



# **CURSO DE DATA SCIENCE PROYECTO FINAL**

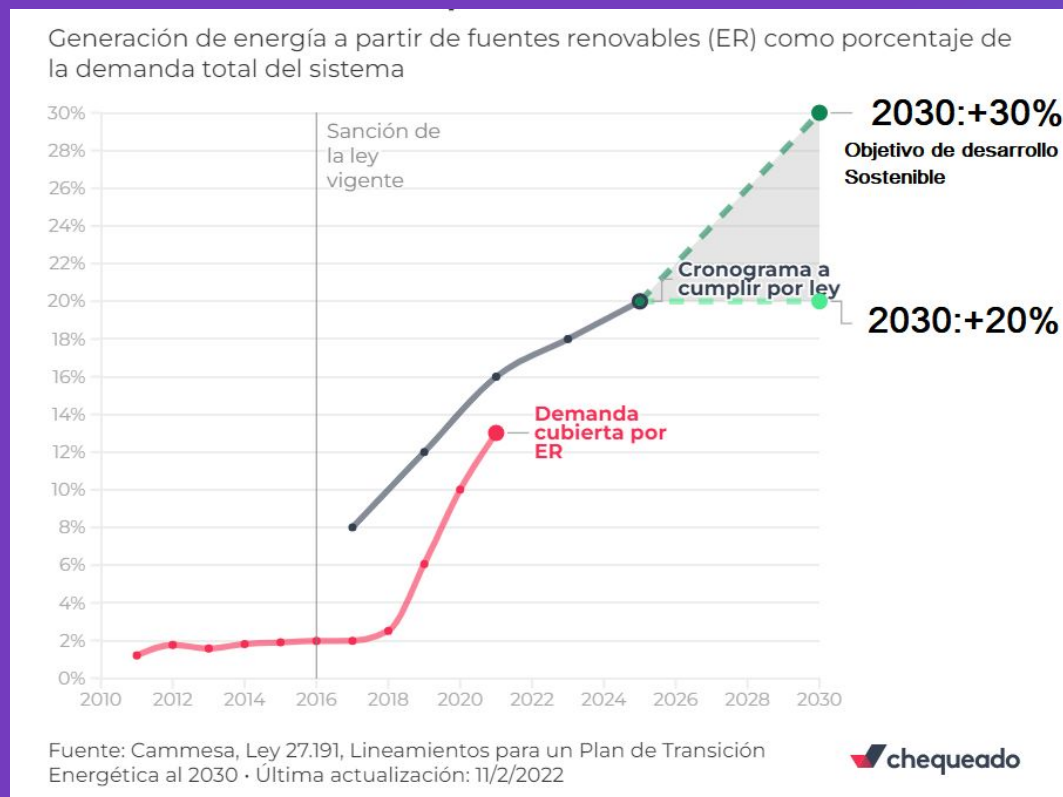
**Objetivo : Predecir la contribución de las fuentes de energías renovables a la satisfacción de la demanda eléctrica del país**

