

## LES MEMOS DU CEEME

### Le Machine Learning au cœur de stratégies de trading et d'optimisation intraday en Allemagne.



Le « machine learning » ou apprentissage automatique constitue un champ d'étude particulièrement prometteur de l'intelligence artificielle, notamment appliqué au domaine du trading et de l'optimisation. Ce Mémo décrit une application concrète réalisée par le CEEME en collaboration avec la BU GEM et les perspectives associées à ces nouvelles approches. Bonne lecture.

Olivier Lecointe, Directeur du CEEME

Stimulés par la croissance des énergies renouvelables, les marchés *intraday* de l'électricité occupent une place de plus en plus importante dans les salles de marché des grands groupes énergétiques. Le but de l'*intraday* est de permettre aux producteurs de satisfaire leurs engagements de fourniture d'électricité lorsque la production et/ou la consommation dévie des prévisions faites en *day-ahead* (les plans de production devant être soumis sur les bourses de l'électricité au plus tard la veille à midi).

#### Le marché *intraday* allemand : un terrain fertile pour les stratégies basées sur le *machine learning*.

Le marché *intraday* allemand comporte plusieurs caractéristiques qui le rendent particulièrement bien adapté au développement de stratégies de trading et d'optimisation automatisées.

Il s'agit tout d'abord d'un marché très physique, peu spéculatif. Les mouvements de prix résultent souvent de fondamentaux mesurables, les volumes en jeu sont relativement faibles et les acteurs nombreux.

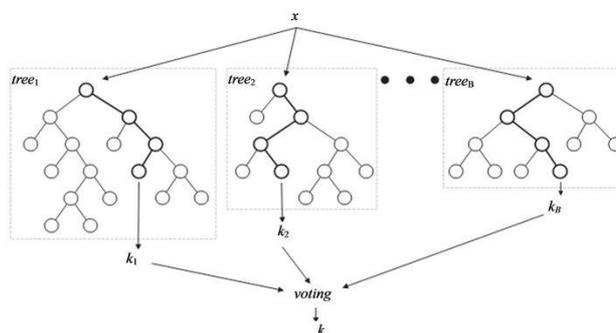
Il existe en revanche un grand nombre de facteurs qui peuvent être à l'origine d'importants mouvements de prix *intraday*, parmi lesquels : les changements de prévisions météorologiques, les indisponibilités non-programmées d'unités de production, l'état du réseau de transport électrique, la flexibilité du parc de production, etc... Il est parfois difficile de juger de l'impact qu'aura un événement particulier tant le nombre de facteurs en jeu est grand et tant ces différents facteurs sont interdépendants les uns des autres.

Dans ce contexte, le *machine learning* est utile pour détecter et exploiter des situations de marché dans lesquelles les prix *intraday* diffèrent de leur comportement attendu, et qu'il ne serait pas possible de détecter à l'aide d'approches statistiques classiques, car le nombre de variables pouvant influencer de manière concomitante sur la formation des prix est trop important.

#### Analyser le marché intraday grâce à un algorithme de *machine learning*.

Le *machine learning* permet de déceler des *combinaisons de facteurs* dans lesquelles on observe statistiquement un phénomène que l'on cherche à détecter. Dans notre cas, il s'agit de détecter des situations de marché dans lesquelles les prix s'écartent sensiblement de la valeur prédite par un modèle. Ce modèle est construit par *apprentissage* des relations entre l'ensemble des variables observées ces derniers mois ou années.

Lorsque le prix de marché dévie du prix prédit par le modèle, on peut alors prendre une position dans le marché, en anticipant que le prix de marché va évoluer vers le prix prédit par le modèle. En systématisant l'approche, on peut générer un profit substantiel lorsque le système parvient à prévoir correctement l'évolution des prix avec un taux d'erreur suffisamment bas pour couvrir les frais de transaction et le différentiel entre les prix d'achat et de vente (bid-ask spread).



On peut utiliser des arbres de décisions pour classer une situation de marché en opportunité d'achat ou de vente. Sur le graphe ci-dessus, on illustre comment combiner différents arbres pour augmenter la puissance prédictive. Les *Boosted Trees* cherchent à trouver une combinaison linéaire optimale d'arbres de décision, tandis que les *Random Forests* utilisent pour chaque arbre un échantillon aléatoire de l'ensemble d'apprentissage.



## Une « forêt » de décision pour atteindre un taux de classification correcte de 57%.

Il existe un large panel d'algorithmes et il convient de choisir l'approche qui sera la mieux adaptée au problème posé. Nous avons opté pour des méthodes basées sur la juxtaposition d'un grand nombre d'arbres de décision (*boosted trees*, *random forests*, ...). Ces méthodes sont particulièrement recommandées pour les applications de *data mining*, dans lesquelles on tente d'extraire de la connaissance de larges bases de données. Le principe consiste à constituer un large ensemble d'arbres de décision, chacun de ceux-ci donnant une indication quant à la décision à prendre. En combinant ces indications, on arrive à créer un « comité » dont la puissance prédictive est importante.

