

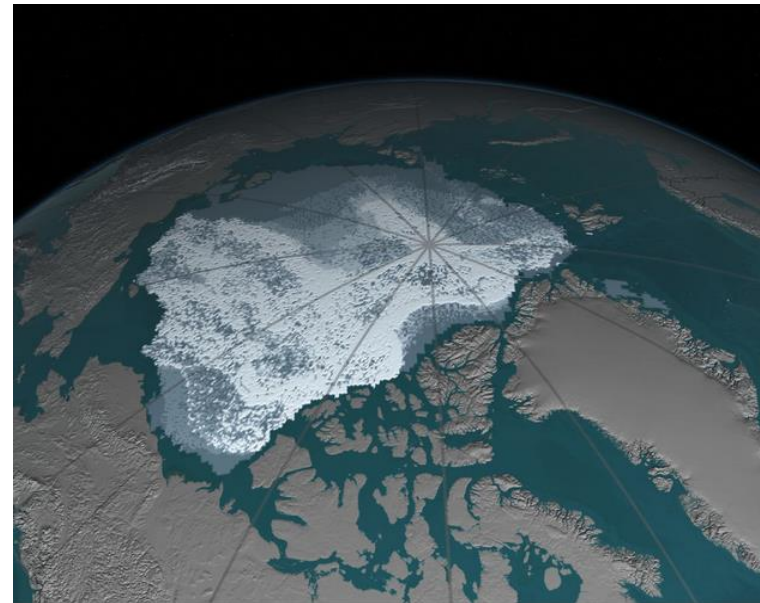


Anàlisis de la relació entre les emissions de gasos hivernacle i l'augment de la temperatura mitjançant algoritmes i tecnologies de Big Data

Miquel Freixes Faya

Canvi Climàtic

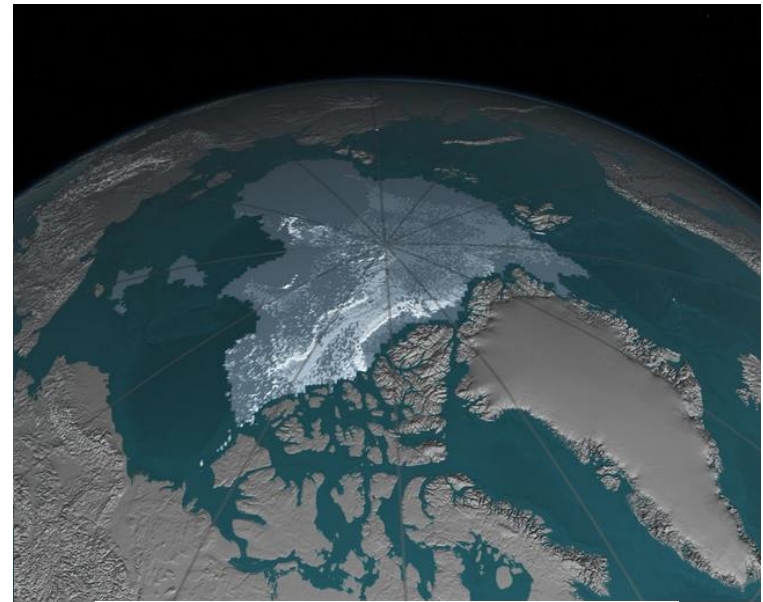
- Augment de la temperatura de la Terra
- Principal causa: Contaminació causada per l'ésser humà
- Efectes catastròfics



Setembre 1984 NASA

Canvi Climàtic


- Augment de la temperatura de la Terra
- Principal causa: Contaminació causada per l'ésser humà
- Efectes catastròfics



Setembre 2016 NASA



Objectius

- Analitzar les prediccions de temperatura, fetes per models de ML, a partir de les emissions de gasos de cada país:
 - Crear un conjunt de dades que relacioni països amb les seves temperatures i les seves emissions
 - Analitzar models més utilitzats actualment en problemes de regressió i implementar-los correctament
 - Trobar la millor configuració de cada un d'ells
 - Determinar el millor
- 



Conjunt de dades



Camp	Tipus	Exemple	Significat
Data	Date	DD-MM-YYYY	Primer dia de cada mes
Temperatura Mitjana	Float	1.714	Temperatura amb una confiança del 95%
País	String	Bèlgica	Nom de cada país
Incertesa	Float	0.11	Error del 5% de la temperatura

Conjunt de dades temperatures



Conjunt de dades

Camp	Tipus	Exemple	Significat
País	String	Bèlgica	Nom de cada país
Any	Int	1999	Any de l'emissió
Valor	Float	393126.946	Mitjana de Kilotones produïdes en un any
Categoria	String	Co2	Tipus de gas

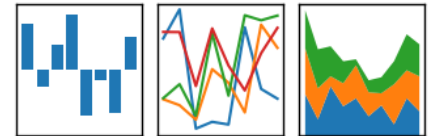
Conjunt de dades emissions gasos

Eines de pre-processament

- Spark
 - Obtenir anàlisis eficaçes en entorns distribuïts
 - Treballa en memòria
- Pandas
 - Llibreria de codi obert
 - Estructures flexibles i ràpides



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$





Pre-Processament de dades

Camp	Tipus	Exemple
País	String	Bèlgica

Any	Kt per any
2012	100931.59
2011	104945.71

Pre-Processament de dades

Camp	Tipus	Exemple
País	String	Bèlgica

**Get
Dummies**

Any	Kt per any
2012	100931.59
2011	104945.71

Pre-Processament de dades

Camp	Tipus	Exemple
País	String	Bèlgica

**Get
Dummies**

Àustria	Alemanya	UK	Croàcia
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Any	Kt per any
2012	100931.59
2011	104945.71

