

Projet Nova 5.0 – Synthèse d'avancement (Trot attelé autostart)

État actuel (intégré et testé)

- **Verrou SG/SP et critères fiables** : La version Nova v4.9 a introduit un « **verrou SG/SP** » qui garantit qu'aucun cheval **incontournable** ne soit omis des sélections, éliminant ainsi 100% des oublis de chevaux clés observés précédemment. Ce verrouillage des favoris s'est traduit par une **hausse d'environ +30 points du taux de réussite** des paris prudents (succès passant d'environ 40% à 70% sur l'échantillon test). De manière générale, les **critères SG/SP** (sélection Gagnant/Placé prudente) ont été durcis et sont **opérationnels** : seuls les chevaux réguliers, fiables et à cote raisonnable sont retenus, conformément aux enseignements des versions Nova 4.5 à 4.9.
- **Données historiques intégrées (100 courses)** : Environ **100 courses de trot autostart** ont été intégrées dans la base historique pour analyser le comportement du modèle Nova 5.0. Cette analyse a mis en évidence des **profils de chevaux prédictifs** pour la réussite en pari Simple Placé – par exemple, un cheval **ultra-régulier** (plus de 4 podiums sur ses 5 dernières sorties, aucune faute) se place dans plus de 70% des cas. Ces constats ont guidé le réglage des filtres actuels. Les premières métriques de performance sont prometteuses : en misant de façon très sélective (2-3 paris max par jour), Nova 5.0 a approché un **rendement stable à deux chiffres (~+10% de ROI)** sur cet échantillon test, ce qui est cohérent avec l'objectif visé (rentabilité durable en Simple Placé).
- **Pipeline data (parsing PDF, Parquet, visu/scoring)** : Le **processus d'extraction des données** est enclenché. Des bulletins PDF de résultats/course ont été **parsés automatiquement** afin d'alimenter une base de données structurée (fichiers **Parquet** pour une exploitation analytique efficace). Les informations clés de chaque course (classement, cotes, conditions, etc.) sont ainsi centralisées. Des **premières visualisations** et analyses de **scoring** ont été réalisées pour valider empiriquement les critères Nova : un tableau synthétique des **profils de chevaux rentables** a par exemple été établi, confirmant les typologies attendues (cheval ultra-régulier fiable, cheval bien

placé avec forme ascendante, opportuniste en seconde ligne “aspiré”, etc. conformément à Nova v4.5+).

À faire / Priorités

- **Extraction ferrures, cotes, météo, drivers** : Intégrer les **variables cruciales manquantes** pour affiner la sélection. Il s’agit notamment des informations de **ferrure** (déferrage complet/partiel des chevaux), des **cotes PMU** (et leurs variations en temps réel), des **conditions de piste/météo**, ainsi que des changements de **driver** (driver/jockey en sulky). Le besoin est justifié par l’analyse : par exemple, une **piste lourde** peut faire chuter le taux de réussite d’environ 15 pointsfile-7dmuzir7rn9v7vwu3yhv3c, et un cheval **ferré** alors qu’il performe habituellement déferré (ou un changement de driver de dernière minute) est un **signal négatif** important pour un pari placéfile-7dmuzir7rn9v7vwu3yhv3c. De même, une **forte variation de cote en live** peu avant le départ (baisse significative ou hausse brutale) indique une information de dernière minute à considérerfile-7dmuzir7rn9v7vwu3yhv3c. La **priorité** est donc d’extraire ces données (via scraping ou API des sources appropriées) et de les intégrer dans le dataset Nova afin de **raffiner le modèle** et d’éviter tout biais ou omission.
- **Modélisation (baseline XGBoost, calibration des odds)** : Commencer le développement du **modèle prédictif**. On prévoit d’entraîner un modèle de type **XGBoost** (arbres de décision boostés) en baseline, en utilisant les données historiques enrichies. L’un des objectifs sera de **calibrer les probabilités** de réussite estimées par le modèle avec les **cotes** du marché (odds) afin d’identifier les écarts de valorisation (value bets) et de s’assurer que les scores du modèle correspondent à des probabilités réalistes. Cette étape passera par de la validation croisée et possiblement du tuning (paramètres du modèle, ajustement pour éviter le sur-apprentissage) dès qu’un volume de données suffisant incluant les nouvelles variables sera disponible.
- **Sélection automatique des paris & plans de mise** : Développer un module d’**automatisation des sélections** quotidiennes. Sur la base des scores du modèle et des règles d’exclusion Nova (verrou SG/SP, filtres fiabilité, etc.), le système devra **choisir automatiquement les chevaux à parier** chaque jour (par exemple 1 ou 2 chevaux “gagnants” SG et 2 ou 3 chevaux “placés” SP maximum, suivant les opportunités identifiées). En parallèle, définir un **plan de mise optimisé** : par exemple, une mise fixe par pari placé ou une mise proportionnelle au degré de confiance (gestion de bankroll type Kelly modérée ou palier de mise prédéfini). L’objectif est de maximiser le **ROI** tout en contrôlant