**Integrantes : Eugenia Aciar, Eileen Corbalán, Dolores Corva, Julia Panei, Evelin Paez**

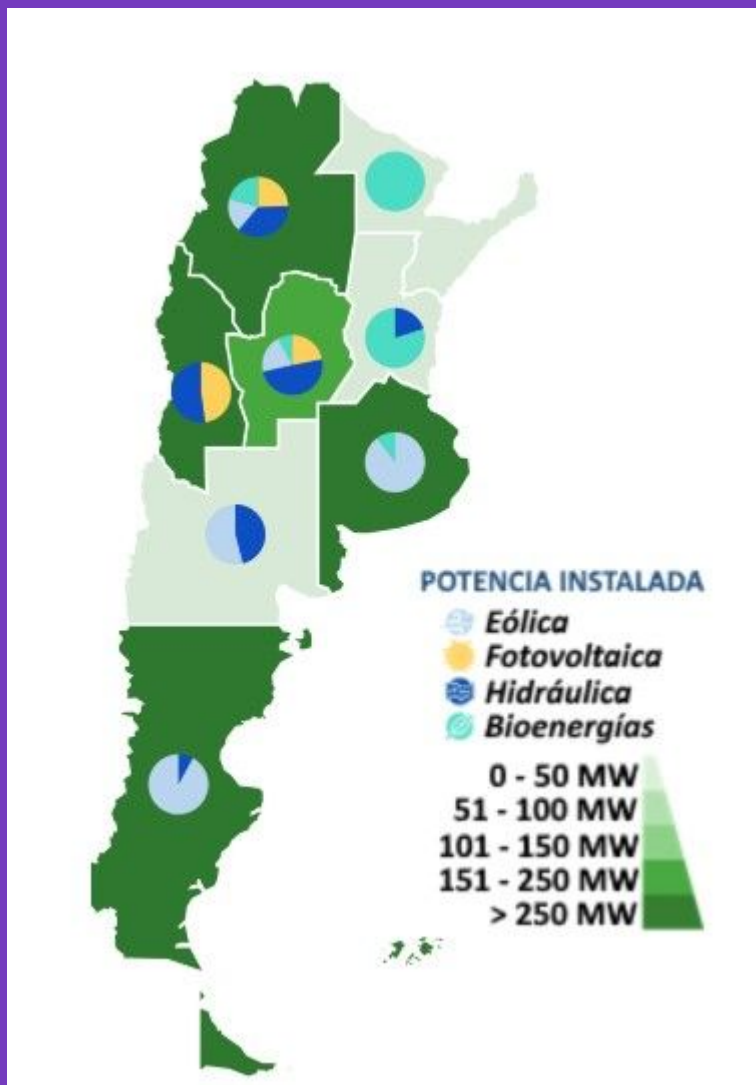


# ¿ POR QUÉ ELEGIMOS ESTE TEMA ?

- ✓ Uno de los Objetivos de Desarrollo Sostenible planteados en la Agenda 2030 de Naciones Unidas para Argentina es “ **tener una mayor inversión en fuentes de energías renovables que contribuyan al sistema eléctrico así como innovar en soluciones energéticas del futuro** ”. Este objetivo está en **consonancia con la estrategia de sustentabilidad de YPF**



# ACERCA DEL DATASET...

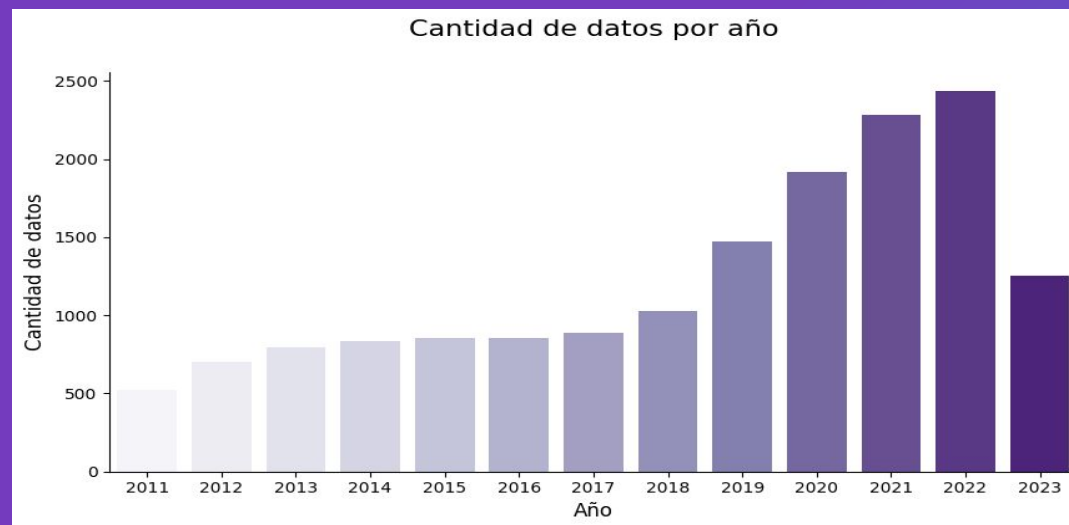


- ✓ Fue obtenido de la página de CAMMESA (Compañía Administradora del Mercado Mayorista Eléctrico Sociedad Anónima)
- ✓ Contiene la información de la energía generada por cada energía renovable (Hidro < 50 MW; Eólico, Solar, Biomasa, Biodiesel), por región y por provincia, la central que las genera y es de tipo mensual con registros desde 2011 hasta junio de 2023



# ANÁLISIS EXPLORATORIO DE DATOS I

¿Cuántos datos  
tenemos por año?



