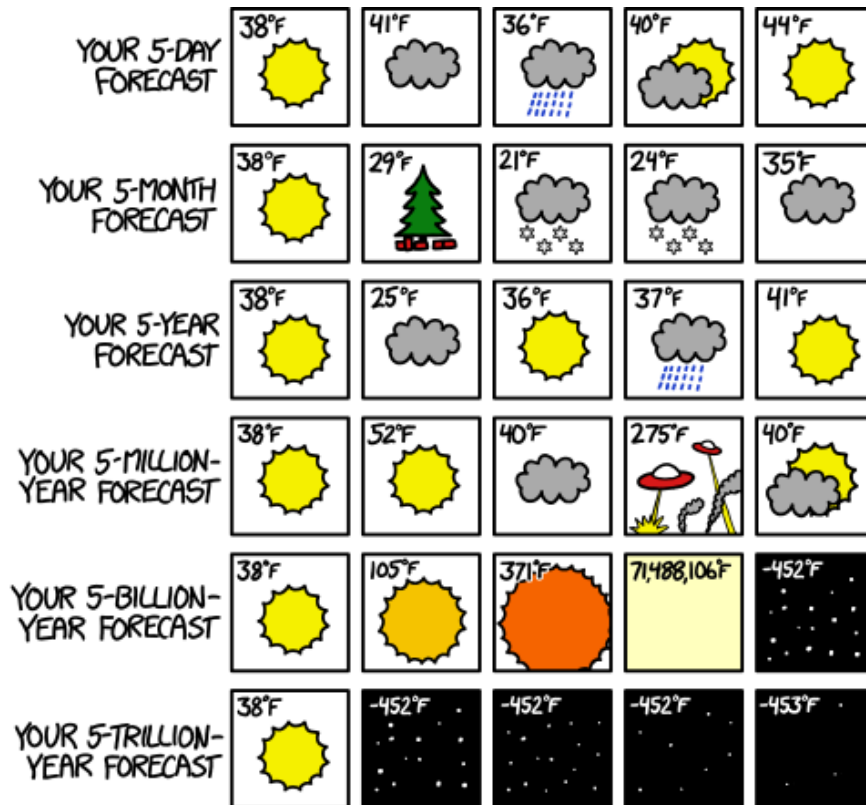




Time series forecasting



Your 5 day forecast

Source: [xkcd](https://xkcd.com/217/)

Objectifs du module

Une introduction de base à l'analyse et la prévision des séries temporelles

- Comprendre les défis de l'analyse de données de séries chronologiques
- Comprendre en quoi les étapes de prédiction de séries chronologiques sont différentes des autres prédictions de machine learning
- Utiliser des modèles de séries chronologiques linéaires conventionnels pour faire des prévisions univariées
- Utiliser une bibliothèque de prévisions de séries chronologiques pour les prévisions univariées

Démarche pédagogique

- Durée du projet : 3 jours
- Work alone but exchange as much as possible between you!
- Produire vos propres scripts et mémos individuels pour terminer le projet

Contexte

Dans le monde numérique actuel, nous collectons des quantités incroyables de données. Tout ce que nous créons dans le monde numérique est horodaté: de nos interactions, clics, achats en ligne, à nos moniteurs de santé comme les «fitbits», à notre position GPS à partir de nos téléphones, à notre «selfie» le plus récent... Plus les capteurs deviennent bon marché et plus nous les intégrons dans nos vies. Nous collectons ainsi de nombreux points de données pour construire des séries chronologiques de notre vie et de notre société. Si certaines données que vous avez actuellement ne sont PAS une série chronologique, elles le seront très certainement dans un avenir proche.

Par extension, les tentatives de prédictions sur l'avenir se multiplient, aussi qualifiées de prévisions. Ce podcast «[Les prévisions sont toujours fausses \(mais nous en avons besoin quand même\)](#)» explique à quoi ressemblent les prévisions modernes. Dans la prévision, la terminologie est particulièrement fournie et un temps certain peut être requis afin de comprendre l'ensemble des méthodes utilisées, leurs spécificités utiles pour différents types de données. Le but de ce module est de vous faire utiliser python pour faire des prévisions robustes et de commencer à appréhender la terminologie utilisée dans la prévision des séries temporelles.

Etape 1 (1 jour) - L'analyse de séries temporelles à but de prédiction

Objectifs de l'activité

- Découvrir ou redécouvrir les notions de base derrière l'analyse de séries temporelles.
- Initier sa propre boîte à outils d'analyse de données de type série temporelle.

Compétences

- Développer une approche méthodologique lors de l'analyse de séries temporelles
- Utiliser des librairies classiques à ce titre
- De la théorie, de la théorie et... du code!

Consignes

- Effectuer les tâches définies dans le jupyter notebook 2023_Campus_ML6-NB1.ipynb. Ce NB vous servira de guide pour construire votre méthodologie afin d'appréhender au mieux toute série temporelle.
- Réaliser un logigramme des étapes à effectuer afin d'appréhender au mieux toute série temporelle sur laquelle s'appuyer pour faire des prévisions.

Livrables

- Le NB complété et déposé sur l'espace github prévu à cet effet:
 - 👉 NB 2023_Campus_ML6-NB2.ipynb renommé avec vos Nom et Prénom mentionné → 2023_NB1_FrançoisRobert.ipynb.

Ressources (implement with your own!)

- <https://www.kaggle.com/code/prashant111/complete-guide-on-time-series-analysis-in-python>
- ...

Si vous avez fini, un peu de préparation pour demain : [Une petite vidéo!](#)

Etape 2 (2 jours) - ARIMA et compagnie

Objectifs de l'activité

- Comprendre en quoi les étapes de prévision des séries temporelles sont différentes des autres prédictions d'apprentissage automatique.
- Comprendre et utiliser un modèle classique de séries temporelles linéaires, et ses déclinaisons, pour faire des prévisions univariées.

