
Fouille des séries temporelles sur Weka

Travail 1 : Installation du package timeseriesForecasting sur Weka

1. Lancer weka (version 3.8.0 et +),
2. Sur la fenêtre principale, aller sur le menu "Tools->Package Manager",
3. Choisir "timeseriesForecasting" dans la liste des packages puis cliquer sur "install",
4. Vérifier l'ajout de l'onglet "Forecast" dans l'explorateur Weka

Travail 2 : Test des fonctionnalités de Forecast

1. Ouvrir la base "airline.arff" du répertoire Data de l'application Weka. Cette base représente l'évolution du nombre de voyageur d'une compagnie aérienne du mois de janvier 1949 jusqu'au mois de décembre 1960.
2. Aller sur l'onglet "Forecast",
3. Identifier les différents paramètres utilisés par cet outil à savoir :
 - **Target Selection** : La liste des attributs à analyser,
 - **Number of time units to forecast** : Le nombre de données à prévoir,
 - **Time stamp** : l'attribut utilisé comme axe temporel,
 - **Periodicity** : la période séparant deux données successives,
 - **Advanced configuration -> Base learner** : pour choisir la méthode de régression utilisée (Linéaire, SMOReg, ...)
 - **Periodic attributes** : pour ajouter des attributs temporels (trimestre, semestre, année,...)
 - **Evaluation** : pour choisir les métriques d'évaluation,
 - **Output** : pour choisir les données affichés dans les résultats textuels et graphiques.
4. Mettre le paramètre **Number of time units to forecast** à 6 pour prévoir 6 mois.
5. Lancer l'exploration en cliquant sur "Start".
6. Donner le nombre de passagers de la compagnie pour la période 01/1961-06/1961
7. Vérifier ces données sur le graphique de l'onglet "Train future pred"

Travail 3 Exploration de la base "Diabetes"

1. Ouvrir le site UCI puis télécharger la base "Diabetes".
2. Analyser la structure des fichiers de la base.
3. Nous souhaitons prévoir les doses d'insulines (code 33) pour les deux jours suivants la période des prélèvements en utilisant l'outil Forecast sur Weka. On vous demande de
 - Transformer la base en un fichier Weka avec un champs "DateTime" comme axe temporel,
 - Créer un fichier arff pour le code 33 uniquement,
 - Utiliser la méthode d'apprentissage "SMOReg" pour prévoir les deux jours suivants la période des prélèvement.