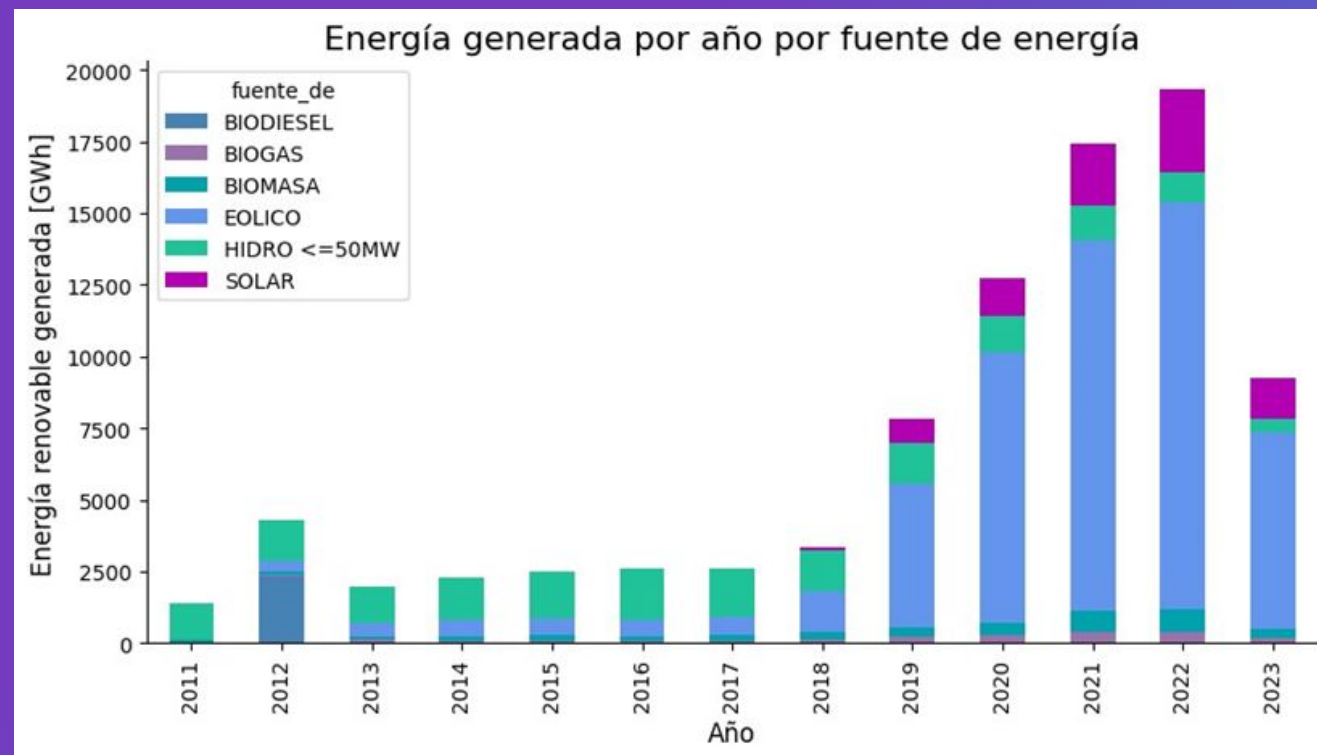
¿Cuánta energía  
eléctrica (Gwh) es  
producida por  
fuentes de  
energías  
renovables?





