



UNIVERSIDAD DEL BÍO-BÍO
FACULTAD CIENCIAS EMPRESARIALES

SOLICITUD INSCRIPCIÓN ACTIVIDAD DE TITULACIÓN

1. Identificación Alumno (S).

| | |
|-----------|------------------------------------|
| Nombre | : Francisco Fernando Muñoz Coletti |
| Dirección | : O'Higgins 1500 Quillón |
| Teléfono | : +56 9 4290 8358 |
| E-Mail | : franmucol16@gmail.com |
| Carrera | : Ingeniería Civil Informática |
| Depto. | : Sistemas de Información |

2. Evaluadores

2.1 Profesor guía

Nombre : Pedro Gerónimo Campos Soto

Firma:

3. Proyecto

3.1 Título Que Identificará La Actividad De Titulación.

Análisis de comentarios usuarios de la red social Twitter en temáticas de **contingencia nacional**.

3.2 Objetivos Generales Y Específicos De La Actividad De Titulación.

3.2.1 Objetivos General

Categorizar opiniones de usuarios de Twitter en **el territorio Chileno** sobre temas de interés nacional mediante Machine Learning.

3.2.1 Objetivos Específicos

- 1 Revisar literatura relacionada a análisis de textos en redes sociales y uso de la API de Twitter.
- 2 Identificar y filtrar usuarios y sus respectivos Tweets pertenecientes a territorios específicos dentro de Chile mediante la **API de Twitter**.
- 3 Diseñar y entrenar **modelo de clasificación** de temas de interés nacional sobre Tweets obtenidos
- 4 Implementar **interfaz gráfica** para visualizar los resultados.

3.3 Justificación Del Proyecto Propuesto.

Las redes sociales en la actualidad poseen un muy alto flujo de información, sea formal o informal. Alrededor del 60.1% de la población mundial actualmente utiliza internet(4.7 billones de personas). De estos usuarios, el 92.12% utiliza plataformas de redes sociales activamente. (**Kemp. S., 2021**). Si se observa exclusivamente en Chile, un 83.5% de su población(16 Millones) son usuarios activos en redes sociales. (**Alvino. C., 2021**).

El uso de dichas redes va desde el ocio, pasando por columnas de **opinión hasta incluso** denuncias ciudadanas sobre problemáticas que se viven día a día en nuestro país. Estas plataformas se han convertido actualmente en la voz del pueblo chileno y su uso se intensificó aun más post 18 de Octubre de 2019 (en comparación a meses anteriores) (**Ferreira Sandoval. M., 2020**). Sin embargo esta información queda registrada, mas no organizada para el posterior

estudio de la misma por las dificultades que se presentan a la hora de clasificarla: Cuentas fraudulentas, información falsa, el alto costo que significaría tener un equipo encargado de recolectar y clasificar dichas opiniones, entre otras.

Es por todo lo anteriormente mencionado que en este proyecto de tesis se propone: Desarrollar un sistema automatizado basado en Machine Learning para obtener y clasificar una mayor cantidad de opiniones de usuarios de la red social Twitter con respecto a temas de interés nacional y social acorde a territorios específicos del país. Adicionalmente esto permitiría facilitar y reducir costos a empresas dedicadas a la investigación de opinión (por ejemplo: CADEM, IPSOS, CEP, Etc.)

3.4 Actividades a realizar

- 1 Revisar literatura sobre análisis de textos en redes sociales, manejo de la API de Twitter mediante Tweepy, obtención de lugares geográficos mediante formato GeoJSON, Opinion Mining, Python 3 y Deep Learning.
- 2 Recopilar datos requeridos desde API de Twitter (Streaming y Rest) y almacenarlos en una base de datos local.
- 3 Generar reglas para filtrar a los potenciales usuarios fraudulentos o cuentas de entidades
- 4 Desarrollar módulo para clasificar los Tweets en las distintas regiones del país.
- 5 Encontrar y entrenar el modelo de Machine Learning a utilizar.
- 6 Realizar un procesamiento de lenguaje natural en los Tweets mediante modelo de machine learning.
- 7 Graficar los resultados obtenidos posterior al procesamiento realizado.

