March 11th 2024

Electricity Price Explanation

Ensemble Learning

Hongyang YE | Irene SUNNY | Wenjing ZHAO | Zheng WAN

Github link

Content

01 Introduction

Introduction

Problem
Several factors including weather, energy, and commercials affect the price of electricity making estimation a challenge

Goal
Learn a model that outputs from these explanatory variables a good estimation for the daily price variation of electricity futures contracts in France and Germany

Input Data
35 columns
(ID, DAY ID, COUNTRY
Weather measures etc)

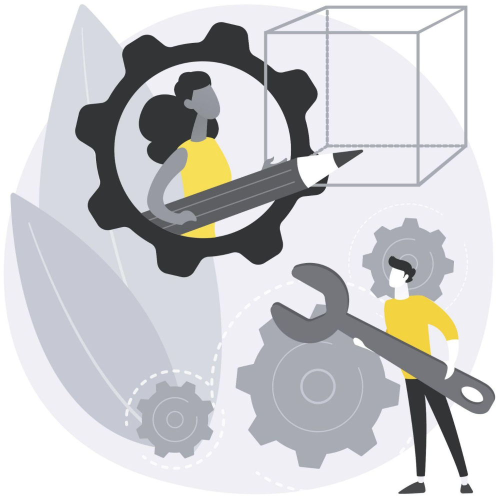
Output Data
ID, Target (Daily price variation for futures of 24H electricity baseload),

Benchmark
Simple Linear Regression

Approach based on
Ensemble Models

Source: ENS Data Challenge

A Classic Supervised Learning Problem



02
EDA & Preprocessing

EDA & Preprocessing

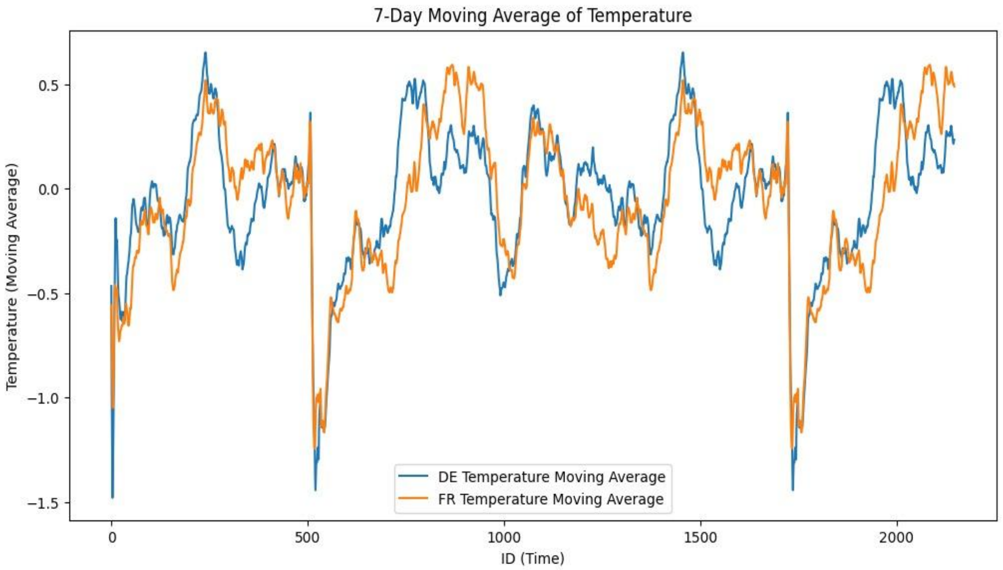
Key observations and the data pre processing solutions to the same are as below