# ANÁLISIS EXPLORATORIO DE DATOS I

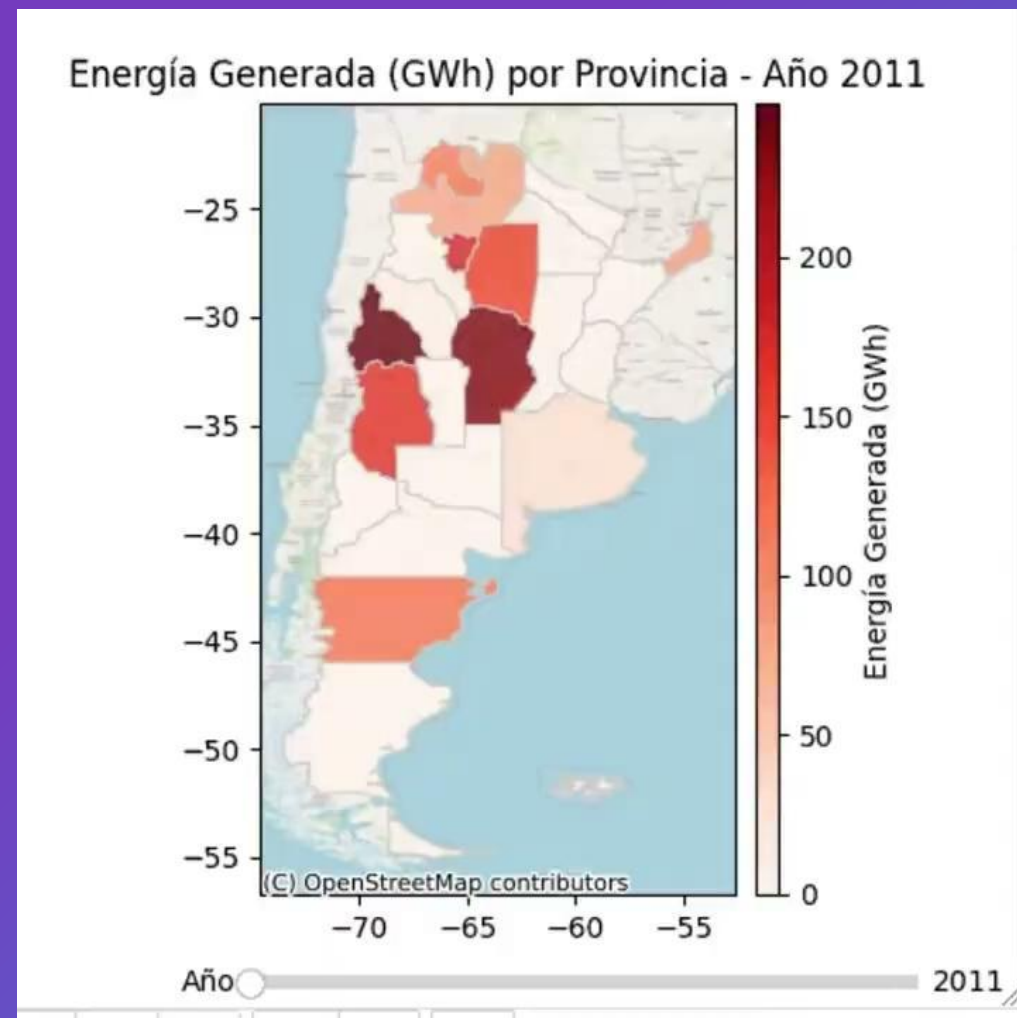
¿Cuánta energía generó cada fuente por año?





# ANÁLISIS EXPLORATORIO DE DATOS I

¿Cuánta energía  
proveniente de energías  
renovables generó cada  
provincia para cada año?







# ANÁLISIS EXPLORATORIO DE DATOS II

Data set Inicial



	AÑO	CENTRAL	CENTRAL DESCRIPCIÓN	MAQUINA	FUENTE DE ENERGÍA	REGIÓN	PROVINCIA	MES	ENERGÍA GENERADA [Gwh]	Nueva Generación
0	2011	AESP	AES PARANA	AESPCC01	BIODIESEL	LITORAL	BUENOS AIRES	ene-11	0	Resto
1	2011	AESP	AES PARANA	AESPCC02	BIODIESEL	LITORAL	BUENOS AIRES	ene-11	0	Resto
2	2011	AMEGHI	F. AMEGHINO	AMEGHI	HIDRO <=50MW	PATAGONIA	CHUBUT	ene-11	13.8	Resto
3	2011	ARAUEO	ARAUCO EOLICO	ARAUEO	EOLICO	NOROESTE	LA RIOJA	ene-11	0.1	Resto
4	2011	CACVHI	C.H.CACHEUTA VI	CACVHI	HIDRO <=50MW	CUYO	MENDOZA	ene-11	0	Resto
5	2011	CADIHI	CADILLAL	CADIHI	HIDRO <=50MW	NOROESTE	TUCUMAN	ene-11	0.6	Resto
6	2011	CALEHI	LA CALERA	CALEHI	HIDRO <=50MW	CENTRO	CORDOBA	ene-11	0.9	Resto
7	2011	CARRHI	CH CARRIZAL	CARRHI	HIDRO <=50MW	CUYO	MENDOZA	ene-11	5.4	Resto
8	2011	CASSHI	CASSAFOUSTH	CASSHI	HIDRO <=50MW	CENTRO	CORDOBA	ene-11	4.1	Resto
9	2011	CEJEHI	CRUZ DEL EJE	CEJEHI	HIDRO <=50MW	CENTRO	CORDOBA	ene-11	0	Resto

