lesiones	corresponde a daños físicos de una persona	Art. 151, 152.
Contra el derecho a la propiedad privada	Daño al bien ajeno	Daños materiales	Art. 204.
Contra el derecho a la igualdad	Intimidación	Amenazas	Art. 154.
	Actos de odio		Art 177.

Fuente: (Ministerio de Justicia Derechos Humanos y Cultos, 2014)

Finalmente, el fútbol, en Ecuador, es uno de los eventos deportivos con mayor afluencia de espectadores, convirtiéndose en un espacio de entretenimiento para familias, jóvenes y adultos quienes asisten con la intención de apoyar a su equipo preferido y pasar un momento ameno.

“En el año 2016 la asistencia de aficionados a los partidos de fútbol del campeonato nacional “Serie A” fue de 1.602,428 personas, con 256 partidos jugados” (Federación Ecuatoriana de Fútbol; Mifutbolecuador, 2016b).

En el transcurso de los años 2013 al 2016 según los medios nacionales de comunicación impresos de mayor circulación han existido diferentes enfrentamientos entre aficionados, los cuales han dejado consecuencias que atentan contra la vida del ser humano. Derivando actos como lesiones, heridas, daños materiales, consecuencias legales tipificadas como contravenciones según el código integral penal e incluso la pérdida de vidas humanas.

Estos actos se convierten en un gran problema para la sociedad. En consecuencia, el pleno de la asamblea nacional constituyente en el año 2015 interviene realizando reformas esenciales en la Ley del Deporte, Educación Física y Recreación, promulgada en el año 2010 según el registro oficial; en la cual expone que al Estado le corresponde proteger, promover y coordinar el deporte y la actividad física como actividades para la formación integral del ser humano preservando principios de universalidad, igualdad, equidad, progresividad, interculturalidad, solidaridad y no discriminación.

Incorporando en este cuerpo legal en su TÍTULO XIII LA VIOLENCIA EN ESCENARIOS Y EVENTOS DEPORTIVOS en el cual de manera expresa en su Art. 156 manifiesta:

“El Ministerio Sectorial, tendrá la función de emitir los criterios técnicos, regulaciones, procesos de prevención y controles que se requieran para prevenir la violencia en escenarios y eventos deportivos, así como, las obligaciones de los propietarios de las instalaciones, organizaciones deportivas, dirigentes, deportistas, árbitros, autoridades, trabajadores de los escenarios y público asistente. Para lo cual contará con la asistencia técnica de la Policía Nacional.”(Ministerio del Deporte, 2010)

Acorde a la Ley expuesta en el inciso anterior, en su respectivo reglamento declara en el TÍTULO IV DE LA SEGURIDAD en su Art. 81 manifiesta:

“La seguridad en los estadios de fútbol, es requisito fundamental para la utilización del mismo; entendiéndose como tal lo señalado en Art. 82.- “El estadio será un lugar seguro para todos los usuarios, sean estos espectadores, protagonistas del partido, funcionarios oficiales, representantes de los medios de comunicación social, trabajadores u otros, incluyendo a los elementos de la seguridad ciudadana.” (Federación Ecuatoriana de Fútbol, 2009)

Por lo que, considerando el incremento de incidentes en los últimos 4 años, el incremento de usuarios en redes sociales en el país y las leyes nacionales que buscan asegurar los escenarios deportivos (y el cuerpo legal punitivo que regula hechos violentos) la creación de una aplicación que utilice tecnologías nuevas para el análisis de texto del contenido de redes sociales para producir reportes oportunos para los asistentes a estos escenarios es pertinente y necesaria.

1.2 Marco Teórico

Aprendizaje de Máquina

Según González (2014) el aprendizaje de máquina o *machine learning*, es la creación de sistemas de aprendizaje autónomos es decir que una máquina pueda ser capaz de identificar patrones completos dentro una gran cantidad de datos sin la intervención de un ser humano. La máquina aprende de un algoritmo y su sistema debe ir mejorando con el tiempo y debe ser capaz de predecir comportamientos futuros.

Aprendizaje de Máquina Supervisado

De acuerdo con Mccrea (2014) el aprendizaje supervisado, quiere conseguir una función de predicción avanzada, mediante el uso de algoritmos y de esta manera optimizar el proceso al momento de recibir una gran cantidad de datos, lo que conlleva a tener una mayor precisión al momento de dar una predicción.

Aprendizaje de Máquina No Supervisado

De acuerdo con Mccrea (2014) en su estudio “teoría de Aprendizaje de Máquina y sus Aplicaciones” manifiesta que el aprendizaje sin supervisión, consiste en encontrar relaciones entre una gran cantidad de datos. Se asignan datos y luego se busca patrones existentes entre los datos, en este aprendizaje no hay entrenamientos previos.

El Problema de Clasificación

Según Castillo (2016) menciona que existen limitantes en la clasificación de datos, lo cual ocasiona lentitud en los procesos y esto se debe a que su algoritmo elemental de entrenamiento crea una gran cantidad de vectores de soporte.

Support Vector Machine SVM (Máquina de soporte vectorial)

En la actualidad para el procesamiento de grandes cantidades de datos, las máquinas de soporte vectorial son una técnica efectiva. Según Carmona (2014) nos explica que las máquinas de soporte vectorial mapean puntos de entrada en un espacio de características, para luego hallar un hiperplano que los separe y maximice el margen entre cada una de las clases, la formulación matemática de las Máquinas de Vectores Soporte varía dependiendo de la naturaleza de los datos.

Un hiperplano puede ser escrito de la siguiente manera:

$$a_1 x_1 + a_2 x_2 + \dots + a_p x_p = b$$

Ecuación # 1. Fórmula de un hiperplano.

Maldonado y Weber (2012) indica que un hiperplano óptimo que separa un conjunto de datos, requiere encontrar el par (w, b) para clasificar correctamente los vectores, es decir para un espacio de hipótesis dado por un conjunto de funciones:

$$f_{w, b} = \text{signo}(w^t * x_i + b)$$

Ecuación # 2. Fórmula de un hiperplano sujeta a un conjunto.

se impone la siguiente restricción:

$$\underset{i = 1, \dots, m}{Min} |w^T * x_i + b| = 1$$

Procesamiento de Lenguaje Natural (PLN)

El procesamiento del lenguaje natural (PLN), es una de las tareas fundamentales de la inteligencia artificial. (Vásquez et al., 2009)

El PLN consiste en la manipulación del lenguaje natural utilizando equipos de computación ya que estos mediante lenguajes de programación permitirán la manipulación del lenguaje natural y así poder comprenderlo de mejor manera. El procesamiento del lenguaje natural permite desarrollar programas o modelos que ayuden a comprender y analizar los mecanismos humanos. (Vásquez et al., 2009)

Arquitectura de un sistema de PLN

La arquitectura de un sistema de PLN se sustenta en una definición del lenguaje natural por niveles, estos son: fonológico, morfológico, sintáctico, semántico, y pragmático.

- a) **Nivel Fonológico:** trata de cómo las palabras se relacionan con los sonidos que representan.
- b) **Nivel Morfológico:** trata de cómo las palabras se construyen a partir de unas unidades de significado más pequeñas llamadas morfemas.
- c) **Nivel Sintáctico:** trata de cómo las palabras pueden unirse para formar oraciones, fijando el papel estructural que cada palabra juega en la oración y que sintagmas son parte de otros sintagmas.
- d) **Nivel Semántico:** trata del significado de las palabras y de cómo los significados se unen para dar significado a una oración, también se refiere al significado independiente del contexto, es decir de la oración aislada.
- e) **Nivel Pragmático:** trata de cómo las palabras se usan en distintas situaciones y de cómo el uso afecta al significado de las oraciones. Se reconoce un subnivel recursivo y discursivo, que trata de cómo el significado de una oración se ve afectado por las oraciones inmediatamente anteriores.



Figura # 6. Arquitectura de un Sistema de Procesamiento de lenguaje Natural

Fuente: (Vásquez et al., 2009)

Mongo Db

MongoDB es una base de datos no relacional, esta característica permite que, en lugar de guardar los datos en registros, guarda los datos en documentos. (MongoDB, 2009)

Python

Python es un lenguaje de programación de alto nivel, preparado para realizar cualquier tipo de trabajo. Además, es un lenguaje interpretado, lo que significa que no se necesita compilar el código fuente para poder ejecutarlo lo que ofrece ventajas de programación imperativa, funcional, procedural y reflexiva. (Python Software Foundation, 2014)

Django

Django es un framework web de alto nivel escrito en Python que fomenta el desarrollo rápido y el diseño limpio y pragmático. (Django Software Foundation, 2015)

1.2.1 Marco Conceptual

Early Warning Systems (Sistemas de Alerta Temprana)

Estos se basan en el monitoreo y recolección de información sobre posibles peligros en un determinado lugar o zona, lo que permitirá una respuesta inmediata al peligro y así reducir riesgos. Existen sistemas de alerta temprana para peligros geofísicos y biológicos naturales, emergencias sociopolíticas complejas, riesgos industriales y riesgos para la salud personal, entre muchos otros. (Poolman, 2011)

- **Aplicación de los Sistemas de Alerta Temprana**

Estos sistemas se pueden incluir en donde se pronostique una amenaza a determinada población por lo tanto existen al menos dos diferentes formas de realizar alertas una es mediante la recolección manual de datos, y la segunda es mediante la aplicación de estrategias automáticas de recolección. (Vargas et al., 2016)

- **¿Objetivo de un SAT?**

“Es el pronóstico temprano de una posible amenaza a una población vulnerable, mediante la recolección de información la cual ayudará una mejor toma de decisiones” (Vargas et al., 2016).

Data Mining (Minería de datos)

Según Vilches y Broitman (2007) explica que la minería de datos consiste en la extracción de información desconocida o ignorada, la cual puede ser potencialmente útil, si bien es importante descubrir nueva información de un conjunto de datos, también es muy importante tener claro cuál es el significado de lo que se está buscando para poder interpretar de una manera más eficiente la información.

▪ **Aplicaciones de la minería de datos**

La minería de datos se puede aplicar en todas las áreas o actividades humanas que generen datos como por ejemplo: (Riquelme et al., 2006)

1. Comercio y banca
2. Medicina y farmacia
3. Seguridad y detección de fraudes
4. Astronomía
5. Ciencias sociales

▪ **Técnicas utilizadas para la minería de datos**

La elección de la técnica viene determinada por dos condiciones, el tipo de dato y el objetivo que se desea obtener, a continuación, se presentara una pequeña lista de las técnicas que se pueden utilizar: (Aluja, 2001)

1. Técnica de «*clustering*». – parten de una medida de proximidad entre individuos para luego buscar grupos de individuos parecidos entre sí.
2. Redes bayesianas. - permiten establecer relaciones casuales y efectuar predicciones para esto se debe partir del conocimiento de un experto o puede ser inferido a partir de los datos recolectados
3. Árboles de decisión. - parten de datos históricos y permiten obtener de forma visual las reglas de decisión bajo las cuales opera un consumidor.
4. Redes neuronales. – radica en el aprendizaje secuencial, basándose en los datos de entrada para luego realizar una predicción. Esta técnica es no lineal

Red Social

Son comunidades virtuales cuya finalidad es permitir a los usuarios relacionarse, comunicarse, compartir contenido e intereses comunes con muchos más usuarios, el formar parte de una red social nos la apertura de reencuentros con antiguas amistades o crear nuevas relaciones. (Urueña et al., 2011)

Existen varias maneras de clasificar las redes sociales, de acuerdo a su público objetivo, sujeto principal de la relación o según su plataforma: (Enciclopedia de Característica, 2017)

- **Público Objetivo**

Horizontales. - se dirige a todo el público de Internet.

Verticales. – Se basan en un eje temático, es por ello que se crean comunidades dentro de las redes sociales, son enfocadas a un tema específico.

- **Sujeto Principal**

Humanas. – relaciones entre usuarios, según sus gustos, perfil social o aficiones.

De Contenido. – la relación se da acorde a la información compartida por los usuarios.

De interés. - permiten realizar marketing de objetos lugares o marcas.

- **Según su Plataforma**

Web. – Redes sociales basadas en la estructura tradicional de internet.

Mataversos. – Redes sociales que exigen al usuario una conexión tipo cliente- servidor.

1.2.2 Estado del Arte

Según Coronel (2011), en su estudio titulado “Redes Sociales Marcan Tendencias en la Comunicación: Ecuador”, *Twitter* en Ecuador se convirtió en un medio de difusión en línea de gran importancia al momento de brindar información ya que se conoce de manera inmediata los hechos que están ocurriendo en un determinado lugar, lo que no hacen los medios masivos de información como lo es la radio y televisión. Los comunicadores sociales que mantenían sus cuentas en esta plataforma se están convirtiendo en fuentes de información para las personas que pasan tras una pantalla de computador o dispositivo móvil.

No obstante, *Twitter* es un instrumento mediante los cuales se informa y se comunica de forma masiva; a los miembros de una sociedad o de una comunidad, hoy en día se ha convertido una de las plataformas de comunicación más utilizadas a nivel mundial, por su simplicidad y acceso rápido. (Rodríguez & Ugalde, 2012)

Adicionalmente, según Amirkhanyan y Meinel (2016) en su publicación titulada, “Analysis of the Value of Public Geotagged Data from Twitter from the Perspective of Providing Situational Awareness” indica que *Twitter* suministra una interfaz de programación de aplicaciones (API) para desarrolladores, la cual permite obtener sus datos públicos. La API es un conjunto de métodos de comunicación claramente definidos entre varios componentes de software y estos permiten la extracción de *tweets* mediante criterios ya establecidos, los cuales pueden ser palabras clave, nombre, lugares, nombre de lugares, etc. En su investigación el principal objetivo que se plantearon fue el análisis de los datos geoestadísticos sociales públicos en tiempo real y se enfrentaron a grandes desafíos como agrupación de *tweets* en tiempo real, visualización y filtrado de los mismos, además, nos mencionan herramientas de análisis sociales como TweetDeck, Twitonomy, Hootsuite, Geofeedia, etc.; Las cuales son de pago y proporcionan una poderosa funcionalidad de análisis y visualización de datos.

Vargas, Tovar y Villanueva (2016) en su estudio “The SAT (Early Warning Systems)” explica que los Sistemas de Alerta Temprana (SAT) permiten anticipar posibles amenazas hacia una o más poblaciones, por medio del monitoreo de una o varias señales físicas, cabe mencionar que el objetivo primordial de los SAT es la prevenir la pérdida de vidas humanas y de bienes materiales a través de la recolección de información.

Por otro lado según Adan et al. (2012) en su estudio “Social Media Alert and Response to Threats to Citizens (SMART-C)” indica que actualmente las redes sociales son un mecanismo de información y recomienda que se la debe aprovechar para casos de emergencia. Por lo cual en su estudio nos presenta un sistema que tiene como objetivo el permitir una comunicación bidireccional entre una emergencia y el público en general mediante dispositivos y modalidades que incluyen teléfonos celulares, mensajes MMS, mensajes de texto, blogs, *Twitter*, Facebook, etc. Adicional a esto, nos manifiesta el mayor desafío de su investigación, el cual está relacionado con los mecanismos para modelar la información semántica a nivel de un evento, la veracidad de la información y limpieza de datos.

Ahora bien Asensio (2015) en su estudio aplicación de técnicas de minería de datos en redes sociales/web”, indica que “si se realiza un análisis correcto de la información obtenida de las redes sociales, se puede obtener una valiosa información la cual permitirá la apertura de cualquier investigación. Adicional a esto nos menciona que *Twitter* proporciona su propia API (Application Programming Interface) oficial. La cual permite la comunicación entre diferentes componentes de software, permitiendo controlar una cuenta y la recuperación de información.

Existen dos maneras de extraer información desde *Twitter*: la API Rest y la Streaming API.

- **API Rest.** Permite realizar todas las acciones a las que tenemos acceso desde la página web o las aplicaciones. Proporciona acceso a la información ya existente en *Twitter* en el momento de hacer la llamada. Accedemos a los datos por un sistema en forma de caja negra realizando peticiones GET y POST. (Asensio, 2015)
- **Streaming API.** Recibiremos información creada posteriormente a la petición de datos. Se inicia con *Twitter* abriendo una conexión entre su servidor y nuestro sistema, y enviará por ella *tweets* que sean publicados a partir de ese momento, siempre que cumplan los filtros indicados al inicio de la conexión hasta su cierre. Es decir, es un proceso en tiempo real. (Asensio, 2015)

De igual manera, según Anguita y Lorenzo (2014) en su estudio “Extracción, Análisis y Visualización de Información Social desde *Twitter*”, indica que *Twitter* se caracteriza por tener la gran mayoría de su contenido accesible de forma pública, e incluso provee diferentes métodos para acceder a él.

Al momento de tener una respuesta de *Twitter* se recibe un bloque prefijado de información. Este bloque siempre contendrá los mismos registros de información. A continuación, se analiza la información disponible en los modelos definidos por *Twitter*.

Información contenida en las respuestas de tipo usuario

- Fecha de creación de la cuenta.
- Descripción provista por el propio usuario.
- Recuento de *tweets* marcados como favoritos por el usuario desde el momento de creación de la cuenta.
- Número de personas que siguen al usuario.
- Número de personas a las que sigue el usuario.
- Flag que determina si el usuario ha activado la opción de mostrar su ubicación al resto de la red.
- Idioma marcado como predeterminado por el usuario.
- Número de listas públicas de las que el usuario es miembro.
- Localización física establecida por el usuario en formato texto.
- Flag que determina si ésta cuenta está marcada como privada por su dueño.
- Número de *tweets* publicados por el usuario.
- Zona horaria de la cuenta (configuración elegida por el usuario).
- Url dada por el usuario relacionada con su perfil.
- **Información contenida en las respuestas de tipo *tweet***

- Texto del *tweet*.
- Fecha de creación del *tweet*.
- Coordenadas geográficas del *tweet*, indicadas mediante latitud y longitud.
- Identificador del usuario que ha escrito el *tweet*, que es único en la red.
- Número de veces que este *tweet* ha sido marcado como favorito.
- Flag que determina si el *tweet* contiene algún enlace y/o contenido multimedia.
- Idioma en el que está escrito el *tweet*, según el procesamiento propio de *Twitter*.
- Identificador de lugar con al que el *tweet* está asociado. Esta información podría no estar presente y no corresponde necesariamente con el lugar donde se publicó.
- Entidades determinadas por *Twitter* que están presentes en el contenido del *tweet*.

