

Projet 8 du parcours
Data Analyst
OpenClassRooms
Janvier à Mai 2019

Étude et modélisation d'une éolienne en fonctionnement optimal

- ▶ Les éoliennes sont basées sur une technologie qui à connu un bel essor ces dernières années, la recherche d'énergies vertes étant devenu un enjeu mondial majeur.
- ▶ Cependant posséder une eolienne est couteux et demande de l'entretien ainsi que des connaissances des facteurs qui influencent la production.
- ▶ Engie propose de mettre à disposition certaines de ses éoliennes, mais doit s'assurer du bon usage qui en est fait.

PROBLÉMATIQUE BUSINESS, RÉFLEXIONS PRÉLIMINAIRES

INFORMATIONS SUR LES ÉOLIENNES ET SUR LES DONNÉES

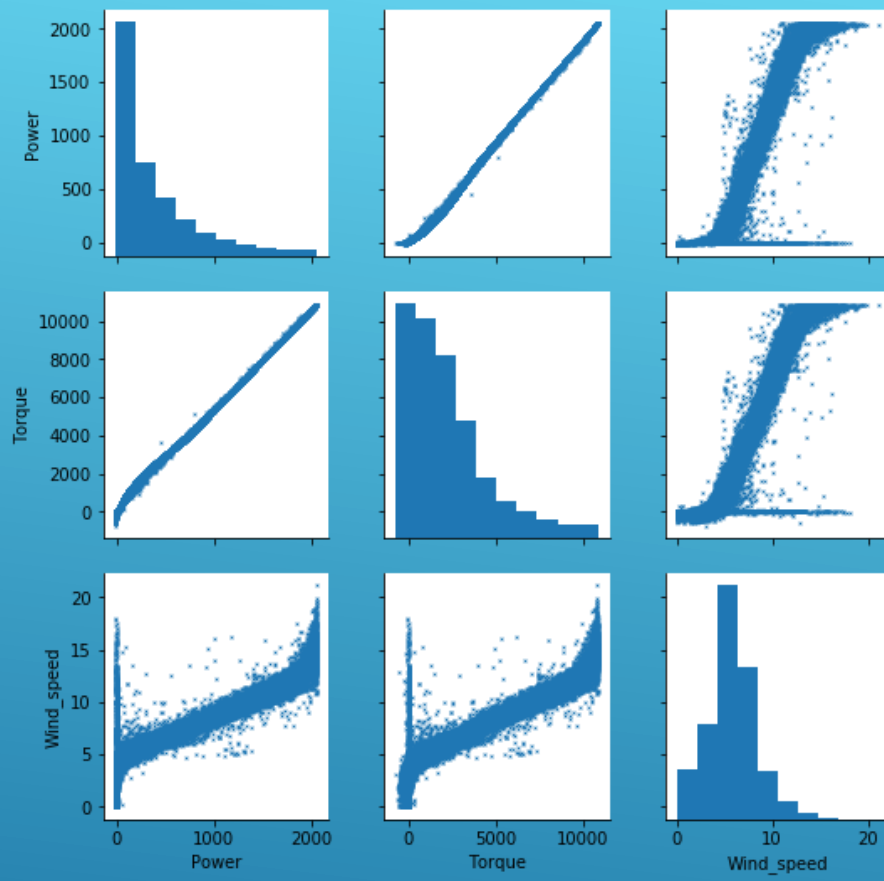
- ▶ Site : La haute borne, département de la Meuse, 411m d'altitude
- ▶ Modèle : MM82 de Senvion
- ▶ Données libres fournies par Engie, 4 éoliennes
- ▶ 3 ans (2013-16)
- ▶ ~200 000 lignes par éoliennes
- ▶ 4 variantes des variables => on garde average !