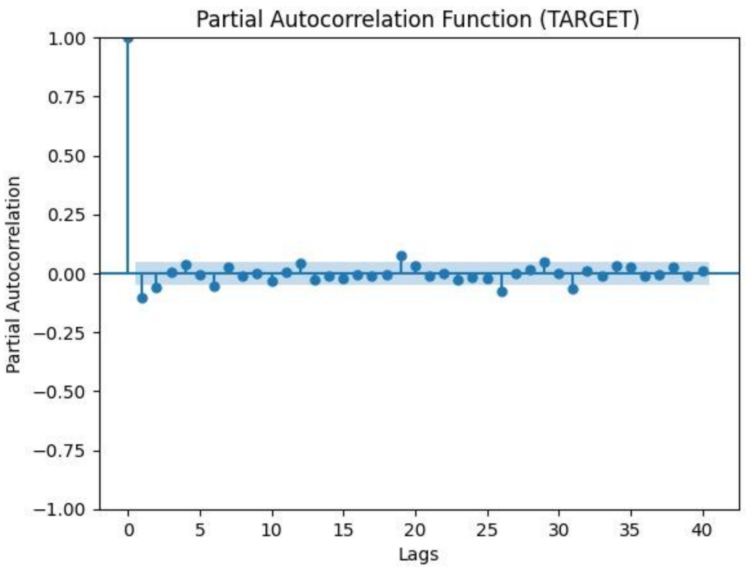
01. Missing Values	Considers both backward and forward values, and assigns the average or a single valid neighbor if only one exists. In cases of no neighbors, use a default value of zero.
02. Scale Difference	Performs normalization using a StandardScaler to improve model convergence and reduced bias
03. Multicollinearity Amongst Features	Performs PCA using a PCA object with n_components set to 0.95. This retains 95% of the variance in the data while reducing the number of dimensions.

Detailed EDA here

03
Feature Engineering

Feature Engineering





- From EDA, we captured the sequential trends in our dataset, therefore, time-series specified models might have a good performance.
- We checked the trends with different features to validate our conclusions, and check the autocorrelation function and partial autocorrelation function.
- Based on the results, we add 1 lag of each feature as new features to our dataset.