Pre-Processament de dades

Camp	Tipus	Exemple
País	String	Bèlgica

**Get
Dummies**

Àustria	Alemanya	UK	Croàcia
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Any	Kt per any
2012	100931.59
2011	104945.71

Interpolate

Pre-Processament de dades

Camp	Tipus	Exemple
País	String	Bèlgica

**Get
Dummies**

Àustria	Alemanya	UK	Croàcia
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Any	Kt per any
2012	100931.59
2011	104945.71

Interpolate

Any	Mes	Kt Mes
2011	1	104945.71
2011	2	100664.2
2011	3	105762.34
...
2012	1	100931.59

Pre-Processament de dades

Camp	Tipus	Exemple	Significat
Data	Int	1999	Any de la fila
Co2	Float	393126.946	Kt de Co2 d'un mes
N2O	Float	453722.447	Kt de N2O d'un mes
CH4	Float	126826.946	Kt de CH4 d'un mes
Mes	Int	1-12	Mes de la fila
Països	Int	1 o 0	Indica a quin país pertany
Av. Temp	Float	7.5	Mitjana de °C del mes

Pre-Processament de dades

Camp	Tipus	Exemple	Significat
Data	Int	1999	Any de la fila
Co2	Float	393126.946	Kt de Co2 d'un mes
N2O	Float	453722.447	Kt de N2O d'un mes
CH4	Float	126826.946	Kt de CH4 d'un mes
Mes	Int	1-12	Mes de la fila
Països	Int	1 o 0	Indica a quin país pertany

Entrenament d'algoritmes

- Scikit-Learn
 - Fàcil d'utilitzar
 - Gran varietat de models



Entrenament d'algoritmes

- Scikit-Learn
 - Fàcil d'utilitzar
 - Gran varietat de models
- Keras:
 - Treballa sobre TensorFlow
 - Entrenament Xarxes Neuronals i *Deep Learning*



Entrenament d'algoritmes

- Scikit-Learn
 - Fàcil d'utilitzar
 - Gran varietat de models
- Keras:
 - Treballa sobre TensorFlow
 - Entrenament Xarxes Neuronals i *Deep Learning*
- Optimitzar:
 - *Grid SearchCV*
 - *Randomized SearchCV*



Entrenament d'algoritmes

- Scikit-Learn
 - Fàcil d'utilitzar
 - Gran varietat de models
- Keras:
 - Treballa sobre TensorFlow
 - Entrenament Xarxes Neuronals i *Deep Learning*
- Optimitzar:
 - *Grid SearchCV*
 - *Randomized SearchCV*
- Evaluar:
 - *Mean Absolut Error*





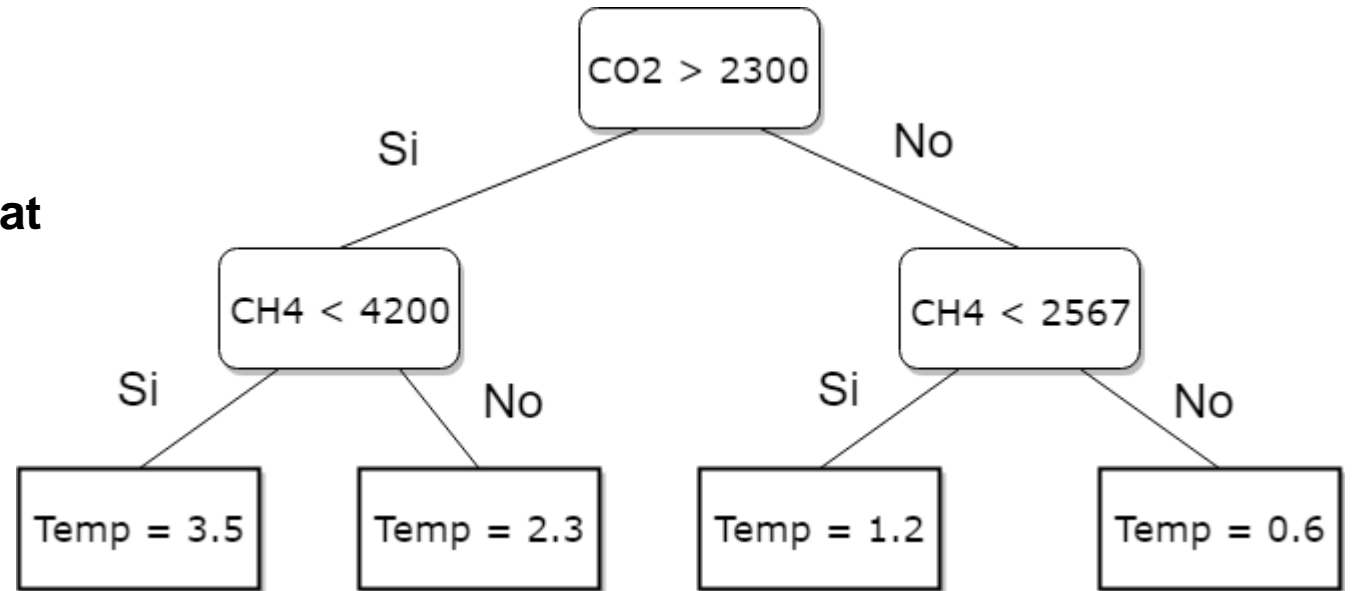
Arbres de Decisió



- Models que fan créixer arbres binaris
- S'escull l'entrada més rellevant a cada node
- Paràmetres editables:
 - **Màxima profunditat**