le risque, en appliquant systématiquement la discipline préconisée (pas de sur-pari en cas de perte isolée, etc., cf. méthode Nova).

- **Évaluation du ROI / tests croisés** : Mettre en place une **évaluation rigoureuse** de la performance du système Nova 5.0. Cela inclut des **back-tests** supplémentaires sur des données historiques (y compris hors échantillon initial) et des **tests croisés** (cross-validation temporelle, par exemple par trimestre ou par hippodrome) pour mesurer la robustesse des résultats. Seront mesurés le **taux de réussite** (pourcentage de paris gagnants/placés rentrés), le **ROC** des prédictions, et bien sûr le **ROI net** généré par la stratégie. Cette phase de validation permettra d'identifier d'éventuels ajustements à apporter avant de déployer Nova en conditions réelles. L'ambition est de confirmer que la version 5.0 atteint ou dépasse le **ROI +10%** visé de façon statistiquement significative et soutenable à long terme, avant de passer à l'industrialisation.

Ressources premium achetées (rappel)

- **Trotstats Avancé (550 variables)** : Base de données professionnelle offrant ~550 variables par course/cheval, couvrant des aspects pointus du trot (historiques complets, temps partiels, caractéristiques des hippodromes, etc.). Cette ressource riche servira à **étendre le jeu de données** Nova. Il conviendra d'identifier parmi ces nombreuses variables celles qui apportent le plus de **pouvoir prédictif** pour les intégrer dans le modèle (par exemple via une analyse d'importance des features ou de corrélations avec la réussite SP/SG). Trotstats permettra potentiellement de capturer des éléments difficiles à obtenir manuellement (indices de forme, performances relatives par terrain, etc.), renforçant ainsi la **précision des sélections**.
- **ParisTurf Live (infos temps réel)** : Abonnement à un flux **temps réel** d'informations courses (proposé par ParisTurf). Il donne accès aux **cotes en direct**, aux détails de dernière minute sur les partants (ferrure **live**, non-partants, changements de driver, pronostics, météo/état du terrain actualisé). En exploitant ParisTurf Live, Nova 5.0 pourra être **connecté en temps réel** : par exemple, ajuster les sélections peu avant le départ si un cheval prévu défermé reste ferré ou si une averse survient, ou encore détecter un mouvement de cote anormal. Pour en tirer 100%, il faudra intégrer ce flux dans le pipeline (API ou web scraping) de manière fiable, afin que le modèle *scoring* et les recommandations de paris soient toujours à jour des dernières infos terrain.
- **Ebook "Méthode Calvete" (forme, repos, déferrage)** : Un **guide théorique** rédigé par l'expert Michel Calvete, axé sur l'évaluation de la **forme** des chevaux, l'impact des périodes de **repos** et de la stratégie de **déferrage** sur les performances. Cette ressource contient de nombreuses **règles empiriques**

utiles (par ex. comment interpréter un cheval qui court défermé des quatre pieds après une longue pause, ou l'importance de la fraîcheur physique sur la performance). L'intégration de ces connaissances peut se faire soit via de **nouveaux critères** dans Nova 5.0 (check-list qualitative, pénalités/bonus), soit en créant des **variables dérivées** (par ex. nombre de jours de repos depuis la dernière course, % de réussite d'un cheval défermé vs ferré, etc.) qui serviront d'entrées au modèle. Le but est de **tirer profit à 100% des enseignements** de Calvete pour compléter l'approche data par une expertise turf éprouvée.

