

**fcfm**

Ingeniería Matemática

FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

# Procesamiento de datos textuales de EthicApp

Presentación del dataset

**Camilo Carvajal Reyes**

**18 de mayo,  
2023**



# Índice de contenidos

## 1 Resumen

## 2 Datos

- Datos textuales de EthicApp
- Características caso Julieta

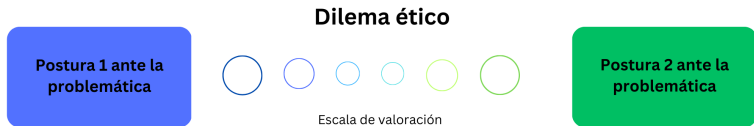
## 3 Procesamiento de texto de EthicApp

- Procesamiento de lenguaje natural
- Uso de NLP en estudio de moralidad
- Análisis global de respuestas
- Predicción de respuestas
- Modelos a utilizar

## 4 Referencias

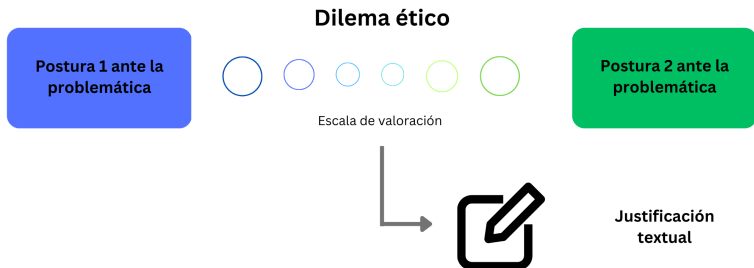
# Resumen

En el marco de actividades ética en cursos iniciales de la FCFM, estudiantes evalúan en una escala de 1 a 6 las respuestas a una **dilema**, a través de la aplicación *EthicApp* [1]. En seguida escriben una **justificación** a tal decisión. Este texto puede contener **información relevante** de la decisión y su estudio es importante para los equipos docentes y el área de ética. Este análisis se dificulta por la **gran cantidad de respuestas**.



# Resumen

En el marco de actividades ética en cursos iniciales de la FCFM, estudiantes evalúan en una escala de 1 a 6 las respuestas a una **dilema**, a través de la aplicación *EthicApp* [1]. En seguida escriben una **justificación** a tal decisión. Este texto puede contener **información relevante** de la decisión y su estudio es importante para los equipos docentes y el área de ética. Este análisis se dificulta por la **gran cantidad de respuestas**.

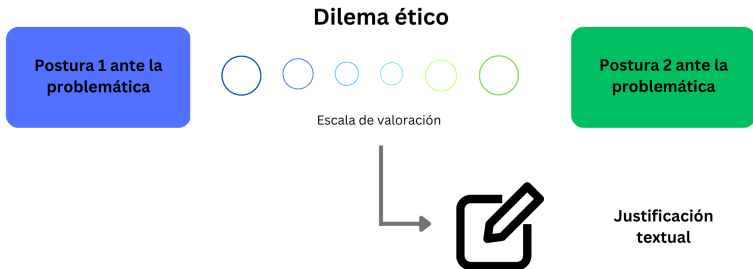




# Resumen

Se plantea la utilización de algoritmos de procesamiento de lenguaje natural para:

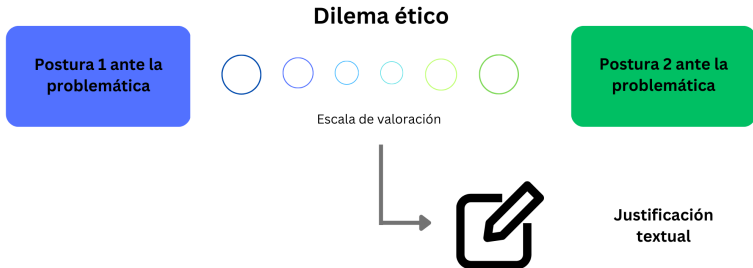
- 1 evaluar la progresión de competencia ética de los estudiantes con menor inversión humana,
- 2 estudiar las capacidades de algoritmos de texto de modelar ética.



