



Plan de travail prévisionnel

I. Motivation

Pendant de ce projet de preuve de concept, je souhaite travailler sur des données séries temporelles univariées afin d'essayer d'améliorer les performances accessibles des modèles « classiques » comme les modèles Auto ARIMA, PROPHET. Des recherches m'ont permis de découvrir que dernièrement d'autres méthodes plus récentes existent afin d'améliorer les performances.[1], [2] Ces articles sont publiés à partir de 2019, il s'agit donc d'une nouvelle approche. Le principe consiste à appliquer des méthodes pour produire des prévisions probabilistes précises, basée sur la formation d'un modèle de réseau récurrent auto-régressif sur un grand nombre de séries temporelles connexes. Mon projet serait donc d'utiliser cette méthode sur des données plus réelles, moins structurées afin de pouvoir analyser la performance sur ce type de données.

II. Choix de la base de données

Pour utiliser des données les plus réelles possibles, plusieurs bases de données sont disponibles sur le datasets du package darts. Le thème choisi est la prévision du prix actuel de l'Energie en Espagne sur plusieurs villes. Plus de quatre années de données énergétiques et météorologiques sont fournies dans 1 fichiers csv. Les productions sont en mégawatts MW et les prix sont en Euros. L'idée est d'utiliser les informations fournies pour répondre à certaines questions, essentiellement pour voir s'il est possible d'améliorer les résultats obtenus avec les modèles classiques. J'utiliserai trois bibliothèques de séries chronologiques différentes pour essayer d'accomplir cette tâche. Les données sont disponibles dans la bibliothèque [DARTS/datasets](#),

III. Planning prévisionnel

Ce projet se décompose en plusieurs étapes, allant de la collecte des données jusqu'à la création de plusieurs modèles de prédictions. Pour chaque étape je prévois :

- Collecte et nettoyage des bases de données : 2 jours
- Création de nouvelles features si nécessaires et compilation de toutes les données : 3 jours
- Développement des modèles (baseline, modèles classiques, modèles récentes) : 4 jours
- Rédaction des livrables : 3 jours

Cela représente au total environ 2 semaines de travail.

IV. Références

[1] ICLR 2020 · Boris N. Oreshkin, Dmitri Carpow, Nicolas Chapados, Yoshua : << NBEATS: Neural basis expansion analysis for interpretable time series forecasting >>

[2] 13 Apr 2017 · David Salinas, Valentin Flunkert, Jan Gastha : << DeepAR, a methodology for producing accurate probabilistic forecasts >>