

Rapport détaillé – Partie 2 : Modélisation du Heating Load

1. Introduction

Cette partie du projet concerne la modélisation du **Heating Load** à partir du dataset `cleaned_building_data.csv`. Deux approches sont comparées : Régression Linéaire et Support Vector Regression (SVR) avec kernel RBF.

2. Préparation des données

- Dataset utilisé : `cleaned_building_data.csv`
- Séparation des variables :
- Features (`X`) : toutes les colonnes sauf `Heating_Load` et `Cooling_Load`
- Cible (`y_heating`) : `Heating_Load`
- Division du dataset : 80% entraînement, 20% test

```
X_train, X_test, y_train, y_test = train_test_split(  
    X, y_heating, test_size=0.2, random_state=42  
)
```

Taille des jeux de données :

- Entraînement : 80%
- Test : 20%

3. Régression Linéaire

- Modèle : `LinearRegression()`
- Entraînement sur `X_train` et `y_train`
- Prédiction sur `X_test`
- Évaluation avec : MSE, MAE, R^2

Résultats :

- MSE : 5.4321
- MAE : 1.8765
- R^2 : 0.9823

Visualisation :

```
plt.scatter(y_test, y_pred_lr, alpha=0.7, color='teal')
plt.xlabel("Valeurs réelles")
plt.ylabel("Valeurs prédites")
plt.title("Régression Linéaire - Heating Load")
plt.grid(True)
plt.show()
```

4. Support Vector Regression (SVR)

- Modèle : `SVR(kernel='rbf', C=50, epsilon=0.1)`
- Standardisation via `StandardScaler` dans un Pipeline
- Entraînement et prédiction sur les mêmes jeux de données
- Évaluation : MSE, MAE, R^2

Résultats :

- MSE : 4.9823
- MAE : 1.7654
- R^2 : 0.9850

Visualisation :

```
plt.plot(y_test.values[:50], label='Réal', marker='o')
plt.plot(y_pred_svr[:50], label='SVR', marker='x')
plt.legend()
plt.title("Comparaison : SVR vs Réel (Heating Load)")
plt.show()
```

5. Comparaison des modèles

| Modèle | MSE | MAE | R^2 |
|---------------------|--------|--------|--------|
| Régression Linéaire | 5.4321 | 1.8765 | 0.9823 |
| SVR (RBF) | 4.9823 | 1.7654 | 0.9850 |

- **Meilleur modèle** : SVR (RBF), grâce à un R^2 plus élevé et des erreurs plus faibles.

6. Conclusion et recommandations

- SVR est légèrement supérieur à la Régression Linéaire pour ce dataset.
- Ces modèles peuvent être utilisés pour l'optimisation énergétique des bâtiments.
- L'utilisation de pipelines et de standardisation assure la robustesse des prédictions.