

Time Series avec ARIMA et Deep Learning

Objectif général :

Réaliser une présentation synthétique et claire mettant en valeur les travaux effectués sur les séries temporelles, en couvrant un éventail allant des modèles classiques comme ARIMA aux approches avancées basées sur le Deep Learning.

L'objectif est de conjuguer rigueur analytique, visualisations attrayantes et modélisation efficace. En outre, les étudiants devront démontrer leur capacité à généraliser les compétences acquises pour traiter des problématiques liées au NLP et aux techniques de Deep Learning avancées, telles que les transformers.

Le travail peut être réalisé en groupe (quatre, trois ou deux personnes) ou individuellement. La taille du groupe influencera la portée des tâches à effectuer :

- Un groupe de quatre personnes doit inclure une analyse approfondie intégrant jusqu'à quatre techniques.
- Un groupe de trois personnes doit explorer trois techniques pertinentes.
- Un groupe de deux personnes se concentrera sur deux techniques clés.
- Pour les travaux individuels, une technique principale doit être choisie, mais avec une analyse détaillée.

La qualité et la quantité du travail attendu doivent refléter l'engagement, la collaboration, et la maîtrise des techniques pertinentes, en lien avec les exigences du projet.

À cette fin, je propose un plan comme source d'inspiration ainsi que des consignes adaptées à un groupe de quatre personnes. Vous êtes libre d'ajouter ou de retirer des éléments pour refléter votre propre vision, à condition que la présentation reste professionnelle et intègre les éléments essentiels témoignant de votre maîtrise.

Consignes Générales :

1. Présentation PowerPoint :

- Limiter l'utilisation de texte long. Privilégiez les figures, schémas explicatifs, graphiques et points concis.
- Chaque slide doit être visuellement clair et bien organisé.

2. Organisation de la présentation :

- Respectez la structure fournie ci-dessous.

- Les diapositives doivent refléter vos travaux de manière synthétique et professionnelle.

Structure de la Présentation :

Partie 1 : Résumé des Travaux Pratiques Effectués (indépendante du nombre dans le groupe)

Présentez les étapes principales abordées durant les travaux pratiques, organisées comme suit :

- 1. Time Series : de ARIMA à Deep Learning – Objectifs et Défis**
Résumez les objectifs des travaux pratiques et les défis rencontrés dans l'analyse des séries temporelles. (1 slide)
- 2. Premiers Pas en Time Series**
Introduction au concept de séries temporelles et aux étapes de base pour les analyser.
- 3. Découverte des Jeux de Données à Travers des Exemples Simples :**
 - **3-1)** Préparation des données et visualisation (1 slide).
 - **3-2)** Division des données en *training* et *test* (1 slide).
 - **3-3)** Modèles classiques : AR, ARMA et ARIMA (1 à 2 slides).
 - **3-4)** Hyperparamètres, ACF, PACF, et Grid Search (1 à 2 slides).
 - **3-5)** Évaluation des modèles (1 slide).
- 4. Deep Learning : ANN, RNN, et LSTM**
 - **4-1)** Introduction aux concepts de Deep Learning appliqués aux séries temporelles (1 à 3 slides).
 - **4-2) Google Stock Analysis (synthèse).**
 - **4-2-1)** Modélisation avec un réseau de neurones artificiels (ANN) (1 slide).
 - **4-2-2)** Modélisation avec un réseau récurrent (RNN) (1 slide).
 - **4-2-3)** Modélisation avec un réseau à mémoire à long terme (LSTM) (1 slide).

Partie 2 : Projets

Plan Général des Projets : Analyse de Séries Temporelles et NLP Avancé

Objectifs Généraux :

- Maîtriser l'analyse et la modélisation des séries temporelles et du NLP.
- Concevoir des architectures adaptées à chaque type de modèle.
- Comprendre les bonnes pratiques, les limites, et les réglages des hyperparamètres.

Projet 1 : Analyse de Séries Temporelles et Modèles Avancés

1. Choix et Exploration d'un Dataset *(indépendante du nombre dans le groupe)*

- Identifier un dataset pertinent pour les séries temporelles (similaire à Google stock avec une complexité suffisante)
- Prétraiter les données : gestion des valeurs manquantes, anomalies, normalisation.
- Analyse exploratoire (visualisations, tendances, saisonnalités).

2. Modélisation ARIMA *(indépendante du nombre dans le groupe)*

- Techniques traditionnelles pour les séries temporelles.
- Paramétrage optimal (p, d, q) via ACF/PACF.
- Analyse des résultats et comparaison avec les approches avancées.

