

Rapport - Data pas science

1. Notre démarche pour comprendre et traiter le sujet

Pour bien appréhender un sujet aussi complexe que l'optimisation de portefeuille alignée avec un scénario en dessous de 2°C, nous avons commencé par une phase de compréhension approfondie. L'objectif était de poser des bases solides avant de passer à l'analyse et au traitement des données. Voici comment nous avons structuré notre travail :

Étape 1 : Lecture de la Documentation et Questions Préliminaires

Les deux premières heures ont été consacrées à une exploration collective des documents mis à disposition. Cette phase de lecture nous a permis :

- De comprendre les objectifs et les contraintes du problème.
- Identifier les concepts clés (empreinte carbone, scénarios climatiques, indicateurs financiers).
- De clarifier le résultat attendu en termes d'optimisation et de modélisation.

Durant cette étape, nous avons pris le temps de nous poser des questions essentielles, comme :

- Quels sont les critères prioritaires pour un portefeuille en dessous de 2°C ?
- Quelles données devons-nous utiliser et comment les traiter ?
- Quelles approches mathématiques ou financières sont adaptées à ce contexte ?

Pour organiser ces réflexions, nous avons utilisé un tableau blanc, où nous avons résumé les idées principales sous forme de schémas et de bullet points. Cela nous a permis d'avoir une vue d'ensemble claire avant de passer à l'action.

Étape 2 : Répartition des tâches selon les profils

Pour travailler efficacement, nous avons tiré parti de nos profils variés et de nos appétences personnelles. L'équipe s'est divisée en deux pôles :

1. Analyse de la Data : Les membres ayant une affinité pour les aspects techniques se sont concentrés sur la collecte, le nettoyage et la structuration des données climatiques et financières. Cela comprenait des tâches comme :
 - Identifier les sources fiables pour les données carbone des entreprises.
 - Préparer les données pour qu'elles soient exploitables dans des modèles d'optimisation.
2. Finance et Modélisation : Ceux ayant une appétence pour la finance se sont penchés sur les stratégies d'allocation et les modèles mathématiques adaptés à l'optimisation sous contraintes climatiques. Cela impliquait :
 - L'analyse des métriques financières.
 - La définition des critères d'exclusion ou de pondération dans le portefeuille.

Cette répartition a permis d'avancer de manière efficace, tout en combinant nos compétences complémentaires.

2. Données ESG

Une fois notre méthodologie définie et les tâches réparties, nous avons concentré nos efforts sur l'analyse des bases de données regroupant des critères ESG (Environnementaux, Sociaux et de Gouvernance).

2.1 Base (climate risk report) et Base RSI

Pour approfondir notre travail sur les critères ESG, nous avons analysé deux bases de données principales :

- Base Climate Risk Report : Contenant des informations détaillées sur les risques climatiques des entreprises.
- Base RSI : Regroupant une multitude de données sur les critères ESG des entreprises.

Ces bases comprenaient une large quantité de variables, chacune apportant des informations précieuses mais nécessitant une analyse approfondie pour en déterminer la pertinence.

Analyse et Nettoyage des Variables

Nous avons passé en revue chaque variable présente dans ces bases de données afin de réfléchir à son importance pour répondre à notre problématique. Ce processus nous a permis :

- De sélectionner uniquement les variables qui avaient une véritable pertinence dans le contexte de l'optimisation de portefeuille sous 2°C.
- De réduire le nombre total de variables à une quinzaine, en éliminant les données redondantes ou peu utiles.

Création d'une variable clé : L'augmentation du Score ESG

En complément de cette sélection, nous avons également créé une nouvelle variable que nous avons appelée "Augmentation du Score ESG". Cette variable a été conçue pour capturer l'évolution des performances ESG des entreprises au fil du temps :

- Si le score ESG d'une entreprise était plus faible auparavant et a augmenté récemment, cela montre que l'entreprise s'implique davantage dans des initiatives en faveur de la durabilité et de la responsabilité.
- En revanche, une stagnation ou une baisse du score ESG pourrait indiquer un manque d'effort ou un recul dans les engagements pris par l'entreprise.

Cette variable nous a permis d'ajouter une dimension dynamique à notre analyse, en ne regardant pas uniquement les performances actuelles mais aussi leur évolution.