Según la publicación de Comercio y Justicia (2017) nos menciona sobre el uso de las redes sociales para prevenir el fraude en los seguros, ya que en los últimos años ha existido un incremento de fraudes, circunstancia que ha puesto en alerta a las compañías y es por ello que han creado departamentos antifraudes de gran sofisticación, dichos departamentos utilizan herramientas tecnológicas, medios de comunicación digital y social para el cruce de datos y analizar los posibles fraudes.

De acuerdo con Sakaki et al. (2010), usaron a los usuario de *Twitter* como sensores sociales para detectar terremotos. Ellos investigaron la interacción de un evento real, como cuando ocurre un terremoto las personas hace muchas publicaciones en *Twitter* relacionadas con el terremoto, lo que permite detectar la ocurrencia de este siniestro rápidamente. Para esto diseñaron un clasificador de tweets basado en palabras clave en un tweet, el número de palabras y su contexto.

En base a la documentación mencionada anteriormente se puede evidenciar la importancia de la información que se genera en las redes sociales, dicha información al ser procesada de manera adecuada puede proporcionar resultados oportunos al momento de que existan percances, como nos muestra Sakaki et al. (2010) y Amirkhanyan et al. (2016) en sus publicaciones.

Además, en la investigación realizada se puede apreciar la importancia que tiene en la actualidad la red social *Twitter* como medio de información; y, algunos pasos que se deben seguir para extraer información desde la misma. Así mismo, nos manifiestan los posibles desafíos que se pueden encontrar al analizar, procesar y almacenar datos de *Twitter*.

CAPÍTULO II

MÉTODO

2.1 Análisis

En esta sección se describe el análisis, diseño e implementación del prototipo desarrollado para la generación de alertas tempranas utilizando minería de texto en redes sociales.

Para prevenir altercados en estadios de fútbol. Tomando en cuenta que el caso de estudio para este módulo es el campeonato de fútbol del Ecuador, la sección de análisis se agregó información sobre la tendencia de uso de redes sociales en el país. Las palabras clave que se utilizan para recopilar *tweets* desde *Twitter* y la estructura de cada *tweet*. Para el diseño de este módulo se ha tomado en cuenta una arquitectura de tres capas debido a que se lleva un desarrollo por niveles y en caso que exista algún error o cambio, solo será necesario cambiar el nivel en cuestión sin comprometer funcionamiento del resto del sistema.

2.1.1 Tendencias de uso de Smartphone y Redes Sociales en Ecuador

Es de gran importancia conocer el crecimiento del uso de redes sociales y *smartphones* en el Ecuador debido a que esto permite establecer un número aproximado de usuarios que actualmente usan *Twitter* y los posibles usuarios a futuro. De acuerdo al Instituto Nacional de Estadística y Censos desde el año 2013 al 2016 el número de personas que utilizan *smartphones* ha crecido en un 7%, dando un valor anual de 1'170.045 personas, las cuales se vinculan a una red social, como se puede apreciar en la figura # 7.

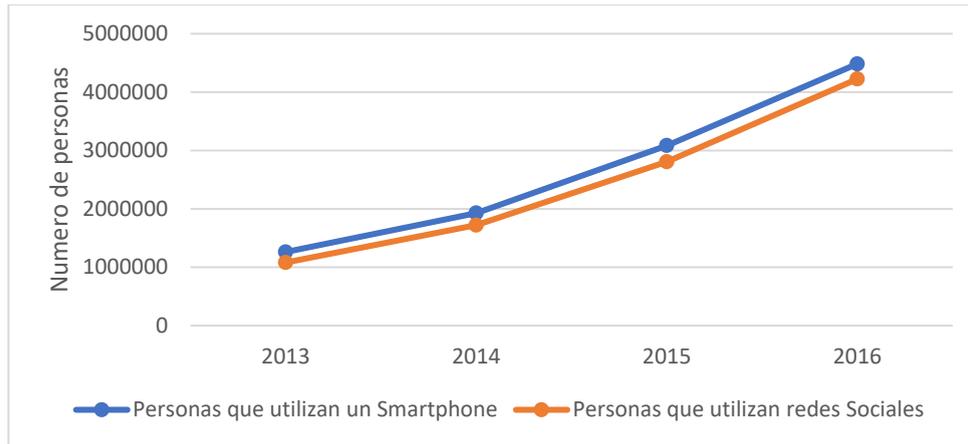


Figura # 7. Tendencia de personas que hacen uso de smartphone y redes sociales en Ecuador

Fuente: (Instituto Nacional de Estadística y Censos, 2017)

Además, se realizó un análisis de asistencia de espectadores a eventos deportivos de fútbol de la Serie A campeonato nacional Banco del Pacífico, durante los años 2014, 2015 y 2016, teniendo como resultado, que existió un ausentismo de personas a los estadios en el año 2016, esto con relación a los años 2014 y 2015. Como se puede apreciar en la figura # 8, el ausentismo coincide con los incidentes suscitados en el año 2015, y que pueden relacionarse como consecuencia de ellos, los cuales se pueden observar de manera más detallada en la figura # 1.



Figura # 8. Tendencia de asistencia de espectadores en un partido de fútbol Serie A

Fuente:(Federación Ecuatoriana de Fútbol)

2.1.2 Palabras Clave de los Equipos que conforman el campeonato nacional ecuatoriano de fútbol Serie A 2017

Se realizó la búsqueda de las palabras claves o key words las cuales son utilizadas por los hinchas para identificar a cada equipo en las redes sociales, específicamente *Twitter*. Este detalle de palabras clave se las puede apreciar en la tabla # 7, posteriormente estas palabras clave nos permitirán realizar la extracción de los tweets que las contengan.

Tabla # 7. Palabras clave de los equipos de fútbol del campeonato ecuatoriano

Equipo	Palabras clave
Barcelona	u'BarcelonaSCweb', u'BarcelonaSC', u'BSC', u'SociosBSC', u'Idolo', u'vamosidolo', u'Vamosidolo', u'MonumentalBP', u'ALoBarcelona', u'UnsoloIdolo', u'FamiliaAmarilla', u'alobarcelona', u'Toreros', u'monumental', u'FiestaMonumental', u'zona15norte', u'ZonaNorte', u'prensabsc', u'BarcelonaBSC', u'SurOscuraGye', u'SurOscura', u'SurOscura'
Clan Juvenil	u'Clan_juvenil', u'ClanJuvenil', u'EsteEsMiClan', u'YYoSoyDelClan', u'VamosClan', u'ClanJuvenilTV', u'ClanJuvenilCD'
Delfín	u'DelfinSC', u'El2017seráCetáceo', u'cetaceo'
Deportivo Cuenca	u'DCuencaOficial', u'Dcuenca', u'HinchaRojo', u'VamosRojos', u'dCuenca'
El Nacional	u'elnacionalec', u'VamosBiTri', u'marearaja', u'SociosNacional', u'LosNachosEC', u'EINacionalBitri', u'Elnacho'
Emelec	u'Emelec', u'CSEmelec', u'VamosEmelec', u'Bombillo', u'emelec', u'emelexistatv', u'azul', u'Hincha_Azul', u'emelec_max', u'losemelexista', u'BocaDelPozoBDP', u'Prensa_Emelec', u'Mundo_Emelec', u'Web_Emelec', u'Hinchade_Emelec', u'BDP_emelec', u'vamosazules', u'electricos'
Fuerza Amarilla	u'FuerzaAmarillaSC', u'FuerzaAmarilla', u'VamosAurinegros', u'FuerzaAmarillaEsLaProvincia', u'TodoSomosFuerza', u'TodosConFuerza', u'fuerza_amarilla'
Guayaquil City	u'GuayaquilCityFC', u'ElEquipoDeLaCiudad', u'GYECITYFC', u'GuayaquilCityFC'
Independiente del valle	u'Independiente Valle', u'Independiente del Valle', u'IndependientedelValle', u'YoSoyIDV', u'Independiente', u'IDV_EC', u'IDV'
Liga de Quito	u'LDU_Oficial', u'LDU', u'EstadioRodrigoPaz', u'YoVeoLigaTV', u'LIGATV', u'SiempreJuntoALIGA', u'muerteblanca', u'LDU2017', u'somos_liga', u'DINOSAURIOSLDU', u'ACTITUDALBA', u'365LDU', u'albo75', u'TodoPorLDU', u'LDUENG', u'vamoligacarajo', u'Reydecopas', u'SuperHinchaLDU'
Macará	u'Macara_Oficial', u'Macará', u'Macara', u'YoCreoEnTiMacará', u'ClubMacara', u'UltrasMacara', u'UltrasMacara', u'somoslosultras'
Universidad Católica	u'UCatolicaEC', u'UC', u'blogChatolei', u'chatolei', u'UCatólicaEc', u'UCatolicaEc'

2.1.3 Datos de *Twitter* que pueden ser Extraídos

Para realizar la extracción de datos de *Twitter* es necesario conocer la estructura de los datos que se pueden obtener. Los campos que contiene un *tweet* se muestran en la tabla # 8.

Tabla # 8. Atributos que contiene un tweet

Atributos que contiene un <i>tweet</i>		
Atributos del Tweet	Atributos del Usuario (User)	Atributos del evento
id	id	hashtag
coordinates	contributors_enabled	
created_at	created_at	
favorite_count	default_profile	
entities	default_profile_image	
user_mentions	description	
in_reply_to_screen_name	entities	
in_reply_to_status_id	favouriytes_count	
in_reply_to_user_id	follwers_count	
lang	friends_count	
place	geo_enabled	
retweet_count	lang	
text	listed_count	
	location	
	name	
	profile_background_color	
	profile_background_image_url	
	profile_image_url	
	profile_link_color	
	profile_use_background_image	
	protected	
	screem_name	
	statuses_count	
	time_zone	
	url	
	utc_offset	
	verified	

Fuente: (Caballero, 2012)

Cotelo et.al. (2015) muestran en la figura # 9 los objetos más destacables que se pueden extraer de *Twitter* y de igual manera las posibles relaciones que pueden existir entre cada uno de estos objetos.

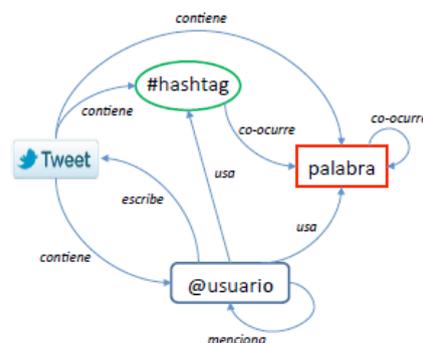


Figura # 9. Posibles relaciones de *Twitter*

Fuente: (Cotelo et al., 2015)

2.2 Diseño del Sistema Propuesto

2.2.1 Esquema de la Arquitectura

La arquitectura planteada para el desarrollo del módulo de generación de alertas tempranas se la puede apreciar en la figura # 10, la cual está compuesta por una solución de 3 capas: la primera capa de presentación, seguida por la capa de negocios y finalmente, la capa de datos.

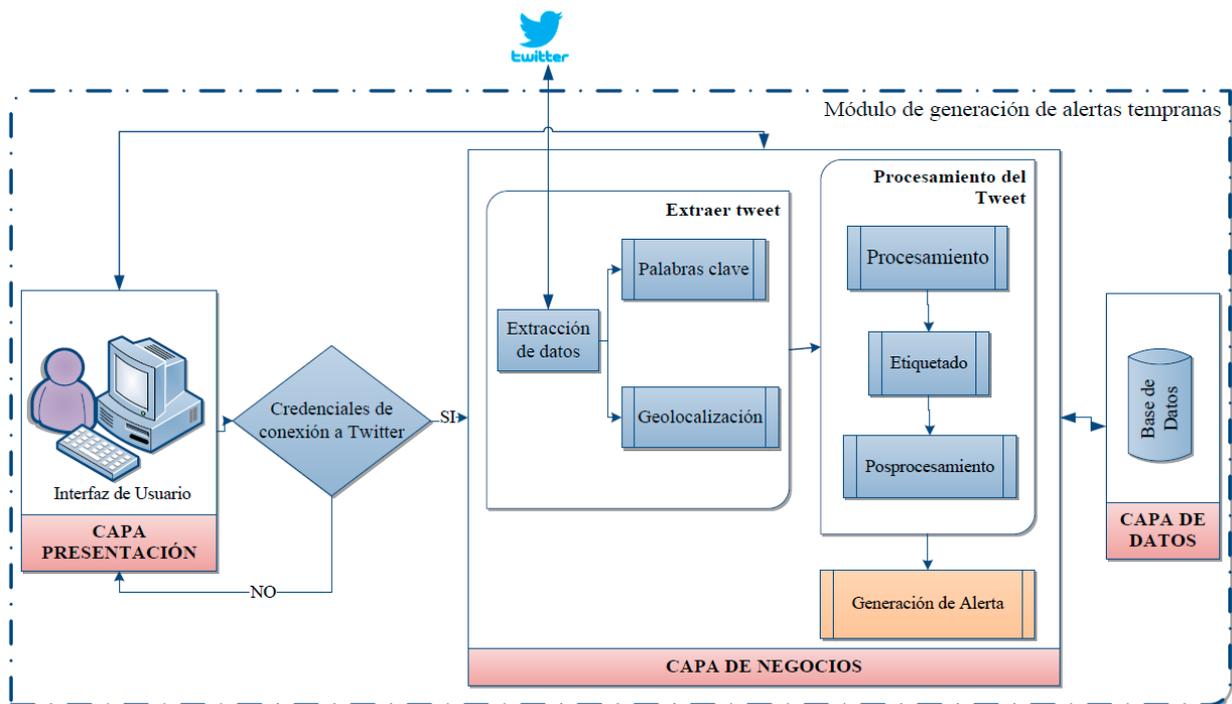


Figura # 10. Arquitectura para el desarrollo del módulo de generación de alertas tempranas

2.2.2 Descripción del Funcionamiento del Módulo de Alertas Tempranas

El funcionamiento del módulo de alertas tempranas puede sintetizarse de la siguiente manera:

1. En primera instancia la capa de presentación proporciona una interfaz visual para el usuario, en la cual se seleccionará los equipos y tiempo de monitoreo del encuentro deportivo, además, se podrán visualizar los *tweets* extraídas.
2. La capa de negocios es la encargada de conectarse a *Twitter* por medio de su *Streaming API*. Para realizar esta conexión, se debe cumplir con un protocolo de autenticación OAuth (*Open Authorization*), lo que proporciona acceso/extracción de *tweets* en tiempo real. La extracción se realizará de dos formas: palabras claves (Keywords) y geolocalización (geolocation).

Previo al procesamiento de los *tweets* se realiza el entrenamiento del algoritmo de clasificación. Para lo cual se utilizaron 52,300 *tweets* referentes a política, salud, deportes y educación. El procesamiento de los *tweets* se lo realiza mediante PLN (Procesamiento de Lenguaje Natural).

3. Por último, la capa de datos se encargará de realizar el almacenamiento, modificación, consulta y borrado de los datos.

A continuación, se describe de manera más detallada la arquitectura y el funcionamiento del módulo de alertas tempranas.

2.2.2.1 Registro en la Plataforma para Desarrolladores de *Twitter*

El registro en la plataforma para desarrolladores de *Twitter*¹, este es el primer paso que se debe realizar, para que el usuario generar una serie de claves las cuales sirven para autenticarse e ingresar a *Twitter*.

Las claves que obtenemos de la plataforma son cuatro: Consumer Key (API Key), Consumer Secret (API Secret), Access Token y Access Token Secret.

Con las llaves (*Keys*), la autenticación se la realiza de forma segura mediante el protocolo OAuth (Open Authorization), permitiendo al usuario mantener una conexión con *Twitter* sin compartir toda su información.

¹ <https://apps.twitter.com>

Para registrarse en la plataforma de desarrolladores es un requisito indispensable ser parte de la comunidad de *Twitter*. Un dato importante es que se debe tener un número de teléfono celular asociado a la cuenta de *Twitter*, debido a temas de seguridad de la aplicación. Además, la plataforma solicita el nombre de la aplicación, descripción y URL de acceso.

2.2.3 Capa de Presentación

La capa de presentación está conformada por una interfaz gráfica la cual es responsable de la interacción del sistema con el usuario y viceversa. Permite la presentación de datos y recibe las peticiones del usuario como se muestra en la figura # 11. Esta capa se comunica exclusivamente con la capa de negocio.

Inicio / INICIO

Equipo Local
Fuerza Amarilla

Equipo Visitante
Universidad Catolica

Hora de Inicio
11 : 15

Tiempo de Monitoreo (Horas):
1

Empezar

Zona Segura!

Alerta!!!

Zona de Peligro!!!!

Tweets Recolectados

Show 10 entries

Search:

Usuario	Texto del Tweet
No matching records found	

Showing 0 to 0 of 0 entries (filtered from NaN total entries)

Previous 1 2 3 4 5 ... Next

Figura # 11. Interfaz gráfica del usuario

2.2.4 Capa de Negocio

En esta capa se establecen los requerimientos solicitados por el usuario, análisis de los *tweets* y presentación de resultados como se puede observar en la figura # 12.

Esta se comunica con la capa de presentación para mostrar resultados y con la de datos para solicitar a los gestores el almacenamiento o recuperación de información.