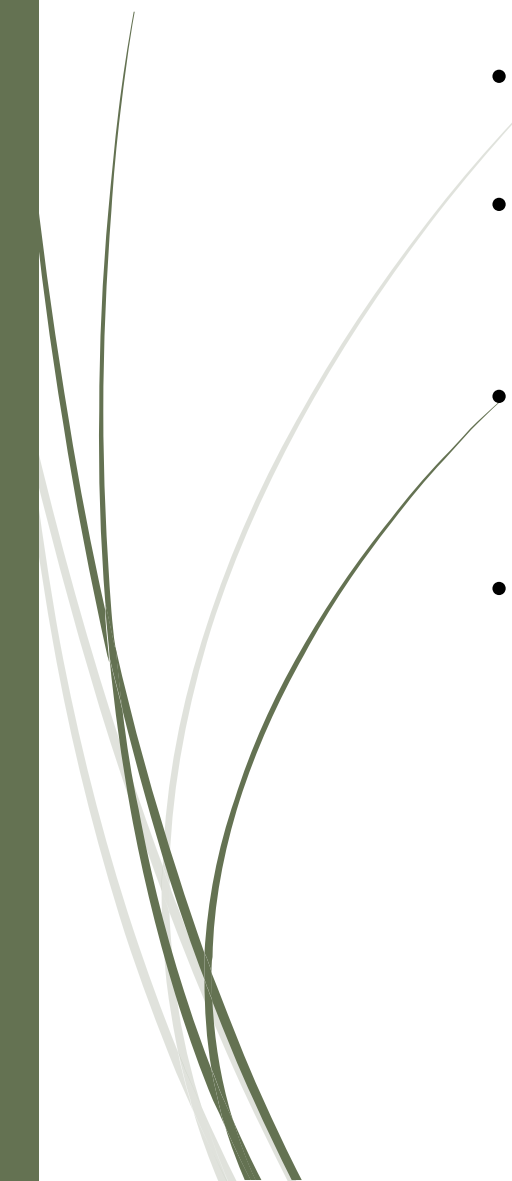
Arbres de Decisió

- Models que fan créixer arbres binaris
- S'escull l'entrada més rellevant a cada node
- Paràmetres editables:
 - **Màxima profunditat**



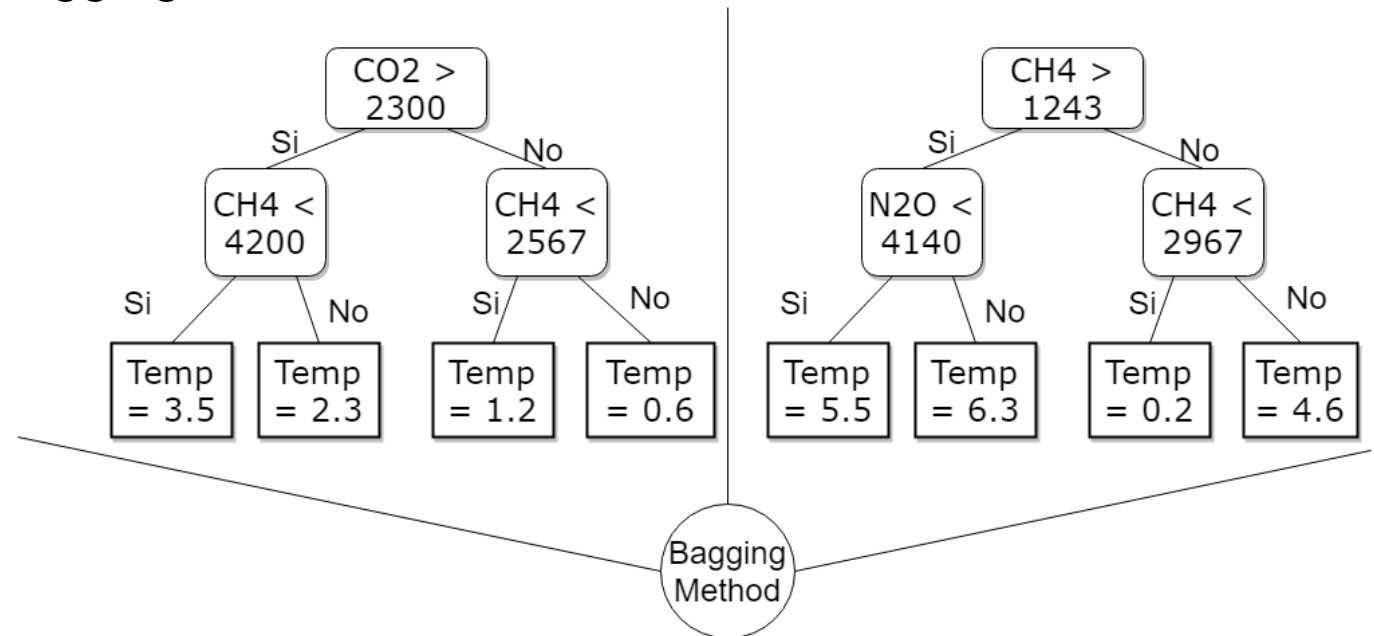


Random Forest

- Algoritme molt utilitzat en l'actualitat
 - Molts algorismes simples funcionen millor que un de sol.
 - Crea arbres de decisió i els combina amb un algoritme com el *Bagging*
 - Paràmetres:
 - **N_estimators**
- 

Random Forest

- Algoritme molt utilitzat en l'actualitat
- Molts algoritmes simples funcionen millor que un de sol.
- Crea arbres de decisió i els combina amb un algoritme com el *Bagging*
- Paràmetres:
 - **N_estimators**