- **Excel MasterPro (700 courses testées, ROI 107%)** : Un fichier Excel "master" qui a servi de modèle sur **700 courses de trot autostart**, avec un **ROI cumulé de +107%** sur ces tests. C'est une **référence précieuse** pour le projet Nova 5.0. Il faudra analyser cette approche MasterPro pour comprendre les **facteurs de succès** qui ont contribué à un tel rendement. Par exemple, MasterPro utilise-t-il des critères spécifiques ou une gestion de mise particulière ? En comparant, point par point, la méthode Nova et celle de MasterPro, on peut identifier des améliorations potentielles. L'objectif est de **capitaliser sur ce benchmark** : soit en validant que Nova 5.0 couvre déjà les mêmes points forts, soit en intégrant des aspects qui lui manqueraient. En somme, MasterPro sert de **mètre-étalon** pour viser une performance au moins équivalente sinon supérieure avec Nova.

Recommandations

- **Exploitation optimale des ressources** : Pour **profiter à 100%** des ressources premium, il est recommandé de planifier leur intégration méthodique. Par exemple, utiliser Trotstats pour effectuer une **sélection des variables** les plus pertinentes (via analyses de contribution au modèle ou en testant l'ajout de groupes de variables progressivement) afin de ne pas surcharger le modèle inutilement. De même, formaliser les enseignements de l'eBook Calvete en une **check-list automatisée** ou en features mesurables (p. ex. créer un indicateur de "fraîcheur" basé sur les jours de repos, ou un indicateur de performance associée au défermé). Concernant ParisTurf Live, il faudra développer des **connecteurs fiables** (scripts ou API) pour injecter en continu les données dynamiques dans Nova : c'est crucial pour que la sélection finale du jour tienne compte de tout élément de dernière minute (changement de terrain, info ferrure, fluctuation de cote, etc.). Enfin, exploiter MasterPro signifie **comparer régulièrement les résultats** : on peut, par exemple, appliquer Nova 5.0 sur les mêmes 700 courses pour voir si le ROI obtenu est similaire, et analyser les écarts course par course pour détecter où Nova diffère (et pourquoi).
- **Priorités de développement** : À ce stade du projet, la priorité #1 est d'**enrichir la base de données** avec les facteurs manquants (ferrures, météo, drivers, etc.)

et de finaliser le **préprocessing** de ces données. En parallèle, il serait bénéfique de lancer une première version du **modèle XGBoost** avec les données déjà disponibles pour obtenir un **baseline** de performance. Même si toutes les features ne sont pas encore intégrées, ce prototype permettra de valider la pipeline d'entraînement, de déboguer le flux de bout en bout et d'observer les premières tendances (même qualitatives). Par ailleurs, il faut prioriser la **traduction des règles Nova** en code (par ex. un module qui, après le scoring brut du modèle, applique les filtres SG/SP, les seuils de cote "jouable", etc.). Cela garantira que la **logique métier** (l'expertise turf encapsulée dans Nova) est bien reflétée dans les résultats automatiques. En somme, **d'abord les données et la validation du modèle**, puis l'orchestration de la sélection automatisée.

- **Préparation du pipeline Nova 5.1 (industrialisation live)** : Anticiper dès maintenant l'**industrialisation** de Nova pour la version 5.1, qui correspondra à la mise en production avec scores **live**. Il s'agit de concevoir une architecture robuste où chaque étape est automatisée et orchestrée : ingestion quotidienne des données (idéalement via des APIs officielles ou des scrapers stables pour PMU/LeTROT/ParisTurf), mise à jour des fichiers Parquet ou de la base de données, exécution du modèle de prédiction sur les partants du jour, application des règles Nova, puis **génération des recommandations** (chevaux à jouer, type de pari, mise conseillée). Ce pipeline devra tourner de manière fiable chaque jour de course, avec des **délais réduits** (par ex. calcul des sélections finalisé quelques heures avant les premières courses). Il est conseillé de mettre en place des **indicateurs de monitoring** en production : suivi du % de réussite quotidien, du ROI cumulé, détection des écarts anormaux par rapport aux attentes, afin de pouvoir ajuster rapidement si la performance réelle s'écarte des prévisions. En parallèle, préparer la **scalabilité** et la maintenance du système (code propre, documentation, éventuellement conteneurisation ou déploiement cloud) pour qu'il soit exploitable par l'équipe et pérenne. L'objectif final est que Nova 5.1 puisse fonctionner en **routine opérationnelle**, fournissant des **sélections fiables en temps réel**, tout en conservant l'esprit R&D agile pour continuer d'améliorer le modèle au fil du temps.

