



## Résumé

Le marché FOREX est un marché extrêmement complexe et dont l'évolution dépend à la fois des facteurs macro et micro économiques ainsi que du comportement des acteurs actifs sur le marché. Dans ce projet de stage, on adopte le problème de classification de la tendance d'une paire de devises auquel on applique plusieurs modèles d'intelligence artificielle. Une large panoplie de variables ont été extraites de la base de données initiale et divers indicateurs de Trading ont été utilisés pour aboutir à de meilleures performances.

## Introduction

Nous présentons dans ce poster des méthodes d'intelligence Artificielle (classification) utilisées dans le cadre de la prédiction de la tendance des prix du marché Forex, le but étant de réaliser des trades selon les modèles établies. La stratégie adoptée est de placer un trade toutes les heures à un montant fixe modifiable par l'utilisateur. Nous nous concentrons en particulier sur la paire EUR/USD, la même procédure peut être suivie pour les autres paires de monnaies. La classification se fera en se basant sur une fenêtre, également modifiable, de 48 heures caractérisées par les indicateurs de Trading et autres variables extraites de la base de données initiale.

## Objectifs

**Automatisation de l'extraction des données:**  
Création d'un algorithme permettant le téléchargement automatique des données Forex à travers un API de Trading.  
**Extraction de nouvelles variables:**  
Création des indicateurs de Trading sensibles aux changements de la tendance et ajout de variables diverses potentiellement utiles pour la prédiction.  
**Elaboration des modèles de classification:**  
Création et réglage des modèles Machine Learning utilisées pour la classification.  
**Evaluation et comparaison des modèles:**  
Comparaison des performances des modèle selon différents indicateurs.  
**Backtesting sur des données courantes:**  
Evaluation et comparaison des modèles sur des données récentes selon le profit réalisé.  
**Automatisation de l'algorithme de Trading:**  
Réalisation d'un programme qui place automatiquement des trades selon la prédiction de l'algorithme de classification le plus éfficient.

## Base de Ronnées

Les données utilisées pour le "training" des modèles provient de l'interface en ligne offerte par la banque suisse Dukascopy tandis que les données courantes qui serviront pour le backtesting proviennent de l'API Alpha Vintage qui est un API de données Stock et FOREX.  
**Dataset:** EURUSD par heure OHLC entre 01/01/2013 et 03/01/2019, soit 37537 données horaires.  
**Training set:** Entre 01/01/2013 - 01/01/2018.  
**Validation set:** 01/01/2018 - 01/06/2018  
**Test Set:** 02/06/2018 - 03/01/2019  
Il est à noter que le nombre observations portant un label de valeur 1 est de 12387 tandis que le nombre d'observations ayant un label 0 est de 12598.  
Il n y aura donc pas de problèmes liés à la prépondérance d'une classe par rapport à l'autre et la précision sera un indicateur fiable pour les modèles.

## Méthodologie

Nous avons décidé de ne travailler que sur les données de clôture pour chaque heure et de baser tous les indicateurs sur cette variable là. Avant toutes choses, on a procédé à la création d'une colonne "return" qui représente la variation de la valeur du prix de clôture. Ensuite se fait la création de notre variable cible ou "label", elle explicite si le prix a augmenté ou diminué durant l'heure qui suit. C'est cette variable-là qu'on va essayer de prédire.  
**Label :**  
1 : Si la valeur augmente l'heure suivante.  
0 : Sinon

	date	open	high	low	close
	2019-01-03 23:00:00	1.13930	1.13964	1.13895	1.13899
	2019-01-04 00:00:00	1.13901	1.14054	1.13893	1.14052
	2019-01-04 01:00:00	1.14052	1.14118	1.14027	1.14040
	2019-01-04 02:00:00	1.14041	1.14076	1.13974	1.13976
	2019-01-04 03:00:00	1.13975	1.13986	1.13881	1.13978

	date	return	label
	2019-01-03 23:00:00	-0.000043	1.0
	2019-01-04 00:00:00	0.000530	1.0
	2019-01-04 01:00:00	0.000843	0.0
	2019-01-04 02:00:00	-0.000425	0.0
	2019-01-04 03:00:00	-0.000617	0.0

