

Nom : Petrov

Prenom : Evgeni

Rapport de veille :

Selectionner le modele de prédiction pour les series
temporelle adapté aux besoins.

Sommaire :

Introduction – 2

Etat de l’art – 2

Evaluation des modeles – 3

Choix du modele – 5

Conclusion – 6

Bibliographie - 7

Introduction :

Le présent rapport vise à fournir un état de l'art des méthodologies et des outils existants pour la prévision de la consommation électrique à court terme dans le contexte de la région des Hauts de France. L'objectif de cette veille est de guider la sélection d'outils de modélisation appropriés pour répondre aux besoins de notre entreprise.

Etat de l'art :

Pour pouvoir faire une prédiction sur une série temporelle il existe de nombreux modèles permettant de le faire. La qualité de la prédiction dépend d'une multitude de facteurs, tels que la longueur de la série, le nombre d'hyperparamètres ou encore l'expérience du data-scientist derrière le tuning du modèle.

De manière générale il est très important de connaître la série temporelle sur laquelle on veut faire une prédiction, il faut connaître les choses comme le domaine (finance, alimentaire etc.), le pays etc., un exemple de l'influence de ces informations serait les jours fériés qui ne sont pas les mêmes en Algérie et en France. Parmi les modèles pris en compte dans cette étude (ARMA, ARIMA, SARIMA, VARMAX, Prophet, Xgboost) on constate que certains sont conçus pour des séries temporelles univariées, tandis que d'autres sont plus adaptés aux séries temporelles multivariées[2].

Par exemple, ARMA, ARIMA, SARIMA, sont des modèles qui peuvent être utilisés pour des séries temporelles univariées. Ils sont particulièrement efficaces pour modéliser et prédire des séries temporelles avec une seule variable cible, comme la consommation électrique[3]. En revanche, VARMAX, Prophet et XGBoost sont des modèles conçus pour gérer des séries temporelles multivariées, ce qui signifie qu'ils sont capables de prendre en compte plusieurs variables explicatives, telles que la

température, les jours fériés, ou d'autres facteurs exogènes, qui peuvent influencer la série temporelle cible[4][5].

La sélection du modèle approprié dépendra donc du type de données dont vous disposez et de la complexité de la série temporelle que vous souhaitez modéliser. De plus, il est essentiel de prendre en compte d'autres facteurs tels que la disponibilité de données exogènes, la précision requise pour les prévisions, la facilité d'interprétation du modèle, et la capacité de généralisation aux nouvelles données. Cette analyse de l'état de l'art vise à éclairer le choix du modèle le plus adapté pour notre cas particulier de prévision de la consommation électrique dans la région des Hauts de France.

Evaluation des modeles :

Pour choisir le modele qui correspondrait a notre contexte de prédiction d'une consommation d'électricité sur une semaine il faut avant tout définir les criteres d'évaluation pour differents modeles afin de pouvoir selectionner celui qui serait le plus en phase avec nos besoins[7].

Type de série : Dans notre cas il est important de choisir un modele qui puisse potentiellement traiter des series temporelles multivariés car meme si la consommation d'électricité dans le temps est une serie univarié nous aurons besoin d'integrer la temperature journaliere a notre étude car elle a un impact non negligeable sur la consommation electrique d'un foyer.

Longuer de la série : La quantité de données est toujours un point clé lorsque on veut faire une prédiction et meme si dans notre cas nous avons un échantillon de données d'une grande taille il est tout de meme pertinent d'évaluer le modele selon ce critere pour une utilisation anterieur.

Interprétabilité : L'interprétabilité, ou à quel point le modèle est facilement interprétable, est un critère important dans notre cas, car on devra pouvoir expliquer notre prédiction à notre manager afin que cette dernière puisse l'aider à prendre des décisions éclairées.

Complexité : Ce critère reflète la capacité du modèle à saisir des relations complexes à l'intérieur d'une série. Dans notre cas, c'est une prévision à court terme (1 semaine) qui nous est demandée, il faut donc choisir un modèle assez complexe pour qu'il puisse saisir des variations rapides et fréquentes de notre série temporelle.

Période de prédiction : La durée de la période qu'on cherche à prédire est également un facteur clé pour une prédiction sur une série temporelle. En effet, selon les modèles, leur performance varie en fonction de la durée de la période. Dans notre cas, on cherche une prédiction sur une période courte (1 semaine), il nous faut donc choisir un modèle qui est performant selon ce critère.

