

Hackathon
QMI/LFIS/SESAMm
05-06 Février 2021

Sujet : Prédiction de volatilité court-terme avec données de marché et données alternatives.

Le but du projet est de prédire la volatilité intrinsèque réalisée.

Les prédictions de volatilité sont utilisées dans le cas de la valorisation d'options ainsi que dans la construction d'allocations jouant la prime *low vol.*

Cet exercice est un exercice de régression dans lequel les prédictions très éloignées de la réalité seront pénalisées, la métrique utilisée étant la *mean squared error*.

Chaque observation correspond à un close d'une journée de trading.

Objectif : Le but de ce sujet est de prévoir la volatilité intrinsèque sur 22 jours à l'aide des indicateurs de marché de la phase 1 :

Matériel : Durant la phase 1 vous disposerez des fichiers suivants :

- ref_train_x.csv : les variables explicatives sur les observations d'entraînement.
- ref_train_y.csv : les valeurs de la volatilité intrinsèque réalisée sur les observations d'entraînement.
- ref_test_x.csv : les variables explicatives sur les observations de test. Vous pourrez soumettre vos prédictions durant la première phase de la compétition et recevoir automatiquement le score obtenu (voir ci-dessous le processus d'évaluation).

Évaluation : Le fichier ref_valid_x.csv contient les variables explicatives sur les observations de validation. Vous ne pourrez pas recevoir votre score automatiquement sur ces échantillons durant la première phase de la compétition. Les organisateurs préviendront samedi matin les participants (à partir de 9H30), sur le chat Telegram du Hackathon, pour annoncer la fin de la phase de test et le début de la phase de validation. Vous pourrez alors soumettre vos prédictions sur les échantillons de validation, mais ne recevrez pas votre score. Il sera révélé à l'issue des soumissions de la totalité des équipes.

Phase finale : Après le déjeuner, les 4 meilleures équipes seront sélectionnées. Elle disposeront alors des fichiers suivants et ne pourront faire qu'un nombre limité de soumissions sur l'échantillon final de validation.

- `ref_train_x_final.csv` : les variables explicatives, sur les observations d'entraînement et de test de la première phase avec l'ajout des données alternatives.
- `ref_train_y_final.csv` : les valeurs de la volatilité intrinsèque réalisée sur les observations d'entraînement et de test de la première phase.
- `ref_valid_x_final.csv` : les variables explicatives sur les observations de validation de la phase 2.

Processus d'évaluation :

Votre score final sera proportionnel à l'inverse d'une *mean squared error*.

$$mse = \left(\frac{1}{n}\right) \sum_{i=1}^n (y_{real} - y_{pred})^2$$

Il est ainsi attendu des participants qu'ils soumettent des fichiers de prédiction qui respectent les normes suivantes :

- un float par ligne, correspondant à la prédiction de vol intrinsèque
- le point (".") comme séparateur décimal

Indicateurs de marché :

- **id** : L'identifiant correspondant à chaque action.
- **close_vol_adjusted_252** : La volatilité intrinsèque sur les 252 derniers jours.
- **close_vol_adjusted_22** : La volatilité intrinsèque sur les 22 derniers jours.
- **close_vol_22** : La volatilité réalisée sur les 22 derniers jours
- **close_vol_252** : La volatilité réalisée sur les 252 derniers jours
- **market** : Le niveau du marché renormalisé.
- **is_earning_5** : Présence d'earnings dans les 5 prochains jours.
- **is_earning_22** : Présence d'earnings dans les 22 prochains jours.
- **sector** : Un code représentant une classification sectorielle.
- **close_log_return_adjusted** : Le log return sur la journée précédente
- **market_cap** : La capitalisation boursière.
- **close** : Le prix au close renormalisé
- **close_beta** : Le beta.
- **upper_close__bollinger__min_periods_None__nb_periods_44** : La bande haute des bandes de Bollinger calculées à partir du close sur une fenêtre de 44 jours.
- **lower_close__bollinger__min_periods_None__nb_periods_5** : La bande basse des bandes de Bollinger calculées à partir du close sur une fenêtre de 5 jours.