## Feature Extraction et Indicateurs

La variation des prix Forex fait intervenir de nombreux phénomènes aléatoires et il s'agit d'un marché très volatile, Il est alors nécessaire d'utiliser plusieurs variables pour rendre compte de l'état du marché et pour améliorer les performances du modèle. Il est impératif d'introduire de nouvelles variables à la base de données, en particulier les indicateurs:

# Utilisation d'algorithmes d'intelligence artificielle pour la prévision de l'évolution des prix dans le marché FOREX.

BOUGUILA Houssem & EL FERCHICHI Wael

houcembougula@gmail.com | wael.elferchichi@gmail.com

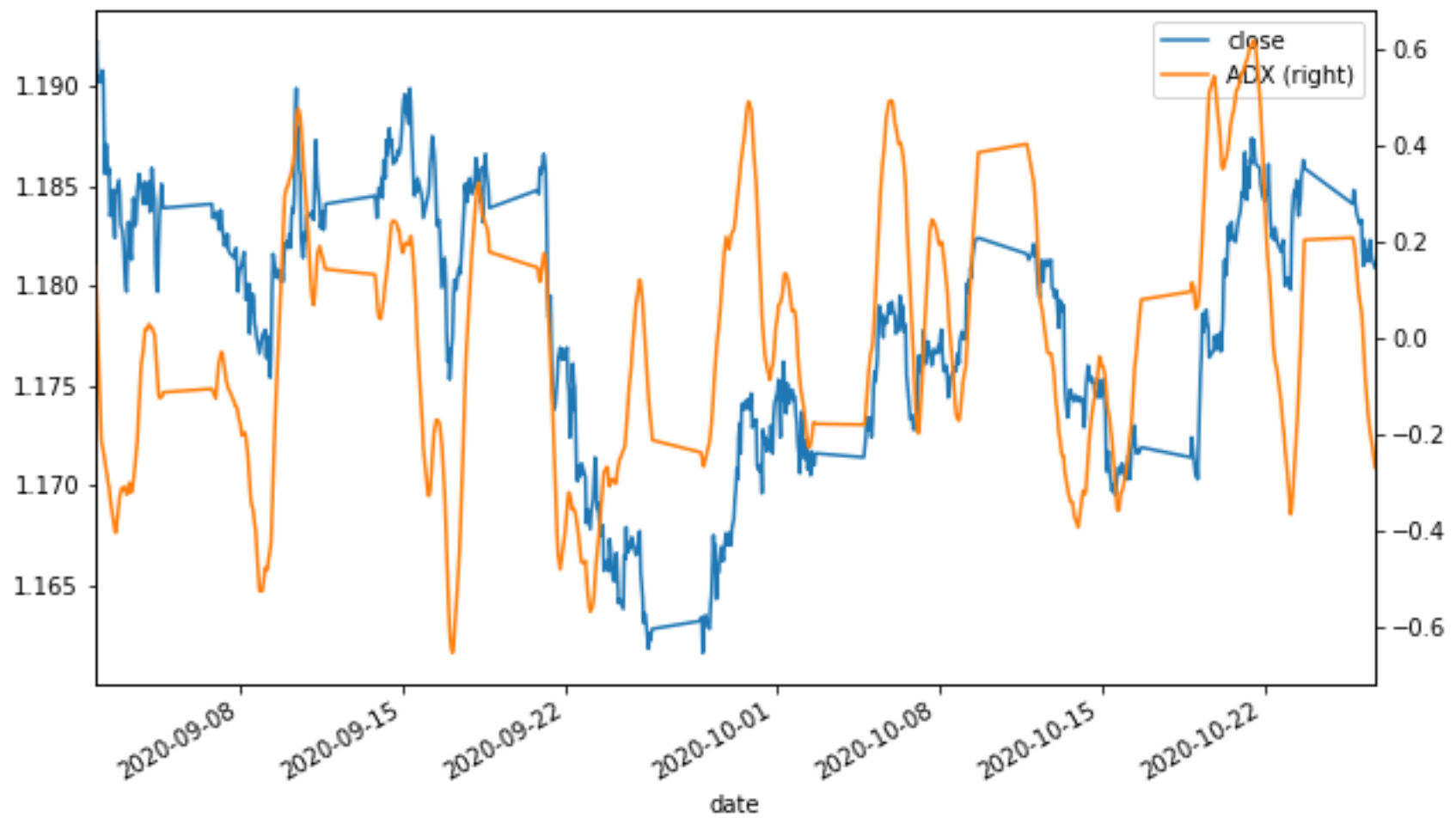


Figure 1: Figure évolution de l'indicateur ADX

**ADX :** Le Average Directional Index est un indicateur couramment utilisé qui nous renseigne sur la force de la tendance du prix. Particulièrement utile dans notre modèle puisqu'il peut potentiellement rendre les prédictions plus sûres (Réduction de la RMSE). Nous avons aussi créé une variable catégorielle qui se base sur les valeurs critiques "traditionnelles" de l'ADX qui permettent de caractériser la force de la tendance.  
**RSI :** Relative Strength Index, également un indicateur de la puissance de la tendance mais qui, en plus, permet de repérer une situation de sur-achat/Vente. Une variable catégorielle se basant sur les différents seuils de la RSI se trouve parmi les colonnes.  
**CCI:** Commodity Channel Index est ce qu'on appelle un oscilateur ou encore variable cyclique, cet indicateur rend compte d'une situation de sur-achat/vente et indique donc que la valeur va regagner une certaine moyenne. Nous avons estimé que cet indicateur est pertinent vu qu'il concerne en premier lieu le changement de la direction de la tendance. Une variable catégorielle qui nous indique si les valeurs sont au dessus du seuil du sur-achat ou en dessous du seuil du sur-vente se trouve également dans la base de données.  