# Ejemplo de caso

En el último control realizado **Julietta** se ve en la posibilidad de copiar una respuesta que fue compartida en el grupo de WhatsApp de su sección. Julietta en esta situación a la que se ve enfrentada en el control debiera

- 1 Usar la información del grupo de WhatsApp
- 6 No usar la información del grupo de WhatsApp





# Índice de contenidos

## 1 Resumen

## 2 Datos

- Datos textuales de EthicApp
- Características caso Julieta

## 3 Procesamiento de texto de EthicApp

- Procesamiento de lenguaje natural
- Uso de NLP en estudio de moralidad
- Análisis global de respuestas
- Predicción de respuestas
- Modelos a utilizar

## 4 Referencias

# Datos textuales de EthicApp

## Contenidos del dataset:

- Identificador de persona
- Género
- Identificador de grupo
- Caso de estudio y opciones
- Número de etapa
- Valoración y texto

Identificador	N. de grupo	N. de etapa	Genero
1420	14	3	F

### Escala de valoración



### Comentario

Adela debería considerar que el sector es de escasez hídrica, por lo que un gran consumo de agua podría tener graves repercusiones en la comunidad, y a la larga podría significar que este producto no se pueda generar en unos años más. Por otro lado, se pueden buscar otras formas de obtener vitamina D para niños y ancianos. Por ello, considero que se debería priorizar preservar el recurso natural.



# Datos textuales de EthicApp

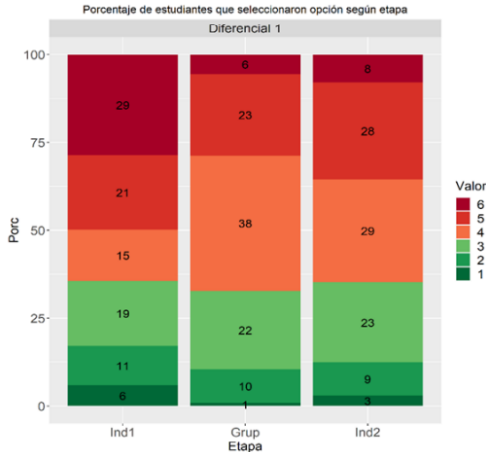
## Cantidad de datos por caso

Caso	Cursos	Cantidad estudiantes	Cantidad grupos
Caso Julieta	1	819	247
Caso Adela	1	237	142
Caso Laura	1	602	335
Caso Alicia	2	1628	549



# Características de caso Julieta

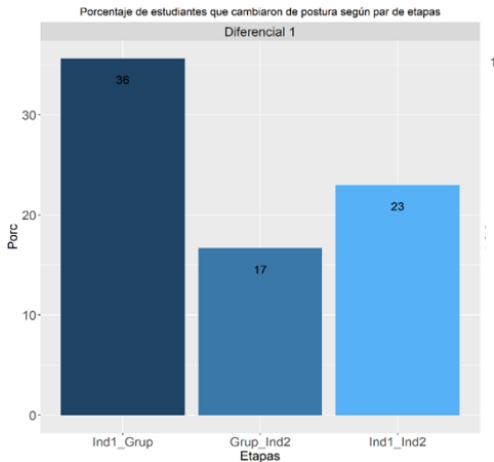
Porcentaje de estudiantes por opción [2].





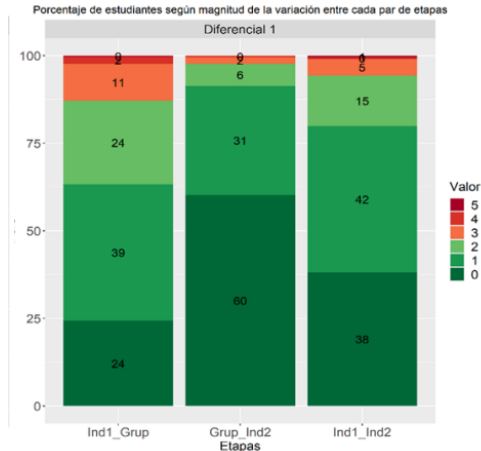
# Características de caso Julieta

Porcentaje de cambios de postura [2].