04

Evaluation

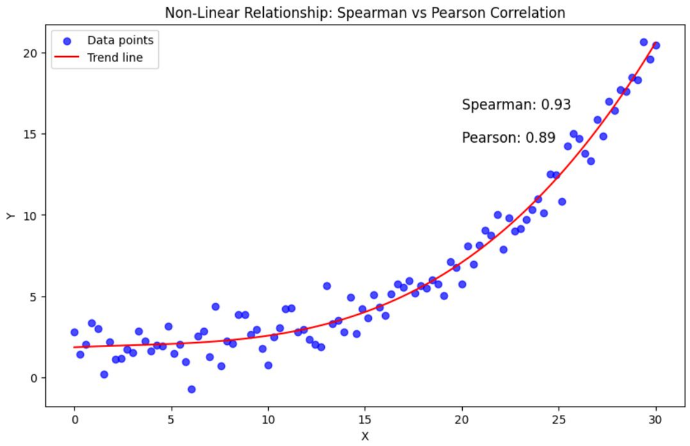
Metric: Spearman Correlation

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

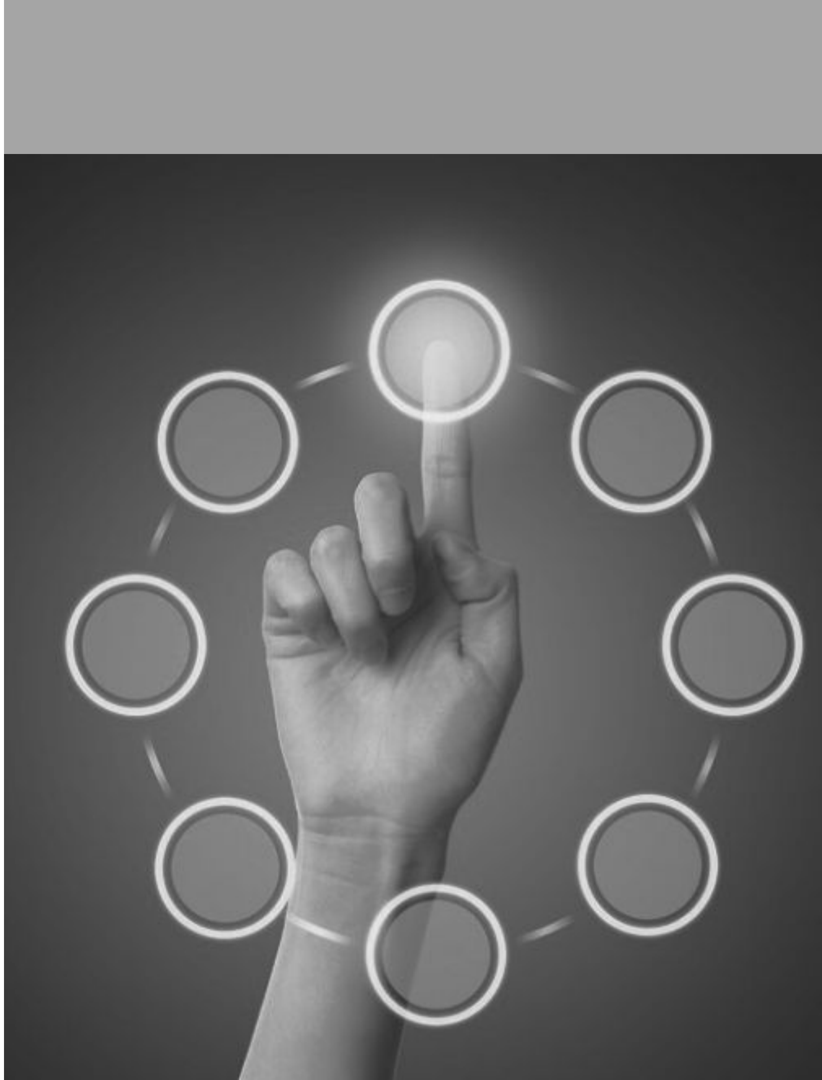
Spearman Correlation (ρ) measures the strength and direction of monotonic associations between two variables, making it suitable for capturing non-linear associations.

Rationale:

- **Captures non-linear relationships:** Electricity prices are influenced by multiple variables with non-linear relationships
- **Less sensitive to outliers:** Electricity prices can contain outliers due to external factors
- **Model assessment:** evaluates and provides insights into how well different models tested captures the underlying relationships between variables and electricity prices



05
Models & Results



1. Linear Regression (Benchmark)

Train: 27.3%

Test: 15.1%

$$Y = b_0 + X_1 b_1 + \cdots + X_n b_n + \varepsilon$$

Rationale: Baseline model to understand relationships between independent & dependent variables

Limitations:

- **Underfitting**: Model's performance is not strong in either the training (27.3%) or test (15.1%) set
- **Linearity Assumption:** assumes linear relationship between independent and dependent variables, causes oversimplification of model as relationship between weather conditions and electricity demand is non-linear
- **Generalisation**: Dropping 'country' column assumes that model findings are equally applicable to both countries, might not account for country-specific factors

Insights:

- Make use of decision tree to capture more complex patterns in data to decrease potential of overfitting
- Addresses problem of non-linearity as decision tree model does not assume linearity and is able to handle non-linear relationships between various features and the target variable (electricity price)

2. Decision Tree

Train: 19.8%

Test: 14.7%

Overview:

Weak supervised learning method that recursively segregates data into branches based on feature values.

Rationale:
- Capable of capturing complex patterns and non-linear relationships that linear regression cannot capture

Methodology and/or Hyperparameter Tuning:
- Tuned with Bayesian optimisation
- Best parameter found: 'max_depth': 83, 'min_impurity_decrease': 0.0212, 'min_samples_leaf': 19, 'min_samples_split': 20

Limitations:
- **Overfitting**: decrease in correlation from train to test set indicates high variance as model captures noise in training data leading to overly complex trees with large depth that do not generalise well to unseen data.
- **Bias**: trees can be biased if training set is imbalanced
- **Axis-aligned split:** at each node, only one feature is considered before making the split, causing inefficiency in capturing relationships between features that are not axis-aligned

3. Moving Average

Train: 51.5%

Test: 46.4%

Overview:

Forecasting technique for time series analysis to predict future values based on the weighted sum of past values to highlight long-term trends instead of short-term changes

Rationale:

- Identify trends by smoothing noise and short-term fluctuations to accurately capture historical trends

Methodology and/or Hyperparameter Tuning:

- Hyperparameter Tuning: experimented with different window sizes to calculate moving average
- Used forward and backward-fill to handle missing values and train a linear regression model using features from moving average

Limitations:

- **Assumes data stationarity:** may not perform well to non-stationary data with differing seasonal trends
- **Difference in model performance:** difference when trained on data from Germany (71.4%) compared to data from France (21.0%) suggests that relationship between features and electricity price vary between countries
- **Complex models:** consider other timeseries/ML models to handle non-linear relationships (e.g. ARIMA, LSTM)
- **CV & Backtesting** to obtain more robust model that can generalise well.