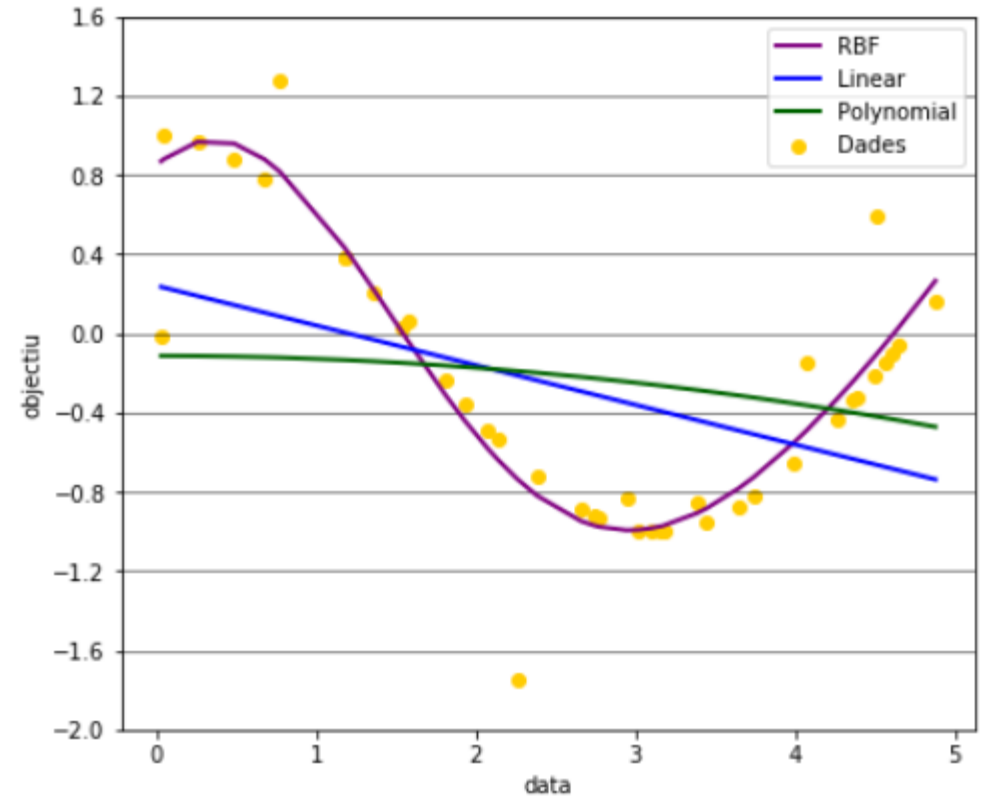
Support Vector Machines



- Aconseguir una funció amb resultats per sota d'una desviació límit
- Paràmetres:
 - **Kernel** -> Com es processen les dades d'entrada


Support Vector Machines

- Aconseguir una funció amb resultats per sota d'una desviació límit
- Paràmetres:
 - **Kernel** -> Com es processen les dades d'entrada