**Variables cycliques :** Date  
Enfin, nous avons remarqué que le comportement du marché est différent selon le jour de la semaine ou encore lorsque le marché est sur le point de fermer ou s'il vient juste d'ouvrir. Nous avons donc jugé judicieux d'ajouter les variables heures/jour/mois encodées sous forme de variables cycliques pour aider le modèle à se positionner et de s'adapter aux changements de comportement dûes à la date.  
**Momentum:** ou encore Taux de Changement est un oscilateur caractérisant l'accélération de la tendance et qui permet dans certains cas de nous prévenir par rapport à un potentiel changement de tendance. Il est à noter que toutes les variables continues ont été normalisées vu que les algorithmes d'optimisation des modèles choisis sont très sensibles aux différences d'échelles entre variables.

## Modèles de Classification

**SVM: Support Vector Machine** La méthode SVM est un algorithme d'apprentissage supervisé qui se base sur un concept de marge maximale dont le système revient à la résolution du problème d'optimisation suivant:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^n \zeta_i$$
$$S/C \ y_i (w^T \phi(x_i) + b) \geq 1 - \zeta_i$$
$$\zeta_i \geq 0, i = 1, \dots, n$$

L'implémentation de l'algorithme s'est faite à travers le framework scikit-learn avec utilisation du noyau RBF (Radial Basis Function) avec un terme de régularisation C=2 et Γ = 'Scale'.  
**Recurrent Neural Network**  
Le réseau de neurones récurrent est une architecture spécialement adaptée à l'étude des données séquentielles (séries temporelles dans notre cas). Nous avons opté pour une architecture mixte à cellule LSTM (Long-Short Term Memory) et perceptron multi-couche qui, d'après de nombreux essais, permettent d'améliorer les performances du modèle. La fonction de coût "categorical\_crossentropy" avec un terme de régularisation a été adopté pour l'optimisation du modèle.

$$C = -\frac{1}{n} \sum_{x_j} [y_j \ln a_j^L + (1 - y_j) \ln (1 - a_j^L)] + \frac{\lambda}{2n} \sum_w w^2$$

La phase de parameter-tuning nous a ammené à choisir l'architecture et les hyperparamètres suivants:

