



AVIS DE SOUTENANCE D'UNE THESE DE DOCTORAT

Le Doyen de la Faculté des Sciences a le plaisir d'informer le public qu'une soutenance de
thèse de Doctorat en

«**Mathématiques, Informatique et Applications**»

aura lieu le 23/12/2023 à 10h à la Faculté des Sciences, Kénitra

La Thèse sera présentée par Mr **JOUILIL YOUNESS**

Sous le thème :

**Comparing The Forecasting Accuracy Metrics For Stationary & Non-Stationary Time
Processes: Machine Learning Algorithms vs Classical Statistical Models**

Devant le jury composé de :

Nom et Prénom	Titre	Etablissement
ALI KACHA	Président	Faculté des Sciences, Kénitra
ELFATINI MOHAMMED	Rapporteur	Faculté des Sciences, Kénitra
REGRAGUI TAKI	Rapporteur	Ecole Supérieure de Technologie, El Jadida
ABARDA ABDALLAH	Rapporteur	Faculté d'Economie et de Gestion, Settat
LAKHNATI GHIZLANE	Examineur	Ecole Nationale des Sciences Appliquées, Agadir
IAOUSSE M'BAREK	Co-Directeur de thèse	Ecole Supérieure de Technologie, Casablanca
MENTAGUI DRISS	Directeur de thèse	Faculté des Sciences, Kénitra



Sujet de thèse :

Résumé: La prévision et l'analyse des séries chronologiques ont attiré une attention écrasante, en particulier avec le développement qui s'est produit ces dernières années, et puisqu'elle est devenue un domaine important, en particulier dans l'analyse des sentiments, la médecine, l'économie et l'industrie. Les avancées récentes en matière de puissance de calcul et, plus important encore, le développement d'algorithmes et d'approches d'apprentissage automatique tels que l'apprentissage en profondeur ont conduit au développement de nouveaux algorithmes d'analyse et de prédiction pour les séries chronologiques.

Par conséquent, l'objectif principal de cette thèse est de fournir un examen approfondi des modèles d'apprentissage automatique les plus efficaces et les plus largement utilisés pour la prévision.

Des simulations numériques et des études expérimentales ont été réalisées pour étudier le potentiel des systèmes classiques et algorithmes d'apprentissage en profondeur pour améliorer les mesures de précision des prévisions. En effet, l'évaluation des simulations, basée sur divers processus temporels de génération de données, soutiennent que l'algorithme KNN a une meilleure précision en termes de prévision par rapport aux modèles candidats. Le MAPE du modèle KNN était d'environ 4.976843, tandis que le modèle de mémoire longue à court terme (LSTM), de régression vectorielle de support (SVR), et de ARIMA avaient des MAPE respectivement de 13.992133, 6.810311, 10.606617. Paradoxalement, à court terme, les mesures de prévision ont les plus petites valeurs pour les processus ARIMA que les modèles concurrents, ce qui signifie que l'algorithme ARIMA est le plus approprié dans le cas de petits ensembles de jeu données univariés. Dans le même sens, l'analyse quantitative a confirmé que les modèles de deep learning, fournissent moins de précision notamment en termes de performances prédictives.

Une extension de cette investigation est recommandée. Ainsi, nous proposons aux futurs chercheurs de fusionner les algorithmes classiques et ceux de DL pour de meilleurs résultats de prévision, de vérifier les méthodes probabilistes, en particulier les simulations de Monte Carlo, et de comparer l'approche classique à l'apprentissage en profondeur en termes de gestion des saisonnalités multiples.

Mots clés: Apprentissage automatique, séries chronologiques univariées, intelligence artificielle, SVR, ARIMA, Deep learning, KNN, LSTM, Prédictions.

Abstract:

Time series forecasting and analysis have attracted overwhelming attention, especially with the development that occurred in recent years, and have become an important field, especially in sentiment analysis, medicine, economy, and industry. Recent advances in computing power and, more importantly, the development of more advanced machine learning algorithms and approaches such as deep learning have led to the development of new analysis and prediction algorithms for time series data.

Accordingly, the main objective of this thesis report is to provide an in-depth examination of the most efficient and widely utilized machine learning models for forecasting.

Both, numerical simulations and experimental case studies have been performed to investigate the potential of classical and Machine learning algorithms to improve the accuracy metrics of forecasting. Indeed, the simulations assessment, based on various time processes of data sets generation and preprocessing, argue that the K - Nearest Neighbors (KNN) algorithm has better accuracy in terms of forecasting than the candidate models, notably in the middle and long term. The MAPE for the KNN was around 4.976843 while the Long short-term memory (LSTM), the Support Vector Regression (SVR), and the classical ARIMA algorithm had a MAPE of 13.992133, 6.810311, and 10.606617 respectively. Paradoxically in the short-term, the forecasting measures have the smallest values for ARIMA processes than the other models which means that the ARIMA algorithm is most appropriate in the case of univariate small data sets, where machine learning algorithms are not yet at their best.

Quantitative analysis has shown that the ML models especially the deep learning algorithms provide the least precision, especially in terms of predictive performance in the short term. An extension of this investigation is recommended. As further improvements, we propose for future researchers to merge classical framework and DL ones for better forecasting results, to check probabilistic methods in particular Monte Carlo Simulations, and compare the classical approach to Deep Learning in terms of handling multiple seasonalities.

Keywords: Machine Learning, Univariate time series, artificial intelligence, KNN, SVR, ARIMA, Deep learning, LSTM, Forecasting.