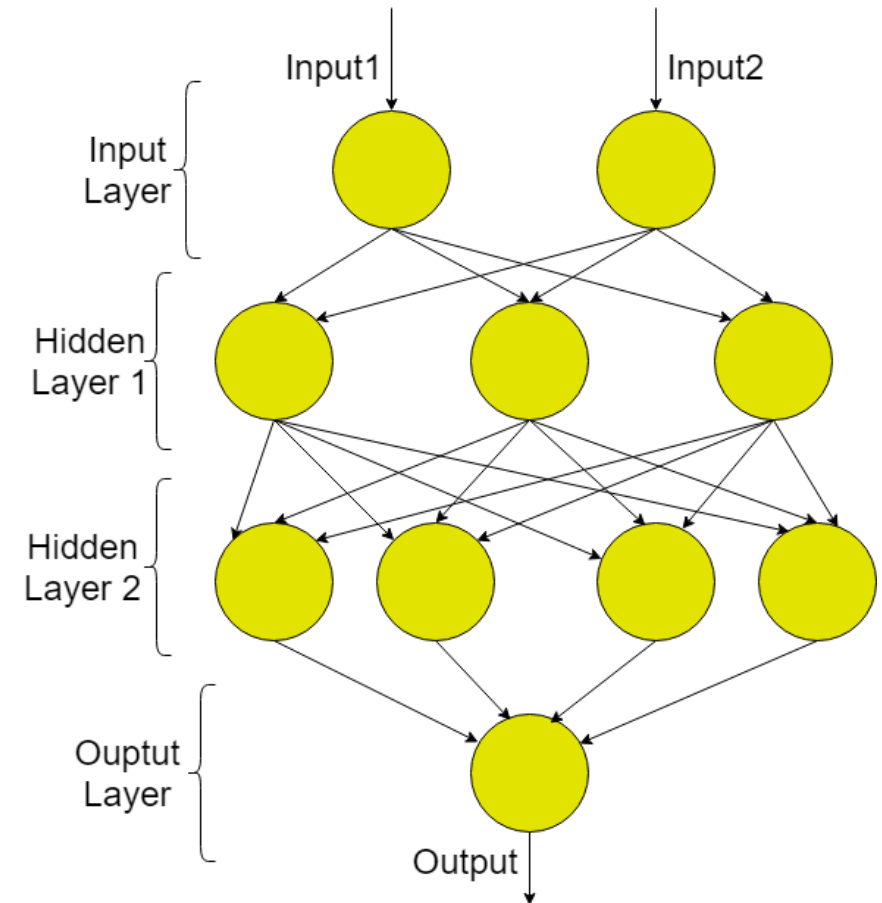


Xarxes Neuronals

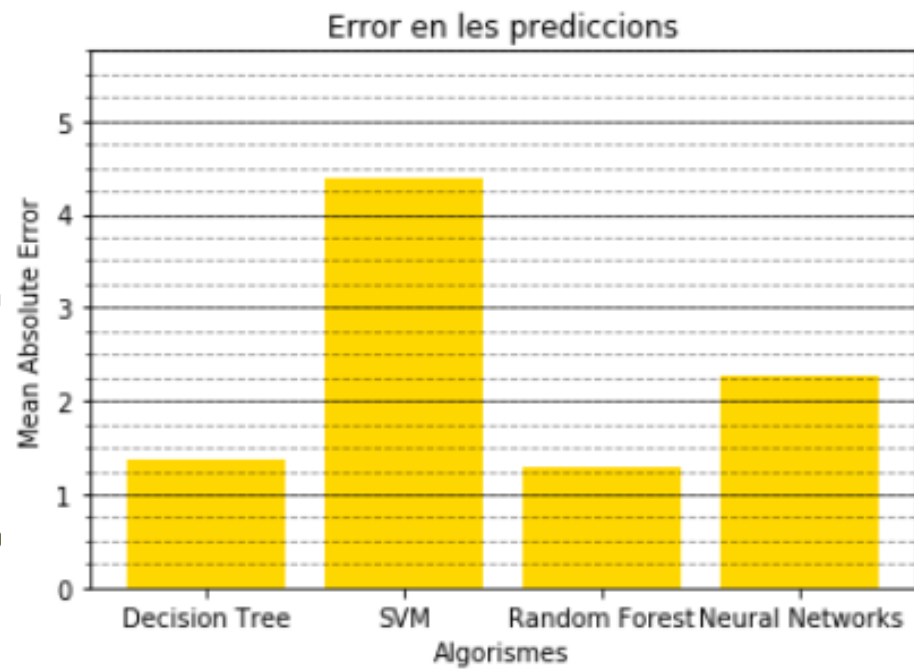
- Algoritme més prometedor actualment
 - Versatilitat, escalabilitat i gran potència de còmput
 - Gran quantitat d'híper-paràmetres
- 

Xarxes Neuronals

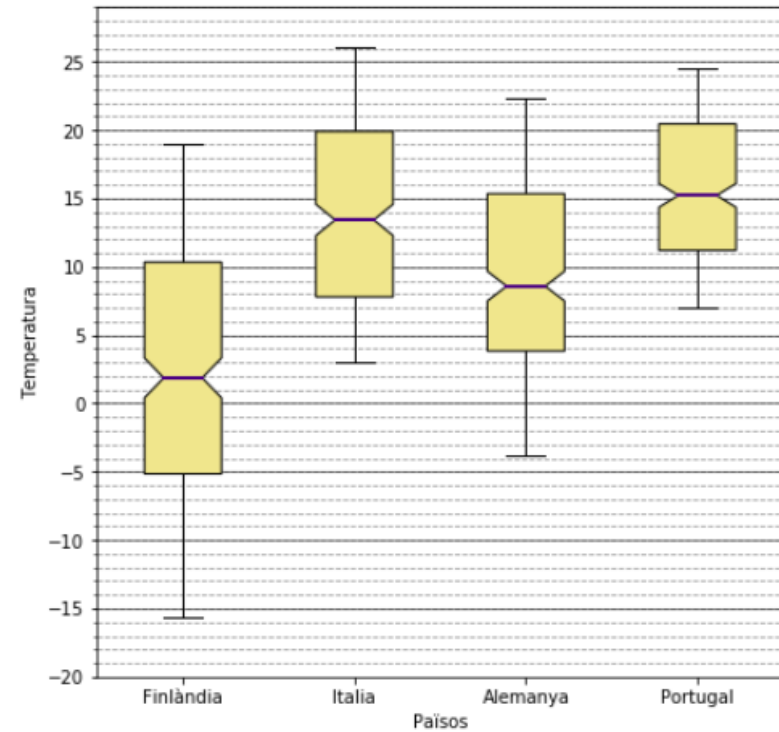
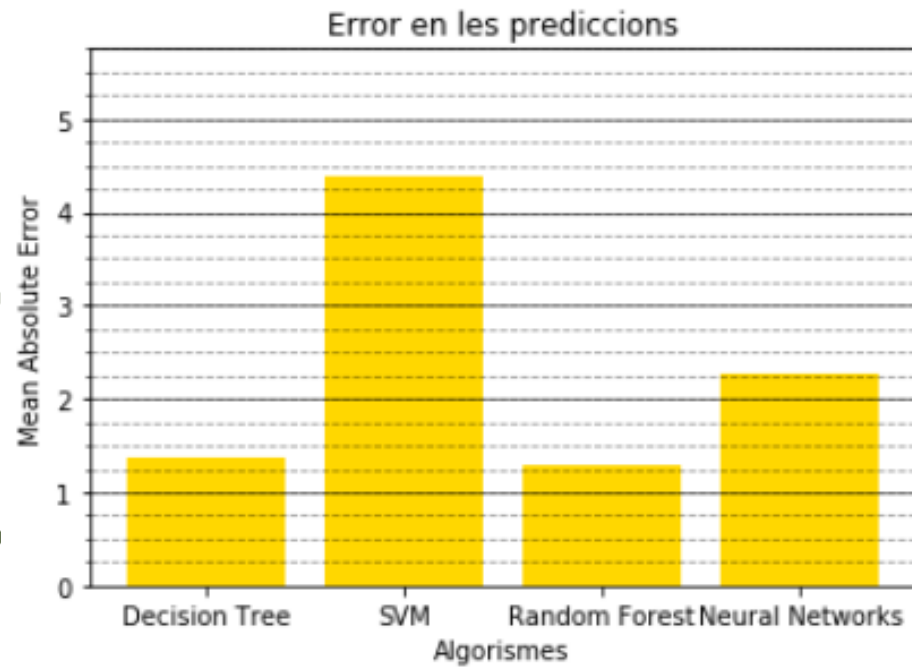
- Algoritme més prometedor actualment
- Versatilitat, escalabilitat i gran potència de càlcul
- Gran quantitat d'híper-paràmetres