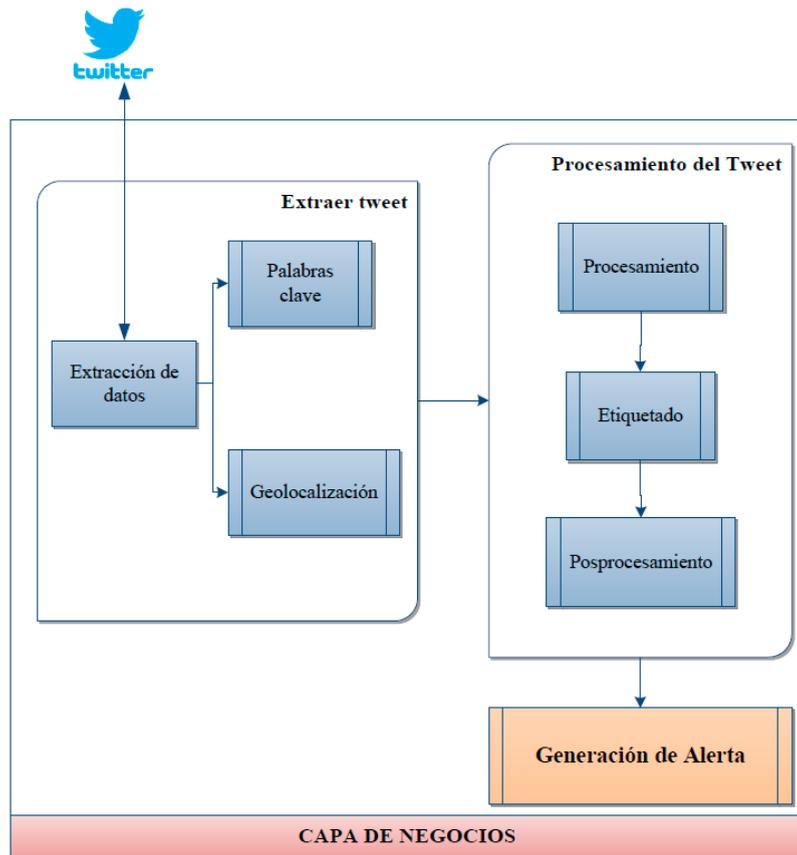


Figura # 12. Capa de negocios

2.2.4.1 Extracción de Datos

Realizado el registro en la plataforma de desarrolladores y adquiriendo las llaves (Keys) de acceso, se puede realizar la extracción de datos desde *Twitter*. Se debe planificar los datos que se quiere extraer de *Twitter*, esta extracción puede ser por palabras claves o geolocalización como se muestra en la figura # 13.

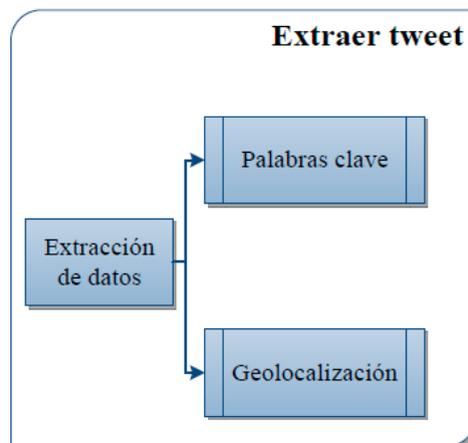


Figura # 13. Proceso para la extracción de tweets

2.2.4.2 Criterios de Extracción de Datos

Extracción mediante palabras claves

Teniendo en consideración que no todos los *tweets* harán referencia al Campeonato Ecuatoriano de fútbol serie A, se debe realizar una búsqueda y extracción mediante palabras clave las cuales son vinculadas a cada club de fútbol que es parte del Campeonato.

- **Inconvenientes**

Al realizar una extracción por palabras clave, estas serán extraídas desde cualquier parte del mundo y no desde un sitio específico (en este caso Ecuador). Lo que ocasiona extraer datos que resultan inservibles para el buen funcionamiento del módulo de alertas tempranas.

- **Solución**

Para evitar el inconveniente anterior, aplicamos filtros correspondientes a: los atributos del *tweet* como nombre del país, nombre de la ciudad y el idioma en que fue escrito; adicional a esto, se agregan dos filtros a los atributos de usuario como son localización y zona horaria.

Aplicando estos cinco filtros se mejora la extracción de los datos para asegurar que sean *tweets* propiamente de Ecuador correspondientes a los equipos que forman parte del Campeonato Ecuatoriano de Fútbol.

Extracción mediante geolocalización

La extracción de geolocalización se la tomará a un kilómetro del estadio de fútbol tomando como punto central el centro de dicho estadio. Para esto utilizamos la herramienta Bounding box, la cual nos permite obtener el punto central y nos da un cuadrante de coordenadas como se muestra en la figura # 14.

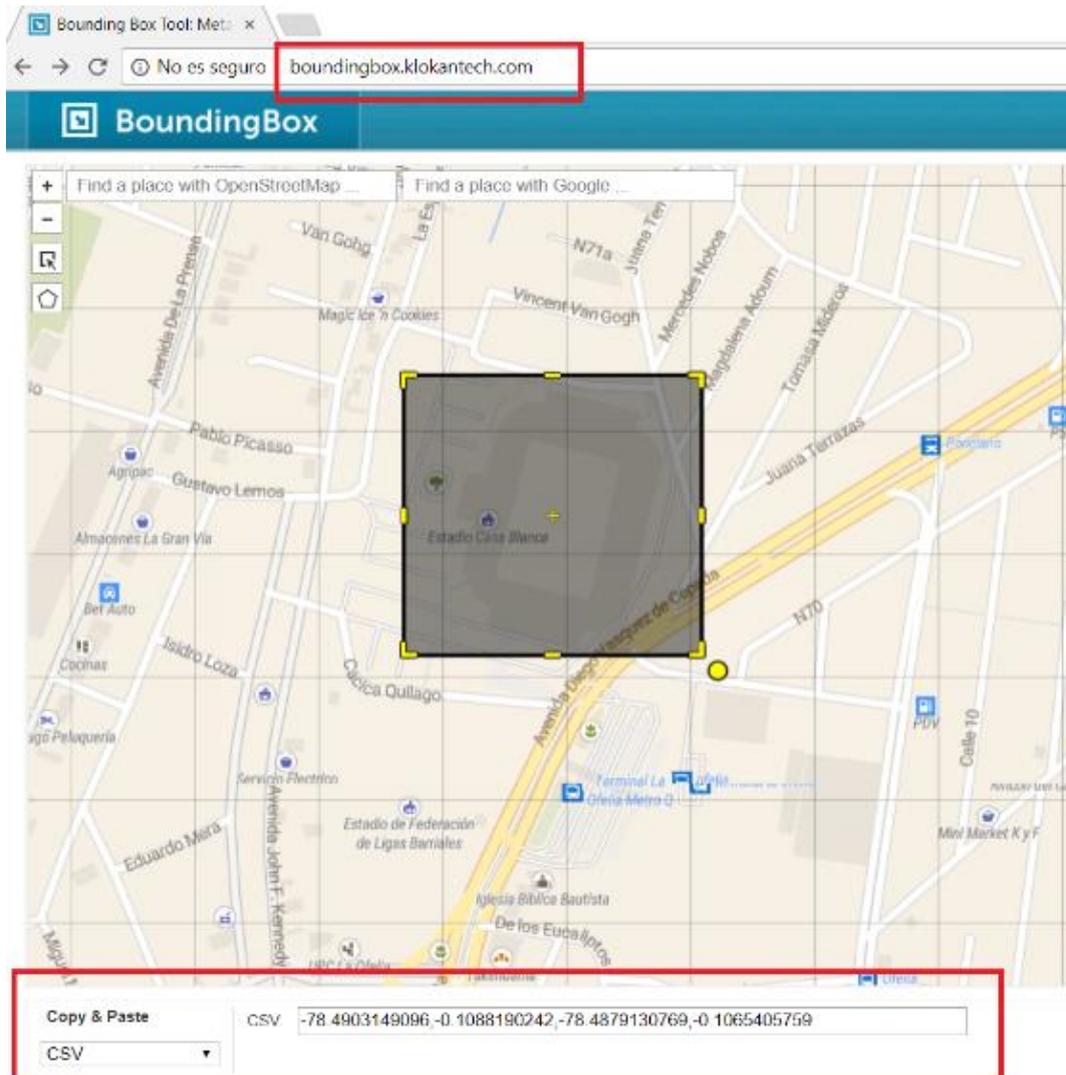


Figura # 14. Coordenadas del estadio de fútbol Rodrigo Paz Delgado de la ciudad de Quito

Fuente: (Klokant Technologies, 2010)

- **Inconvenientes**

El obtener las coordenadas de un cuadrante, como se muestra en la figura # 14, no garantiza que los datos de *Twitter* provengan únicamente de esas coordenadas, ya que al momento de realizar pruebas se obtuvieron datos fuera de estas. Esto se debe a que *Twitter* tiene pre-establecidos cuadrantes a los cuales no se los puede hacer más pequeños como se requiere en este estudio.

- **Solución**

Para evitar el inconveniente descrito anteriormente, se aplicó la fórmula de Haversine la cual permite filtrar por distancia cada *tweet*.

Según Chopde y Nichat (2013) la fórmula de Haversine nos sirve para calcular la distancia entre 2 puntos en la tierra, superficie especificada en longitud y latitud (ecuación # 3).

$$d = 2 r \sin^{-1} \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right)$$

Ecuación # 3. Fórmula de Haversine.

Fuente: (Chopde & Nichat, 2013)

d = es la distancia entre 2 puntos

ϕ_2, ϕ_1 = hace referencia a la latitud

λ_2, λ_1 = hace referencia a la longitud

r = Corresponde al radio de la tierra (6373 km)

2.2.4.3 Procesamiento de Tweets

El tipo de escritura utilizada en *Twitter* representa un desafío grande al momento de su procesamiento, debido a que se trata de una escritura natural, con rasgos de oralidad y en su mayoría con muchos modismos coloquiales, por lo que se debe procesar cada uno de los *tweets* extraídos. Este proceso se lo representa en la figura # 15.

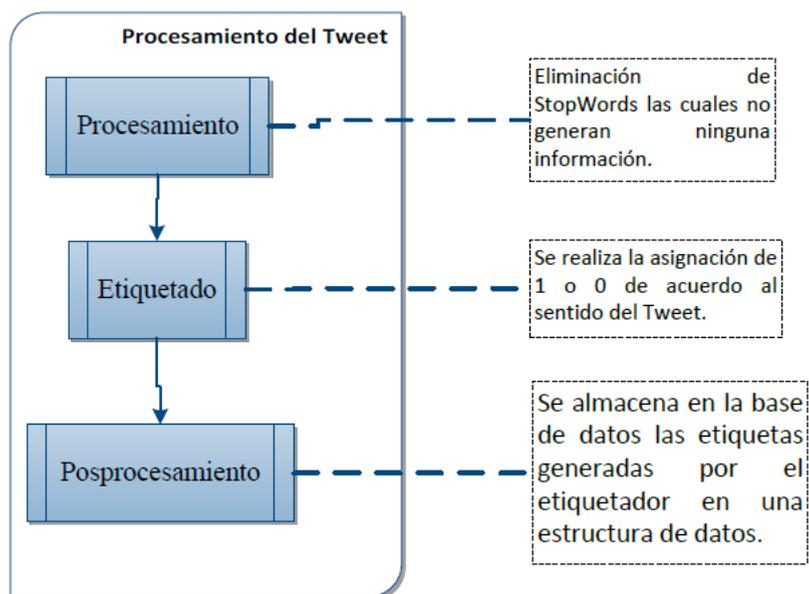


Figura # 15. Procesos para el procesamiento de los tweets

Previo al procesamiento de los *tweets* se construyó un corpus, el cual sirvió para el entrenamiento del algoritmo de clasificación, para lo cual se utilizan tres conjuntos de datos. Uno de ellos con 429 *tweets* del Ecuador, los cuales contienen aspectos políticos, sociales, culturales, educación y deportivos, que tienen una clasificación de polaridad (o sentimiento) positivo y negativo. El segundo conjunto tiene 47,254 *tweets* descargados de la organización TASS (Taller de Análisis Semántico), los cuales contienen aspectos generales, políticos, sociales, economía, educación y sanidad. El tercer conjunto de datos corresponde a *tweets* deportivos con respecto al Campeonato Nacional Ecuatoriano de Fútbol serie A 2017, el cual consta de 4617 *tweets*, los cuales están clasificados como positivos y negativos. El corpus final tiene un total de 52300 *tweets* lo que permite el análisis de sentimientos o polaridad de los *tweets*.

- **Procesamiento**

El procesamiento de cada *tweet* consiste en eliminar las *stop words*, al hablar de stop words o palabras vacías hacemos referencia a todas las palabras que carecen de un significado, estas palabras vacías suelen ser artículos, conjunciones, pronombres preposiciones entre otras a continuación un ejemplo.

- ***Tweet* extraído**

@CSEmelec: Termina el partido en el Estadio Banco del Pacífico Capwell. ¡GANOOOOÓ EL BOMBILLO!

- ***Tweet* procesado**

@CSEmelec: Termina partido Estadio Banco Pacífico Capwell. ¡GANOOOOÓ BOMBILLO!

- **Etiquetado**

Una vez realizado el procesamiento del *tweet* se utiliza el modelo de predicción de sentimiento para etiquetar al *tweet* como se muestra a continuación en la tabla # 9.

Tabla # 9. Etiquetado de los tweets

Text	Polarity
RT @DiarioExtraEc: Liga era ganador durante los primeros 29 minutos y #Emelec le dio a la vuelta al partido	1
QUE DIOS NOS BENDIGA Y NOS PROTEJA Eres mi mayor Alegría 😊👩👧👦 Disculpas por muchas cosas 😞😁😞...	1
@diego_old @jamolestina @BarcelonaSCweb De acuerdo, pero no es un simple solazo. Son niveles de radiación extremos... https://t.co/7txq8vr4Kx	1
RT @SalaDePrensaEc: 38' Un #Barcelona desconocido ante un #Emelec que no deja de atacar, está más cerca la segunda que el empate...	0
RT @LaRedEcuador: 07' ¡GOOOOOOOL DE CLAN JUVENIL! Centro desde el córner y Villacres cabecea a gol #ClanJuvenil 1-0 #Delfín...	1
Esto NO se puede seguir dando nunca más. O quieren que haya un muerto primero? @panchocevallosv https://t.co/EcrTACMasv	0

- **Posprocesamiento**

Una vez realizada la predicción del *tweet* acorde al sentimiento o polaridad de la oración, este dato se envía y almacena en la base de datos.

- **Generación de Alerta**

Para la generación de las alertas se debe tomar en cuenta la relación que existe entre los *tweets* mediante palabras clave y geolocalizados, estos estarán ponderados de manera positiva o negativa de acuerdo a su análisis, además, en los geolocalizados se tomará en cuenta la distancia o radio del cual fueron extraídos. Debido a esto se propone la ecuación # 4, la cual permite normalizar los datos extraídos y analizados desde *Twitter*. Adicional a esto, se ha tomado en cuenta que los *tweets* geolocalizados tengan una ponderación mayor que los datos extraídos por palabras clave ya que estos *tweets* se encuentran dentro del área de incidencia en dónde se lleva a cabo un encuentro deportivo y en dónde un mensaje de alarma tiene que ser considerado con mayor interés.

Esta ecuación nos permitirá realizar una valoración de los *tweets* analizados para presentarlos al usuario mediante un índice al cual se lo describe más adelante.

$$\text{Índice} = \frac{W * RadioGeo * \sum_{i=1}^n N_{geo} + \sum_{i=1}^m N_{Keywords}}{W * RadioGeo * \sum_{i=1}^n N_{geo} + \sum_{i=1}^m N_{Keywords} + W * RadioGeo * \sum_{i=1}^o P_{geo} + \sum_{i=1}^p P_{Keywords}}$$

Ecuación # 4. Fórmula para normalización de datos

W = es la relación de *tweets* geolocalizados y palabras clave.

Para sacar la relación de *tweets* (W) se utilizaron los datos recolectados en 4 fechas completas del campeonato nacional ecuatoriano de fútbol serie a 2017 obteniendo los siguientes resultados. (Ver tabla # 10)

Tabla # 10. Cantidad de datos extraídos de Twitter durante 4 fechas del campeonato de fútbol ecuatoriano

Fecha	Keywords	Geo
Fecha 14	2313	86
Fecha 15	3974	120
Fecha 16	3437	71
Fecha 17	3214	162
Total	12938	439

$$W = \frac{Keywords}{Geo}$$

Ecuación # 5. Relación de *tweets* geolocalizados y key Words

$$W = \frac{12938}{439}$$

$$W = 29$$

Con estos datos obtenidos la relación que se obtiene es la siguiente: por cada tweet geolocalizado, este equivaldrá a 29 *tweets* no geolocalizados.

▪ Índice

Para la generación de alertas se utilizará el siguiente índice, calculado con la ecuación # 4, el cual permitirá identificar en qué zona o estado se encuentra el partido de fútbol monitoreado. El índice de manera gráfica estará representado de la siguiente manera (ver figura # 16).



Figura # 16. Índice de alerta

El proceso de la generación de alerta como se encuentra en la capa de negocios estará comunicado directamente con la capa de presentación la cual ayudará al usuario a identificar el estado del partido.

2.2.5 Capa de Datos

Esta capa será la encargada de almacenar, modificar y consultar los datos extraídos y procesados como se muestra en la figura # 18. En esta capa también se almacenará las características de cada uno de los equipos, por ejemplo:

- En la colección equipos se describe, el nombre del equipo, coordenadas, nombre del estadio.
- En la colección palabras clave se detalla las palabras más frecuentes que son utilizadas para identificar un equipo.
- En las colecciones tweets_geo y tweets_keywords se almacenarán los datos extraídos desde *Twitter*.