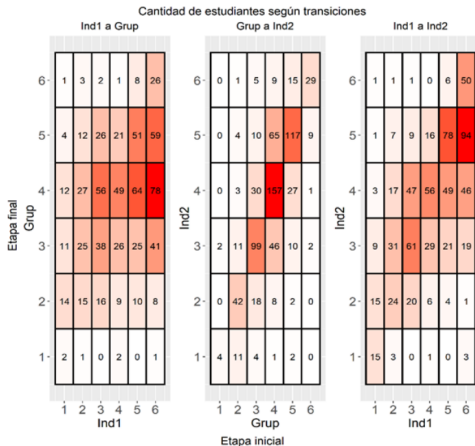
# Características de caso Julieta

Magnitud de cambio de postura por par de etapas [2].



# Características de caso Julieta

Detalle de cambios de postura por par de etapas [2].

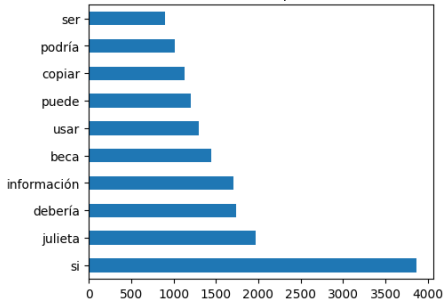




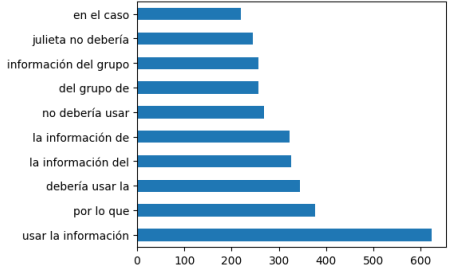
# n-gramas comunes caso Julieta

Preliminarmente mostramos los 1-gramas y 3-gramas más frecuentes del dataset para el caso Julieta.

Tokens más frecuentes para comentarios



Tokens más frecuentes para comentarios





# Índice de contenidos

## 1 Resumen

## 2 Datos

- Datos textuales de EthicApp
- Características caso Julieta

## 3 Procesamiento de texto de EthicApp

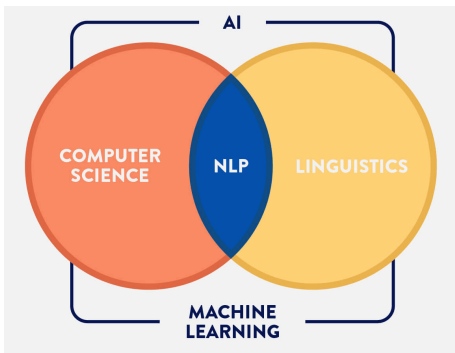
- Procesamiento de lenguaje natural
- Uso de NLP en estudio de moralidad
- Análisis global de respuestas
- Predicción de respuestas
- Modelos a utilizar

## 4 Referencias

# Procesamiento de lenguaje natural

El procesamiento de lenguaje natural (NLP) es una área que incluye tanto la **lingüística** como la **computación**.

El alza en poder de cómputo y desarrollo de aprendizaje profundo a **aumentado sus capacidades**. Actualmente modelos neuronales son utilizados en una gran gama de tareas. Su utilidad para extraer elementos psicológicos (y de imitarlos) está actualmente en estudio.





# Uso de NLP en estudio de moralidad

## Modelos de lenguaje que reflejan la moralidad humana

Varios son los trabajos que buscan la predicción/identificación de elementos morales en texto natural. En esta sección abordaremos algunos de ellos. Primeramente, Garten et al. 2016 busca la detección automática de retóricas morales [3]. Para esto usan el diccionario de fundamentos morales [4] (de la Teoría de fundamentos morales TMF [5]), combinado con word embeddings. Otro trabajo similar es el de Xie et al. que evalúa modelos en la clasificación de dilemas morales en los distintos fundamentos, gracias a lo cual concluyen que modelos de lenguaje tienen ventaja sobre modelos como las representaciones distribucionales [6].