Figure 2: Architecture RNN

Layer (type)	Output Shape	Param #
-----	-----	-----
lstm_12 (LSTM)	(None, 48, 128)	72704
batch_normalization_12 (Bac	(None, 48, 128)	512
lstm_13 (LSTM)	(None, 48, 64)	49408
batch_normalization_13 (Bac	(None, 48, 64)	256
lstm_14 (LSTM)	(None, 64)	33024
batch_normalization_14 (Bac	(None, 64)	256
dense_16 (Dense)	(None, 32)	2080
dense_17 (Dense)	(None, 16)	528
dense_18 (Dense)	(None, 8)	136
dense_19 (Dense)	(None, 3)	27
-----	-----	-----
Total params:	158,931	
Trainable params:	158,419	
Non-trainable params:	512	

Comparaison

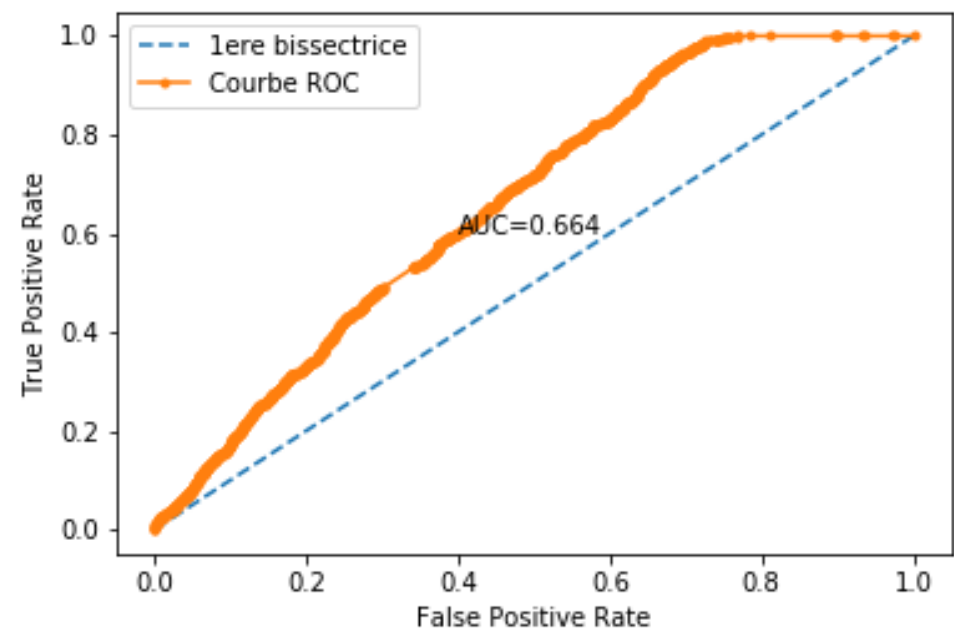


Figure 3: Courbe ROC: Modèle RNN

Mesures	Précision	Sensibilité	F1-Score	RMSE
Modèles				
SVM	0.58	0.17	0.26	0.603
RNN	0.54	0.37	0.445	0.62