2.2 Création d'un indicateur ESG

Au cours de notre analyse, nous avons constaté que les indicateurs ESG actuels présentent certaines limites, notamment dans la manière dont ils évaluent les entreprises fortement polluantes. En partant de cette observation, nous avons réfléchi à une approche différente, visant à mieux intégrer les efforts de transition des entreprises tout en tenant compte de leur impact environnemental actuel.

Limite des Notes ESG Traditionnelles

Dans leur version actuelle, les notes ESG pénalisent directement les entreprises fortement polluantes, comme Total, en leur attribuant des scores très bas. Bien que ce soit logique pour refléter leur empreinte carbone élevée, cette approche entraîne des conséquences :

- Ces entreprises, qui ont un rôle majeur à jouer dans la transition énergétique, sont désavantagées lorsqu'il s'agit d'attirer des investissements.
- Les efforts qu'elles déploient pour réduire leur pollution ou investir dans des technologies vertes ne sont pas toujours suffisamment valorisés.

Une Approche Alternative : Encourager les Efforts de Transition

Nous avons proposé de construire un indicateur ESG légèrement différent, basé sur deux axes principaux :

1. Quantifier la Pollution Actuelle : L'impact environnemental direct de l'entreprise reste un critère essentiel. Cependant, il doit être pondéré de manière à ne pas exclure les acteurs majeurs de certains secteurs, comme l'énergie ou l'industrie lourde, qui sont cruciaux pour une transition globale.
2. Valoriser les Efforts de Transition : Une entreprise qui réduit progressivement son empreinte carbone ou investit dans des initiatives durables doit voir ses efforts récompensés dans sa note ESG. Par exemple :
 - Total est actuellement perçue comme une entreprise à faible score ESG, mais si elle investit massivement dans les énergies renouvelables ou développe des technologies pour capturer et stocker le carbone, cela devrait améliorer sa note.

Un Indicateur qui Encourage la Transition

Avec cette approche, l'objectif est double :

- Récompenser l'action : Les entreprises qui fournissent des efforts concrets pour réduire leurs émissions ou investir dans des solutions durables voient leur score ESG augmenter.
- Éviter la stigmatisation excessive : Plutôt que d'exclure systématiquement les entreprises polluantes, cette méthode incite les investisseurs à soutenir celles qui s'engagent dans un processus de transition, car ce sont elles qui peuvent avoir le plus grand impact sur la réduction globale des émissions.

En intégrant cette logique dans nos modèles, nous pensons qu'il est possible de mieux aligner les portefeuilles sur les objectifs climatiques tout en favorisant un changement positif dans les secteurs qui en ont le plus besoin.

Pour ce faire, nous avons utilisé le WAS (Weighted Average Score), qui est un indicateur ESG corrigé du secteur. En effet, une note ESG de 7 pour une entreprise du secteur de l'énergie n'a pas la même signification qu'une note de 7 pour une compagnie pétrolière. Le WAS permet d'ajuster les notes ESG en tenant compte des spécificités sectorielles, rendant ainsi la comparaison plus équitable.

Nous avons également inclus d'autres éléments dans notre réflexion, comme la Climate Value at Risk (Climate VaR) et notre variable d'augmentation du score ESG, qui mesure les efforts de transition d'une entreprise.

Finalement, notre indicateur aboutit à une note finale comprise entre 0 et 100, où 100 représente la meilleure note possible en termes de performance ESG ajustée.

3. Données financière

Un **indice boursier** a une composition dynamique qui évolue dans le temps. En général, les actions avec les meilleurs scores ESG sont aussi celles qui ont une ancienneté plus importante. Cela s'explique par le fait qu'une politique ESG efficace nécessite une structure d'entreprise solide et une trésorerie suffisante pour maintenir des investissements constants.

Ainsi, nous avons décidé de **supprimer les actions trop récentes**, c'est-à-dire celles ajoutées à l'indice après le 21 septembre 2021. Pour garantir la cohérence de notre période de backtesting et d'optimisation, le dataset commence à cette même date. De plus, nous avons éliminé les actions qui sont entrées et ressorties de l'indice sur une courte période.

Création des métriques glissantes

Nous avons créé différentes métriques sur **quatre temporalités glissantes** : 3 mois, 6 mois, 1 an et 2 ans. Les métriques calculées incluent :

- Les **rendements**,
- Les **volatilités**,
- Les **ratios de Sharpe**,
- Les **ratios de Sortino**,
- La **Value at Risk (VAR)** basée sur la méthode variance-covariance.