ANALYSE ET NETTOYAGE DES DONNÉES

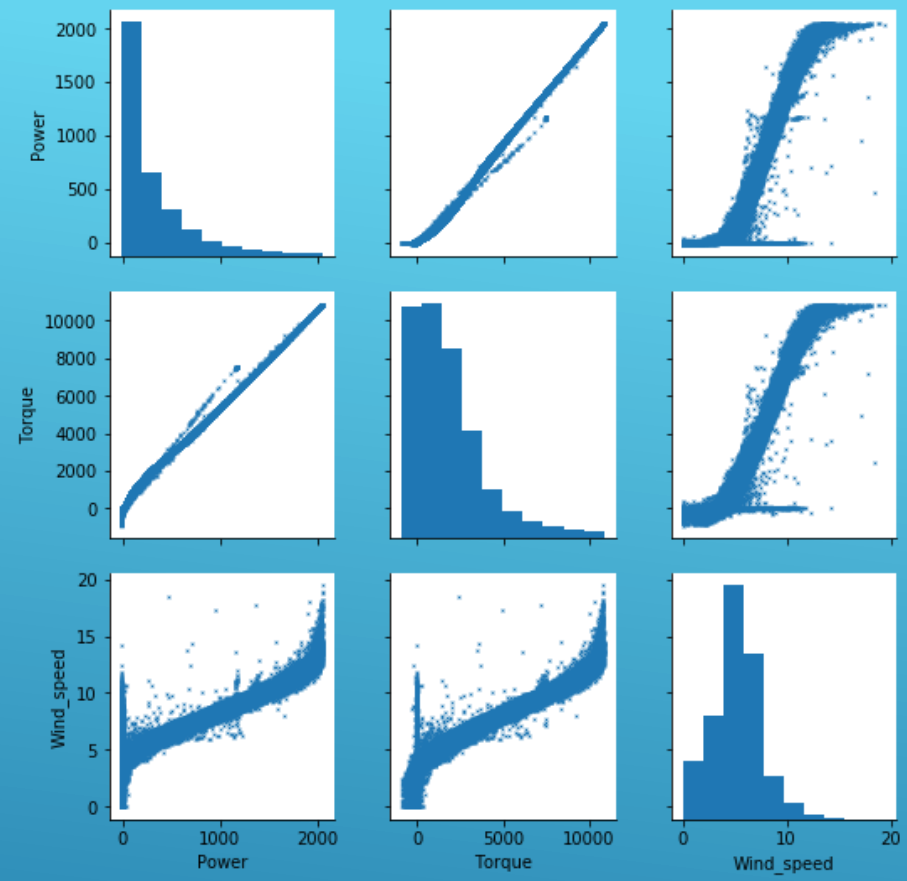
- ▶ Choix des variables utiles:

L'angle de la pale, la puissance générée, la vitesse du générateur, le torque, la vitesse du vent, la direction du vent ainsi que la température

- ▶ Enlever lignes vides / aberrantes
- ▶ Séparation de chaque éolienne
- ▶ Représentations graphiques