Après un long processus itératif d'exploration de données et de modélisations, nous sommes parvenus à prédire correctement le sens de l'évolution des prix dans environ 57% des cas. Cela peut paraître faible à première vue, mais c'est assez remarquable dans un contexte d'analyse de marché.

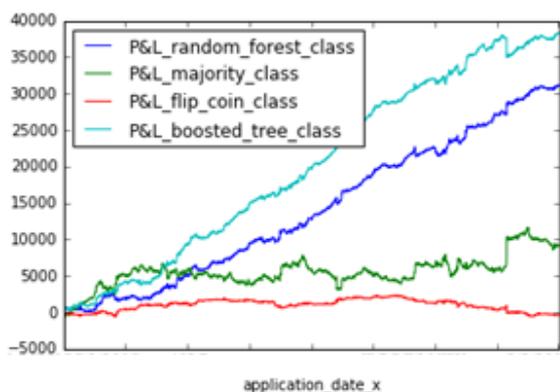
### Construction d'une stratégie et estimation de ses performances

Les deux courbes en bleu de l'illustration ci-dessous montrent les résultats que *pourraient atteindre* les stratégies sur une période de 12 mois, lorsqu'un mégawattheure (MWh) est acheté ou vendu tous les quarts d'heure via les produits les plus liquides.

Les résultats présentés sont obtenus par une méthode dite *k-fold cross-validation* (avec  $k=12$ ). En d'autres termes, on met en place un processus itératif dans lequel 11 mois de données sont utilisés pour apprendre un modèle, qui sera appliqué sur le 12<sup>ème</sup> mois restant, n'ayant pas participé à l'apprentissage du modèle. En prenant à chaque itération un mois différent pour l'application du modèle et en agrégeant ensuite les résultats, on construit une estimation de ce que la stratégie pourrait produire sur 12 mois de données non-connues.

### Présentation des résultats

La courbe bleu foncé représente les résultats d'un apprentissage de type *Random Forest*, tandis que la courbe bleu clair correspond à ceux de la méthode *Boosted Trees*. Ces deux méthodes, bien que semblables, diffèrent dans la manière d'agréger le grand nombre d'arbres de décision construits pour produire un résultat final.



A titre de comparaison, il est intéressant de confronter les résultats obtenus avec ceux d'une règle consistant à acheter ou vendre systématiquement, en choisissant *a posteriori* l'option la plus lucrative des deux (*Majority Rule*, en vert dans le graphe). En rouge, l'approche *Flip-coin* décide, à chaque quart d'heure, d'acheter ou de vendre à pile ou face.

Le meilleur modèle (basé sur les *Boosted Trees*) réalise un gain moyen d'environ 0,80 €/MWh, pour des frais de transaction moyens estimés aux alentours de 0,20 €/MWh (non pris en compte dans cette simulation). Il convient donc de considérer les résultats présentés comme une borne supérieure de ce qu'il est possible d'atteindre, si le marché était parfaitement liquide et les frais de transaction nuls.

### Que nous apprennent ces modèles ?

Le projet présenté ici nous a montré qu'il semblait possible de mettre en place des stratégies de trading et d'optimisation profitables, dans le marché *intraday* allemand, en se basant sur des modèles construits par un algorithme de *machine learning*. Ces modèles sont actuellement testés en production pour déterminer s'ils délivrent des résultats satisfaisants en situation réelle.

Un autre enseignement que l'on peut tirer de ce type d'approche est de nature *exploratoire* : quelles sont les variables qui influencent le plus les prix *intraday* ? quelles sont celles qui n'ont que peu d'impact ?

En analysant les fréquences d'occurrence des différentes variables dans les arbres (ainsi que leur position respective dans ces arbres), on peut les classer par ordre décroissant d'influence relative.

Dans notre exemple, on constate ainsi que le prix *day-ahead* horaire est la variable la plus importante et celle qui sera la plus souvent utilisée, *en conjonction avec d'autres*, pour déterminer la direction attendue des prix.

Dans les positions suivantes, on retrouve par ordre décroissant d'influence :

- La charge résiduelle
- La quantité de production éolienne prévue en *day-ahead*
- L'évolution des prix *intraday* observés les 15 minutes précédant la décision
- La déviation totale (par rapport au *day-ahead*) des prévisions de production renouvelable (vent+solaire) 4 heures avant l'expiration du produit
- La quantité de production solaire prévue en *day-ahead*
- L'heure du produit considéré
- Les déviations de production renouvelable annoncées l'heure précédant la décision
- Le temps restant jusqu'à expiration du produit
- Le spread observé entre le prix *intraday* et le prix *day-ahead*, 3 heures avant la décision
- ...

Au-delà, l'étude révèle un des aspects les plus intéressants du *machine learning* : la capacité à apprendre et véritablement comprendre, à partir des données, la dynamique des prix *intraday*.

Contacts CEEME : Olivier Martin et Marcelo Espinoza