Compétences

- Définir une méthodologie typique utilisée lors de la réalisation de prévisions de séries temporelles.
- Utiliser la librairie pmdarima...du code, du code, du code!

Consignes

- Effectuer les tâches définies dans le jupyter notebook 2023_Campus_ML6-NB2.ipynb. Ce NB vous guidera à travers le processus d'utilisation des modèles de base ARIMA sur les données de séries temporelles pour faire des prévisions.

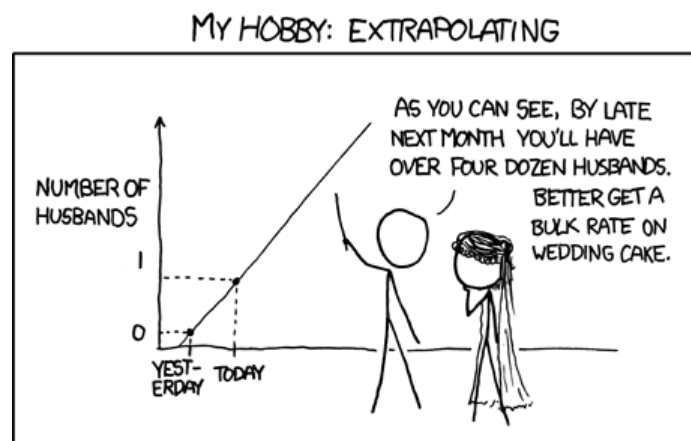
Livrables

- Un mémo qui explique (dans vos propres mots) comment le fractionnement des trains / tests et la validation croisée des données de séries chronologiques doivent être effectués.
- Un logigramme des étapes à effectuer afin d'appréhender au mieux toute série temporelle sur laquelle s'appuyer pour faire des prévisions (Choose your tool!).
- NB 2023_Campus_ML6-NB2.ipynb fini et déposé sur l'espace github crée à cet effet avec vos Nom et Prénom mentionné tels que 2023_NB2_FrançoisRobert.ipynb.

Ressources

- Former les fractionnements des tests et la validation croisée dans les séries chronologiques :
<https://stats.stackexchange.com/questions/14099/using-k-fold-cross-validation-for-time-series-model-selection>
<https://robjhyndman.com/hyndsight/tscv/>
- Forecasting: Principles and Practice livre (Rob Hyndman) :
<https://otexts.com/fpp2/>
<https://otexts.com/fpp2/non-seasonal-arima.html>
<https://otexts.com/fpp2/transformations.html>
- Documentation de Pmdarima:
<http://alkaline-ml.com/pmdarima/>

Etape 3 (1 jour) - L'arrivée d'un "Prophet"



L'extrapolation est une forme de prévision ...

Source : [XKCD](#)

Objectifs de l'activité

- Utiliser un outil de prévision de séries chronologiques automatisé pour faire des prévisions univariées
- Faire des évaluations robustes de l'exactitude des prévisions du modèle
- Comprendre quels sont les avantages et les inconvénients de l'utilisation d'une librairie de prévisions automatisées comme Prophet
- Réaliser un projet de prévision sur un ensemble de données réelles, en prenant des données brutes, en définissant votre objectif, en évaluant votre modèle et en prenant des mesures pour l'améliorer

Compétences

- Savoir utiliser Facebook Prophet pour faire des prévisions
- Développer son esprit critique concernant les librairies accessibles

Consignes

- Effectuer les tâches définies dans le jupyter notebook 2023_Campus_ML6-NB2.ipynb. Ce NB vous guidera à travers le processus d'utilisation des modèles de base ARIMA sur les données de séries temporelles pour faire des prévisions.
- Parcourir les ressources proposées et ajouter les vôtres selon besoins/envies.

Livrables

- NB 2023_Campus_ML6-NB3.ipynb fini et déposé sur l'espace github créée à cet effet avec vos Nom et Prénom mentionné tels que 2023_NB3_FrançoisRobert.ipynb.

Ressources

- Documentation du prophète
<https://facebook.github.io/prophet/>
- Document académique décrivant le prophète
<https://peerj.com/preprints/3190/>
- Pourquoi le prophète existe - "Time series are not sexy"
<https://www.youtube.com/watch?v=pOYAXv15r3A>
- Blog sur le réglage des hyperparamètres dans Prophet
<https://towardsdatascience.com/implementing-facebook-prophet-efficiently-c241305405a3>
- Prophète appliqué à la prévision des données énergétiques
<https://www.kaggle.com/robikscube/time-series-forecasting-with-prophet>

Pour aller plus loin

- Résumer les différents types de modèles de séries chronologiques et leurs différences
https://www.youtube.com/watch?v=d4Sn6ny_5LI
- Essayer d'ajouter des variables exogènes (telles que les prévisions météorologiques) à l'un de vos modèles. Améliorent-ils vos prévisions de modèle? Ceci est possible dans les modèles ARIMA.
https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html
- Essayer de construire un LSTM et comparer ses prédictions à vos autres modèles
<https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-net-work-python/>
<https://machinelearningmastery.com/multi-step-time-series-forecasting-long-short-term-memory-networks-python/>
<https://medium.com/@cdabakoglu/time-series-forecasting-arima-lstm-prophet-with-python-e73a750a9887>

Etape 3 Suite (Optionnel) - Écrire votre propre librairie python de forecasting

Objectifs de l'activité

- Créer une librairie avec une classe (simple mais puissante) permettant de faire du forecasting.

Livrables

- Une librairie fonctionnelle intégrant la comparaison d'au moins deux modèles de prévisions et 2 approches de CV.