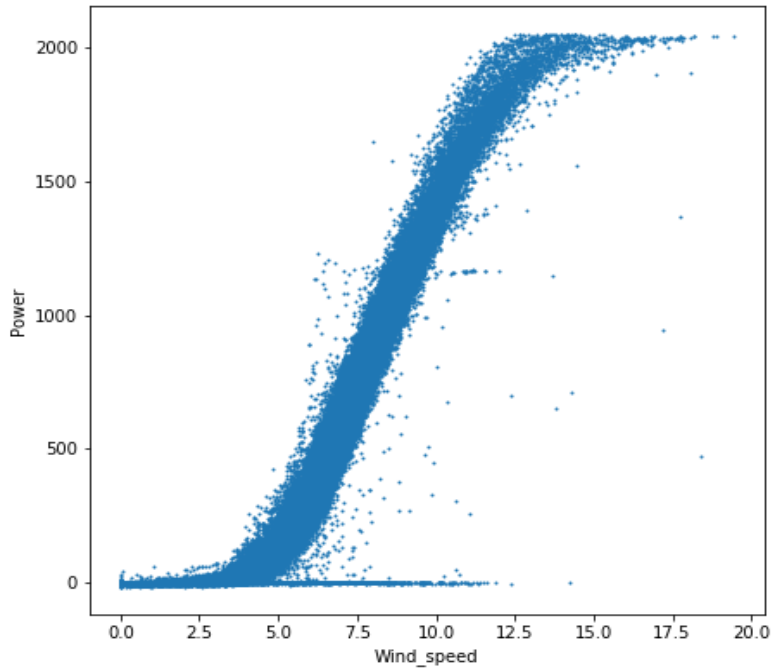


Eolienne 80736



Eolienne 80721

PAIRPLOT DE DEUX ÉOLIENNES

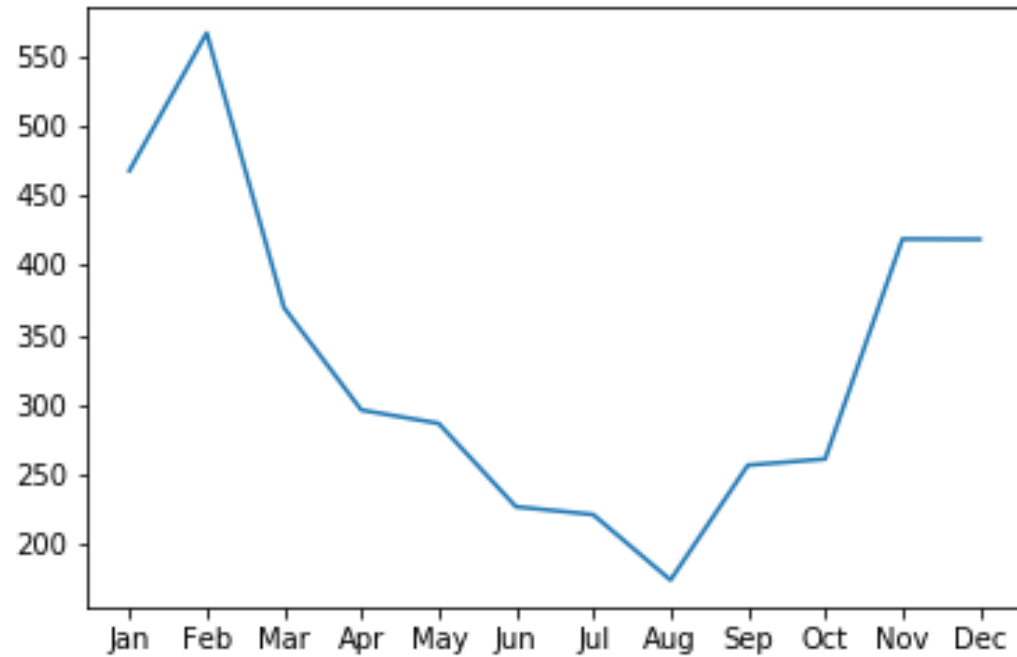


- ▶ Léger effet de seuil dû à un bridage aux environs de 10m/s
- ▶ Mesures de puissance = 0 malgré le vent
- ▶ Variance élevée dans le couple

COURBE DE PUISSANCE 80721

1. Bas de la courbe : filtre sur base d'analyse visuelle: Vitesse > 3,8 & Puissance < 8 ; Puissance < 10 & Angle pale > 15°
2. Milieu de courbe : effet du bridage. Filtre sur base d'analyse visuelle : Vitesse > 9.8 & Puissance < 1190 ; Vitesse > 10.5 & Puissance < 1420
3. Toute la courbe : filtre percentiles .96 et .05 sur les groupes : vitesses de vent "discretisée" , directions de vent "discretisée" et mois de l'année.