Sacamos  
columnas,dejamos  
un dataset por  
energía y aplicamos  
Label Encoder



Data set luego de EDA II



	anio	mes	region	prov	SOLAR	energia_gen
0	2011	1	3	14	True	0.0
1	2011	2	3	14	True	0.0
2	2011	3	3	14	True	0.1
3	2011	4	3	14	True	0.2
4	2011	5	3	14	True	0.2



# MACHINE LEARNING: APRENDIZAJE SUPERVISADO

¿Cuánta energía eléctrica se generará a partir de cada fuente de energía renovable?



Eólico



Fotovoltaico



Bioenergías



Hidráulico Renovable

Como la variable que queremos predecir , es una variable numérica continua decidimos probar con distintos **MODELOS DE REGRESIÓN**. Para esto, en primer lugar se crearon distintos datasets correspondientes a cada fuente de energía y luego se convirtieron todas las variables categóricas a variables numéricas.



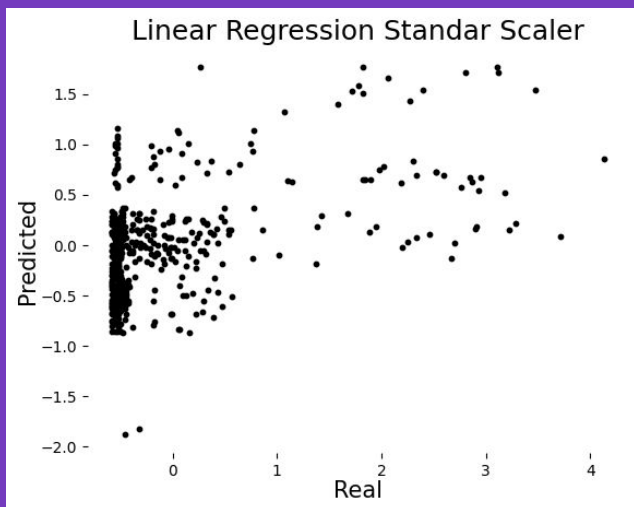


# MACHINE LEARNING: APRENDIZAJE SUPERVISADO

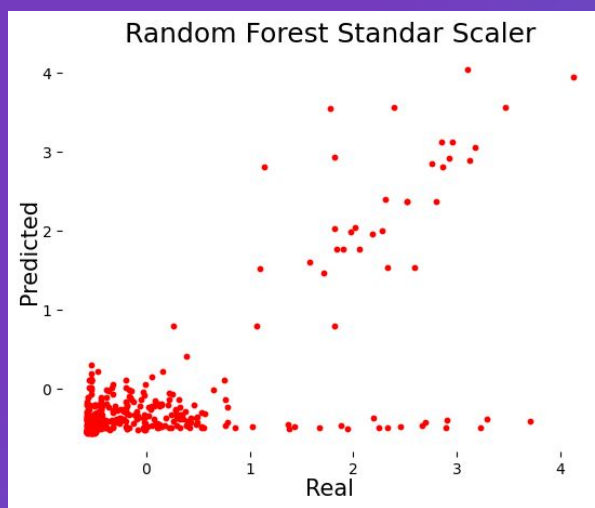
## ¿Cuánta energía eléctrica se generará a partir de la energía solar?