La capa de data se conecta únicamente con la capa de negocios y en esta investigación estará representada como se muestra en la Figura # 17.

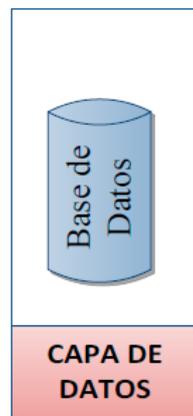


Figura # 17. Capa de datos

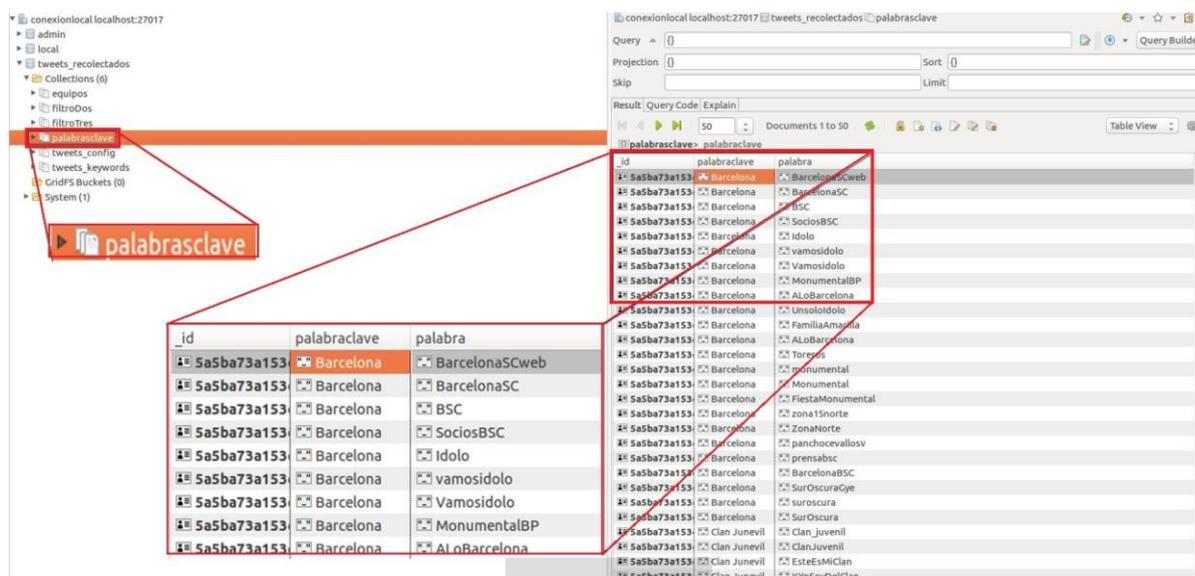


Figura # 18. Almacenamiento de los tweets en la base de datos

2.3 Implementación de Prototipo

En esta sección describimos las tecnologías utilizadas para el desarrollo del prototipo descrito en los apartados anteriores.

2.3.1 Tecnologías Usadas

- **Python**

Para la realización de esta investigación se optó por usar Python como lenguaje de programación, debido a que es fácil de entender y utilizar, además Python cuenta con un gran número de librerías, lo que permite una implementación eficiente y eficaz al momento de desarrollar. Entre las librerías y paquetes más relevantes que se utilizaron son: Tweepy, NumPy, nltk, y pymongo.

- **MongoDB**

Tomando en cuenta que los datos extraídos de *Twitter* se encuentran en formato JSON (JavaScript Object Notation - Notación de Objetos de JavaScript) se utilizó MongoDB, ya que es una base de datos no relacional que utiliza este tipo de objetos.

- **Django**

Django es un framework de desarrollo web de código abierto escrito totalmente en Python, debido a esto se lo ha utilizado para el desarrollo web y es fácil de entender. Además,

hace especial hincapié en el re-uso, la conectividad y extensibilidad de componentes, desarrollo rápido y el principio “No te repitas” (DRY, del inglés *Don't Repeat Yourself*).

- **JavaScript**

Se utilizó JavaScript para integrar funcionalidad a la página web lo cual permite que esta sea más interactiva, dinámica y atractiva hacia el usuario.

2.3.2 Resumen

A continuación, ver tabla # 11, se presenta el software y librerías utilizadas, las cuales son aquellas utilizadas en el desarrollo y pruebas del módulo de alertas tempranas, para un óptimo funcionamiento de este módulo se sugiere utilizar la misma configuración.

Tabla # 11. Resumen de software y librerías

Software utilizado	
Sistema operativo	Ubuntu (16.4.3)
Lenguaje de programación	Python (3.6.3)
Librerías utilizadas	
Extracción de datos	Tweepy (3.5.0)
Procesamiento de datos	Nltk (3.2.5)
	Pandas (0.21.1)
	Pip (9.01)
	Numpy (1.13.3)
Almacenamiento de datos	Pymongo (3.6.0)
Presentación de datos	Django (1.9.7)

CAPÍTULO III

RESULTADOS

3.1 Experimentación

Antes de realizar pruebas de toda la aplicación desarrollada, se realizó una validación empírica del algoritmo de clasificación, para lo cual se analizaron 565 *tweets*. Estos *tweets* fueron validados por un grupo de más de 25 personas que colocaron su apreciación de la polaridad: positiva o negativa de cada *tweet*. Estos 565 *tweets* no fueron parte del conjunto de entrenamiento del modelo y al ser analizados por el modelo de PLN obtenemos un error de clasificación del 3.4%. Las figuras # 19 y 20 muestran el porcentaje de cada categoría tanto para el ejercicio realizado por personas como para el modelo de PLN, respectivamente.

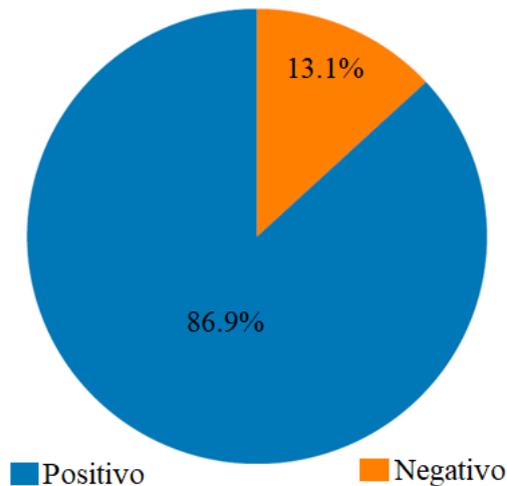


Figura # 19. *Tweets* analizados y ponderados por varias personas

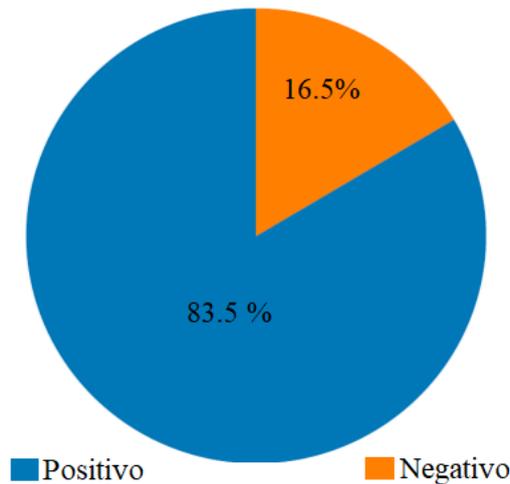


Figura # 20. *Tweets* analizados mediante el modelo PLN

3.2 Experimento Uno: Extracción, análisis y almacenamiento de *Tweets*

Partido Analizado: Macara VS Universidad Católica

El prototipo permite seleccionar el encuentro deportivo que se va a analizar a través de dos componentes de selección para elegir el equipo local y visitante como se muestra en la figura # 21. La selección se la realiza de acuerdo al encuentro establecido por la federación ecuatoriana de fútbol, de igual manera el horario de inicio y el tiempo de análisis es configurable para cada encuentro.

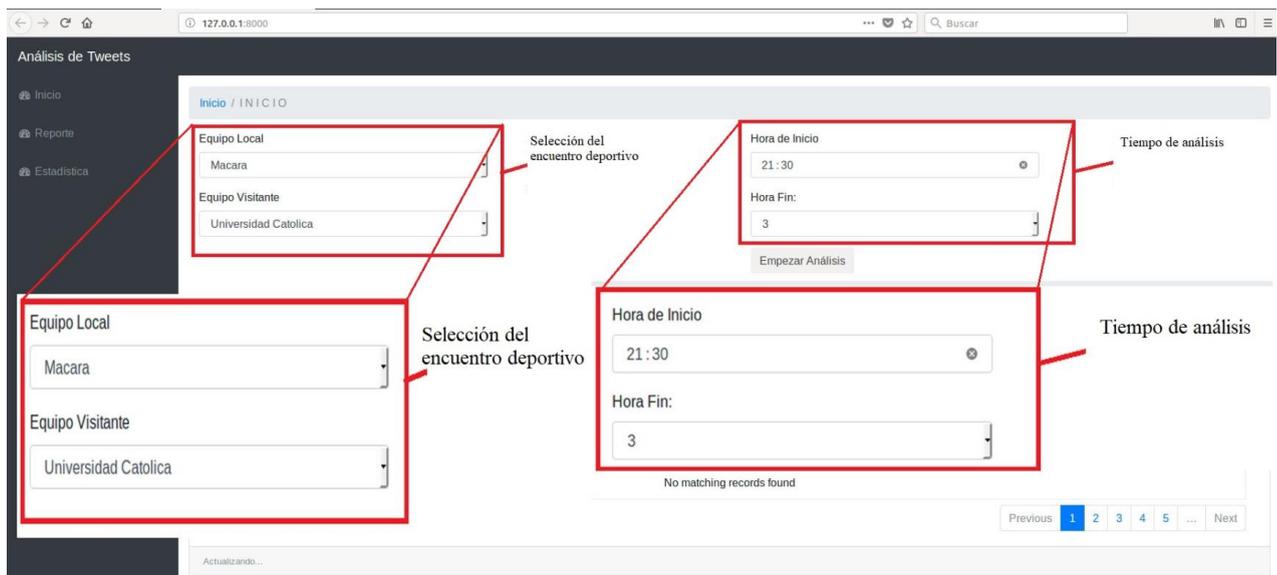


Figura # 21. Elección del encuentro para hacer analizado Macara VS Universidad católica

Para la realización de las pruebas de extracción se envió un *tweet* desde una cuenta de *Twitter* como se muestra en la figura # 22, este *tweet* tiene la palabra clave macará y tiene polaridad positiva. Este mensaje debe ser extraído, analizado y almacenado en la base de datos.

Se realiza una extracción exitosa del *tweet* enviado en la figura anterior y a continuación en la figura # 23 se muestra la predicción del *tweet* de acuerdo a su contexto, adicional a esto se muestra el porcentaje de los *tweets* analizados. Este porcentaje se lo determina con la ecuación # 4 la que permite conocer en el nivel del índice que se encuentra el encuentro deportivo.



Figura # 22. Mensaje para Twittear mediante palabra clave

```
jd1@jd1: ~/Escritorio/tesis_estadios/PLN
[u'Rojas, Fuenzalida y Aued... los mejores de la UC']
Predicción del Tweet-----> 1
Total Tweets Geo Negativo: 0
Total Tweets Key Negativos: 2
Total Tweets Geo Positivos: 0
Total Tweets Key Positivos: 23
Total de Tweets: 25
Porcentaje de los Tweets analizados: 0.08

[u'pruebas exitosas extraccion de datos desde Twitter palabra clave Macara']
Predicción del Tweet-----> 1
Total Tweets Geo Negativo: 0
Total Tweets Key Negativos: 2
Total Tweets Geo Positivos: 0
Total Tweets Key Positivos: 24
Total de Tweets: 26
Porcentaje de los Tweets analizados: 0.0769230769231
```

Figura # 23. Análisis positivo de un *tweet* mediante palabra clave

En la figura # 24 se puede constatar que se ha extraído y almacenado en la base de datos de manera exitosa el *tweet* escrito en la figura # 22. El *tweet* se lo almacena en la colección de *tweets_keywords* ya que se realizó las pruebas mediante palabras clave.

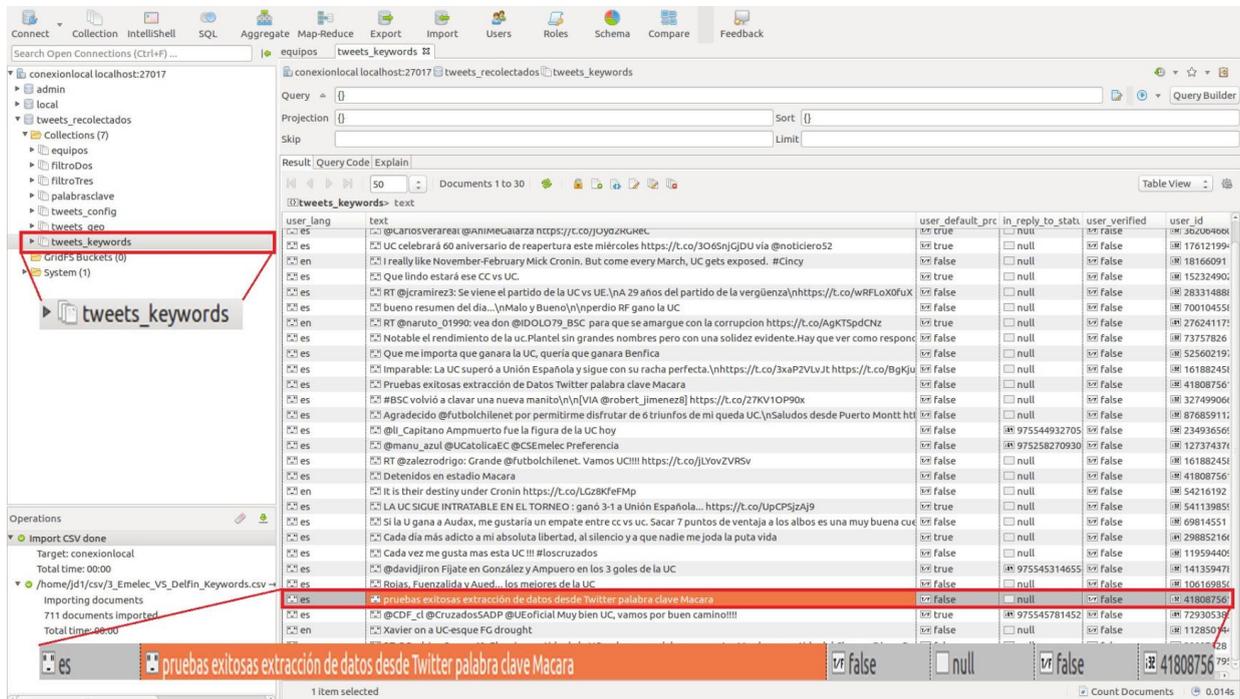


Figura # 24. Almacenamiento de los *tweets* en la base de datos mediante palabra clave

Además, se envió un *tweet* geolocalizado como se puede apreciar en la figura # 25, en el cual se comparte la ubicación desde donde es enviado el *tweet* con sus respectivas coordenadas.



Figura # 25. Mensaje para Twitrear mediante geolocalización

El *tweet* georreferenciado es extraído y analizado como se puede apreciar en la figura # 26, además se toma en cuenta la distancia de donde fue tuiteado ya que si sobrepasa el rango de un kilómetro este no deberá ser analizado.

```
jd1@jd1: ~/Escritorio/tesis_estadios/PLN
Distancia del tweet ----->0.276866914179
[u'Tweets negativo heridos en estadio geo localizado']

Predicción del Tweet----->0
Total Tweets Geo Negativo: 1
Total Tweets Key Negativos: 6
Total Tweets Geo Positivos: 1
Total Tweets Key Positivos: 45
Total de Tweets: 53
Porcentaje de los Tweets analizados: 0.321100917431
<----- Zona Segura ----->
```

Figura # 26. Análisis positivo de un *tweet* mediante geolocalización

El almacenamiento de cada *tweet* georreferenciado se almacena en la base de datos, la cual tiene una colección propia para este tipo de *tweets* denominada *tweets_geo* como se puede apreciar en la figura # 27.

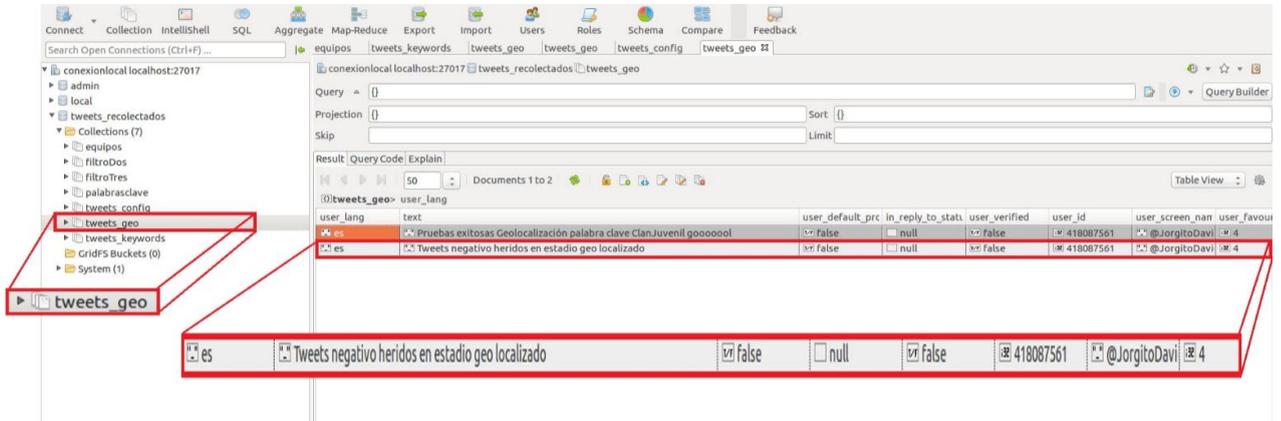


Figura # 27. Almacenamiento del *tweet* en la base de datos mediante geolocalización

Como se ha mencionado anteriormente, el módulo de generación de alertas tiene una capa de presentación en la cual se puede apreciar cada uno de los *tweets* extraídos de *Twitter*, el equipo local y de visitante, la hora de finalización del análisis y el índice en que se encuentra el partido analizado. Esa interfaz de presentación se la puede apreciar en la figura # 28.