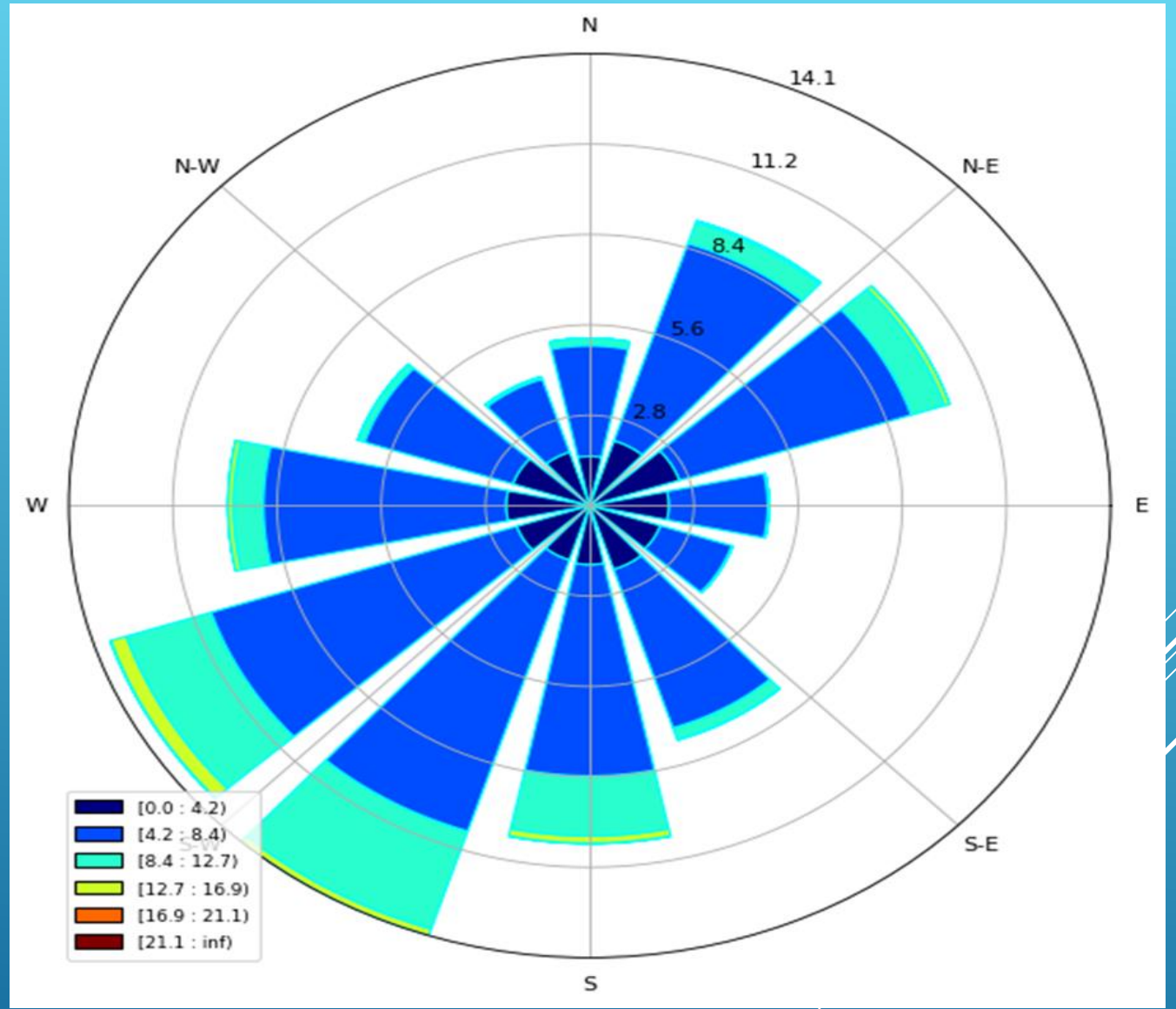
MÉTHODES UTILISÉES



Moyenne de la production de puissance en fonction du mois de l'année.