3. Moving Average

Without Feature Engineering

	Model	DE_Train_Score	FR_Train_Score	Overall_Score
0	dt	0.284955	0.077280	0.175760
1	bagging_ridge	0.491419	0.158929	0.313878
2	extra_trees	0.199631	0.133809	0.156962
3	rf	0.291290	0.186128	0.225199
4	bagging_knn	0.134548	0.102892	0.107535
5	bagging_svr	0.412729	0.245307	0.320654
6	bagging_linear	0.491704	0.156625	0.314120
7	adaboost	0.348586	0.102280	0.196787
8	gradient_boosting	0.294186	0.265961	0.268494

With Feature Engineering

	Model	DE_Score	FR_Score	Overall_Score
0	dt	0.522973	0.063219	0.272966
1	bagging_ridge	0.768485	0.232256	0.526632
2	extra_trees	0.718865	0.099351	0.410986
3	rf	0.671317	0.061491	0.366640
4	bagging_knn	-0.242485	0.123651	-0.042472
5	bagging_svr	-0.075682	0.089942	-0.035366
6	bagging_linear	0.772510	0.228310	0.525331
7	adaboost	0.560786	0.106137	0.296425
8	gradient_boosting	0.650112	0.046904	0.354568
9	xgboost	0.698499	0.084659	0.398867

Comments:

- Overall, Spearman Correlation across models increased on train dataset with the addition of prev_day features with the exception of Bagging-KNN & Bagging-SVR
- The decrease in Spearman Correlation for these two models are most likely due to the addition of prev_day features as the designs of the models are not suitable for time-series data
- Following which, we included these new features into the following ensemble methods: bagging, boosting, and stacking to further boost Spearman correlation score on the test set

4. Ensemble - Bagging

Ridge	SVR	Linear
Train: 31.4%	Train: 32.1%	Train: 31.4%
Test: 16.7%	Test: 17.3%	Test: 16.7%

Overview:

Obtains multiple samples from training dataset and train a model for each sample with the final prediction averaged across predictions of each individual model to reduce variance

Rationale:

- **Reduce variance:** tackle models with high variance such as decision trees to improve accuracy & overfitting

Methodology and/or Hyperparameter Tuning:

- Use of base models including Ridge Regression, SVR, and Linear Regression, includes feature selection
- Hyperparameter tuning performed with Optuna on 'max_depth', 'min_samples_split', 'min_samples_leaf', 'max_features', 'max_leaf_nodes', 'min_impurity_decrease', 'max_samples', and 'ccp_alpha'
- Cross-Validation: evaluates performance of a model through training it on some subsets while validating it on others to ensure that the model generalises well on unseen data (prevent overfitting)

Limitations & Insights:

- SVM performed better than the other models as it captures underlying patterns better but the overall correlation similarity indicates that the choice of base model does not greatly improve results for this dataset
- Overfitting: drop in correlation from train to test set suggests that model does not generalise well to new data
- Model Complexity: choice of base models might not be complex enough for the structure of the dataset

5. Ensemble – Boosting

AdaBoost	GradBoost	XGBoost
Train: 38.2%	Train: 52.8%	Train: 74.5%
Test: 32.8%	Test: 35.9%	Test: 40.5%

Overview:

Sequential ensemble of models that iteratively correct errors from preceding models with weights determined by accuracy. Predictions aggregated via weighted sum or majority vote.