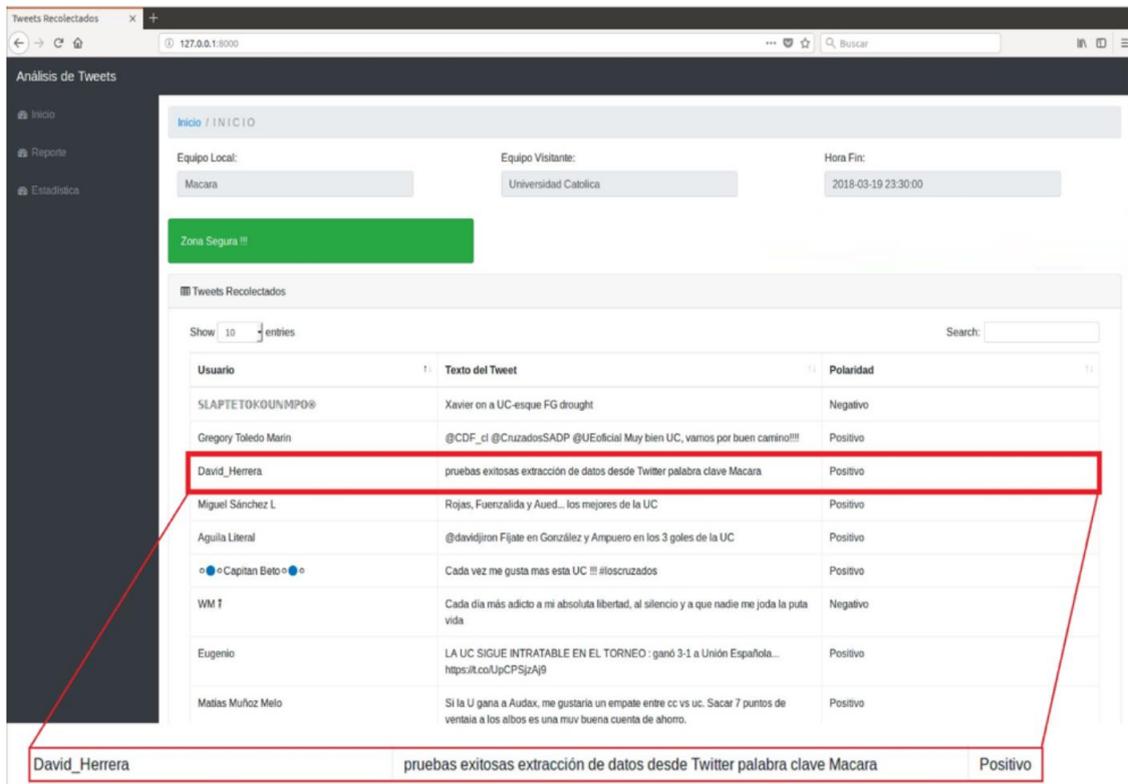


Figura # 28. Presentación del *tweet* en el módulo de generación de alertas

3.3 Experimento Dos: Pruebas de funcionamiento del índice para la generación de alertas

En este experimento se puso a prueba el funcionamiento de cada uno de los índices, los cuales son: zona segura en el rango de 0 a 0.49, zona de alerta en el rango de 0.50 a 0.74 y zona de peligro en el rango de 0.75 a 1. Para cada uno de estos índices se tomó en cuenta tres encuentros deportivos: Macará versus Universidad Católica, Macará versus El Nacional y Macará versus Barcelona.

Encuentro deportivo N° uno: Macara VS Universidad Católica (Índice: Zona Segura)

Mientras el índice no supere el límite establecido de 0.49 este se ubica dentro de una zona segura como se puede apreciar en la figura # 29, es decir que existe un gran número de *tweets* positivos dentro del encuentro deportivo.

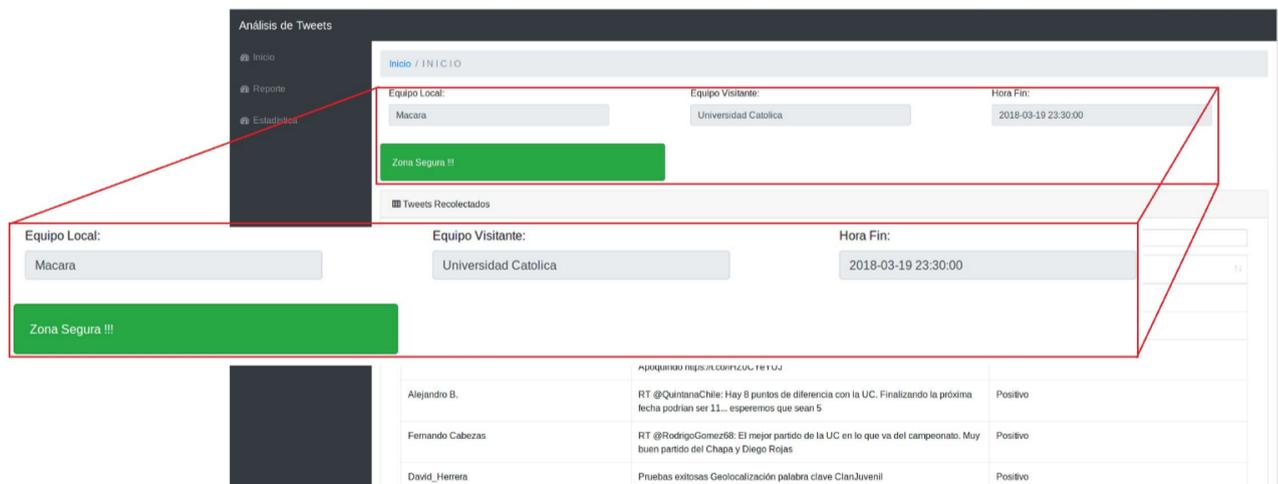


Figura # 29. Presentación de índice zona segura

Al final se obtiene un reporte de los *tweets* analizados como se puede apreciar en la figura # 30. En este reporte se detalla el número total de los *tweets* analizados y el porcentaje del índice con el cual culminó el encuentro deportivo. Adicional a esto en la figura # 31 se presenta una gráfica de los *tweets* que se dieron durante el tiempo de análisis del encuentro deportivo.

Análisis de Tweets

Inicio

Reporte

Estadística

Reporte / REPORTE

Análisis Anteriores

Identificador	Fecha Inicio	Fecha Fin	Equipo Local	Equipo Visitante	Número de Tweets	Porcentaje	Índice
1	2018-03-18 20:30:00	2018-03-18 20:43:00	Macara	Universidad Catolica	65	0.17	Zona Segura

Actualizando...

1	2018-03-18 20:30:00	2018-03-18 20:43:00	Macara	Universidad Catolica	65	0.17	Zona Segura
---	---------------------	---------------------	--------	----------------------	----	------	-------------

Figura # 30. Reporte de índice zona segura

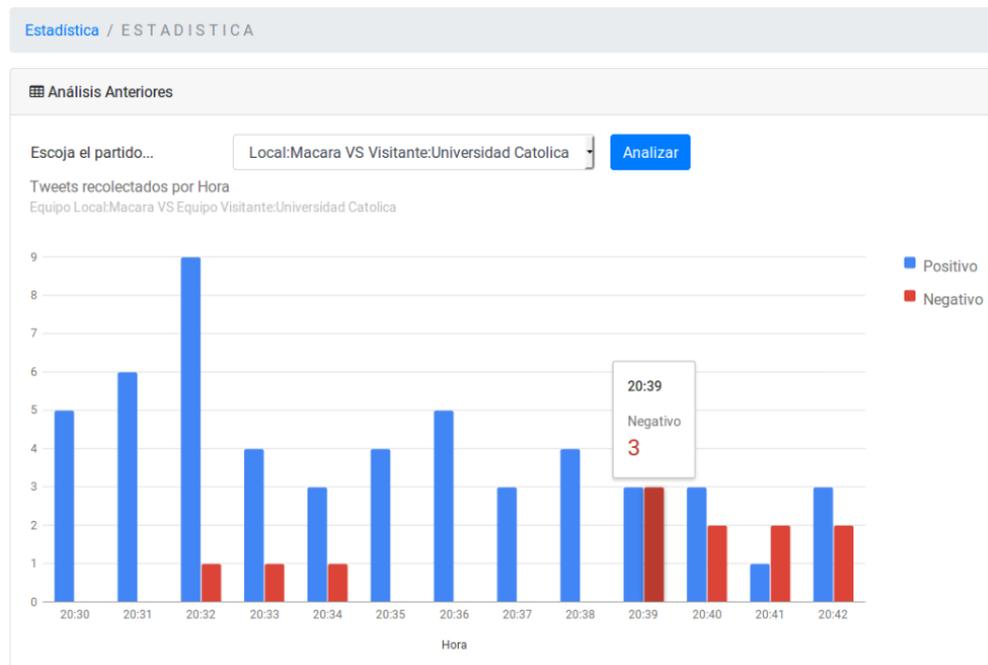


Figura # 31. Gráfica de tweets índice zona segura

Encuentro deportivo N° dos: Macara VS Universidad Católica (Índice: Zona de Alerta)

Cuando el índice es superior a 0.49 y menor a 0.74 el encuentro deportivo se encontrará en una zona de alerta como se puede observar en la figura # 32.

The screenshot shows the 'Análisis de Tweets' interface. On the left is a navigation menu with 'Inicio', 'Reporte', and 'Estadística'. The main area is titled 'Inicio / INICIO' and contains search filters: 'Equipo Local: Macara', 'Equipo Visitante: El Nacional', and 'Hora Fin: 2018-03-20 01:37:00'. Below the filters is a yellow box labeled 'Zona de Alerta!!!'. Underneath is a section 'Tweets Recolectados' with the same filters. A table of tweets is shown below:

Usuario	Tweet	Estado
La SuperMaquina Gris	RT @blascomososo: @calims30 @elnacionalec Me imagino que se dieron cuenta la tontera que publicaron y ya la borraron!!!	Positivo
AGURTO	ya volvimos todos a lima o falta alguien?	Negativo
David_Herrera	Realizando pruebas exitosas Macara	Positivo

Figura # 32. Presentación de índice zona de alerta

Al observar el reporte final de los *tweets* analizados, ver la figura # 33, el número total de los *tweets* es de 35. Los cuales dan como resultado un porcentaje de 0.68 el cual se encuentra en una zona de alerta. Adicional a esto se muestra la figura # 34 la cual representa una gráfica de los *tweets* que se dieron en el encuentro deportivo.

The screenshot shows the 'Análisis de Tweets' interface in the 'Reporte / REPORTE' section. It displays a table of previous analyses:

Identificador	Fecha Inicio	Fecha Fin	Equipo Local	Equipo Visitante	Número de Tweets	Porcentaje	Indice
1	2018-03-18 20:30:00	2018-03-18 20:43:00	Macara	Universidad Catolica	65	0.17	Zona Segura
2	2018-03-19 21:37:00	2018-03-20 01:37:00	Macara	El Nacional	35	0.68	Zona de Alerta

Below the table is a loading indicator 'Actualizando...' and a larger table showing the same data for the selected report:

2	2018-03-19 21:37:00	2018-03-20 01:37:00	Macara	El Nacional	35	0.68	Zona de Alerta
---	---------------------	---------------------	--------	-------------	----	------	----------------

Figura # 33. Reporte de índice zona de alerta



Figura # 34. Gráfica de tweets índice zona de alerta

Encuentro deportivo N° tres: Macara VS Barcelona (Índice: Zona de Alerta)

La zona más crítica será cuando los *tweets* analizados superen el rango de 0.74, debido a esto el índice se encontrará en una zona de peligro como se muestra en la figura # 35, esto se dará cuando exista una gran cantidad de *tweets* negativos.

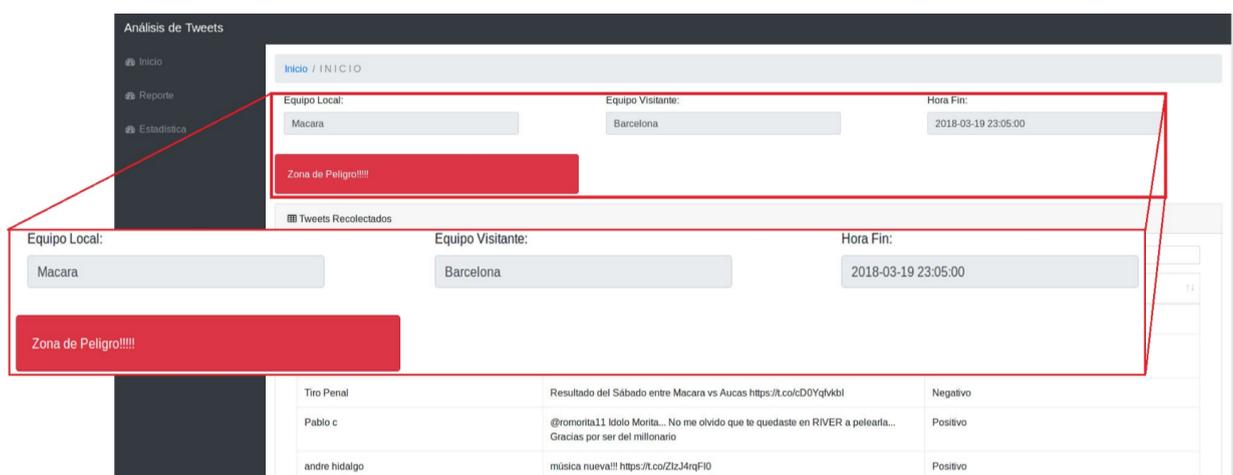


Figura # 35. Presentación de índice zona de peligro

Como se puede observar en el reporte final de los *tweets* analizados de la figura # 36, existe un total de 176 *tweets* cuya polaridad dan como resultado un porcentaje de 0.76 el cual

está dentro del rango de zona de peligro. Adicional a esto se muestra la figura # 37 la cual representa una gráfica de los *tweets* que se dieron en el encuentro deportivo.

Análisis de Tweets

Inicio

Reporte

Estadística

Reporte / REPORTE

Análisis Anteriores

Identificador	Fecha Inicio	Fecha Fin	Equipo Local	Equipo Visitante	Número de Tweets	Porcentaje	Índice
1	2018-03-18 20:30:00	2018-03-18 20:43:00	Macara	Universidad Catolica	65	0.17	Zona Segura
2	2018-03-18 21:37:00	2018-03-18 21:53:00	Macara	El Nacional	35	0.68	Zona de Alerta
3	2018-03-19 22:05:00	2018-03-19 22:24:00	Macara	Barcelona	176	0.76	Zona de Peligro

Figura # 36. Reporte de índice zona de peligro

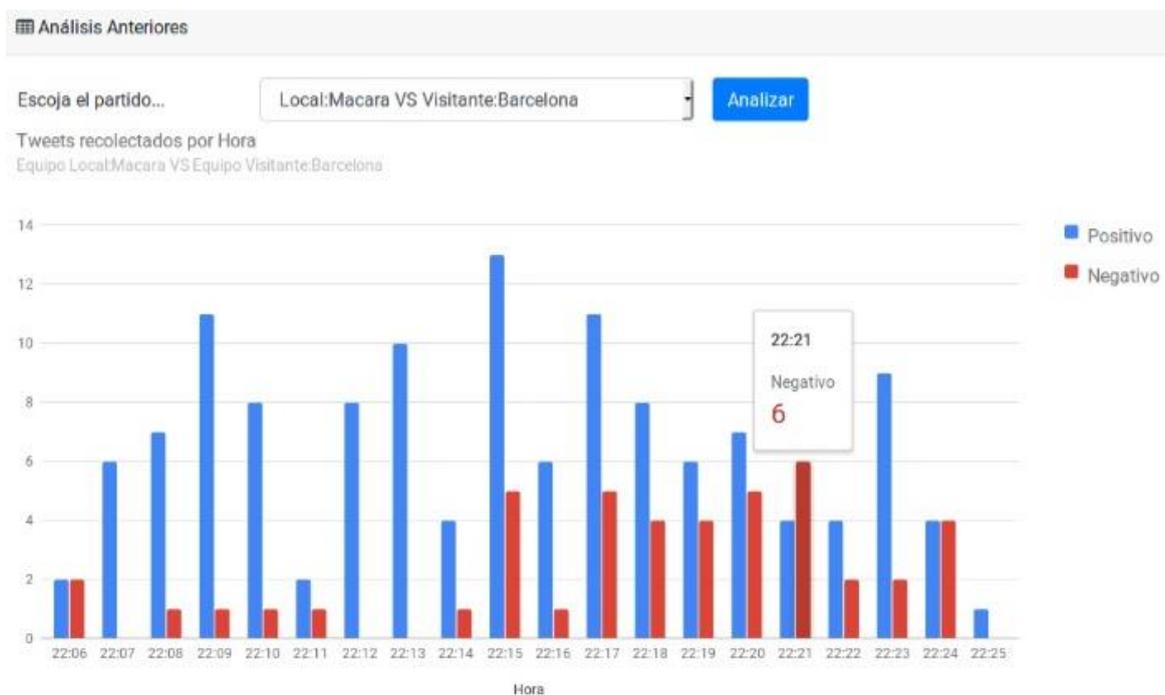


Figura # 37. Gráfica de *tweets* índice zona de peligro

3.4 Experimento Tres: Análisis de tres encuentros deportivos de fútbol del campeonato nacional ecuatoriano Serie A

En este experimento se pone a prueba el módulo de generación de alerta temprana en conjunto. Para esto se analizan tres partidos del campeonato nacional ecuatoriano: Independiente del Valle versus Liga de Quito, Liga de Quito versus Emelec y Emelec versus Delfín.