TFM: (cuidado/daño, equidad/trampa, lealtad/fraude, autoridad/subversión, santidad/degradación y libertad/opresión)

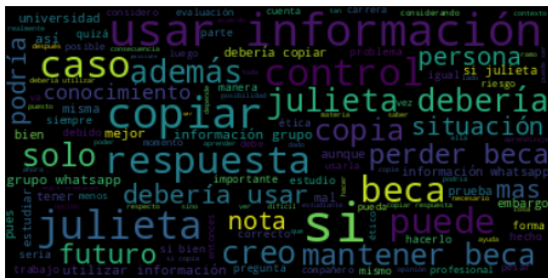
# Uso de NLP en estudio de moralidad

## Estudios de la moralidad usando NLP

Por otro lado, Kennedy et al. buscan la predicción de preocupaciones morales propias a un individuo usando evidencias de lenguaje moral escritas por este [7] (estados de Facebook de usuarios que hayan contestado el cuestionario de fundamentos morales). Se utilizaron distintas técnicas de procesamiento de lenguaje para predecir los puntajes obtenidos por los usuarios, para cada una de las dimensiones morales planteadas en la TMF [5]. Se destacan la variedad de métodos para vectorizar texto testeados, incluyendo *latent dirichlet allocation* (LDA) [8], *word embeddings*, conteo de ocurrencias del DFM [4] y BERT [9], que corresponde a un modelo de lenguaje profundo. Es este último que obtiene mejores resultados. Finalmente, tanto conteos de diccionario como LDA se usaron para interpretar que elementos lingüísticos específicos explicaban cada fundamento por separado.

## Análisis global de respuestas

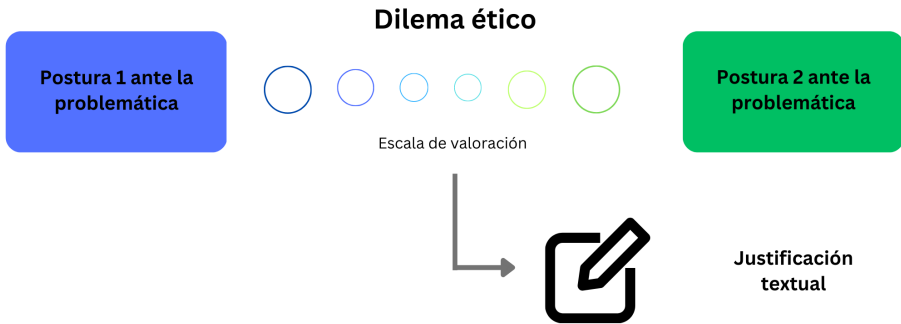
Una exploración básica del texto nos da una noción inicial del tipo de justificación usada.



Se pretende llevar a cabo un estudio frecuencias y correlaciones de variables como postura, etapa de respuesta, dilema moral y elementos semánticos empleados.

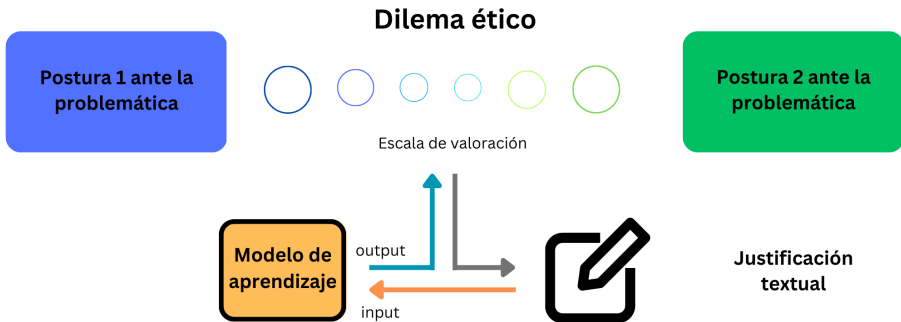
# Predicción de respuestas de estudiantes

Se propone la utilización de modelos para la predicción de valoración de la problemática, usando el texto. Hacer esto con modelos interpretables nos dará una idea de que **elementos lingüísticos** se usaron para escoger tal opción. Por otro lado se plantea usar modelos profundos con fines exploratorios.



