

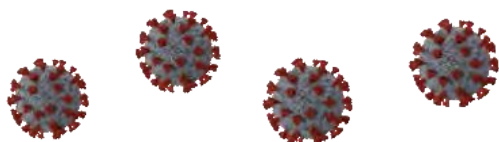


# Aprendizaje automático aplicado a la pandemia del virus Covid-19 en Argentina

Carolina Cardoso, Lorena Talamé, Matías Amor  
Universidad Católica de Salta

## RESUMEN

Este proyecto se desarrolló con el interés de analizar los casos registrados sobre Covid-19 en nuestro país, publicados por fuentes oficiales. Se experimentó con redes neuronales con el fin de predecir casos positivos de la enfermedad, y para encontrar similitudes entre algunos distritos de nuestro país se plantearon relaciones difusas.



## LINEA DE INVESTIGACION

El proyecto se desarrolló en las siguientes etapas:

- Revisión de la literatura referente a las técnicas de aprendizaje automático aplicadas al problema, en reuniones periódicas del equipo.
- Descarga y almacenamiento de los datos, provenientes de fuentes oficiales (Ministerio de Salud de Nación Argentina.)
- Pre-procesamiento de los datos: filtrado, limpieza y selección de datos para entrenamiento y para control del modelo.
- Aplicación de técnicas y algoritmos adecuados para el problema del proyecto
- Validación y evaluación de los modelos obtenidos.

Se utilizó el lenguaje de programación Python y diferentes librerías para machine learning como scikit-learn, scikit-fuzzy y keras.

## RECURSOS HUMANOS

El equipo de trabajo está integrado por tres docentes de la carrera de Ingeniería en Informática. Se espera incorporar al proyecto alumnos interesados en la temática.

## CONTEXTO

Este proyecto continúa la línea de investigación que el Grupo de Análisis de Datos de la Facultad de Ingeniería de la Universidad Católica de Salta viene desarrollando en minería de datos.

## INTRODUCCION

Si bien, el análisis de datos relacionados a diferentes enfermedades fue explorado ampliamente en la comunidad de investigadores de grandes volúmenes de datos, la exploración de los datos de las personas contagiadas con el nuevo coronavirus nos coloca en una situación especial. No solo por los descubrimientos que se van haciendo día a día por parte de los científicos, sino también por la rápida propagación mundial del virus.

En este proyecto se intentó encontrar similitudes entre algunos distritos argentinos a partir de las cantidades de casos diagnosticados con la enfermedad y con el aprendizaje de relaciones difusas. También se intentó construir una red neuronal para predecir la cantidad de casos positivos de Covid-19 en una provincia argentina..

## RESULTADOS

Los datos se obtuvieron del archivo público con fecha de actualización el 31 de octubre de 2020, que tiene más de un millón de registros y 25 atributos con información básica de las personas registradas: edad, sexo, provincia de residencia, fecha de inicio de síntomas, entre otros.

Se realizaron dos tipos de tareas, por un lado, se experimentó con los datos de seis distritos argentinos con el objeto de encontrar similitudes a partir de la definición de relaciones difusas, y por otro, con redes neuronales para la predicción del número de casos positivos de Covid en una provincia.