

Cahier des Charges

AutoTrade : Un Système Intelligent d'Analyse Multi-Source pour le Trading Prédictif

"Sentiments, données, profits : votre trading réinventé"

Version 1.0

L'équipe du projet:

Fatima Zahra Oubella	Amine Firdawsi
Nassira Nguadi	Youness Aabibi

Encadré par: Mr Masrour tawfik

5 janvier 2025

Table des matières

1	Contexte	2
2	Problématiques Identifiées	2
2.1	Surcharge d'Information	2
2.2	Contraintes Temporelles	3
2.3	Biais Comportementaux	3
3	Objectifs du Projet	4
3.1	Analyse Multi-Source	4
3.2	Intelligence Artificielle	4
3.3	Automatisation Décisionnelle	4
4	Architecture du Système : Couche Sources de Données	5
4.1	API AlphaVantage	5
4.2	Flux RSS	5
4.3	Données Reddit	5
4.4	Données Temporelles	6
5	Architecture du Système : Couche Pipeline d'Analyse	7
5.1	Analyse de Sentiment	7
5.2	Classification des Sentiments	7
5.3	Modèles de Séries Temporelles	8
6	Architecture du Système : Couche d'Orchestration	9
6.1	Orchestrateur de Score	9
6.2	Agent de Trading	9
7	Livrables Attendus	10
7.1	Système Complet	10
7.2	Visualisations	11
7.3	Rapports	12

1 Contexte

Le paysage actuel des marchés financiers a subi une transformation radicale par rapport au siècle précédent. Cette évolution est caractérisée par trois changements fondamentaux :

1. L'émergence du trading haute fréquence

Les transactions s'effectuent désormais en millisecondes, créant un environnement où la rapidité d'exécution devient un avantage concurrentiel crucial. Cette évolution a fondamentalement modifié la dynamique des marchés, rendant obsolètes les approches traditionnelles de trading.

2. La démocratisation des données de marché

L'accès aux données de marché s'est considérablement élargi, permettant à un plus grand nombre d'acteurs de participer aux marchés financiers. Cette démocratisation a augmenté la complexité des interactions de marché et la nécessité d'outils d'analyse sophistiqués.

3. L'impact des médias sociaux

La diffusion instantanée des nouvelles via les plateformes sociales a créé un nouveau paradigme où l'information circule et influence les marchés en temps réel, nécessitant des capacités d'analyse et de réaction rapides.

2 Problématiques Identifiées

2.1 Surcharge d'Information

La problématique de la surcharge d'information se manifeste à travers plusieurs aspects critiques :

— Volume de données hétérogènes

Le système doit traiter des données provenant de multiples sources :

- Actualités financières traditionnelles via l'API AlphaVantage
- Flux RSS de différentes sources d'information financière
- Données de sentiment social issues de Reddit et autres plateformes
- Séries temporelles de prix et volumes des actifs financiers

— Complexité d'intégration

L'intégration de ces sources diverses nécessite :

- Une normalisation des formats de données
- Une gestion des différentes échelles temporelles
- Une validation et nettoyage des données en temps réel

— Nécessité d'automatisation

L'analyse manuelle devient impossible en raison :

- Du volume croissant de données à traiter
- De la vitesse requise pour l'analyse
- De la complexité des corrélations à identifier

2.2 Contraintes Temporelles

Les contraintes temporelles imposent des exigences strictes sur le système :

- **Réactivité du marché**

Le système doit gérer :

- Des réactions du marché en millisecondes
- Des variations de prix instantanées
- Des opportunités d'arbitrage éphémères

- **Traitement en temps réel**

Nécessité de :

- Analyser les flux de données en continu
- Maintenir des performances constantes
- Optimiser les temps de latence

2.3 Biais Comportementaux

Les biais comportementaux représentent un défi majeur :

- **Impact émotionnel**

Les décisions humaines sont affectées par :

- La peur et l'avidité dans les moments de volatilité
- Le stress lié à la pression temporelle
- Les biais de confirmation dans l'analyse des données

- **Limitations cognitives**

Les traders humains font face à :

- Une capacité limitée de traitement de l'information
- Des difficultés à maintenir l'objectivité
- Une fatigue cognitive lors de longues sessions de trading

3 Objectifs du Projet

3.1 Analyse Multi-Source

Le système vise à intégrer et analyser efficacement :

- **Données de sentiment**

À travers :

- L'API AlphaVantage pour les actualités en temps réel
- Le modèle FinBERT pour l'analyse des textes financiers
- VADER pour l'analyse des sentiments sociaux
- Les flux RSS pour la couverture médiatique générale

- **Données techniques**

Incluant :

- Les séries temporelles de prix
- Les indicateurs techniques calculés
- Les volumes de transaction
- Les patterns de marché identifiés