# Predicción de respuestas de estudiantes

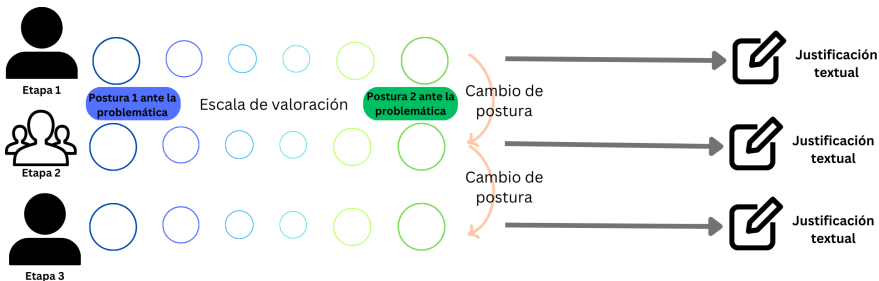
Se propone la utilización de modelos para la predicción de valoración de la problemática, usando el texto. Hacer esto con modelos interpretables nos dará una idea de que **elementos lingüísticos** se usaron para escoger tal opción. Por otro lado se plantea usar modelos profundos con fines exploratorios.



# Predicción de cambio en respuesta

Similarmente, queremos usar modelos similares para predecir cambios en las valoraciones de una etapa a otra. Esto nos permite estudiar que **elementos** son **comunes** en un futuro **cambio de postura o valoración**.

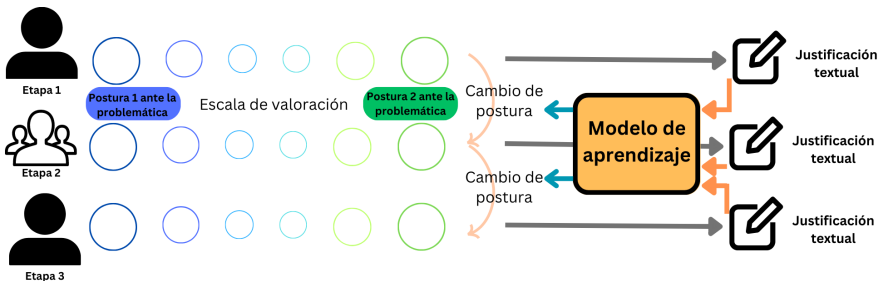
## Dilema ético



# Predicción de cambio en respuesta

Similarmente, queremos usar modelos similares para predecir cambios en las valoraciones de una etapa a otra. Esto nos permite estudiar que **elementos** son **comunes** en un futuro **cambio de postura o valoración**.

## Dilema ético





# Modelos con interpretabilidad

- **Topic modelling** (Latent Dirichlet allocation - LDA)

Es una técnica que agrupa de manera no supervisada los textos. Genera una distribución palabra - tópico latente (oculto) y tópico - palabra.



# Modelos con interpretabilidad

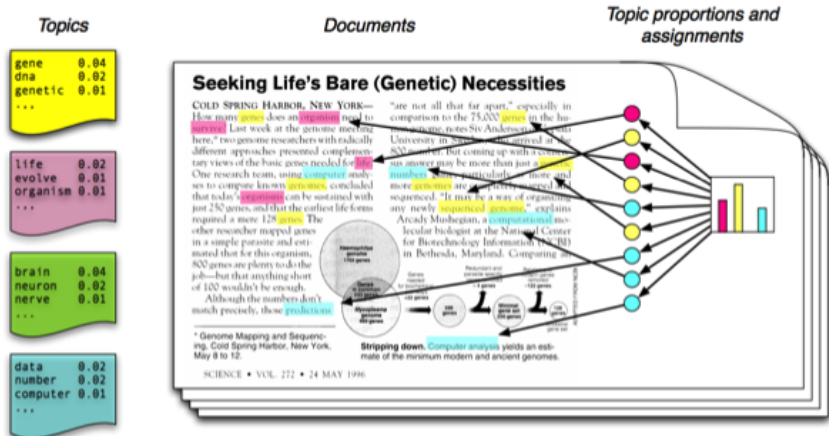


Figure source: Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.



# Modelos con interpretabilidad

- **Topic modelling** (Latent Dirichlet allocation - LDA)

Es una técnica que agrupa de manera no supervisada los textos. Genera una distribución palabra - tópico latente (oculto) y tópico - palabra.

- **Naive-Bayes:**

Modelo de clasificación que asigna a cada elemento (palabra) una probabilidad de pertenecer a una clase. Las probabilidades se suman para la predicción final.

# Modelos de aprendizaje profundo

Word embeddings:  
modelos que  
permiten vectorizar  
palabras, basados en  
su co-ocurrencia.