▪ **Partido Uno:**

El primer encuentro analizado fue el de **Independiente del Valle VS Liga de Quito**, el cual se desarrolló el 12-11-2017. Para este encuentro se configuró un tiempo de análisis de cuatro horas: una hora antes del partido, dos horas en el cual se desarrolla el partido y una hora después del partido. Esto se puede apreciar en la figura # 38.

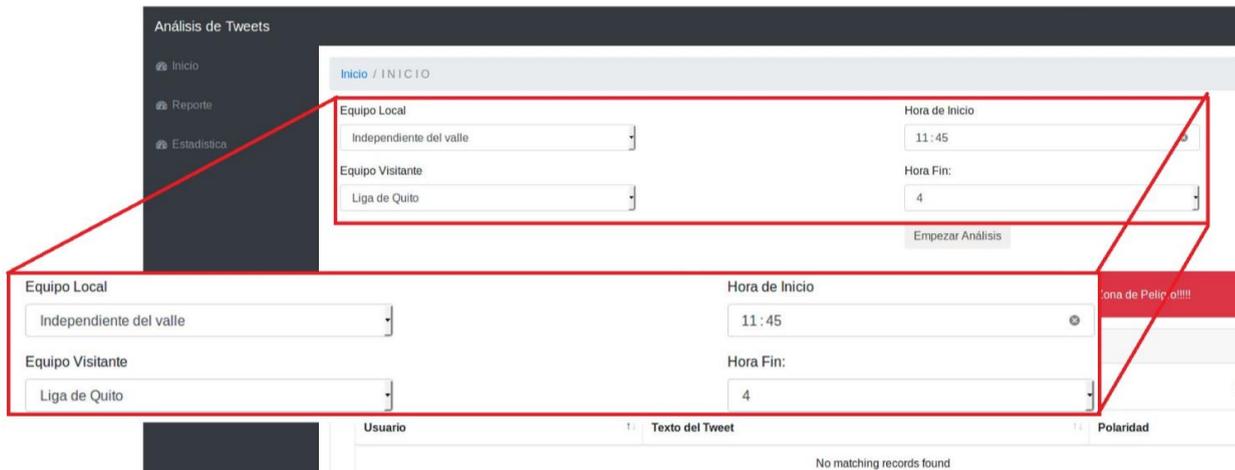


Figura # 38. Análisis del encuentro Independiente del Valle VS Liga de Quito

El encuentro deportivo terminó con la recolección y análisis de 1080 *tweets* dando como resultado un índice de zona segura con un valor de 0.20, como se puede observar en la figura # 39. En la figura # 40 se presenta la gráfica estadística de los *tweets* analizados durante el encuentro deportivo.

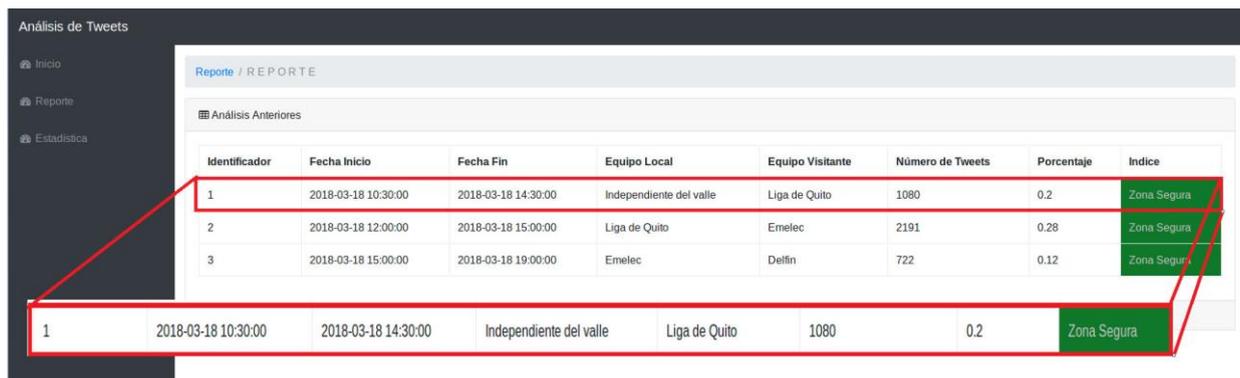


Figura # 39. Reporte del índice del encuentro Independiente del Valle VS Liga de Quito

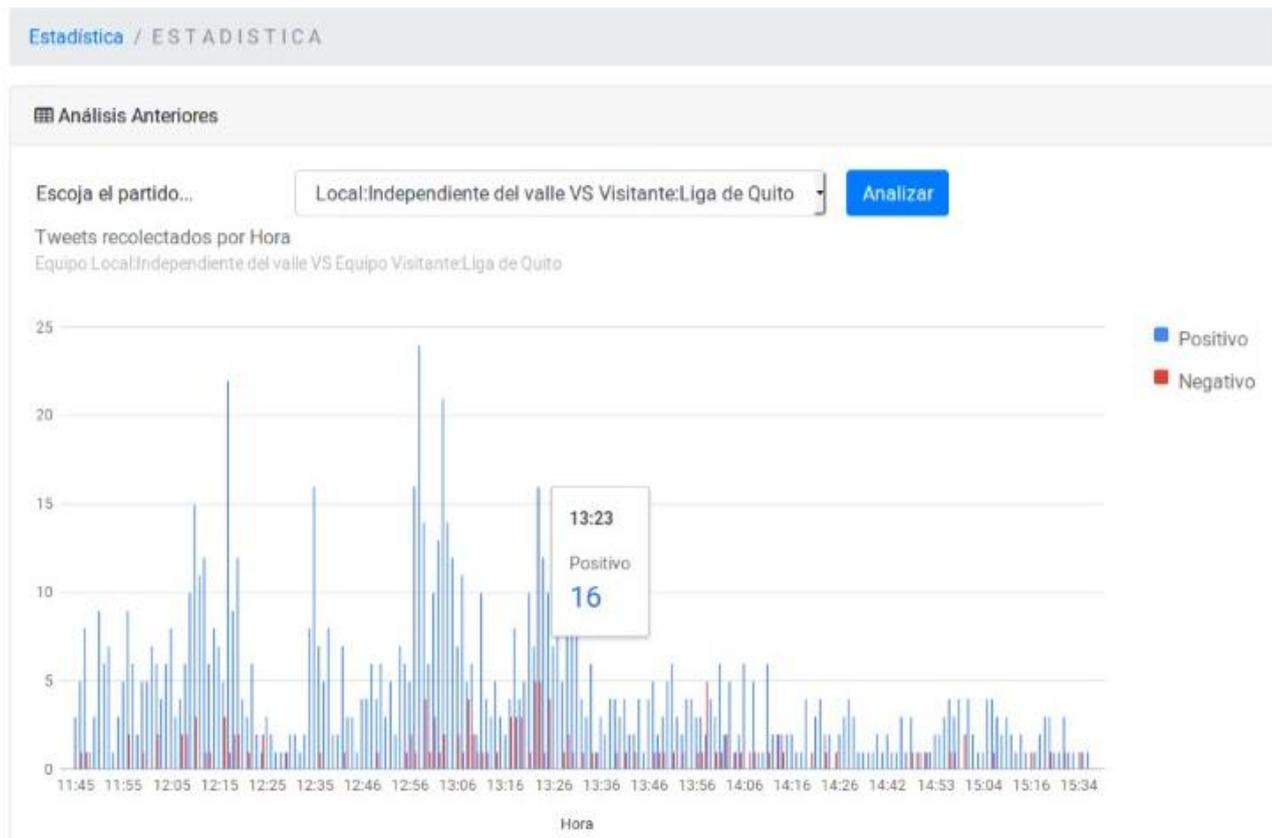


Figura # 40. Gráfica estadística de los tweets analizados en encuentro Independiente del Valle VS Liga de Quito

▪ **Partido Dos:**

El segundo encuentro analizado fue **Liga de Quito VS Emelec**, el cual se desarrolló el 26-11-2017. Para este encuentro se configuró un tiempo de análisis de cuatro horas: una hora antes del partido, dos horas en el cual se desarrolló el partido y una hora después del partido. Esto se lo puede apreciar en la figura # 41.

Análisis de Tweets

Inicio / INICIO

Equipo Local: Liga de Quito
Equipo Visitante: Emelec
Hora de Inicio: 10:52
Hora Fin: 4
Empezar Análisis

Equipo Local: Liga de Quito
Equipo Visitante: Emelec
Hora de Inicio: 10:52
Hora Fin: 4

No matching records found

Figura # 41. Análisis del encuentro Liga de Quito VS Emelec

El encuentro deportivo terminó con la recolección y análisis de 2191 *tweets* dando como resultado un índice de zona segura con un valor de 0.28, como se puede observar en la figura # 42. En la figura # 43 se presenta la gráfica estadística de los *tweets* analizados en el encuentro deportivo.

Análisis de Tweets

Inicio
Reporte
Estadística

Reporte / REPORTE

Análisis Anteriores

Identificador	Fecha Inicio	Fecha Fin	Equipo Local	Equipo Visitante	Número de Tweets	Porcentaje	Índice
1	2018-03-18 10:30:00	2018-03-18 14:30:00	Independiente del valle	Liga de Quito	1080	0.2	Zona Segura
2	2018-03-18 12:00:00	2018-03-18 15:00:00	Liga de Quito	Emelec	2191	0.28	Zona Segura
3	2018-03-18 15:00:00	2018-03-18 19:00:00	Emelec	Delfin	722	0.12	Zona Segura

2 | 2018-03-18 12:00:00 | 2018-03-18 15:00:00 | Liga de Quito | Emelec | 2191 | 0.28 | Zona Segura

Figura # 42. Reporte del índice del encuentro Liga de Quito VS Emelec

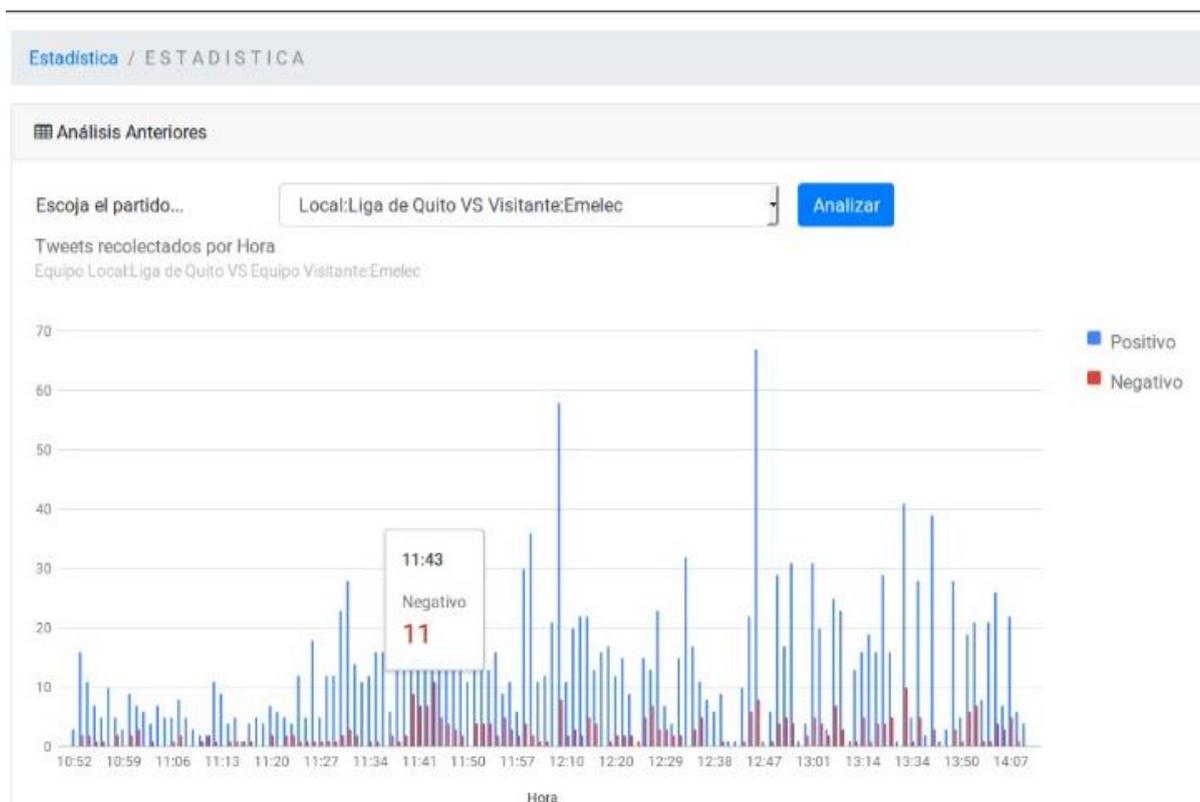


Figura # 43. Gráfica estadística de los *tweets* analizados en encuentro Liga de Quito VS Emelec

▪ **Partido Tres:**

El tercer encuentro analizado fue la final del campeonato nacional ecuatoriano que se disputó entre el **Emelec VS Delfín**, el cual se desarrolló el 17-12-2017. Para este encuentro se configuró un tiempo de análisis de cuatro horas: una hora antes del partido, dos horas en el cual se desarrolló el partido y una hora después del partido. Esto se puede apreciar en la figura # 44.

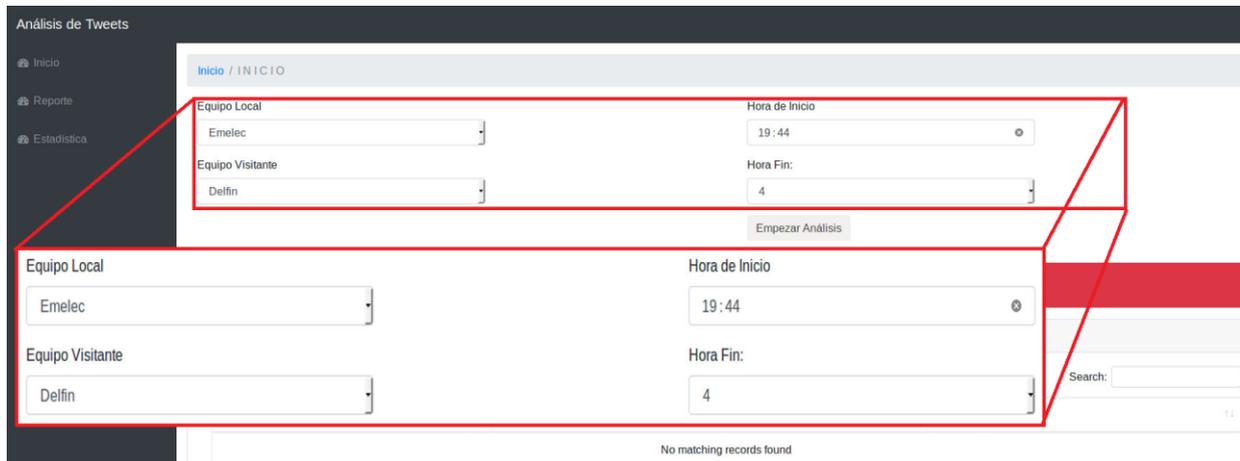


Figura # 44. Análisis del encuentro Emelec VS Delfín

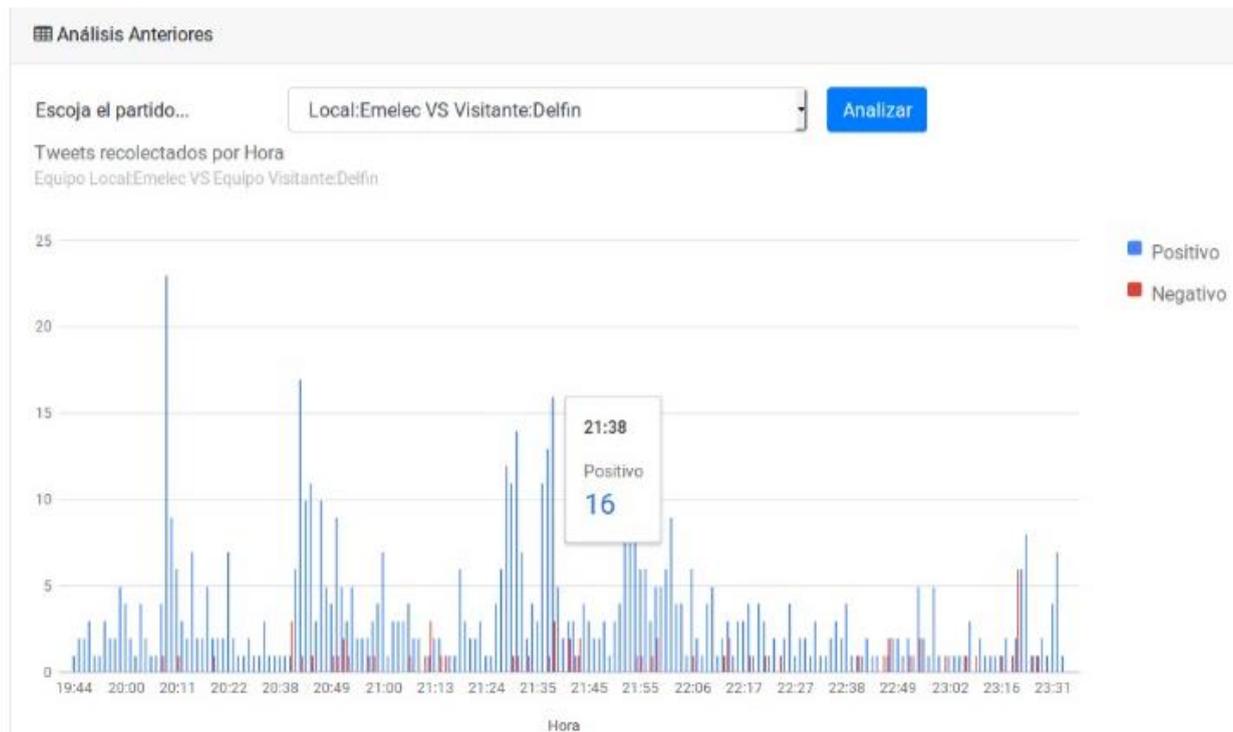
El encuentro deportivo finalizó con la recolección y análisis de 722 *tweets* dando como resultado un índice de zona segura con un valor de 0.12, como se puede observar en la figura # 45. En la figura # 46 se presenta la gráfica estadística de los *tweets* analizados en el encuentro deportivo.