3.2 Intelligence Artificielle

Le développement de l'IA se concentre sur :

- **Modèles prédictifs**

Comprenant :

- LSTM univarié et multivarié pour l'analyse temporelle
- MLP univarié pour les prédictions de prix
- Modèles de sentiment spécialisés

- **Optimisation**

Avec :

- Calibration continue des modèles
- Adaptation aux conditions de marché
- Validation en temps réel des prédictions

3.3 Automatisation Décisionnelle

L'automatisation des décisions s'appuie sur :

- **Moteur de décision**

Offrant :

- Trois options claires : Achat, Vente, Attente
- Règles de trading paramétrables
- Gestion intégrée des risques

- **Transparence**

Assurant :

- Traçabilité complète des décisions
- Documentation des facteurs d'influence
- Métriques de performance claires

4 Architecture du Système : Couche Sources de Données

4.1 API AlphaVantage

L'intégration de l'API AlphaVantage constitue une source primaire de données et comprend :

- **Système de rotation des clés API**
Implémenté via la classe `APIKeyRotator` :
 - Gestion automatique des rotations entre multiples clés API
 - Protection contre les limites de taux d'appel (rate limiting)
 - Suivi des clés défaillantes
 - Logique de nouvelle tentative et gestion des erreurs
 - Temps d'attente configuré à 12 secondes entre les requêtes
- **Analyse des actualités**
Réalisée par la classe `StockNewsAnalyzer` :
 - Système de cache avec durée configurable (1 heure par défaut)
 - Analyse de la distribution des scores de sentiment
 - Suivi temporel du sentiment
 - Comparaison inter-titres du sentiment
 - Identification des articles les plus pertinents

4.2 Flux RSS

Le système de flux RSS comprend :

- **Sources configurées**
Intégration de sources majeures :
 - Yahoo Finance (<https://finance.yahoo.com/news/rssindex>)
 - MarketWatch (<http://feeds.marketwatch.com/marketwatch/topstories/>)
 - CNBC (<https://www.cnbc.com/id/10000664/device/rss/rss.html>)
- **Fonctionnalités de traitement**
Le système offre :
 - Filtrage par date
 - Filtrage par mots-clés
 - Extraction de contenu complet
 - Gestion des flux multiples

4.3 Données Reddit

L'analyse des données Reddit s'effectue via le système `RedditParser` qui comprend :

- **Extraction et structuration**
Capacités incluant :
 - Organisation hiérarchique des commentaires

- Nettoyage du texte
- Sauvegarde au format JSON
- Gestion des métadonnées des posts
- **Analyse de sentiment**
Utilisation de VADER avec :
 - Calcul des scores de polarité
 - Catégorisation du sentiment
 - Analyse des distributions de sentiment
 - Visualisation des résultats

4.4 Données Temporelles

L'analyse des séries temporelles s'appuie sur deux modèles principaux :

- **Modèle LSTM**
Implémenté en deux variantes :
 - Version univariée utilisant uniquement les prix de clôture
 - Version multivariée intégrant prix, volumes et autres indicateurs
 - Fenêtres temporelles configurables
 - Normalisation des données via MinMaxScaler
- **Modèle MLP**
Configuration comprenant :
 - Approche univariée
 - Prétraitement des données temporelles
 - Prédiction basée sur les fenêtres historiques
 - Sauvegarde au format pickle

5 Architecture du Système : Couche Pipeline d'Analyse

5.1 Analyse de Sentiment

Le système d'analyse de sentiment comprend trois composants majeurs :

- **Traitement AlphaVantage**

Fonctionnalités incluant :

- Distribution des scores de sentiment
- Suivi temporel du sentiment
- Comparaison entre différents titres
- Identification des articles influents

- **Analyse FinBERT**

Caractéristiques :

- Modèle pré-entraîné sur des textes financiers
- Traitement par lots pour l'efficacité
- Support GPU
- Scores de confiance pour chaque prédiction

- **Analyse VADER**

Spécialisée pour :

- Contenu des médias sociaux
- Calcul des scores composites
- Analyse des tendances temporelles
- Visualisation des distributions

5.2 Classification des Sentiments

Le système utilise plusieurs échelles de classification :

- **Échelle AlphaVantage**

Cinq niveaux de sentiment :

- Très Haussier
- Modérément Haussier
- Neutre
- Modérément Baissier
- Très Baissier

- **Échelle FinBERT et VADER**

Classification ternaire :

- Positif
- Neutre
- Négatif

5.3 Modèles de Séries Temporelles

L'analyse temporelle s'appuie sur :

- **Prédiction LSTM**

Caractéristiques techniques :

- Architecture à mémoire long-terme
- Support des dépendances temporelles étendues
- Optimisation pour les séries financières
- Validation continue des prédictions