- **upper_close__bollinger__min_periods_None__nb_periods_22** : La bande haute des bandes de Bollinger calculées à partir du close sur une fenêtre de 22 jours.
- **lower_close__bollinger__min_periods_None__nb_periods_22** : La bande basse des bandes de Bollinger calculées à partir du close sur une fenêtre de 22 jours.
- **upper_close__bollinger__min_periods_None__nb_periods_5** : La bande haute des bandes de Bollinger calculées à partir du close sur une fenêtre de 5 jours.
- **lower_close__bollinger__min_periods_None__nb_periods_44** : La bande basse des bandes de Bollinger calculées à partir du close sur une fenêtre de 44 jours.
- **stocha_close__bollinger__min_periods_None__nb_period_44** : Position relative de la valeur de la série actuelle par rapport à ses limites supérieures et inférieures de Bollinger calculées à partir du close sur une fenêtre de 44 jours.
- **width_close_adjusted__bollinger__min_periods_None__nb_periods_44** : Largeur des bandes de Bollinger calculées à partir du close ajusté sur une fenêtre de 44 jours.
- **lower_close_adjusted__bollinger__min_periods_None__nb_periods_5** : La bande basse des bandes de Bollinger calculées à partir du close ajusté sur une fenêtre de 5 jours.
- **width_close_adjusted__bollinger__min_periods_None__nb_periods_5** : Largeur des bandes de Bollinger calculées à partir du close ajusté sur une fenêtre de 5 jours.
- **lower_close_adjusted__bollinger__min_periods_None__nb_periods_44** : La bande basse des bandes de Bollinger calculées à partir du close ajusté sur une fenêtre de 44 jours.
- **upper_close_adjusted__bollinger__min_periods_None__nb_periods_44** : La bande haute des bandes de Bollinger calculées à partir du close ajusté sur une fenêtre de 44 jours.
- **close_adjusted** : Valeur du close ajusté.
- **close_adjusted__rsi__nb_periods_44** : Relative strength index calculé à partir du close ajusté sur une fenêtre de 44 jours.
- **close_adjusted__trend__min_periods_None__nb_periods_44** : Tendance du close ajusté sur une fenêtre de 44 jours.
- **close_adjusted__rsi__nb_periods_22** : Relative strength index calculé à partir du close ajusté sur une fenêtre de 22 jours.
- **close_adjusted__rollingmean__min_periods_None__window_22** : Moyenne mobile du close ajusté sur une fenêtre de 22 jours.
- **close__rollingmean__min_periods_None__window_44** : Moyenne mobile du close sur une fenêtre de 44 jours.

- **close_rollingmean_min_periods_None_window_5** : Moyenne mobile du close sur une fenêtre de 5 jours.
- **close_trend_min_periods_None_nb_periods_44** : Tendence du close sur une fenêtre de 44 jours.
- **close_macdivergence_div_period_9_long_period_52_short_period_24** : Moyenne mobile 24 jours de l'indicateur technique de macd pour une série de close (période courte de 9 jours et période longue de 52 jours))

Indicateurs de données alternatives :

- **nb_sources_esg_misc_neg_listv3_rollingmean_min_periods_None_window_22** : Moyenne mobile 22 jours du nombre de sources citant un événement ESG de type misc (tous les risques qui ne peuvent pas être associés clairement à l'une des trois catégories E, S et G)
- **webhose_valence3_single_negative_entity_mention_weighted_rollingmean_min_periods_None_window_44** : Moyenne mobile 44 jours du sentiment négatif pondéré par le nombre de mentions de l'entité cible.
- **webhose_valence3_single_negative_entity_ent_conf_weighted_rollingmean_min_periods_None_window_22** : Moyenne mobile 22 jours du sentiment négatif pondéré par la probabilité de détection de l'entité cible.
- **volume_article_rollingmean_min_periods_None_window_44** : Moyenne mobile 44 jours du volume d'article concernant la société cible.
- **score_esg_embeddings_gouv_neg_listv3_rollingmean_min_periods_None_window_22** : Moyenne mobile 22 jours du score gouvernance utilisant des modèles d'embeddings.
- **webhose_valence3_single_negative_entity_return_7_rollingmean_min_periods_None_window_22** : Moyenne mobile 22 jours du sentiment différencié (sentiment à date - sentiment à t-7)
- **bruit_esg_env_neg_listv3_rollingmean_min_periods_None_window_44** : Indicateur composite reflétant le volume de publication sur le concept d'environnement pour la journée donnée
- **nb_sources_esg_gouv_neg_listv3_rollingmean_min_periods_None_window_22** : Moyenne mobile 22 jours du nombre de sources citant un événement ESG de type gouvernance.
- **nb_sources_esg_env_neg_listv3_rollingmean_min_periods_None_window_44** : Moyenne mobile 44 jours du nombre de sources citant un événement ESG de type environnement.