**🎯 PRINCIPE DU MODÈLE ENSEMBLISTE TOP-2**

Imagine que tu veux prévoir la météo de demain. Au lieu de faire confiance à UN SEUL météorologue, tu demandes à PLUSIEURS experts, puis tu combines leurs avis pour avoir une meilleure prévision. C'est exactement le principe de ton modèle !

**📚 ÉTAPE 1 : LES 4 EXPERTS (MODÈLES)**

Ton système entraîne **4 modèles différents**, comme 4 experts avec des méthodes différentes :

**1️⃣ XGBoost (L'analyste technique)**

* Regarde les 7 derniers jours
* Analyse les tendances, les moyennes
* Dit : "D'après mes calculs complexes, demain = 5 millions"

**2️⃣ Prophet (Le saisonnier)**

* Connaît les cycles : "chaque début de mois c'est pareil"
* Repère les patterns hebdomadaires, mensuels, annuels
* Dit : "Selon la saison et le jour de la semaine, demain = 4.8 millions"

**3️⃣ ARIMA (Le statisticien classique)**

* Suit l'évolution progressive des données
* Dit : "Selon la tendance récente, demain = 5.2 millions"

**4️⃣ Moyenne Mobile (Le pragmatique)**

* Fait simplement la moyenne des 7 derniers jours
* Dit : "La moyenne des 7 derniers jours = 4.9 millions"

**🏆 ÉTAPE 2 : SÉLECTION DES TOP-2**

Maintenant, on va tester ces 4 experts sur des données qu'ils n'ont pas vues (le jeu de TEST).

**Exemple concret :**

**Période de test : Mars à Juin 2025 (80 jours)**

| **Modèle** | **MAPE (% d'erreur)** | **Classement** |
| --- | --- | --- |
| XGBoost | **12.5%** | 🥇 1er |
| Prophet | **15.8%** | 🥈 2ème |
| ARIMA | 28.3% | 3ème |
| Moyenne Mobile | 35.7% | 4ème |

✅ **On garde SEULEMENT les 2 meilleurs : XGBoost + Prophet**

**Pourquoi ?**

* Les 2 meilleurs sont plus fiables
* Combiner trop de modèles (dont des mauvais) dégrade la qualité
* C'est comme prendre l'avis des 2 meilleurs médecins, pas des 10 médecins dont 8 sont médiocres

**⚖️ ÉTAPE 3 : LA PONDÉRATION (LE CŒUR DU SYSTÈME)**

Maintenant qu'on a nos 2 experts, on ne va pas leur donner le même poids ! Si XGBoost est meilleur que Prophet, il doit avoir plus d'influence.

**Comment ça marche ? (Validation Croisée)**

On découpe la période de TEST en **fenêtres de 10 jours** :

PÉRIODE TEST (80 jours) :

├── Split 1 : jours 1-10

├── Split 2 : jours 11-20

├── Split 3 : jours 21-30

├── Split 4 : jours 31-40

├── Split 5 : jours 41-50

├── Split 6 : jours 51-60

├── Split 7 : jours 61-70

└── Split 8 : jours 71-80

**Pour CHAQUE split, on calcule des poids différents**

**📊 Exemple concret - SPLIT 1 (jours 1-10) :**

**Jour 5 - Valeur réelle = 5.0 millions**

| **Modèle** | **Prédiction** | **Erreur** | **MAPE** |
| --- | --- | --- | --- |
| XGBoost | 4.9M | 0.1M | 2% |
| Prophet | 5.3M | 0.3M | 6% |

**Sur les 10 jours du split 1 :**

* XGBoost : MAPE = **10%** (très bon)
* Prophet : MAPE = **18%** (correct)

**Calcul des poids (méthode inverse) :**

1. **Inverse des erreurs :**
   * XGBoost : 1/10 = 0.10
   * Prophet : 1/18 = 0.056
2. **Total :** 0.10 + 0.056 = 0.156
3. **Poids normalisés :**
   * XGBoost : 0.10 / 0.156 = **64%** ⭐
   * Prophet : 0.056 / 0.156 = **36%**

**Interprétation :** XGBoost compte pour 64%, Prophet pour 36%. C'est comme si dans un comité de vote, XGBoost avait 64 voix et Prophet 36 voix.

**📊 Les 8 ensembles différents**

Chaque split donne des poids différents car les performances varient :

| **Ensemble** | **Période** | **XGBoost** | **Prophet** |
| --- | --- | --- | --- |
| CV0 | Jours 1-10 | 64% | 36% |
| CV1 | Jours 11-20 | 58% | 42% |
| CV2 | Jours 21-30 | 70% | 30% |
| CV3 | Jours 31-40 | 55% | 45% |
| CV4 | Jours 41-50 | 62% | 38% |
| CV5 | Jours 51-60 | 68% | 32% |
| CV6 | Jours 61-70 | 60% | 40% |
| CV7 | Jours 71-80 | 65% | 35% |

