

# Matías Rojas

✉ [matirojasga@gmail.com](mailto:matirojasga@gmail.com)  
🌐 <https://github.com/matirojasg>  
in <https://www.linkedin.com/in/matirojasg/>

## Educación

- Marzo 2015 - **Magíster en Ciencias de la Computación**, *Universidad de Chile*, Reconocimiento de entidades médicas anidadas en diagnósticos de la lista de espera en hospitales públicos. Distinción máxima.  
Abril 2022
- Marzo 2015 - **Ingeniería Civil en Computación**, *Universidad de Chile*, Reconocimiento de entidades anidadas en diagnósticos de la lista de espera en hospitales públicos. Distinción máxima.  
Abril 2022

## Experiencia Laboral

- Enero 2022 - **Data Scientist**, Grupo de Procesamiento de Lenguaje Natural en Centro de Modelamiento Matemático, Universidad de Chile. Santiago, Chile.  
Actualidad
- Diciembre 2019 - **Data Engineer**, Worldalytics. Santiago, Chile, Diseño e implementación de una arquitectura de big data montada en Google Cloud Platform, junto con rediseño y optimización de procedimientos analíticos asociados a la contactabilidad del cliente y mejora de campañas segmentadas para una empresa de retail.  
Septiembre 2020
- Enero 2019 - **Data Engineer**, Enternet S.A. Santiago, Chile, Desarrollo de una aplicación que maneja la configuración correcta de varios dispositivos instalados en clientes (IOT) en un registro.  
Febrero 2019

## Publicaciones

- 2022 Matías Rojas, Jocelyn Dunstan, and Fabián Villena. Clinical Flair: A Pre-Trained Language Model for Spanish Clinical Natural Language Processing. In *Proceedings of the 4th Clinical Natural Language Processing Workshop NAACL*, 2022.
- 2022 Matias Rojas, Casimiro Pio Carrino, Aitor Gonzalez-Agirre, Jocelyn Dunstan, and Marta Villegas. Assessing the limits of straightforward models for nested named entity recognition in Spanish clinical narratives. In *Proceedings of the 13th International Workshop on Health Text Mining and Information Analysis (LOUHI)*, pages 14–25, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.
- 2022 Matias Rojas, Felipe Bravo-Marquez, and Jocelyn Dunstan. Simple yet powerful: An overlooked architecture for nested named entity recognition. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2108–2117, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.
- 2022 Matias Rojas, Jose Barros, Kinan Martin, Mauricio Araneda-Hernandez, and Jocelyn Dunstan. PLN CMM at SocialDisNER: Improving detection of disease mentions in tweets by using document-level features. In *Proceedings of The Seventh Workshop on Social Media Mining for Health Applications, Workshop & Shared Task*, pages 52–54, Gyeongju, Republic of Korea, October 2022. Association for Computational Linguistics.
- 2022 Matías Rojas, Jose Barros, Mauricio Araneda, and Jocelyn Dunstan. Flert-matcher: A two-step approach for clinical named entity recognition and normalization. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2022)*, A Coruña, Spain, September 20, 2022, volume 3202 of *CEUR Workshop Proceedings*, 2022.

- 2022 Jose Barros, Matias Rojas, Jocelyn Dunstan, and Andres Abeliuk. Divide and conquer: An extreme multi-label classification approach for coding diseases and procedures in Spanish. In *Proceedings of the 13th International Workshop on Health Text Mining and Information Analysis (LOUHI)*, pages 138–147, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.
- 2022 Pablo Báez, Felipe Bravo-Marquez, Jocelyn Dunstan, Matías Rojas, and Fabián Villena. Automatic extraction of nested entities in clinical referrals in spanish. *ACM Trans. Comput. Healthcare*, volume 3. Association for Computing Machinery, apr 2022.
- 2022 Claudio Aracena, Fabián Villena, Matias Rojas, and Jocelyn Dunstan. A knowledge-graph-based intrinsic test for benchmarking medical concept embeddings and pretrained language models. In *Proceedings of the 13th International Workshop on Health Text Mining and Information Analysis (LOUHI)*, pages 197–206, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.
- 2021 Fabián Villena, Pablo Báez, Sergio Peñafiel, Matías Rojas, Inti Paredes, and Jocelyn Dunstan. Automatic support system for tumor coding in pathology reports in spanish. *SSRN Electronic Journal*. Elsevier BV, 2021.
- 2020 Pablo Báez, Fabián Villena, Matías Rojas, Manuel Durán, and Jocelyn Dunstan. The Chilean waiting list corpus: a new resource for clinical named entity recognition in Spanish. In *Proceedings of the 3rd Clinical Natural Language Processing Workshop*, pages 291–300, Online, November 2020. Association for Computational Linguistics.