Probamos con 3 modelos que usan la regresión como base sobre los datos **ESCALADOS** (MinMax) y **NORMALIZADOS** (StandardScaler)

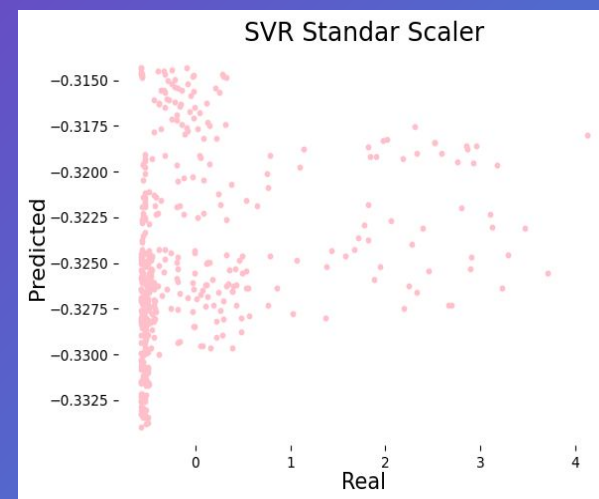
### LINEAR REGRESSION



### RANDOM FOREST



### SUPPORT VECTOR MACHINE





# MACHINE LEARNING: APRENDIZAJE SUPERVISADO

## Evaluación de los modelos a partir de las métricas

Para evaluar el desempeño de los modelos elegimos calcular los errores y el coeficiente de determinación para cada uno de los modelos y compararlos

**ERROR MEDIO  
ABSOLUTO(MAE)**

```
mae_linear_regression_y_min_max_scaler: 0.11542078316432124
mae_linear_regression_y_standar_scaler: 0.5684455961955582
mae_random_forest_y_min_max_scaler: 0.06815434187314041
mae_random_forest_y_standar_scaler: 0.33509516315711857
mae_svr_y_min_max_scaler: 0.1786426704347987
mae_svr_y_standar_scaler: 0.5235273108160509
```

**ERROR MEDIO  
CUADRADO(MSE)**

```
mse_linear_regression_y_min_max_scaler: 0.027293162500233822
mse_linear_regression_y_standar_scaler: 0.6620086132207417
mse_random_forest_y_min_max_scaler: 0.018279365076248145
mse_random_forest_y_standar_scaler: 0.4428835352797319
mse_svr_y_min_max_scaler: 0.042481787510920896
mse_svr_y_standar_scaler: 0.915913267132385
```

**COEFICIENTE DE  
DETERMINACIÓN (r2)**

```
r2_linear_regression_y_min_max_scaler: 0.20339455323864963
r2_linear_regression_y_standar_scaler: 0.20339455323864908
r2_random_forest_y_standar_scaler: 0.4670712292271675
r2_random_forest_y_min_max_scaler: 0.4670712292271675
r2_svr_y_min_max_scaler: -0.23991579645883743
r2_svr_y_standar_scaler: -0.10213293722714067
```



# MACHINE LEARNING: APRENDIZAJE SUPERVISADO

## Interpretación de las métricas

ERROR MEDIO  
ABSOLUTO(MAE)

```
mae_linear_regression_y_min_max_scaler: 0.11542078316432124  
mae_linear_regression_y_standar_scaler: 0.5684455961955582  
mae_random_forest_y_min_max_scaler: 0.06815434187314041  
mae_random_forest_y_standar_scaler: 0.33509516315711857  
mae_svr_y_min_max_scaler: 0.1786426704347987  
mae_svr_y_standar_scaler: 0.5235273108160509
```

ERROR MEDIO  
CUADRADO(MSE)

```
mse_linear_regression_y_min_max_scaler: 0.027293162500233822  
mse_linear_regression_y_standar_scaler: 0.6620086132207417  
mse_random_forest_y_min_max_scaler: 0.018279365076248145  
mse_random_forest_y_standar_scaler: 0.4428835352797319  
mse_svr_y_min_max_scale: 0.042481787510920896  
mse_svr_y_standar_scaler: 0.915913267132385
```

COEFICIENTE DE  
DETERMINACIÓN (r2)

```
r2_linear_regression_y_min_max_scaler: 0.20339455323864963  
r2_linear_regression_y_standar_scaler: 0.20339455323864908  
r2_random_forest_y_standar_scaler: 0.4670712292271675  
r2_random_forest_y_min_max_scaler: 0.4670712292271675  
r2_svr_y_min_max_scale: -0.23991579645883743  
r2_svr_y_standar_scaler: -0.10213293722714067
```

Aprobado  
por  
GridSearch  
CV

Al considerar las 3 métricas, se busca un equilibrio entre la PRECISIÓN EN LA PREDICCIÓN y la CAPACIDAD EXPLICATIVA del modelo. En este caso, un buen rendimiento sería un modelo con un MAE bajo, un MSE bajo y un r2 alto. Es por esto que para el caso de la energía solar, el mejor modelo es RANDOM FOREST con los valores escalados con el método MinMax.