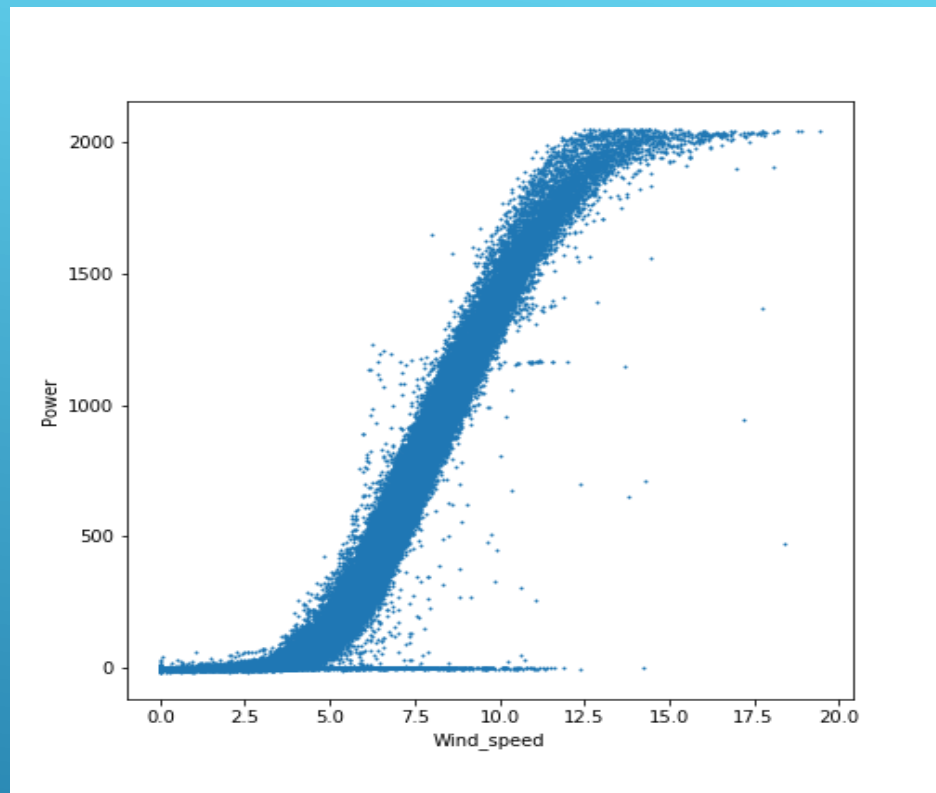
INFLUENCE DU MOIS

ROSE DES VENTS

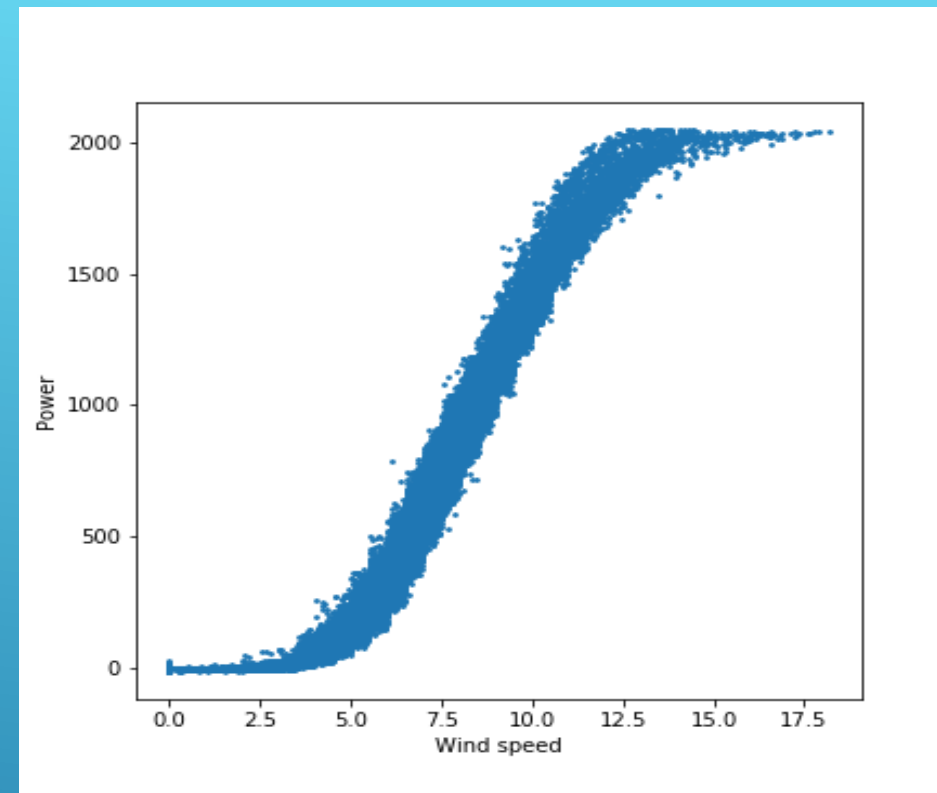


- ▶ Utilité de créer ces groupes et pertinence d'y inclure le mois et la direction du vent
- ▶ Précision bien plus grande pour les percentiles, évite d'éliminer beaucoup trop de lignes

FILTRAGE EN COURS...