**Pourquoi ces variations ?**

* Parfois XGBoost est excellent (70%), parfois moins (55%)
* Selon la période, un modèle peut mieux capturer les patterns
* C'est normal : aucun modèle n'est parfait tout le temps

**🎬 ÉTAPE 4 : PRÉDICTION FINALE**

Maintenant on teste ces 8 ensembles sur la période de VALIDATION (Juillet+)

**Exemple de prédiction pour le 1er juillet :**

**Prédictions individuelles :**

* XGBoost : 5.2 millions
* Prophet : 4.8 millions

**Prédictions des 8 ensembles :**

|  |  |  |
| --- | --- | --- |
| **Ensemble** | **Calcul** | **Prédiction** |
| CV0 | 0.64×5.2 + 0.36×4.8 | 5.06M |
| CV1 | 0.58×5.2 + 0.42×4.8 | 5.03M |
| CV2 | 0.70×5.2 + 0.30×4.8 | 5.08M |
| CV3 | 0.55×5.2 + 0.45×4.8 | 5.02M |
| CV4 | 0.62×5.2 + 0.38×4.8 | 5.05M |
| CV5 | 0.68×5.2 + 0.32×4.8 | 5.07M |
| CV6 | 0.60×5.2 + 0.40×4.8 | 5.04M |
| CV7 | 0.65×5.2 + 0.35×4.8 | 5.06M |

**Évaluation sur 5 jours de validation :**

**Valeurs réelles : [5.1M, 4.9M, 5.3M, 5.0M, 4.8M]**

| **Modèle** | **MAPE** |
| --- | --- |
| XGBoost (seul) | 8.5% |
| Prophet (seul) | 12.3% |
| Ensemble CV2 | **6.2%** ✅ |
| Ensemble CV5 | 6.8% |
| Ensemble CV7 | 7.1% |

**🏆 GAGNANT : Ensemble CV2** (XGBoost 70% + Prophet 30%)

**✅ AVANTAGES**

**1️⃣ Robustesse**

* Si XGBoost se trompe un jour, Prophet peut compenser
* Exemple : XGBoost prédit 6M (erreur +20%), Prophet prédit 4.8M (erreur -4%)
* Ensemble (70/30) : 5.56M (erreur +11%) → **erreur réduite**

**2️⃣ Adaptabilité**

* Les 8 ensembles s'adaptent à différents contextes
* On choisit le meilleur selon la période de validation
* C'est comme avoir 8 stratégies et garder la plus performante

**3️⃣ Réduction de variance**

* Un modèle seul peut être instable
* L'ensemble lisse les prédictions extrêmes
* Exemple : XGBoost fait un pic à 7M, Prophet dit 4.5M → Ensemble dit 6.1M (plus réaliste)

**4️⃣ Meilleure performance**

* Généralement, l'ensemble bat les modèles individuels
* Dans ton exemple : Ensemble (6.2%) vs XGBoost seul (8.5%)
* **Gain de 2.3 points de MAPE !**

**5️⃣ Sécurité**

* Si un modèle "déraille" (bug, données aberrantes), l'autre limite les dégâts
* C'est un filet de sécurité

**❌ INCONVÉNIENTS**

**1️⃣ Complexité**

* Au lieu de gérer 1 modèle, tu en gères 4 (entraînement) + 8 (ensembles)
* Plus difficile à expliquer à un non-technique
* Temps de calcul plus long

**2️⃣ Sur-optimisation possible**

* Si tu as peu de données de validation (5 jours), tu peux "tomber sur le bon par hasard"
* Risque : le CV2 est bon sur ces 5 jours mais mauvais après

**3️⃣ Maintenance**

* Chaque modèle du Top-2 doit être maintenu
* Si Prophet plante, tout l'ensemble est cassé
* Plus de dépendances (bibliothèques)

**4️⃣ Interprétabilité**

* Difficile d'expliquer : "Pourquoi 5.06M ?"
* Réponse : "Parce que 64% de XGBoost qui dit 5.2M + 36% de Prophet qui dit 4.8M"
* Un manager préfère : "XGBoost dit 5.2M car la tendance des 7 derniers jours..."

**5️⃣ Pas toujours meilleur**

* Parfois, le meilleur modèle individuel bat l'ensemble
* Exemple : si XGBoost est VRAIMENT excellent (MAPE 3%), l'ensemble peut le dégrader
* Dans ton cas, si XGBoost seul fait 8.5% et l'ensemble 6.2%, c'est gagnant. Mais si XGBoost faisait 4%, l'ensemble pourrait faire 5% (dégradé)

**🎯 RÉSUMÉ POUR UN PROFANE**

**C'est comme organiser un concours de prévision :**

1. **4 candidats participent** (XGBoost, Prophet, ARIMA, Moyenne)
2. **On garde les 2 meilleurs** (exemple : XGBoost et Prophet)
3. **On leur donne des rôles selon leur force** :
   * XGBoost = 64% de la décision
   * Prophet = 36% de la décision
4. **On teste 8 répartitions différentes** (64/36, 70/30, 55/45...)
5. **On garde la meilleure répartition** pour le futur