Méthode de sélection

Notre méthode de sélection est innovante et s'inspire de la méthode de Laplace pour identifier les meilleurs investissements. Voici les étapes principales :

1. **Dernière valeur de chaque métrique** : Nous utilisons uniquement la valeur la plus récente pour chaque métrique.
2. **Regroupement sectoriel** : Les actions sont regroupées dans des datasets par secteur d'activité.

3. Classement par métrique :

- Pour les métriques positives (Rendements, Ratios de Sharpe, Ratios de Sortino), les actions sont classées par **ordre décroissant**. La première action reçoit le meilleur rang (1), et ainsi de suite.
- Pour les métriques de risque (Volatilité, Value at Risk), les actions sont classées par **ordre croissant** afin de privilégier les moins risquées.

4. **Classement général sectoriel** : Pour chaque dataset sectoriel, nous créons un classement général en additionnant les rangs des différentes colonnes de classement. Les actions sont ensuite triées par **ordre croissant** de leur classement général.

Sélection finale

Nous sélectionnons les **50 % des meilleures actions** de chaque dataset sectoriel sur la base de ce classement général. Les actions retenues sont ensuite regroupées pour constituer un nouvel ensemble, qui servira de base à notre **modèle ESG et d'optimisation**.

3.1 Indice de référence

Pour l'indice de référence, qui nous était imposé, nous avons recalculé celui-ci en appliquant une équipondération sur tous les indices qui le composent. Cette méthode nous a permis d'aligner les données et de travailler de manière cohérente avec les contraintes de l'indice défini.

Un défaut subsistait néanmoins : l'indice de référence ne couvrait pas l'ensemble des actifs initiaux. En effet, nous avons effectué une pré-sélection des actifs afin de traiter les valeurs manquantes (NA) dans les données. Par conséquent, environ 100 actions ne figuraient plus dans notre analyse finale.

4. Modélisation

Enjeux du sujet selon notre compréhension :

- Concevoir un portefeuille répondant à plusieurs critères tout en visant une rentabilité adaptée au contexte des actifs financiers.
- Identifier et sélectionner les données pertinentes pour optimiser la construction du portefeuille.
- Déterminer les bons critères et les combiner de manière cohérente pour constituer un portefeuille performant.

L'utilisation d'un modèle classique de machine learning n'est pas idéale pour ce projet, car il ne s'agit pas de faire des prédictions, mais plutôt de réaliser une allocation optimale avec des contraintes spécifiques. Ces contraintes ne sont pas directement gérables avec les approches classiques.

C'est pourquoi nous avons choisi d'utiliser un algorithme génétique, une méthode particulièrement adaptée pour implémenter ce type de problématique, grâce à sa capacité à explorer efficacement un espace de solutions tout en respectant les contraintes imposées.

C'est quoi un algorithme génétique ? :

Un **algorithme génétique (AG)** est une méthode d'optimisation inspirée du processus de sélection naturelle et de l'évolution biologique. Il est utilisé pour résoudre des problèmes complexes où il n'existe pas de solution analytique facile ou pour lesquels les méthodes traditionnelles d'optimisation ne fonctionnent pas bien.

Fonctionnement de base :

1. **Population initiale** : Un groupe de solutions possibles (appelées "individus" ou "chromosomes") est généré de manière aléatoire.
2. **Évaluation (Fitness)** : Chaque solution (poids) de chaque individu est évaluée en fonction de la fitness qui regroupe les différents critères qui doivent être respectés comme la maximisation du rendement, la cible des 2 degrés etc....
3. **Sélection** : Les meilleures solutions (les plus adaptées) sont sélectionnées pour se reproduire.
4. **Croisement (Crossover)** : Les solutions sélectionnées s'associent pour créer de nouvelles solutions (descendants) en combinant leurs caractéristiques.
5. **Mutation** : De petites modifications aléatoires sont apportées à certains descendants pour introduire de la diversité.
6. **Nouvelle génération** : Les solutions actuelles sont remplacées par les nouvelles solutions créées, et le processus se répète.
7. **Finalité** : L'algorithme tend vers une solution optimale après un certain nombre d'itération.

Dans notre projet, comment ces concepts s'appliquent ?

1. Population initiale

Chaque *individu* est une solution : une liste de **poids** attribués aux entreprises de notre base de données. Ces poids représentent la répartition de notre portefeuille.