The screenshot shows the 'Reporte / REPORTE' screen of the 'Análisis de Tweets' application. It features a sidebar with 'Inicio', 'Reporte', and 'Estadística'. The main area contains a table titled 'Análisis Anteriores' with the following data:

Identificador	Fecha Inicio	Fecha Fin	Equipo Local	Equipo Visitante	Número de Tweets	Porcentaje	Índice
1	2018-03-18 10:30:00	2018-03-18 14:30:00	Independiente del valle	Liga de Quito	1080	0.2	Zona Segura
2	2018-03-18 12:00:00	2018-03-18 15:00:00	Liga de Quito	Emelec	2191	0.28	Zona Segura
3	2018-03-18 15:00:00	2018-03-18 19:00:00	Emelec	Delfin	722	0.12	Zona Segura

A red box highlights the third row of the table. Below the table is a detailed view of the third analysis, showing the same data as the table row.

Figura # 45. Reporte del índice del encuentro Emelec VS Delfín



3.5 Análisis Resultados

Se obtuvieron resultados exitosos en todos los experimentos realizados. En el primer experimento se realizó la extracción de *tweets* mediante palabras clave y geolocalización permitió probar empíricamente el modelo de predicción y analizar el correcto almacenamiento de los *tweets* en una colección de datos acorde a su categoría: *tweets_keywords* y *tweets_geo*. En este experimento se muestra también con éxito la interfaz de presentación y visualización de cada uno de los *tweets* analizados al usuario final.

Para la comprobación del almacenamiento de los *tweets* se *twiteo* una oración con una palabra clave y con un sentido positivo, como se puede observar en la figura # 22. Este *tweet* fue extraído desde *Twitter* y almacenado en la base de datos en la colección “*tweets_keywords*” como se observa en la figura # 24, con estos resultados se logró validar una extracción y almacenamiento exitoso de los *tweets*.

Al realizar las pruebas de funcionamiento de los índices (Segundo experimento) se consiguieron resultados positivos, al obtener reportes y gráficas de cada uno de los *tweets* analizados y almacenados previamente. En este experimento se analizaron cada uno de los

estados del índice que son: zona segura, zona de alerta y zona de peligro. Cabe mencionar que para obtener cada uno de los índices se trabajó con datos fabricados.

Para una mejor comprensión tomaremos como ejemplo el encuentro deportivo N° uno: Macara VS Universidad Católica (Índice: Zona Segura), en este se obtuvo un total de 66 tweets durante su análisis de este total 11 de los *tweets* fueron negativos y 55 positivos, con estos resultados aplicándolos a la ecuación # 4 nos da como resultado 0.17, el cual no sobrepasa la zona segura cuyo rango es de 0 a 0.49 como se muestra en la figura # 16.

En el tercer experimento se puso en funcionamiento el módulo de generación de alertas tempranas, analizando tres encuentros deportivos del campeonato ecuatoriano de fútbol, a cada uno de estos se los analizó durante cuatro horas obteniendo en el primer encuentro 1080 *tweets* entre positivos y negativos dándonos un porcentaje de análisis del 0.20 lo que determina según el índice establecido una zona segura. Para los otros encuentros deportivos se obtuvieron resultados similares del índice y se validó el índice obtenido.

CAPÍTULO IV

DISCUSIÓN

4.1 Conclusiones

En la actualidad las redes sociales son un medio de comunicación de gran relevancia, en Ecuador 4.2 millones de personas interactúan con esta herramienta, según el Instituto Nacional de Censos y Estadísticas (INEC). En virtud de esto, se realiza la presente investigación que recopila y analiza datos e información específicamente de *Twitter* debido a que es un medio de información de eventos de corta duración; y, que se caracteriza por proporcionar un conocimiento circunstancial de lo que está pasando en un determinado lugar u hora.

Para la extracción y análisis de los *tweets* se ha utilizado el lenguaje de programación Python, ya que es una herramienta eficiente y eficaz, que permite una conexión hacia *Twitter* y la base de datos MongoDB, versatilidad del lenguaje que contribuyó al desarrollo de este trabajo.

En la fase experimental se obtuvieron resultados satisfactorios. En primer lugar, se validó empíricamente al modelo de predicción de polaridad de *tweets* que consigue un 3.4% de error para los temas tratados. Además, con la extracción y clasificación exitosa de los *tweets*, mediante palabras clave y geolocalización, se analizó cada uno de estos, los cuales fueron almacenados de manera apropiada en la base de datos, acorde a su clasificación.

Finalmente, para una mejor presentación y comprensión de los datos analizados, se diseñó e implementó un índice, el cual está dividido en tres instancias: zona segura, zona de alerta y zona de peligro. Cabe mencionar que para obtener cada uno de los índices se trabajó con datos fabricados; adicional a ello, se presenta un reporte del porcentaje de la zona con que finaliza el encuentro deportivo y una gráfica en la cual se considera el tiempo y número de *tweets* durante el desarrollo del encuentro deportivo. Todo esto nos permite conocer con anticipación los posibles riesgos a los que se expone una persona que asiste a estos eventos e incluso se obtiene una estadística de los encuentros deportivos que presentan mayor riesgo, para así poder tomar medidas anticipadas de seguridad.

Vale mencionar que en los encuentros deportivos que fueron analizados durante el campeonato ecuatoriano de fútbol 2017 no se encontraron índices que hayan levantado una alerta durante el tiempo de análisis de los mismos. Así también, en revisión documental (en los diarios de mayor circulación en el país) no se encontró evidencia de altercados fuera o dentro de los estadios de los tres encuentros deportivos analizados del campeonato nacional ecuatoriano Serie A 2017.

4.2 Recomendaciones y Trabajos Futuros

En términos generales, los datos recopilados durante el desarrollo de este proyecto de investigación contienen la información del campeonato ecuatoriano de fútbol 2017 lo que limitará la capacidad de este sistema para adaptarse a nuevos campeonatos. Por lo que se recomienda tomar en cuenta la naturaleza de la información levantada y las actualizaciones pertinentes para utilizar este módulo en futuros campeonatos nacionales.

Como se trabajó con una arquitectura de tres capas, se da la flexibilidad de agregar nueva información o nuevos datos de cada uno de los equipos que formen parte del campeonato nacional en la capa de base de datos.

Como trabajos futuros se debe integrar al análisis de datos nuevas redes sociales como Facebook, Instagram, etc.; con el objetivo de mejorar la calidad de datos y mejorar la calidad de las alertas del sistema.

En cuanto al índice propuesto, vale mencionar que este índice es una primera aproximación para levantar alertas y nos sirve para un análisis exploratorio de aquellos acontecimientos que en redes sociales levantan alertas, por lo que se recomienda realizar un análisis propio del índice en un trabajo futuro para afinarlo y evitar falsos positivos o falsos negativos.

Vale mencionar que este módulo de alertas tempranas se lo puede extender a otros eventos como reuniones políticas, conciertos, u otros acontecimientos que aglomeran a muchas personas en un mismo lugar. Como trabajo futuro se podría explorar los cambios a realizar en la aplicación para que sea fácilmente configurable a otros escenarios.

Finalmente, se plantea conectar este módulo de alerta temprana con una aplicación móvil para que las personas puedan recibir notificaciones de alerta en temas de su interés o basados en su geolocalización.

REFERENCIAS

- Adam, N., Eledath, J., Mehrotra, S., & Venkatasubramanian, N. (2012). Social Media Alert and Response to Threats to Citizens. En *Proceedings of the 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*. <https://doi.org/10.4108/icst.collaboratecom.2012.250713>
- Alcazar, J. (2015). Ranking Principales Redes Sociales Ecuador. Recuperado el 7 de noviembre de 2017, a partir de <http://blog.formaciongerencial.com/ranking-redes-sociales-ecuador-mayo-2014/>
- Aluja, T. (2001). La Minería de Datos, entre la Estadística y la Inteligencia Artificial. *Qüestiió*, 25(3), 479–498. <https://doi.org/0210-8054>
- Alvarado, C. (2016). Que es Twitter, para que sirve y como se utiliza. Recuperado el 7 de noviembre de 2017, a partir de <http://www.christiamalvarado.com/marketing-con-twitter/que-es-twitter-para-que-sirve-y-como-se-utiliza/>
- Amirkhanyan, A., & Meinel, C. (2016). Analysis of the value of public geotagged data from twitter from the perspective of providing situational awareness. En *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9844 LNCS, pp. 545–556). https://doi.org/10.1007/978-3-319-45234-0_48
- Asensio, E. (2015). Aplicación de técnicas de minería de datos en redes sociales / web, 50.
- Caballero, C. (2012). *Extracción de información social desde Twitter y análisis mediante Hadoop*.
- Carmona, E. J. (2014). Máquinas de Vectores Soporte (SVM), 1–25. Recuperado a partir de [http://www.ia.uned.es/~ejcarmona/publicaciones/\[2013-Carmona\] SVM.pdf](http://www.ia.uned.es/~ejcarmona/publicaciones/[2013-Carmona] SVM.pdf)
- Castillo, D., Pérez, M., Hernández, L., Morález, R., & Ginori, J. (2016). Algoritmos de aprendizaje automático para la clasificación de neuronas piramidales afectadas por el envejecimiento Machine learning algorithms for classification of pyramidal neurons affected by aging, 8(3), 559–571.
- Chaing, C. (2017). El Campeonato Ecuatoriano de Fútbol. Recuperado el 6 de mayo de 2018, a partir de <http://losdeportesenlasociedadguayaqui.blogspot.com/2017/11/el-campeonato-ecuatoriano-de-futbol.html>
- Chopde, N., & Nichat, M. (2013). Landmark Based Shortest Path Detection by Using A* and Haversine Formula. *GH Raison College of Engineering and ...*, 1(2), 298–302. <https://doi.org/10.1.1.300.5943>
- Comercio y Justicia. (2017). El uso de las redes sociales para prevenir el fraude en los seguros | Comercio y Justicia. Recuperado el 11 de junio de 2018, a partir de <https://comercioyjusticia.info/blog/opinion/el-uso-de-las-redes-sociales-para-prevenir-el-fraude-en-los-seguros/>
- Coronel, G. (2011). Redes Sociales Marcan Tendencias En La Comunicación. *Razón Y Palabra*, 77. Recuperado a partir de www.razonypalabra.org.mx
- Cotelo, J., Cruz, F., Ortega, J., & Troyano, J. (2015). Explorando Twitter mediante la integración

- de información estructurada y no estructurada. *Procesamiento de Lenguaje Natural*, 55, 75–82.
- Defensoría del Pueblo Ecuador. (2014). Eje derecho a la vida e integridad personal. Recuperado el 7 de noviembre de 2017, a partir de <http://www.dpe.gob.ec/derecho-a-la-vida-e-integridad-personal/>
- Django Software Foundation. (2015). Django. Recuperado el 23 de enero de 2018, a partir de <https://www.djangoproject.com/>
- El Universo. (2017). Policía detuvo 53 personas por tener armas en El Nacional vs Liga de Quito. Recuperado el 6 de noviembre de 2017, a partir de <http://www.eluniverso.com/deportes/2017/08/20/nota/6341119/policia-detuvo-53-personas-tener-armas-nacional-vs-liga-quito>
- Enciclopedia de Característica. (2017). Características de las Redes Sociales. Recuperado el 9 de junio de 2018, a partir de <https://www.caracteristicas.co/redes-sociales/>
- Federación Ecuatoriana de Fútbol. (s/f). Centro de Noticias. Recuperado el 6 de noviembre de 2017, a partir de http://www.ecuafutbol.org/web/noticias_seccion.php?sec=11
- Federación Ecuatoriana de Fútbol. (2009). Reglamento de calificación, administración y seguridad de los escenarios para la práctica del fútbol profesional. Guayaquil. Recuperado a partir de <http://www.ecuafutbol.org/web/reglamentos/10-ReglamentoCalificaciondeEstadios.pdf>
- Federación Ecuatoriana de Fútbol. (2017). Campeonato ecuatoriano 2017 copa Banco del Pacífico (calendario de juegos). Recuperado el 6 de noviembre de 2017, a partir de <http://ecuafutbol.org/web/noticia.php?idn=20446&idc=2#.WgD032jWxPa>
- González, A. (2014). ¿Qué es Machine Learning? Recuperado el 7 de noviembre de 2017, a partir de <http://cleverdata.io/que-es-machine-learning-big-data/>
- Guzmán, J. (2017). Calendario de la segunda etapa del campeonato ecuatoriano 2017. Recuperado el 6 de noviembre de 2017, a partir de <http://www.eluniverso.com/deportes/2017/07/14/nota/6279628/calendario-segunda-etapa-campeonato-ecuatoriano-2017>
- Instituto Nacional de Estadística y Censos. (2017). Tecnologías de la Información y Comunicación-TIC. Recuperado el 6 de noviembre de 2017, a partir de <http://www.ecuadorencifras.gob.ec//tecnologias-de-la-informacion-y-comunicacion-tic/>
- Jiménez, G., & Recio, J. (2014). Extracción, análisis y visualización de información social desde Twitter.
- Klokantech. (2010). Bounding Box Tool: Metadata Enrichment for Catalogue Records by Visually Selecting Geographic Coordinates (Latitude / Longitude) for Maps. Recuperado el 16 de enero de 2018, a partir de <http://boundingbox.klokantech.com/>
- Maldonado, S., & Weber, R. (2012). Modelos de selección de atributos para support vector machines.pdf.
- Mccrea, N. (2014). Introducción a la Teoría de Aprendizaje de Máquina y sus Aplicaciones: Un Tutorial Visual con Ejemplos | Toptal. Recuperado el 23 de mayo de 2018, a partir de <https://www.toptal.com/machine-learning/introducción-a-la-teoría-de-aprendizaje-de-máquina-y-sus-aplicaciones-un-tutorial-visual-con-ejemplos/es>

- Mifutbolecuador. (2014). Anuario 2014: Espectadores y taquilla por equipo. Recuperado el 6 de noviembre de 2017, a partir de <https://mifutbolecuador.wordpress.com/2014/12/28/anuario-2014-espectadores-y-taquilla-por-equipo/>
- Mifutbolecuador. (2016a). Anuario 2015: Espectadores y taquilla por equipo. Recuperado el 6 de noviembre de 2017, a partir de <https://mifutbolecuador.wordpress.com/2016/01/02/anuario-2015-espectadores-y-taquilla-por-equipo/>
- Mifutbolecuador. (2016b). Anuario 2016: Espectadores y taquilla por equipo. Recuperado el 6 de noviembre de 2017, a partir de <https://mifutbolecuador.wordpress.com/2016/12/25/anuario-2016-espectadores-y-taquilla-por-equipo/>
- Ministerio de Justicia Derechos Humanos y Cultos. (2014). *Código Orgánico Integral Penal* (1ra. Edi). Quito. Recuperado a partir de http://www.justicia.gob.ec/wp-content/uploads/2014/05/código_orgánico_integral_penal_-_coip_ed._sdn-mjdhc.pdf
- Ministerio del Deporte. (2010). Ley del Deporte, Educación Física y Recreación. Quito: [deporte.gob.ec](http://www.deporte.gob.ec). Recuperado a partir de <http://www.deporte.gob.ec/wp-content/uploads/downloads/2015/03/Ley-del-Deporte.pdf>
- Ministerio del Interior. (2012). Un llamado a terminar con la violencia en los estadios de fútbol- Ministerio del Interior. Recuperado el 6 de noviembre de 2017, a partir de <http://www.ministeriointerior.gob.ec/un-llamado-a-terminar-con-la-violencia-en-los-estadios-de-futbol/>
- MongoDB. (2009). MongoDB. Recuperado el 23 de enero de 2018, a partir de <https://www.mongodb.com/es>
- Poolman, E. (2011). EARLY WARNING SYSTEMS. *policy.practicalaction.org*, (July). Recuperado a partir de <https://policy.practicalaction.org/projects/ews>
- Puruncajas, Á. (2017). Los actos de violencia en el fútbol ecuatoriano en una década. Recuperado el 6 de noviembre de 2017, a partir de <http://www.elcomercio.com/deportes/violencia-futbol-ecuador-barrasbravas-deportes.html>
- Python Software Foundation. (2014). Python.org. Recuperado el 23 de enero de 2018, a partir de <https://www.python.org/>
- Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial*, 10(29), 11–18.
- Rodríguez, D., & Ugalde, O. (2012). Evaluación del uso de la red social Twitter. *ug.edu.ec*. Recuperado a partir de <http://repositorio.ug.edu.ec/bitstream/redug/2410/1/PROYECTO-TWITTER.pdf>
- Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. *Proceedings of the 19th International Conference on World Wide Web*, 851–860. <https://doi.org/10.1145/1772690.1772777>
- Tocci, B. (2016). Redes Sociales: Redes sociales. Recuperado el 6 de noviembre de 2017, a partir de <http://socialesredess.blogspot.com/2016/05/redes-sociales.html>
- Urueña, A., Ferrari, A., Blanco, D., & Valdecasa, E. (2011). Las Redes Sociales en Internet. *Observatorio nacional de las telecomunicaciones y de la SI*, 1–173. Recuperado a partir de

http://www.osimga.gal/export/sites/osimga/gl/documentos/d/20111201_ontsi_redes_sociais.pdf

Vargas, H., Tovar, M., & Villanueva, J. (2016). Los SAT (Sistemas De Alertas Tempranas) The SAT (Early Warning Systems), (26), 21–28.