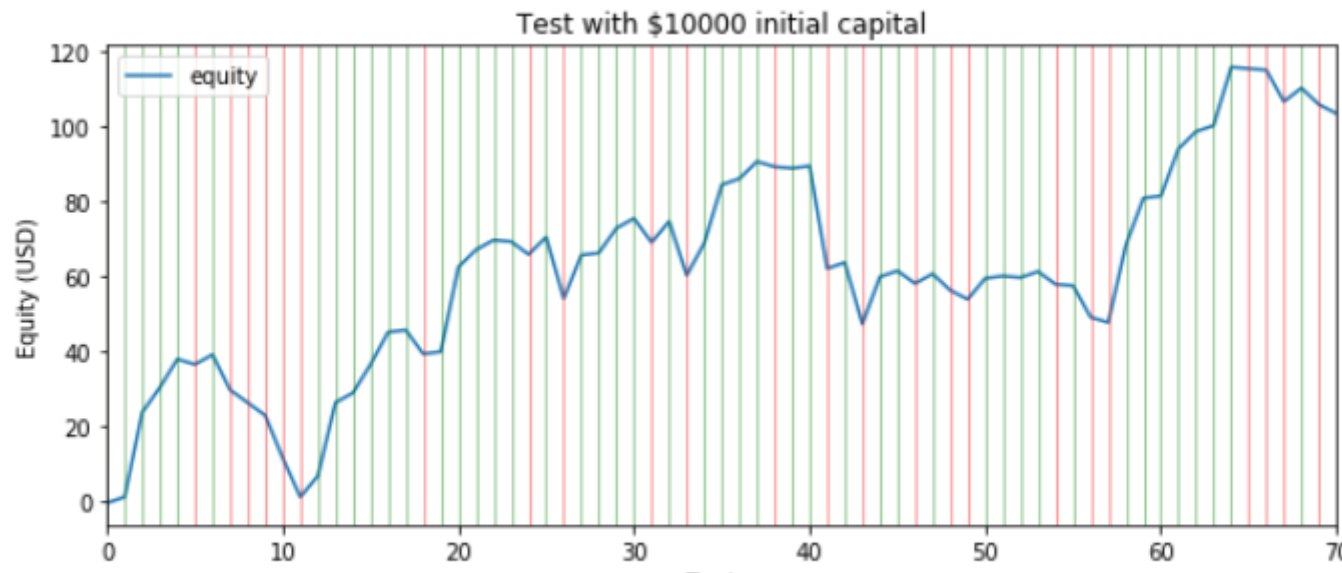
Table 1: Comparaison des modèles

## Back-Testing

Avant de procéder à l'automatisation de l'algorithme, une phase de test ou encore de "back-testing" s'impose. En effet, plusieurs facteurs comme la comission ainsi que le coût de gestion du compte Forex ne sont pas pris en compte dans les modèles créés. Cette partie va nous permettre de cerner les problèmes qui échappent à la modélisation initiale. Ce tableau récapitulatif illustre le test de l'algorithme réalisé sur des données récentes et rend compte des performances concrètes du modèle SVM.

	return	label	pred	won	pnl	equity
66	0.0000	0	1	False	-0.4	115.2
67	0.0008	1	0	False	-8.4	106.8
68	-0.0004	0	0	True	3.6	110.4
69	0.0004	1	0	False	-4.4	106.0
70	0.0002	1	0	False	-2.4	103.6

Profit Net	:	\$103.60
Nombre de prédictions justes	:	41
Nombre de prédictions fausses	:	30
Nombre total de trades	:	71
Précision	:	57.75%
Moyenne par Transaction Gagnée	:	\$6.746
Moyenne par Transaction Perdue	:	-\$5.767
Gain le plus important	:	\$22.600
Perte la plus importante	:	-\$27.400



## Conclusion

Les comparaisons précédentes montrent qu'il y a une similarité entre les performances des deux modèles mais en les évaluant sur plusieurs plages de données, nous avons conclu que les prédictions RNN sont plus stables, ou encore robustes vis à vis du backtesting. Suite à la réalisation de ces modèles et l'évaluation de leur performances, on peut conclure qu'il est irréfutable qu'ils arrivent à reconnaître un signal réel ou encore une certaine relation entre les indicateurs, valeurs horaires et la tendance des paires Forex. Cependant, pour exploiter les prédictions de ces modèles, il faudra développer un algorithme de Trading plus sophistiqué que celui utilisé dans ce projet. Notamment en se basant sur le degré de confiance des prédictions ou encore en essayant de le rendre plus robuste face aux commissions. Le travail a été traité d'un point de vue Machine Learning et on estime qu'en ayant une expérience Forex concrète, d'autres indicateurs peuvent être ajoutés et la conception d'un algorithme plus rigoureux peut être possible.

## References

- [1] A. Baasher and M. Fakhr, "Forex trend classification using machine learning techniques," *Proceedings of the 11th WSEAS International Conference on Applied Computer Science*, pp. 41–47, 01 2011.
- [2] L. Cao and F. Tay, "Financial forecasting using support vector machines," *Neural Computing and Applications*, vol. 10, pp. 184–192, 05 2001.
- [3] S. L. H.-V. S. Alexander Jakob Dautel, Wolfgang Karl Härdle, "Forex exchange rate forecasting using deep recurrent neural networks," *Neural Computing and Applications*, vol. 2, 03 2020.
- [4] A. Geron, *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. Sebastopol, CA: O'Reilly Media, 2017.