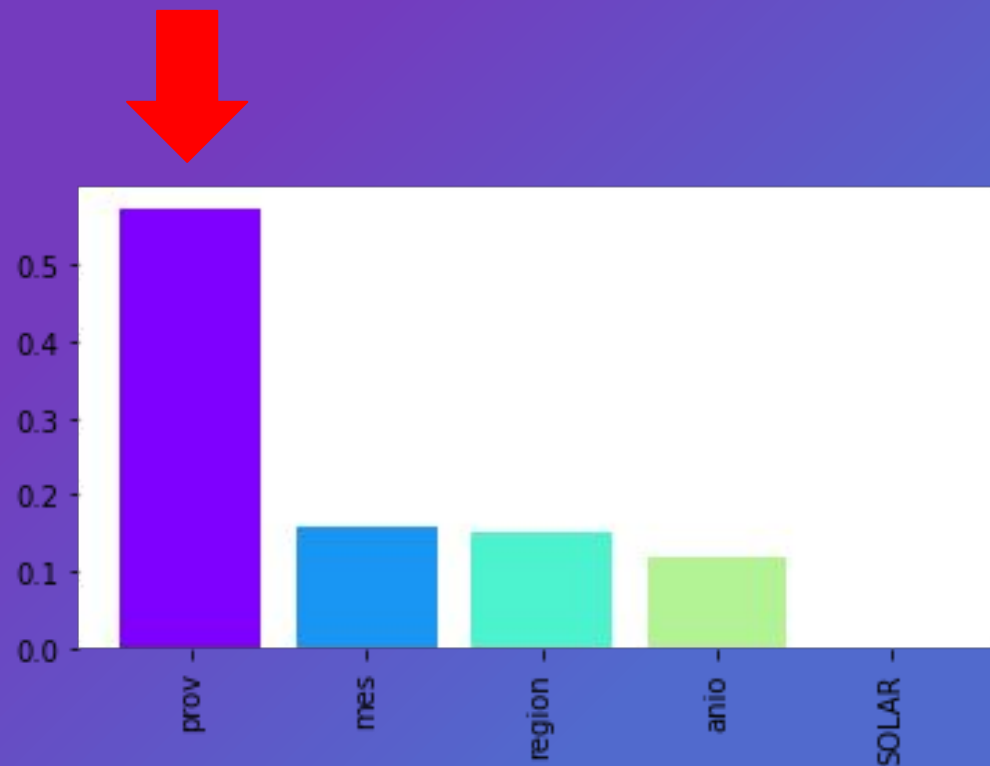


# MACHINE LEARNING: APRENDIZAJE SUPERVISADO

## Importancia de las features

Aplicamos la métrica de **Feature Importance** para evaluar la contribución relativa de cada feature en nuestro modelo de Random Forest.

En este caso se observa que la **provincia** sería la que genera mayor impacto en la capacidad predictiva del modelo.







# CONCLUSIONES

**Retomando nuestra pregunta inicial : ¿Cuánta energía eléctrica se generará a partir de cada fuente de energía?**

- ❑ Con el dataset elegido, con el EDA y con los modelos elegidos pudimos correr un modelo completo para una de las energías (SOLAR) y podría considerarse que los resultados no fueron malos y **SI** responden a nuestra pregunta inicial.
- ❑ Sin embargo nos llevaron a reevaluar las decisiones que fuimos tomando durante todo el proceso y que cambiaríamos para mejorarlo :
  - Dejar más features en los datasets de cada energía
  - Evaluar la distribución y normalidad de cada energía
  - No correr modelos como SVM que consumen mucha capacidad de cómputo
  - Dejar de lado el escalado cuando trabajamos con una sola energía







# CONCLUSIONES GENERALES

- El dataset de CAMMESA es gran calidad y permitiría predecir la contribución de las energías renovables al sistema eléctrico en los próximos años.
- Poseemos los datos históricos de la demanda eléctrica, por cual también se podría modelar y relacionar con los objetivos de desarrollo sostenible
- Los modelos sobre las energías renovables podrían mostrar posibilidades de inversión y desarrollo a nivel provincial.



GRACIAS POR SU ATENCIÓN CHICAS!!