Vásquez, A., Huerta, H., Quispe, J. P., & Huayna, A. M. (2009). Procesamiento de lenguaje natural. *Revista de investigación de Sistemas e Informática*, 6(2), 45–54. Recuperado a partir de <http://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/5923>

Vilches, E., & Escobar, I. (2007). Minería de datos.

Zapata, J. (2013). 38 acusados de rebelión por incidentes afuera del Independiente - Barcelona. Recuperado el 6 de noviembre de 2017, a partir de <http://www.ecuavisa.com/articulo/noticias/actualidad/42333-38-acusados-rebelion-incidentes-afuera-del-independiente>

ANEXOS



Anexo 1

FEDERACION ECUATORIANA DE FUTBOL
R.U.C. 099098665001
COPA BANCO DEL PACIFICO - 2017 - 2017 - División : MAYORES
CALENDARIO DE JUEGO - PRIMERA - A

Emisión: 18/01/2017
Usuario: FGOMEZ
Página: 1

		PRIMERA ETAPA - PRIMERA VUELTA					
FECHAS		AZUAY	EL ORO	GUAYAS	MANABI	PICHINCHA	TUNGURAHUA
PRIMERA	Domingo 29 de enero	D.CUENCA - BARCELONA S.C.		C.S.EMELEC - U.CATOLICA	DELFIN S.C. - RIVER PLATE	EL NACIONAL - INDEPENDIENTE DEL	
						CLAN JUVENIL - L.D.U.QUITO	MACARA - FUERZA AMARILLA S.C.
SEGUNDA	Domingo 05 de febrero		FUERZA AMARILLA S.C. - EL NACION	BARCELONA S.C. - MACARA			
				RIVER PLATE - C.S.EMELEC		U.CATOLICA - CLAN JUVENIL	
						INDEPENDIENTE DEL VALLE - D.CUE	
						L.D.U.QUITO - DELFIN S.C.	
TERCERA	Domingo 12 de febrero	D.CUENCA - FUERZA AMARILLA S.C.		C.S.EMELEC - L.D.U.QUITO			
				RIVER PLATE - INDEPENDIENTE DEL V	DELFIN S.C. - BARCELONA S.C.	EL NACIONAL - U.CATOLICA	
						CLAN JUVENIL - MACARA	
CUARTA	Domingo 26 de febrero		FUERZA AMARILLA S.C. - CLAN JUVE	BARCELONA S.C. - EL NACIONAL		U.CATOLICA - DELFIN S.C.	
						INDEPENDIENTE DEL VALLE - C.S.E	
						L.D.U.QUITO - RIVER PLATE	MACARA - D.CUENCA
QUINTA	Domingo 05 de marzo			C.S.EMELEC - MACARA			
				RIVER PLATE - U.CATOLICA	DELFIN S.C. - FUERZA AMARILLA S.	EL NACIONAL - D.CUENCA	
						CLAN JUVENIL - BARCELONA S.C.	
						L.D.U.QUITO - INDEPENDIENTE DEL V	
SEXTA	Domingo 12 de marzo	D.CUENCA - CLAN JUVENIL	FUERZA AMARILLA S.C. - C.S.EMEL	BARCELONA S.C. - RIVER PLATE		U.CATOLICA - L.D.U.QUITO	
						INDEPENDIENTE DEL VALLE - DELFI	MACARA - EL NACIONAL
SEPTIMA	Viernes 17 de marzo			C.S.EMELEC - D.CUENCA			
				RIVER PLATE - FUERZA AMARILLA S.	DELFIN S.C. - MACARA	U.CATOLICA - INDEPENDIENTE DEL V	
						CLAN JUVENIL - EL NACIONAL	
						L.D.U.QUITO - BARCELONA S.C.	
OCTAVA	Domingo 26 de marzo	D.CUENCA - DELFIN S.C.	FUERZA AMARILLA S.C. - L.D.U.QU	BARCELONA S.C. - U.CATOLICA		EL NACIONAL - C.S.EMELEC	
						CLAN JUVENIL - RIVER PLATE	MACARA - INDEPENDIENTE DEL VAL
NOVENA	Domingo 09 de abril			C.S.EMELEC - CLAN JUVENIL			
				RIVER PLATE - MACARA	DELFIN S.C. - EL NACIONAL	U.CATOLICA - FUERZA AMARILLA S.	
						INDEPENDIENTE DEL VALLE - BARCE	
						L.D.U.QUITO - D.CUENCA	
DECIMA	Domingo 16 de abril	D.CUENCA - U.CATOLICA	FUERZA AMARILLA S.C. - INDEPEND	C.S.EMELEC - BARCELONA S.C.		EL NACIONAL - RIVER PLATE	
						CLAN JUVENIL - DELFIN S.C.	MACARA - L.D.U.QUITO
DECIMA PRIMERA	Domingo 23 de abril			BARCELONA S.C. - FUERZA AMARIL			
				RIVER PLATE - D.CUENCA	DELFIN S.C. - C.S.EMELEC	U.CATOLICA - MACARA	
						INDEPENDIENTE DEL VALLE - CLAN J	

Figura # A1. Calendario de juego primera etapa-primera vuelta serie A, Copa Banco del Pacífico 2017

Fuente: (Federación Ecuatoriana de Fútbol, 2017)



FEDERACION ECUATORIANA DE FUTBOL
R.U.C. 0990986665001
COPA BANCO DEL PACIFICO - 2017 - 2017 - División : MAYORES
CALENDARIO DE JUEGO - PRIMERA - A

Emisión: 18/01/2017
Usuario: FGOMEZ
Página: 2

		PRIMERA ETAPA - SEGUNDA VUELTA					L.D.U.QUITO - EL NACIONAL
FECHAS	AZUAY	EL ORO	GUAYAS	MANABI	PICHINCHA	TUNGURAHUA	
DECIMA SEGUNDA	Domingo 30 de abril	D.CUENCA - RIVER PLATE	FUERZA AMARILLA S.C. - BARCELON	C.S.EMELEC - DELFIN S.C.		EL NACIONAL - L.D.U.QUITO	
						CLAN JUVENIL - INDEPENDIENTE DE	MACARA - U.CATOLICA
DECIMA TERCERA	Domingo 07 de mayo			BARCELONA S.C. - C.S.EMELEC			
			RIVER PLATE - EL NACIONAL		DELFIN S.C. - CLAN JUVENIL	U.CATOLICA - D.CUENCA	
						INDEPENDIENTE DEL VALLE - FUERZ	
						L.D.U.QUITO - MACARA	
DECIMA CUARTA	Domingo 14 de mayo	D.CUENCA - L.D.U.QUITO	FUERZA AMARILLA S.C. - U.CATOLI	BARCELONA S.C. - INDEPENDIENTE D		EL NACIONAL - DELFIN S.C.	
						CLAN JUVENIL - C.S.EMELEC	MACARA - RIVER PLATE
DECIMA QUINTA	Domingo 21 de mayo			C.S.EMELEC - EL NACIONAL			
				RIVER PLATE - CLAN JUVENIL	DELFIN S.C. - D.CUENCA	U.CATOLICA - BARCELONA S.C.	
						INDEPENDIENTE DEL VALLE - MACAR	
						L.D.U.QUITO - FUERZA AMARILLA S.C	
DECIMA SEXTA	Domingo 28 de mayo	D.CUENCA - C.S.EMELEC	FUERZA AMARILLA S.C. - RIVER PLA	BARCELONA S.C. - L.D.U.QUITO		EL NACIONAL - CLAN JUVENIL	
						INDEPENDIENTE DEL VALLE - U.CAT	MACARA - DELFIN S.C.
DECIMA SEPTIMA	Viernes 02 de junio			C.S.EMELEC - FUERZA AMARILLA S.			
				RIVER PLATE - BARCELONA S.C.	DELFIN S.C. - INDEPENDIENTE DEL V	EL NACIONAL - MACARA	
						CLAN JUVENIL - D.CUENCA	
						L.D.U.QUITO - U.CATOLICA	
DECIMA OCTAVA	Domingo 11 de junio	D.CUENCA - EL NACIONAL	FUERZA AMARILLA S.C. - DELFIN S.	BARCELONA S.C. - CLAN JUVENIL		U.CATOLICA - RIVER PLATE	
						INDEPENDIENTE DEL VALLE - L.D.U	MACARA - C.S.EMELEC
DECIMA NOVENA	Domingo 18 de junio	D.CUENCA - MACARA		C.S.EMELEC - INDEPENDIENTE DEL V			
				RIVER PLATE - L.D.U.QUITO	DELFIN S.C. - U.CATOLICA	EL NACIONAL - BARCELONA S.C.	
						CLAN JUVENIL - FUERZA AMARILLA S	
VIGESIMA	Domingo 25 de junio		FUERZA AMARILLA S.C. - D.CUENCA	BARCELONA S.C. - DELFIN S.C.		U.CATOLICA - EL NACIONAL	
						INDEPENDIENTE DEL VALLE - RIVER P	
						L.D.U.QUITO - C.S.EMELEC	MACARA - CLAN JUVENIL
VIGESIMA PRIMERA	Domingo 02 de julio	D.CUENCA - INDEPENDIENTE DEL VA		C.S.EMELEC - RIVER PLATE	DELFIN S.C. - L.D.U.QUITO	EL NACIONAL - FUERZA AMARILLA S	
						CLAN JUVENIL - U.CATOLICA	MACARA - BARCELONA S.C.
VIGESIMA SEGUNDA	Domingo 09 de julio		FUERZA AMARILLA S.C. - MACARA	BARCELONA S.C. - D.CUENCA			
				RIVER PLATE - DELFIN S.C.		U.CATOLICA - C.S.EMELEC	
						INDEPENDIENTE DEL VALLE - EL NAC	
						L.D.U.QUITO - CLAN JUVENIL	

Figura # A2. Calendario de juego primera etapa-segunda vuelta serie A, Copa Banco del Pacífico 2017

Fuente: (Federación Ecuatoriana de Fútbol, 2017)

CALENDARIO DE LA II ETAPA DEL CAMPEONATO NACIONAL				MARCADOR	
Fecha 1: 16 de julio de 2017 Universidad Católica vs. Independiente del Valle Clan Juvenil vs. Liga de Quito Club Sport Emelec vs. El Nacional Fuerza Amarilla vs. Delfin Sporting Club Macará vs. Barcelona Sporting Club Deportivo Cuenca vs. Guayaquil City		Fecha 9: 10 de septiembre de 2017 El Nacional vs. Delfin Sporting Club Independiente del Valle vs. Barcelona Sporting Club Liga de Quito vs. Macará Club Sport Emelec vs. Clan Juvenil Guayaquil City vs. Fuerza Amarilla Deportivo Cuenca vs. Universidad Católica		Fecha 17: 5 de noviembre de 2017 Liga de Quito vs. El Nacional Deportivo Cuenca vs. Independiente del Valle Guayaquil City vs. Barcelona Sporting Club Universidad Católica vs. Fuerza Amarilla Clan Juvenil vs. Macará Club Sport Emelec vs. Delfin Sporting Club	
Fecha 2: 23 de julio de 2017 El Nacional vs. Clan Juvenil Independiente del Valle vs. Macará Liga de Quito vs. Deportivo Cuenca Barcelona Sporting Club vs. Fuerza Amarilla Guayaquil City vs. Club Sport Emelec Delfin Sporting Club vs. Universidad Católica		Fecha 10: 17 de septiembre de 2017 Universidad Católica vs. Guayaquil City Clan Juvenil vs. Deportivo Cuenca Club Sport Emelec vs. Barcelona Sporting Club Fuerza Amarilla vs. Liga de Quito Macará vs. El Nacional Delfin Sporting Club vs. Independiente del Valle		Fecha 18: 12 de noviembre de 2017 Macará vs. Universidad Católica Barcelona Sporting Club vs. Clan Juvenil Independiente del Valle vs. Liga de Quito Fuerza Amarilla vs. Club Sport Emelec El Nacional vs. Guayaquil City Delfin Sporting Club vs. Deportivo Cuenca	
Fecha 3: 30 de julio de 2017 Universidad Católica vs. El Nacional Clan Juvenil vs. Fuerza Amarilla Club Sport Emelec vs. Liga de Quito Guayaquil City vs. Independiente del Valle Macará vs. Delfin Sporting Club Deportivo Cuenca vs. Barcelona Sporting Club		Fecha 11: 24 de septiembre de 2017 El Nacional vs. Fuerza Amarilla Independiente del Valle vs. Clan Juvenil Liga de Quito vs. Universidad Católica Barcelona Sporting Club vs. Delfin Sporting Club Guayaquil City vs. Macará Deportivo Cuenca vs. Club Sport Emelec		Fecha 19: 19 de noviembre de 2017 Deportivo Cuenca vs. El Nacional Club Sport Emelec vs. Independiente del Valle Guayaquil City vs. Liga de Quito Universidad Católica vs. Barcelona Sporting Club Macará vs. Fuerza Amarilla Clan Juvenil vs. Delfin Sporting Club	
Fecha 4: 6 de agosto de 2017 El Nacional vs. Deportivo Cuenca Independiente del Valle vs. Club Sport Emelec Liga de Quito vs. Guayaquil City Barcelona Sporting Club vs. Universidad Católica Fuerza Amarilla vs. Macará Delfin Sporting Club vs. Clan Juvenil		Fecha 12: 29 de septiembre de 2017 Universidad Católica vs. Liga de Quito Clan Juvenil vs. Independiente del Valle Club Sport Emelec vs. Deportivo Cuenca Fuerza Amarilla vs. El Nacional Macará vs. Guayaquil City Delfin Sporting Club vs. Barcelona Sporting Club		Fecha 20: 26 de noviembre de 2017 El Nacional vs. Universidad Católica Fuerza Amarilla vs. Clan Juvenil Liga de Quito vs. Club Sport Emelec Independiente del Valle vs. Guayaquil City Delfin Sporting Club vs. Macará Barcelona Sporting Club vs. Deportivo Cuenca	
Fecha 5: 13 de agosto de 2017 Universidad Católica vs. Macará Clan Juvenil vs. Barcelona Sporting Club Liga de Quito vs. Independiente del Valle Club Sport Emelec vs. Fuerza Amarilla Guayaquil City vs. El Nacional Deportivo Cuenca vs. Delfin Sporting Club		Fecha 13: 6 de octubre de 2017 El Nacional vs. Macará Independiente del Valle vs. Delfin Sporting Club Liga de Quito vs. Fuerza Amarilla Barcelona Sporting Club vs. Club Sport Emelec Guayaquil City vs. Universidad Católica Deportivo Cuenca vs. Clan Juvenil		Fecha 21: 3 de diciembre de 2017 Clan Juvenil vs. El Nacional Macará vs. Independiente del Valle Deportivo Cuenca vs. Liga de Quito Fuerza Amarilla vs. Barcelona Sporting Club Club Sport Emelec vs. Guayaquil City Universidad Católica vs. Delfin Sporting Club	
Fecha 6: 20 de agosto de 2017 El Nacional vs. Liga de Quito Independiente del Valle vs. Deportivo Cuenca Barcelona Sporting Club vs. Guayaquil City Fuerza Amarilla vs. Universidad Católica Macará vs. Clan Juvenil Delfin Sporting Club vs. Club Sport Emelec		Fecha 14: 15 de octubre de 2017 Universidad Católica vs. Deportivo Cuenca Clan Juvenil vs. Club Sport Emelec Barcelona Sporting Club vs. Independiente del Valle Fuerza Amarilla vs. Guayaquil City Macará vs. Liga de Quito Delfin Sporting Club vs. El Nacional		Fecha 22: 10 de diciembre de 2017 Independiente del Valle vs. Universidad Católica Liga de Quito vs. Clan Juvenil El Nacional vs. Club Sport Emelec Delfin Sporting Club vs. Fuerza Amarilla Barcelona Sporting Club vs. Macará Guayaquil City vs. Deportivo Cuenca	
Fecha 7: 25 de agosto de 2017 El Nacional vs. Independiente del Valle Clan Juvenil vs. Universidad Católica Liga de Quito vs. Barcelona Sporting Club Club Sport Emelec vs. Macará Guayaquil City vs. Delfin Sporting Club Deportivo Cuenca vs. Fuerza Amarilla		Fecha 15: 22 de octubre de 2017 El Nacional vs. Barcelona Sporting Club Independiente del Valle vs. Fuerza Amarilla Liga de Quito vs. Delfin Sporting Club Club Sport Emelec vs. Universidad Católica Guayaquil City vs. Clan Juvenil Deportivo Cuenca vs. Macará		Segunda fase • En esta fase se conocerá al posible segundo finalista. • Si Delfin repite el primer lugar será el campeón directo y jugará la Copa Libertadores. • El vicecampeón clasificará como Ecuador 2 a la Copa Libertadores. • En la clasificación final (sumados los puntos de la 2ª etapa) del 3º al 6º lugar clasificarán a la Copa Sudamericana.	
Fecha 8: 3 de septiembre de 2017 Universidad Católica vs. Club Sport Emelec Clan Juvenil vs. Guayaquil City Barcelona Sporting Club vs. El Nacional Fuerza Amarilla vs. Independiente del Valle Macará vs. Deportivo Cuenca Delfin Sporting Club vs. Liga de Quito		Fecha 16: 29 de octubre de 2017 Independiente del Valle vs. El Nacional Universidad Católica vs. Clan Juvenil Barcelona Sporting Club vs. Liga de Quito Macará vs. Club Sport Emelec Delfin Sporting Club vs. Guayaquil City Fuerza Amarilla vs. Deportivo Cuenca			

Figura # A3. Calendario de juego segunda etapa serie A, Copa Banco del Pacífico 2017

Fuente: (Guzmán, 2017)