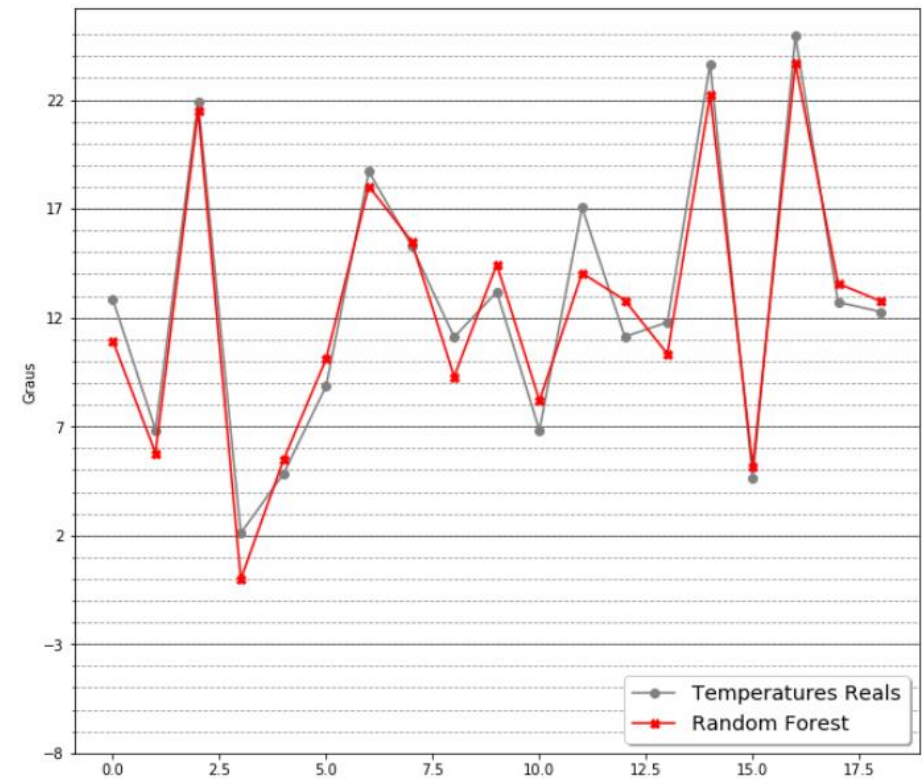
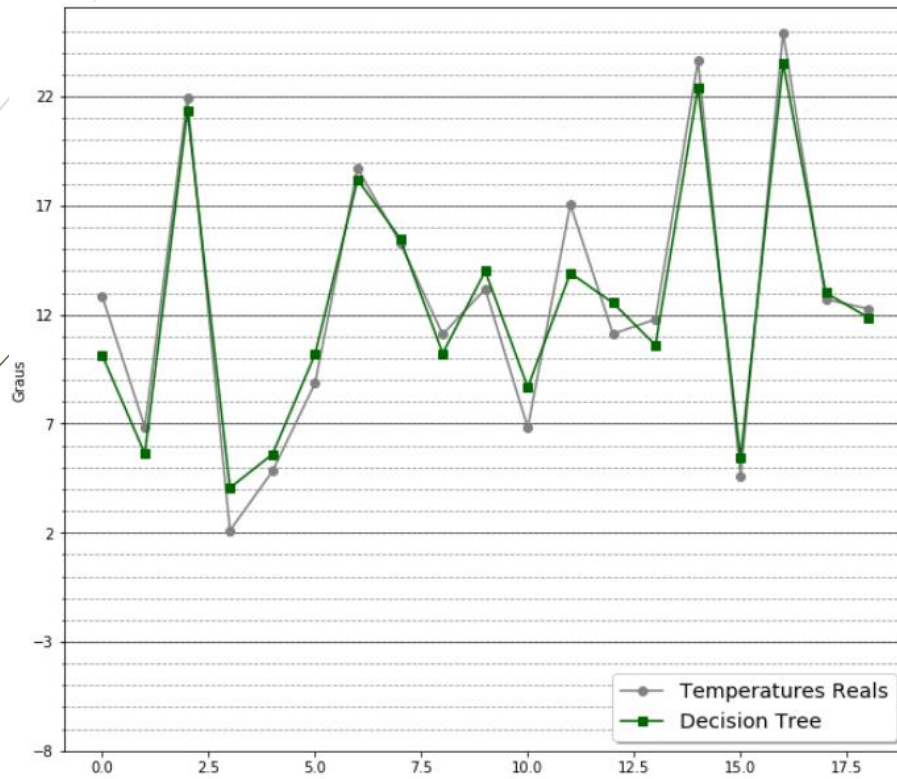
Anàlisi de resultats