Exemple : Pour 5 entreprises, un individu pourrait ressembler à ceci :

[0.3, 0.2, 0.1, 0.25, 0.15]

(Chaque poids doit respecter les contraintes, comme une somme totale de 1).

2. Évaluation (Fitness)

La **fonction de fitness** calcule la qualité de chaque individu en tenant compte de critères, tels que :

- Le score ESG, le respect de la limite de température implicite (max 2°C), la diversification des actifs, et d'autres contraintes financières (par exemple, aucun poids individuel supérieur à 10%).
- Les critères tels que la maximisation du rendement, du ratio de Sharpe et du score ESG sont calculés en utilisant le **produit scalaire** entre les poids du portefeuille et les valeurs associées de chaque actif. Ces calculs permettent d'évaluer la contribution globale de chaque critère pour une solution donnée.

- Ensuite, d'autres contraintes, comme la **température pondérée** de la solution ou le fait qu'un poids dépasse 10 %, sont intégrées sous forme de **pénalités**. Ces pénalités, calculées proportionnellement aux écarts constatés, sont déduites du score de maximisation (rendement, Sharpe, etc.) pour obtenir une **valeur finale de fitness**.
- Ainsi, cette méthode équilibre les objectifs de maximisation des critères financiers et ESG tout en s'assurant que les contraintes définies sont respectées.

3. Sélection des parents

Les individus ayant les meilleures valeurs de fitness sont sélectionnés comme *parents* pour la prochaine étape.

- On prend généralement **la moitié** des meilleurs individus d'une génération.
- Ces parents seront utilisés pour créer de nouveaux individus (*descendants*).
Exemple : Si la population initiale contient 10 individus, on sélectionne les 5 ayant les meilleures valeurs de fitness.

4. Croisement (Crossover)

- Les parents sélectionnés sont combinés pour créer les descendants.
- Le croisement consiste à mélanger les *gènes* (poids) des deux parents.
- On choisit un **point de croisement** aléatoire, puis on combine les deux parents.

Le descendant est ensuite normalisé (les poids doivent toujours respecter les contraintes, comme une somme totale égale à 1).

5. Mutation

De petites modifications aléatoires sont appliquées sur certains individus pour :

- Éviter un blocage dans un optimum local,
- Introduire de la diversité dans la population.

La mutation est appliquée avec une **probabilité faible** (par exemple, 5%).

6. Nouvelle génération

Une nouvelle population est formée en combinant les parents sélectionnés et leurs descendants. Ce processus est répété pendant un certain nombre de générations.

- **Convergence** : Avec chaque génération, l'algorithme tend à améliorer les solutions, convergeant vers un optimum.

5. Ouverture

1. Perspectives d'analyse par pays

- **Perception du risque climatique et sélection ESG par pays** : Une agrégation des variables ESG par pays (en fonction de la domiciliation des actions) pourrait fournir une vue d'ensemble des pays les plus exposés au risque climatique ou ceux ayant les meilleurs scores ESG au sein d'un

indice. Cela ouvrirait la possibilité de sélectionner les actions les plus performantes sur le plan ESG ou les moins risquées, pays par pays.

- **Incorporation de données fiscales et d'aides à la transition écologique** : L'intégration de données supplémentaires sur les politiques fiscales et les aides à la transition écologique par pays pourrait permettre de mieux évaluer la dynamique de transition économique de chaque zone géographique. Cela offrirait une vision plus complète des efforts des pays et des entreprises dans l'adoption de modèles économiques durables et respectueux de l'environnement.

2. Allocation géographique et gestion des risques

- **Accessibilité aux données sur l'allocation géographique des activités** : L'accès à des données détaillées sur l'allocation géographique des revenus des entreprises (par exemple, le chiffre d'affaires par région) pourrait enrichir l'analyse du risque climatique.
- **Couplage avec les risques pays** : En associant ces données avec les évaluations des risques pays, telles que celles fournies par Tac Economics, il serait possible de développer des portefeuilles dynamiques. Ces portefeuilles pourraient désengager les investissements dans certaines zones géographiques jugées trop risquées ou vulnérables face aux enjeux climatiques.

Rendu visuel final

Enfin, nous avons développé un dashboard Grafana afin de visualiser les données de manière dynamique et interactive. Cet outil nous a permis de suivre l'évolution des différents indicateurs en temps réel, facilitant ainsi l'analyse et la prise de décisions.