---

## Docencia - Universidad de Chile

- 2022 **Profesor Laboratorio de Programación Científica para Ciencia de Datos**, Magister en Ciencia de Datos.
- 2021 **Auxiliar Procesamiento de Lenguaje Natural**, Departamento de Ciencias de la Computación.
- 2021 **Ayudante Pensamiento Estadístico**, Departamento de Ciencias de la Computación.
- 2021 **Ayudante Minería de Datos**, Departamento de Ciencias de la Computación.
- 2020 **Auxiliar Proyecto de Ciencias de Datos**, Departamento de Ciencias de la Computación.
- 2020 **Ayudante Lenguajes de Programación**, Departamento de Ciencias de la Computación.
- 2019-2021 **Ayudante Modelamiento y Computación Gráfica (x2)** Departamento de Ciencias de la Computación.
- 2019-2021 **Ayudante Diseño y Análisis de Algoritmos (x3)**, Departamento de Ciencias de la Computación.
- 2019 **Auxiliar Herramientas Computacionales para Ingeniería y Ciencias**, Departamento de Ciencias de la Computación.
- 2018 **Ayudante Teoría de la Computación**, Departamento de Ciencias de la Computación.
- 2017 **Auxiliar Cálculo Avanzado y Aplicaciones (x2)**, Departamento de Ingeniería Matemática.
- 2016-2017 **Ayudante Ecuaciones Diferenciales Ordinarias (x2)**, Departamento de Ingeniería Matemática.

---

## Proyectos Aplicados

- 2022 **Colaboración con Ministerio de Salud**, *Diseño de un sistema para la codificación automática de enfermedades en pacientes de la lista de espera.*

- 2022 **Colaboración con Centro Nacional de Supercomputación de Barcelona**, *Diseño y desarrollo de modelos de extracción automática de entidades anidadas en texto clínicos en Español. Creación de modelos del lenguaje clínicos en Español.*
- 2021-2022 **Reto Corfo CMF**, *Diseño e implementación de un sistema basado en técnicas de aprendizaje profundo para la clasificación automática de reclamos en el mercado financiero.*

## Habilidades

|              |  |
|--------------|--|
| Áreas        | Procesamiento de Lenguaje Natural, Aprendizaje de Máquinas, Aprendizaje Profundo, Extracción de Información, Modelos del Lenguaje. |
| Programación | Python, Java y R.  |
| Librerías    | Pytorch, Tensorflow y Flair.   |
| Idiomas      | Español (nativo), Inglés (nivel intermedio).   |

## Reconocimientos

- 2021 100 Jóvenes Líderes 2021. Expedida por El Mercurio, Revista Sábado y Universidad Adolfo Ibañez.
- 2019 Alumno destacado en Ciencias de la Computación. Universidad de Chile, Santiago, Chile.
- 2018 Alumno destacado en Ciencias de la Computación. Universidad de Chile, Santiago, Chile.

## Referencias

### Jocelyn Dunstan

*Profesora guía tesis de magíster*  
Universidad de Chile  
✉ jdunstan@uchile.cl

### Felipe Bravo

*Profesor guía tesis de magíster*  
Universidad de Chile  
✉ fbravo@dcc.uchile.cl

### Cristian Muenas

*Gerente de Big Data & Analytics*  
*Worldalytics*  
✉ cmuenas@worldalytics.cl