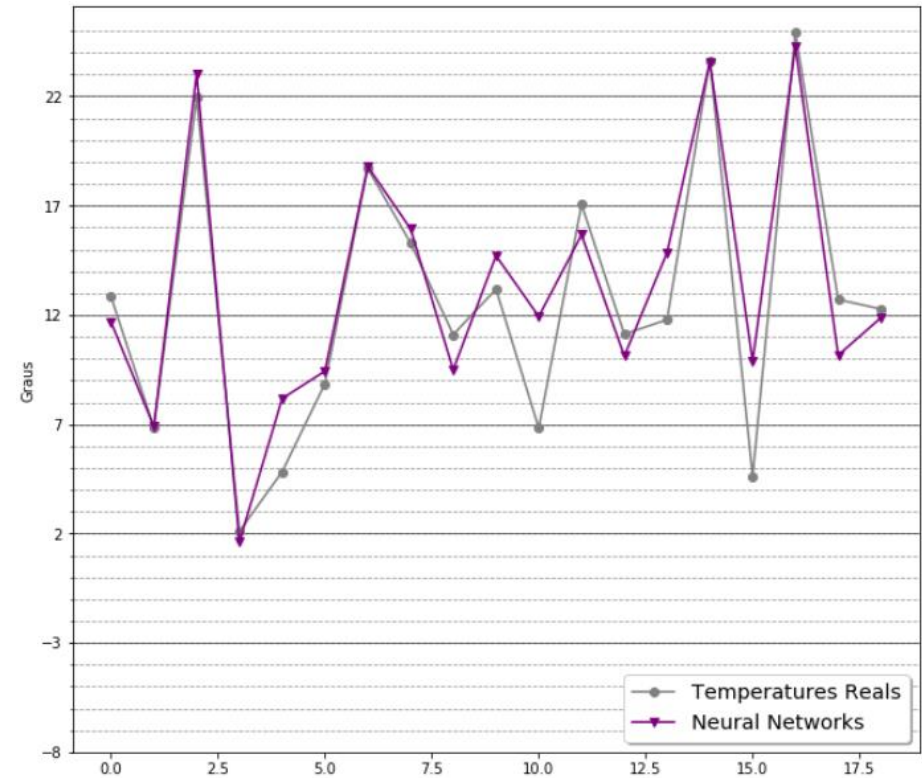
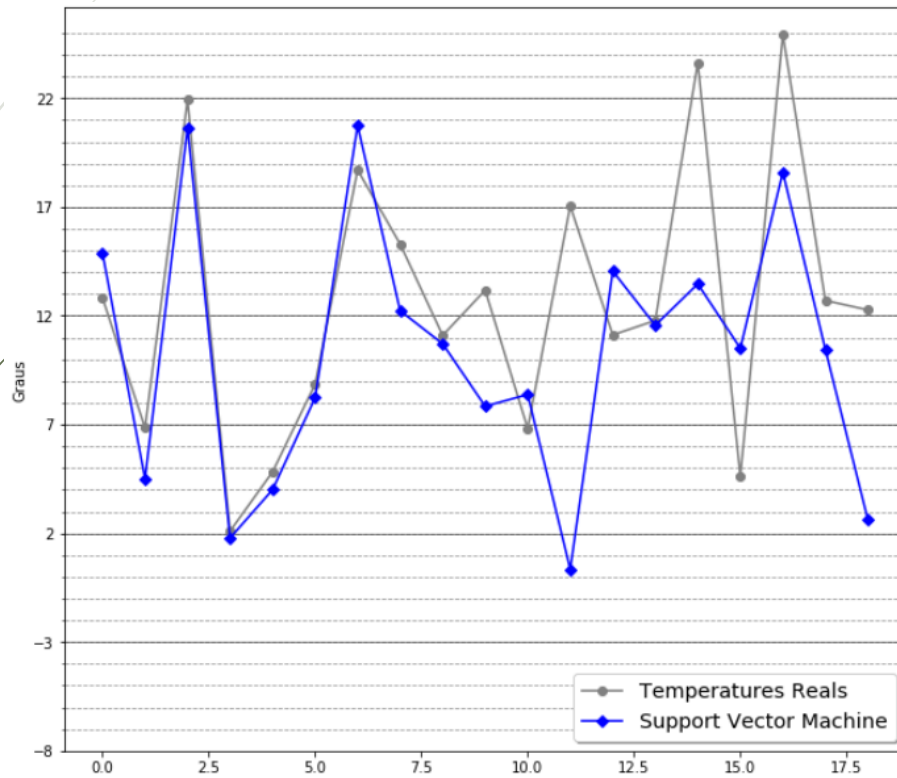
Anàlisi de resultats