- **Approche MLP**

Configuration :

- Réseau neuronal multicouche
- Apprentissage supervisé
- Optimisation des hyperparamètres
- Métriques de performance dédiées

6 Architecture du Système : Couche d'Orchestration

6.1 Orchestrateur de Score

Le système d'orchestration assure :

- **Agrégation des signaux**

Processus incluant :

- Normalisation des différentes sources
- Pondération dynamique des signaux
- Calibration continue des poids
- Validation des scores agrégés

- **Gestion des flux**

Mécanismes pour :

- Coordination des services
- Tolérance aux pannes
- Mise à l'échelle automatique
- Monitoring des performances

6.2 Agent de Trading

L'agent de trading implémente :

- **Logique décisionnelle**

Basée sur :

- Scores agrégés normalisés
- Règles de trading paramétrables
- Seuils de décision adaptatifs
- Gestion des risques intégrée

- **Exécution**

Comprenant :

- Trois états possibles : Achat, Vente, Attente
- Validation des décisions
- Suivi des performances
- Documentation des actions

7 Livrables Attendus

7.1 Système Complet

Code Source

La livraison du code source comprendra :

- **Composants Core**

Incluant :

- `APIKeyRotator` pour la gestion des clés API
- `StockNewsAnalyzer` pour l'analyse des actualités
- `FinancialNewsFetcher` pour la collecte RSS
- `RedditParser` pour l'analyse des médias sociaux
- `FinancialSentimentAnalyzer` avec FinBERT

- **Modèles d'Analyse**

Comprenant :

- Modèles LSTM (univarié et multivarié)
- Modèle MLP pour les prédictions
- Modèles de sentiment entraînés
- Scripts d'entraînement et d'évaluation

- **Infrastructure**

Composants :

- Scripts de déploiement
- Configuration de l'orchestration
- Gestion du cache et des données
- Outils de monitoring

Tests Automatisés

Suite de tests complète :

- **Tests Unitaires**

Couvrant :

- Logique de rotation des clés API
- Fonctions d'analyse de sentiment
- Traitement des séries temporelles
- Gestion des erreurs et cas limites

- **Tests d'Intégration**

Validant :

- Flux complet d'analyse
- Interaction entre composants
- Performance du système
- Scénarios de trading réels

Documentation Technique

Documentation détaillée :

- **Architecture**

Détaillant :

- Diagrammes d'architecture système
- Flux de données
- Interactions entre composants
- Configurations requises

- **APIs**

Documentation des :

- Endpoints et paramètres
- Formats de données
- Exemples d'utilisation
- Gestion des erreurs

7.2 Visualisations

Tableaux de Bord Sentiment

Visualisations en temps réel :

- **Distribution des Sentiments**

Graphiques montrant :

- Histogrammes de distribution par titre
- Courbes de tendance temporelle
- Comparaisons multi-titres
- Statistiques de confiance

- **Analyse Temporelle**

Graphiques incluant :

- Évolution du sentiment sur différentes échelles
- Corrélations prix-sentiment
- Identification des points d'inflexion
- Patterns récurrents

- **Analyse des Sources**

Visualisations pour :

- Comparaison des sources d'information
- Impact relatif des médias
- Fiabilité des sources
- Biais potentiels

Indicateurs Techniques

Visualisations techniques :

- **Prédictions de Prix**

Graphiques pour :

- Prévisions LSTM vs réalité
- Intervalles de confiance
- Erreurs de prédiction
- Performances comparées des modèles

- **Analyses Quantitatives**

Visualisations des :

- Indicateurs techniques clés
- Signaux de trading
- Volatilité et volumes
- Corrélations inter-marchés

7.3 Rapports

Métriques de Performance

Rapports détaillés sur :

- **Performance des Modèles**

Incluant :

- Précision des prédictions LSTM et MLP
- Scores de sentiment par source
- Temps de traitement et latence
- Utilisation des ressources

- **Métriques Commerciales**

Détaillant :

- Taux de succès des signaux
- Performance ajustée au risque
- Analyse des coûts de transaction
- Métriques de Sharpe et Sortino

Analyses de Risque

Rapports de risque :

- **Risques Systémiques**

Analyses des :

- Expositions par marché
- Corrélations de risque
- Stress tests
- Scénarios extrêmes

- **Risques Techniques**

Évaluation des :

- Fiabilité du système
- Points de défaillance potentiels
- Latences critiques
- Dépendances externes

Statistiques d'Utilisation

Rapports d'utilisation :

- **Utilisation du Système**

Suivi des :

- Appels API par service
- Charge du système
- Temps de réponse
- Taux d'erreur

- **Analyse des Patterns**

Identification des :

- Patterns d'utilisation
- Pics de charge
- Comportements anormaux
- Opportunités d'optimisation