Rationale:

- **Reduce bias-variance trade-off:** reduces bias and variance to improve generalisation of model to new data
- **Flexible & Accurate:** can be performed with decision tree to boost accuracy on complex dataset

Methodology and/or Hyperparameter Tuning:

- GridSearchCV over hyperparameters ‘n_estimators’ & ‘learning_rate’
- Cross-Validation: evaluates performance of a model through training it on some subsets while validating it on others to ensure that the model generalises well on unseen data (prevent overfitting)

Limitations & Insights:

- XGBoost has the highest capacity to capture complex patterns in data with its highest performance on train and test sets but the notable decrease might indicate overfitting
- AdaBoost & Gradient Boosting could be explored with deeper depths for base learners to improve results as the current fixed depths may not be optimal

6. Ensemble – Stacking

Train: 77.8%

Test: 39.7%

Overview:

Combination of the predictions of various base models' predictions to reduce overfitting, providing a more robust prediction

Rationale:

- Better performance: combine the strengths of all models to generalize better especially for complex datasets and reduce the impact of outliers

Methodology and/or Hyperparameter Tuning:

- Used train results of all base models including Decision Tree, Random Forest, SVM, Gradient Boosting, XGBoost, AdaBoost, Moving Average, and Bagging model based on ridge regression

Limitations & Insights:

Theoretically, stacking should decrease variance and bias thereby increasing Spearman correlation of the best performing model. The decrease in correlation despite positive individual performances could be due to the following reasons:

- **Overfitting**: model learns noise in training data instead of underlying pattern, causing poor results on new data
- **Unoptimised model aggregation:** weights used to combine base models might be unoptimised in a way that base models that overfit the training data are weighted higher than others

Further exploration is required to correctly identify reasons for the sharp decrease in Spearman correlation.

Summary of Model Performance

Model		**Spearman Correlation**
Linear Regression		15.1%
Decision Tree		14.7%
Bagging	Ridge Regression	16.7%
	SVR	17.3%
	Linear Regression	16.7%
Boosting	AdaBoost	32.8%
	GradBoost	35.9%
	XGBoost	40.5%
Stacking		39.7%
Moving Average		46.4%

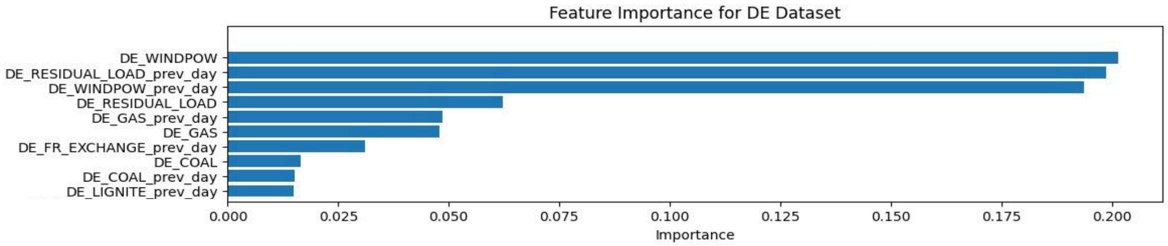
Ranking	Date	User(s)	Public score
1	Sept. 13, 2023, 8:03 p.m.	BenAmara.MohamedAli	0.6083
2	Feb. 22, 2024, 4:36 p.m.	emmanuel2024	0.5507
3	Feb. 22, 2024, 3:26 p.m.	Emmanuel2024-2	0.5347
4	Feb. 20, 2024, 10:30 p.m.	xiaobaoren123	0.5287
5	Sept. 23, 2023, 4:40 a.m.	mb	0.5280
6	May 11, 2023, 12:14 p.m.	rinv	0.5091
7	Nov. 12, 2023, 7 p.m.	Jouini	0.4889
8	Feb. 20, 2024, 10:21 a.m.	ssaji	0.4889
9	Nov. 13, 2023, 1:24 p.m.	Aziz	0.4884
10	Feb. 22, 2024, 12:24 a.m.	saji8953	0.4703
11	**Feb. 18, 2024, 8:54 p.m.**	**wz & IreneSUNNY & HongyangYE & icebluewatermelons**	**0.4641**
12	Feb. 24, 2024, 11:16 a.m.	wz001202	0.4610
13	Dec. 12, 2023, 1:58 p.m.	eti	0.4599
14	Dec. 3, 2023, 1:04 a.m.	chamine	0.4509
15	Feb. 21, 2023, 2:43 p.m.	mica31	0.4022