3.5 Plan De Trabajo A Desarrollar

| Actividad | Rango de fechas |
|--|-------------------------|
| Revisar literatura sobre análisis de textos en redes sociales, manejo de la API de Twitter mediante Tweepy, obtención de lugares geográficos mediante formato GeoJSON, Opinion Mining, Python 3 y Deep Learning. | 03/09/2021 - 20/10/2021 |
| Recopilación de datos requeridos desde API de Twitter (Streaming y Rest) | 15/09/2021 - 15/10/2021 |

| | |
|---|--|
| Generar reglas para filtrar a los potenciales usuarios fraudulentos o cuentas de entidades | 15/10/2021 - 15/11/2021 |
| Desarrollar módulo para clasificar los Tweets en las distintas regiones del país. | 15/11/2021 - 15/12/2021 |
| Encontrar y entrenar el modelo de Machine Learning a utilizar. | 15/12/2021 – 15/04/2022 (Aprox 10 semanas, dejando enero y febrero como vacaciones) |
| Realizar un procesamiento de lenguaje natural en los Tweets mediante modelo de machine learning. | 15/04/2022 - 15/05/2022 |
| Graficar los resultados obtenidos posterior al procesamiento realizado. | 15/05/2022 - 15/06/2022 |

3.6 Trabajos Similares Realizados Previamente.

Ferreira Sandoval, M. Análisis de cambios en comentarios de usuarios de la red social Twitter a partir de la crisis social en Chile. Septiembre, 2020.

Las temáticas a estudiar en este proyecto de tesis son variadas, no solo enfocadas a alguna contingencia con fecha y tiempo (pandemia por ejemplo) sino abordar temas como salud, educación, vivienda, etc. Adicionalmente, el proyecto realizado por el compañero Miguel Ferreria solo analizaba el cambio de actividad dentro de la red social antes y después del estallido, mas no clasificaba las opiniones recopiladas.

3.6 Descripción De Los Aspectos Fundamentales De La Metodología A Utilizar (Software)

Este proyecto genera la necesidad de tener una alta cantidad de filtros, reglas y pruebas para que el estudio sea lo más preciso posible. Es por esto que se utilizará la metodología espiral para el desarrollo del Software, donde se intenta reducir al mínimo los fallos, realizar constantemente una planificación y gestión de estos para definir cuales presentan más riesgos y buscar estrategias para lidiar con ellos. Al final de cada **proceso** espiral se generará un bloque entregable el cual será evaluado en conjunto al profesor guía. ~~Posteriormente, se iniciaría nuevamente un proceso espiral para la siguiente actividad a realizar.~~ Una vez el software esté en las etapas finales, se realizará una última evaluación antes de dejar el software operativo y listo para pasar

a la fase de mantención.

4. Bibliografía

4.1 Referencia bibliográfica

- 1 Ferreira Sandoval, M. (Septiembre, 2020). Análisis de cambios en comentarios de usuarios de la red social Twitter a partir de la crisis social en Chile.
- 2 Vinodhini, G., & Chandrasekaran, R.M. (2012). Sentiment Analysis and Opinion Mining: A Survey(6ªed., Vol.2).
- 3 Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S., Schaub, T., & Wilde, E. (2016). *The GeoJSON Format*.
- 4 Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.

4.2 Linkografía

- 1 Brownlee, J. (2019, 16 septiembre). *A Gentle Introduction to Transfer Learning for Deep Learning*. Machine Learning Mastery. Recuperado el 12 de septiembre de 2021, de <https://machinelearningmastery.com/transfer-learning-for-deep-learning/>
- 2 Education, I. C. (2021, 12 agosto). *Machine Learning*. IBM. Recuperado el 12 de septiembre de 2021, de <https://www.ibm.com/cloud/learn/machine-learning>
- 3 Twitter, Inc. (s. f.). *Getting Started with Our Platform*. Docs | Twitter Developer Platform. Recuperado 8 de septiembre de 2021, de <https://developer.twitter.com/en/docs/getting-started>
- 4 Kemp, S. (2021, 21 abril). *60 percent of the world's population is now online*. We Are Social. <https://wearesocial.com/blog/2021/04/60-percent-of-the-worlds-population-is-now-online>
- 5 Alvino, C. (2021, 3 junio). *Estadísticas de la situación digital de Chile en el 2020–2021*. Branch. <https://branch.com.co/marketing-digital/estadisticas-de-la-situacion-digital-de-chile-en-el-2020-2021/>

LA PRESENTE SOLICITUD DE INSCRIPCIÓN DE LA ACTIVIDAD DE TITULACIÓN
SIGNIFICA UN COMPROMISO DE CUMPLIR LO ESTIPULADO EN ELLA.

.....
FIRMA ALUMNO

FECHA PRESENTACIÓN SOLICITUD:

| RESOLUCIÓN DIRECTOR DE ESCUELA/JEFE DE CARRERA | |
|--|--------------------------|
| APROBADO | <input type="checkbox"/> |
| OBSERVACIONES: | |
| | |
| | |
| | |
| <u>Linkografía</u> | |
| FECHA RESOLUCIÓN: | |