Algunos ejemplos:

- Word2vec
- GloVe
- Td-idf

Se pueden combinar  
con diversos  
modelos de  
clasificación.



# Modelos de aprendizaje profundo

BETO/BERT:

modelo profundo

basado en la

arquitectura

Transformers (como

es el caso de

ChatGPT). Son

pre-entrenados en

grandes corpuses de

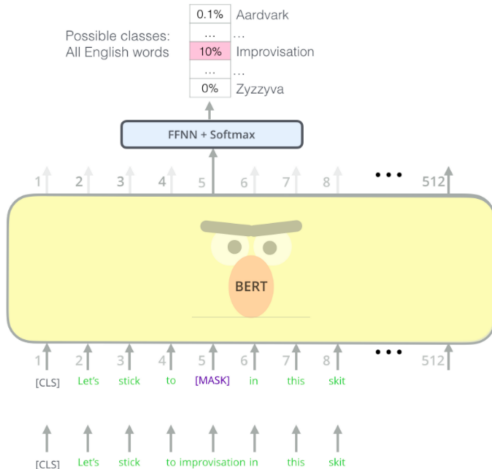
texto y se pueden

usar en variadas

tareas de NLP.

Permiten vectorizar

palabras y texto.





# Referencias I



Alvarez, C., Zurita, G., Hasbún, B., Peñafiel, S., Pezoa, Á., Alvarez, C., Zurita, G., Hasbún, B., Peñafiel, S., Pezoa, Á. (2021). A Social Platform for Fostering Ethical Education through Role-Playing. In Factoring Ethics in Technology, Policy Making, Regulation and AI. IntechOpen. <https://doi.org/10.5772/intechopen.96602>



Ramírez Rivas, P., Guerrero, S., Cerda Maureira, J., Ross, J. P., Flores Mandeville, G. (2022). La formación ética canalizada mediante la tecnología. Experiencia y resultados preliminares del uso de la herramienta web Ethicapp. XXXIV Congreso Chileno de Educación en Ingeniería.



Garten, J., Boghrati, R., Hoover, J., Johnson, K. M., Dehghani, M. (2016). Morality Between the Lines: Detecting Moral Sentiment In Text. Proceedings of IJCAI 2016 Workshop on Computational Modeling of Attitudes.



Graham, J., Haidt, J., Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. Journal of Personality and Social Psychology, 96, 1029–1046. <https://doi.org/10.1037/a0015141>



Haidt, J. (2007). The New Synthesis in Moral Psychology. Science, 316(5827), 998–1002. <https://doi.org/10.1126/science.1137651>



# Referencias II



Xie, J. Y., Hirst, G., Xu, Y. (2020). Contextualized moral inference (arXiv:2008.10762). arXiv. <https://doi.org/10.48550/arXiv.2008.10762>



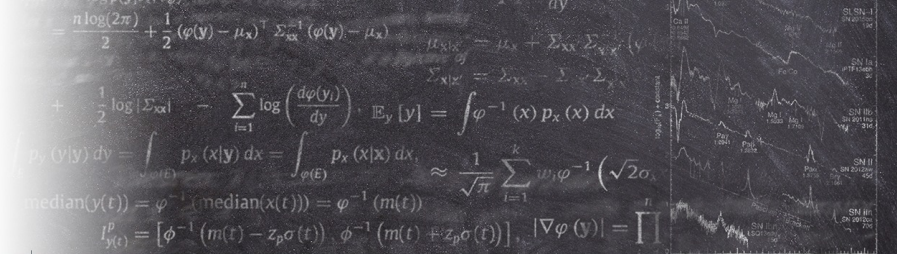
Kennedy, B., Atari, M., Mostafazadeh Davani, A., Hoover, J., Omrani, A., Graham, J., Dehghani, M. (2021). Moral concerns are differentially observable in language. Cognition, 212, 104696. <https://doi.org/10.1016/j.cognition.2021.104696>



Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation. The Journal of Machine Learning Research, 3(null), 993–1022.



Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186. <https://doi.org/10.18653/v1/N19-1423>



Ingeniería Matemática

FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

# Procesamiento de datos textuales de EthicApp

Presentación del dataset

**Camilo Carvajal Reyes**

**18 de mayo,  
2023**