201,905 lignes



140,048 lignes

environ 30% éliminés

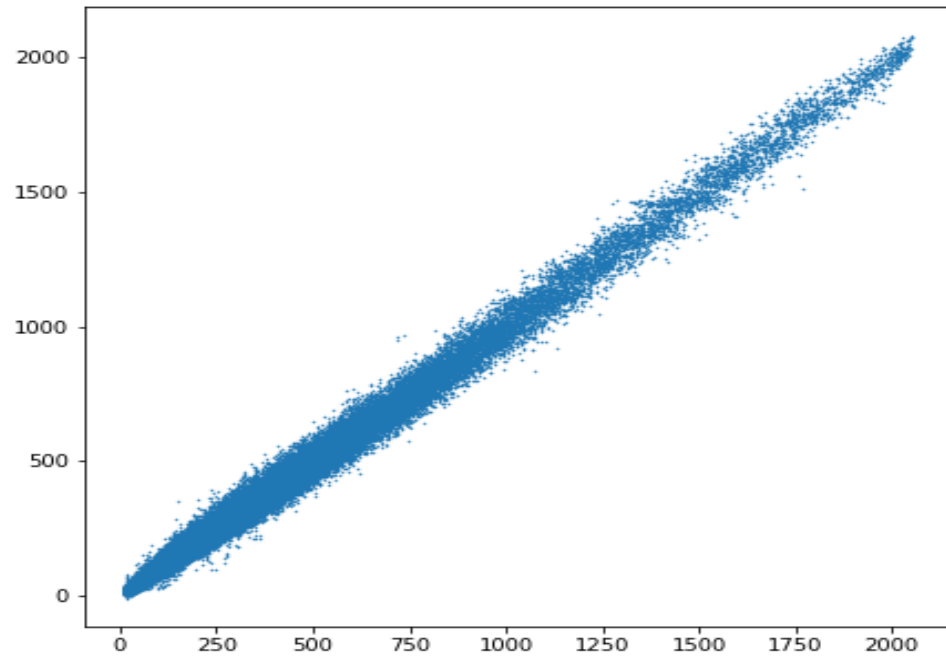
AVANT/APRÈS FILTRAGE

- ▶ Variables explicatives : Mois, vitesse du vent, direction du vent, température
- ▶ Variable à expliquer : la puissance
- ▶ Paramètres RandomizedSearchCV
- ▶ XGBoost

CREATION DU MODÈLE

- ▶ R^2 : 0,992. C'est énorme !
- ▶ Le “tuning” XGBoost.

RÉSULTATS DU MODÈLE

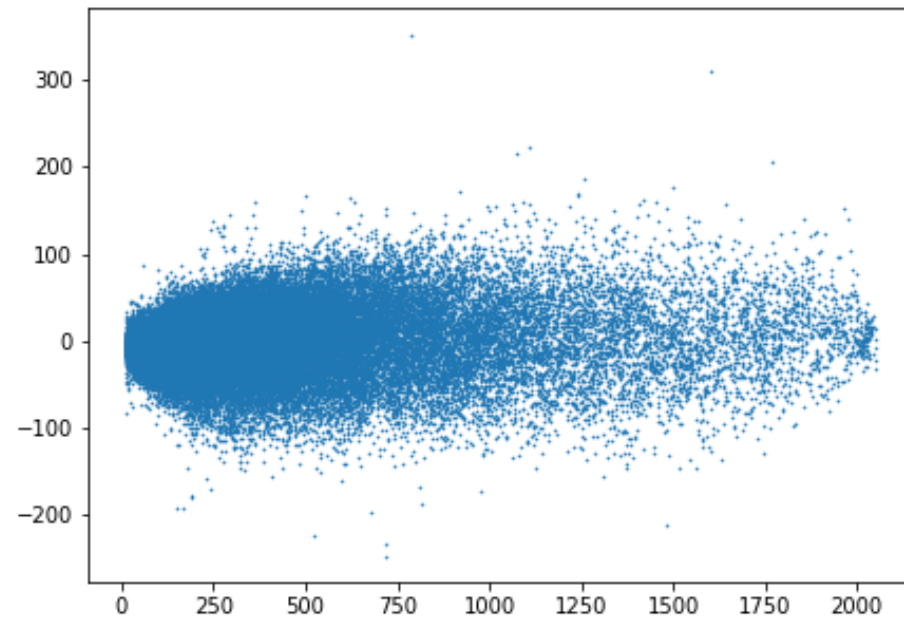


Réalisé sur échantillon
de test

Valeurs observées X
valeurs prédites par la
modèle

Un bon modèle
présente un graphique
linéaire

SCATTERPLOT



Valeurs observées X
(Valeurs observées –
Valeurs prédites par le
modèle = résidu)

Un bon modèle est
1:centré autour de 0
2:variance minimale et
constante

GRAPH DES RÉSIDUS

- ▶ Nous avons un modèle robuste et précis pour la puissance. Il est possible d'adapter pour d'autres variables (Vitesse ou température du générateur,...)
- ▶ Cela peut donc servir d'instrument de contrôle, avec une marge de 10% (5 en-dessous, 5 au-dessus)
- ▶ Si les chiffres ne correspondent pas, il est facile de mettre en évidence un manquement au niveau de l'entretien. Ou en cas d'achat, de démontrer une publicité mensongère.

CONCLUSION

Merci pour votre écoute attentive 😊

DES QUESTIONS ?

Several thin, parallel white lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.