Team's standing in the challenge

06

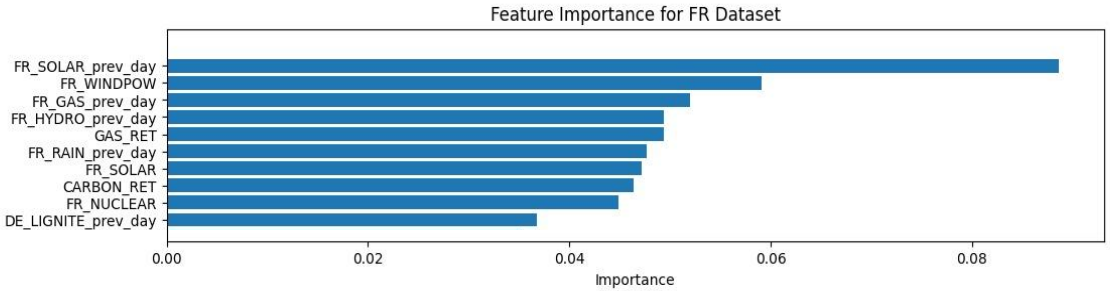
Business Implications

Feature Importance



Germany:

- Top 3 features stand out
- Most critical features are those related to wind power, residual load, as well as the information surrounding the previous day's



France:

- Previous day's data are influential, the level of importance of a single feature is more evenly distributed than that of Germany
- Nuclear power have more influence in France than in Germany due to the country's reliance on nuclear

Features highlighted here are the most important features only for the Spearman Correlation metric used in this challenge. Results of this feature importance might vary for other evaluation methods like MSE.

Insights and Considerations

For Germany, focusing on wind power and previous day's residual load can enhance forecasting accuracy, allowing for **better energy resource allocation and trading strategies**.

In France, the importance of solar power and gas from the previous day suggests a need for **diversified energy strategies** that account for the variability of renewable sources.

Previous day's energy metrics are the most important features for both datasets, highlights that **short term historical data** is important in this modelling energy prices.

Companies can optimize operations, reduce costs, and increase profitability by **tailoring energy procurement** and investment decisions to these insights.

Understanding the distinct energy profiles of each country can inform cross-border energy trade and policy development, promoting a more **efficient and sustainable energy market** in Europe

07
Conclusion

Challenges & Explorations

1. **Balance between accuracy and overfitting**
   Due to the limited scale of the datasets, it is very difficult to achieve balance between the accuracy of our models and avoiding overfitting at the same time.

2. **Sequential features not always working**
   As presented in previous sessions, results are improved in general after sequential features were added. However, some models like SVR and KNN resulted in negative values, indicating their unsuitability for time-series data by nature

3. **Results from ensemble learning methods are not promising**
   Theoretically, results obtained from stacking should be better than all the base models utilized. However, our stacking result did not improve than that of our MA model. Potential reasons could be further explored.

Takeaways

1. **Ensemble Learning as a Robust Approach :** Particularly boosting techniques like XGBoost, GradBoost, and AdaBoost, demonstrated strong potential in improving prediction accuracy by reducing both bias and variance.

2. **Model Evaluation with Spearman Correlation:** This metric was instrumental in capturing non-linear relationships between variables.

3. **Customized Modeling for Country-Specific Insights:** The differences in feature importance between France and Germany highlighted the need for customized models to accurately predict electricity prices in each country.

4. **Continuous Learning and Improvement:** Learnt the importance of iterative modeling, tuning and continuous experimentation

March 11th 2024

ESSEC BUSINESS SCHOOL
CentraleSupélec

Thank you!

08
Appendix

Gantt Chart

	8 Jan - 15 Jan	15 Jan - 30 Jan	1 Feb - 15 Feb	15 Feb - 29 Feb	1 Mar - 4 Mar	11 Mar
Task 1						
• Project Selection & Proposal						
Task 2						
• EDA & Slides						
• EDA Presentation			12 Feb			
Task 3						
• Modelling						
• Oral Defence Slides						
• Oral Defence						

Note: Responsibilities were evenly divided amongst all team members, each member collaborated effectively & efficiently.