En résumé, le projet Nova 5.0 progresse solidement : les bases de la méthode sont en place et validées sur un échantillon initial, et les prochaines étapes visent à enrichir les données et formaliser le modèle pour gagner en efficacité. En suivant ces recommandations et en exploitant pleinement les ressources disponibles, Nova a le potentiel de devenir un **outil semi-industriel performant** pour les courses de trot autostart, combinant puissance de la data et expertise turf, avec à la clé une **rentabilité visée à deux chiffres** sur le long terme.

Synthèse historique 1 an (PMU_Complete.csv) — Socle pédagogique Nova 5

Ce document synthétise tous les enseignements structurels exploitables issus de ton historique d'1 an, **même sans les arrivées**, pour servir de base à la réflexion, à la modélisation et à l'apprentissage de ta méthode Nova 5.

1. Hippodromes majeurs (priorité d'étude et de pondération Nova 5)

- **VINCENNES - GP** domine totalement (14 208 courses). C'est le terrain de jeu clé pour toute méthode sérieuse.
- 9 autres hippodromes majeurs (CAGNES, ENGHEN, GRAIGNES, etc.) totalisent de gros volumes —> valider leurs spécificités et pondérations dans Nova 5.

2. Typologie ferrure/déferrage

- **F4 et D4 = la norme** (59 785 et 52 174 partants). Tous les filtres Nova 5 doivent commencer par là. Les profils plus rares (PADP, DA, PA...) doivent être documentés (impact, rareté, valeur ajoutée ?).
- **Détecter les axes saisonniers/localisés** pour certains profils de ferrures : ex. DP, PADP.

3. Volume de partants/course et structure des réunions

- Beaucoup de courses à 11-14 partants, mais la variabilité existe selon hippodrome/date.
- Nova 5 doit intégrer la notion de "taille de course" comme critère : pas la même analyse sur une course à 8 partants vs 16.

4. Profils des chevaux récurrents

- Top profils de chevaux = “hyperactifs” (certains courent plus de 30 fois/an)
- Nova 5 peut utiliser ces profils pour affiner les stratégies sur la régularité ou le repos/engagement.
- **À nettoyer** : 309 valeurs NAN dans les chevaux —> prévoir un module “data quality” dans Nova 5.

5. Saison & temporalité

- **Mai à août = saison ultra-chargée** (16 000 à 17 000 courses/mois !)
- Nova 5 doit prévoir une adaptation de stratégie selon la saison : surcharge d’offres, plus de pièges, plus de profils “occasionnels” à détecter.

6. Préconisations Nova 5 sur base de l’historique (sans résultats)

- Utiliser cet historique pour **modéliser les profils de courses, la typologie des filtres** (taille, ferrure, hippodrome, saison...), poser des “zones de danger” et des “zones d’opportunité”.
- Proposer dans Nova 5 une pondération ou un module “contexte course” basé sur la densité ou la rareté (ex : pondérer différemment une course à Vincennes vs un petit hippodrome).
- **Documenter les manques** : NAN à nettoyer, absences de résultats, profils “hors-normes”.
- Prévoir des routines d’audit/qualité sur les prochains fichiers quotidiens.

7. Ce que Nova 5 doit apprendre de cet historique

- Où concentrer son attention (Vincennes + gros hippodromes + top ferrures)
- Pourquoi la structure d’une course ou d’un mois change la donne stratégique
- Comment modéliser un “filtre intelligent” avant même de valider les résultats réels
- Anticiper les pics de volume pour ajuster les critères Nova 5 à la saisonnalité

Mémo :

Cette synthèse sert de “boussole” pour structurer la méthode Nova 5 avant même l’intégration des résultats.

Dès que tu auras quelques fichiers quotidiens avec résultats, tu pourras comparer et corriger en direct — la structure de Nova 5 aura déjà un socle solide !