Gestion des valeurs manquantes : Les valeurs manquantes peuvent impacter fortement la qualité de la prédiction. Si notre série temporelle comporte des valeurs manquantes, il va falloir faire des opérations afin de les combler ou bien d'amputer les lignes où les valeurs manquent. Cette dernière démarche constitue une réduction de la série temporelle et donc une potentielle baisse de performance du modèle. Il est préférable d'avoir un modèle robuste pour la gestion des valeurs manquantes afin d'optimiser la qualité des prédictions.

Choix du modele :

Nous avons effectué une évaluation théorique de plusieurs modèles selon les critères établis. Les résultats de cette évaluation sont représentés sous forme de tableau ci-dessous :

	ARMA	ARIMA	SARIMA	VARMAX	Prophet	XGboost
Series univariés	✓	✓	✓	✓	✓	✓
Series multivariés	✗	✗	✓	✓	✓	✓
Series courtes	+	+	++	+++	++	+++
Series longues	++	++	++	+++	+++	+++
Valeurs exogenes	✗	✗	✓	✓	✓	✓
Interpretabilité	++	++	++	+	+++	+
Complexité	+	+	+	++	++	+++
Prévisions cour terme	+	+	++	+++	+++	+++
Prévision long terme	++	++	++	+++	+++	+++
Gestion missing data	✗	✗	✓	✓	✓	✓

D'après les critères établis, nous pouvons directement écarter les modèles ARMA et ARIMA. Ces modèles, bien que puissants pour des prédictions sur des séries temporelles univariées, ne travaillent pas avec des séries temporelles multivariées. Or, nous avons besoin d'inclure la température en tant que variable exogène dans notre prédiction.

Parmi les modèles SARIMA, VARMAX, Prophet et Xgboost, on peut voir que SARIMA perd en puissance de prédiction à court terme face aux trois autres candidats, ce modèle sera donc également mis de côté.

D'après notre évaluation, Prophet semble avoir une performance réduite par rapport à VARMAX et XGBoost sur des séries temporelles courtes, mais équivalente sur des séries longues. On voit également que les 3 modèles sont équivalents sur des prévisions à court et long termes. Dans notre cas, nous avons une série longue et nous

avons besoin d'une prédiction à court terme, on ne peut donc pas trancher vis-à-vis de ces critères pour choisir le modèle.

Prophet parmi les 3 modèles semble avoir la meilleure interprétabilité, ce qui est important pour nous car il sera expliqué facilement. Cependant, il perd en complexité, ce qui veut dire qu'il ne pourra probablement pas capturer les variations fréquentes de consommation d'électricité sur la semaine.

Pour avoir une prédiction de qualité dans notre cas, il serait judicieux de tester les modèles XGBoost, Prophet et VARMAX. En effet, le cadre théorique ne suffit pas à trancher sur le modèle le plus adapté au vu des différents points forts des modèles.

Conclusion

Après étude de l'état de l'art en ce qui concerne les prédictions sur les séries temporelles, ainsi qu'établissement de plusieurs critères pertinents pour notre cas d'étude, nous avons pu conduire une évaluation théorique de plusieurs modèles afin de réduire le nombre de candidats en vue de notre étude pratique.

Cette évaluation a permis d'écarter certains modèles (ARMA, ARIMA, SARIMA) au vu de leur incapacité à atteindre les critères clés de notre cas de prédiction. Elle nous a également permis d'identifier plusieurs modèles potentiellement pertinents qui vont participer dans notre étude pratique.

Bibliographie :

- 1- <https://bilan-electrique-2020.rte-france.com/consommation-sensibilite-a-la-temperature-et-aux-usages/#> (Billan de consommation d'électricité en France 2020, données officielles)
- 2- <https://neptune.ai/blog/select-model-for-time-series-prediction-task> (Site de partage et de travail sur différents modèles de machine learning)
- 3- <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/> (Site de Jason Brownlee, renommé pour les informations sur le machine learning.)
- 4- <https://neptune.ai/blog/arma-vs-prophet-vs-lstm> (Site de partage et de travail sur différents modèles de machine learning))
- 5- <https://machinelearningmastery.com/xgboost-for-time-series-forecasting/> (Site de Jason Brownlee, renommé pour les informations sur le machine learning.)
- 6- <https://www.artefact.com/blog/is-facebook-prophet-suited-for-doing-good-predictions-in-a-real-world-project/> (Article écrit par deux data scientist)
- 7- <https://www.ikigailabs.io/blog/how-to-choose-the-best-model-for-time-series-forecasting-arma-prophet-or-mssa> (Site commercial proposant des cours sur le machine learning)
- 8- <https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/> (Site de Jason Brownlee, renommé pour les informations sur le machine learning.)