Anàlisi de resultats

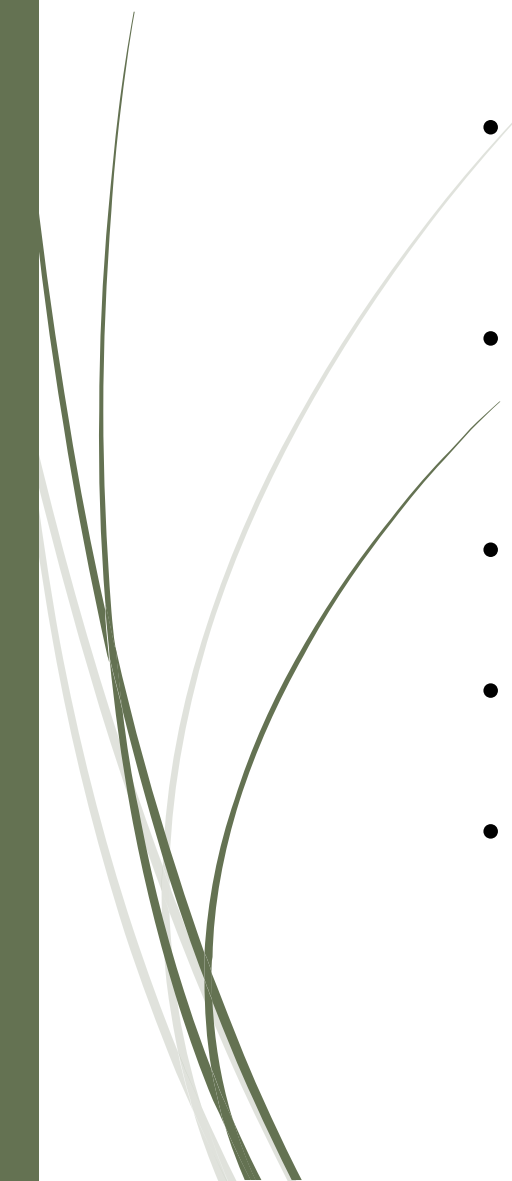


Anàlisis de resultats



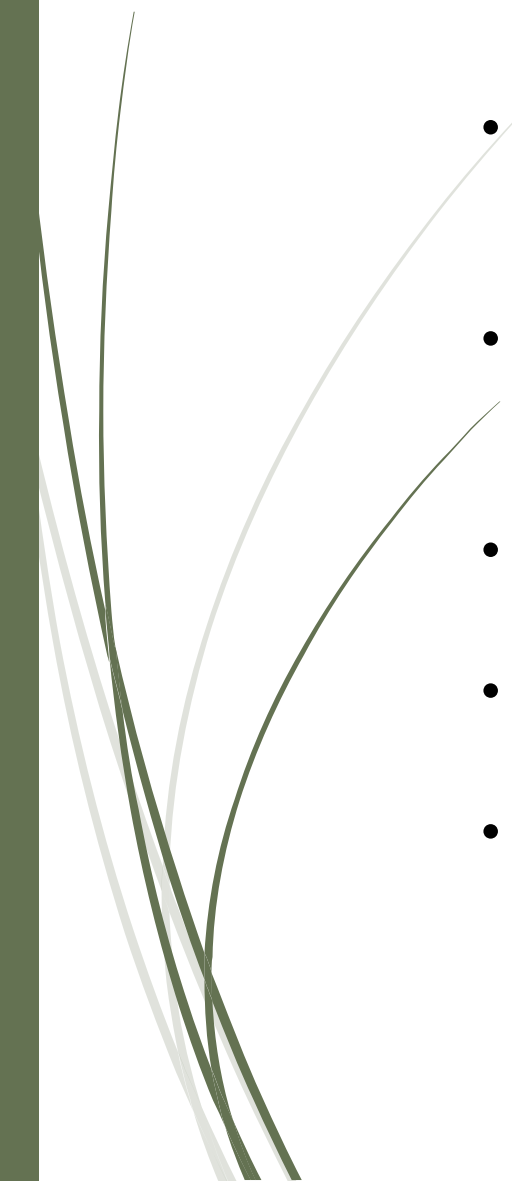


Conclusions

- Crear un conjunt de dades que relacioni països amb les seves temperatures i les seves emissions
 - Analitzar models més utilitzats actualment en problemes de regressió i implementar-los correctament
 - Trobar la millor configuració de cada un d'ells
 - Determinar el millor
 - Relació entre les emissions de gasos i les temperatures
- 

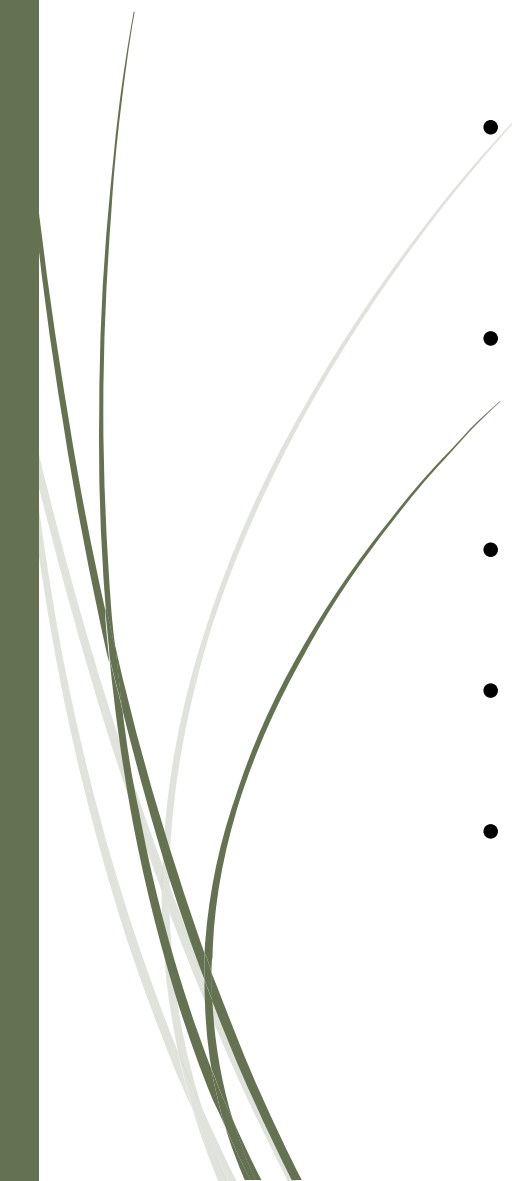


Conclusions

- Crear un conjunt de dades que relacioni països amb les seves temperatures i les seves emissions ✓
 - Analitzar models més utilitzats actualment en problemes de regressió i implementar-los correctament ✓
 - Trobar la millor configuració de cada un d'ells ✓
 - Determinar el millor ✓
 - Relació entre les emissions de gasos i les temperatures
- 



Conclusions

- Crear un conjunt de dades que relacioni països amb les seves temperatures i les seves emissions ✓
 - Analitzar models més utilitzats actualment en problemes de regressió i implementar-los correctament ✓
 - Trobar la millor configuració de cada un d'ells ✓
 - Determinar el millor ✓
 - Relació entre les emissions de gasos i les temperatures ✓
- 



Gràcies per la
vostra atenció



Universitat Autònoma
de Barcelona