3. Modélisation avec Deep Learning

3.1. ANN (Artificial Neural Networks)

- Proposition d'architecture :
 - Couches entièrement connectées adaptées pour séries chronologiques.
 - Utilisation de fonctions d'activation comme ReLU et Sigmoid.
 - Régularisation via dropout pour limiter le surapprentissage.
- Bonnes pratiques :
 - Normalisation des données et gestion des features importantes.
- Limites et problèmes récurrents :
 - Incapacité à capturer les relations temporelles complexes.
- Initialisation & Hyperparamètres :
 - Nombre optimal de couches et de neurones par couche.

3.2 RNN (Recurrent Neural Networks)

- Proposition d'architecture :

- Couches adaptées pour RNN avec empilement si nécessaire.
- Régularisation et ajustement des flux d'information entre les couches.
- Bonnes pratiques :
 - Préparation des séquences et augmentation des données.
- Limites et problèmes récurrents :
 - Défis spécifiques liés à RNN (e.g., vanishing gradient, performances sur séquences longues).
- Initialisation & Hyperparamètres :
 - Configuration optimale des couches et des paramètres (taille des séquences, taux d'apprentissage, etc.).

3.3 LSTM (Long Short-Term Memory)

- Proposition d'architecture :
 - Couches adaptées pour LSTM avec empilement si nécessaire.
 - Régularisation et ajustement des flux d'information entre les couches.
- Bonnes pratiques :
 - Préparation des séquences et augmentation des données.
- Limites et problèmes récurrents :
 - Défis spécifiques liés à LSTM (e.g., vanishing gradient, performances sur séquences longues).
- Initialisation & Hyperparamètres :
 - Configuration optimale des couches et des paramètres (taille des séquences, taux d'apprentissage, etc.).

3.4 GRU (Gated Recurrent Unit)

- Proposition d'architecture :
 - Couches adaptées pour GRU avec empilement si nécessaire.
 - Régularisation et ajustement des flux d'information entre les couches.
- Bonnes pratiques :
 - Préparation des séquences et augmentation des données.
- Limites et problèmes récurrents :
 - Défis spécifiques liés à GRU (e.g., vanishing gradient, performances sur séquences longues).
- Initialisation & Hyperparamètres :
 - Configuration optimale des couches et des paramètres (taille des séquences, taux d'apprentissage, etc.).

3.5 Comparaison des différentes techniques

Projet 2 : NLP et Techniques Avancées

1. Préparation et Transformation des données

- Prétraitement des données textuelles : tokenization, encodage.
- Transformation en séquences adaptées pour les modèles.

2. Modélisation NLP avec RNN, LSTM et GRU

2.1 RNN

- Proposition d'architecture :
 - Utilisation de couches RNN adaptées pour les séquences textuelles.
 - Intégration de mécanismes complémentaires comme l'attention.
- Bonnes pratiques :
 - Représentation vectorielle des mots avec embeddings préentraînés.
- Limites et problèmes récurrents :
 - Défis spécifiques liés à RNN.
- Initialisation & Hyperparamètres :
 - Optimisation des couches RNN, taille des séquences, et autres hyperparamètres.

2.2 LSTM

- Proposition d'architecture :
 - Utilisation de couches LSTM adaptées pour les séquences textuelles.
 - Intégration de mécanismes complémentaires comme l'attention.
- Bonnes pratiques :
 - Représentation vectorielle des mots avec embeddings préentraînés.
- Limites et problèmes récurrents :
 - Défis spécifiques liés à LSTM.
- Initialisation & Hyperparamètres :
 - Optimisation des couches LSTM, taille des séquences, et autres hyperparamètres.

2.3 GRU

- Proposition d'architecture :
 - Utilisation de couches GRU adaptées pour les séquences textuelles.
 - Intégration de mécanismes complémentaires comme l'attention.
- Bonnes pratiques :
 - Représentation vectorielle des mots avec embeddings préentraînés.
- Limites et problèmes récurrents :
 - Défis spécifiques liés à GRU.
- Initialisation & Hyperparamètres :
 - Optimisation des couches GRU, taille des séquences, et autres hyperparamètres.

3. Ajout du Mécanisme d'Attention *(indépendante du nombre dans le groupe)*

- Proposition d'architecture :
 - Couche d'attention appliquée après LSTM ou GRU.
 - Visualisation des scores d'attention pour interprétabilité.
- Bonnes pratiques :
 - Ajustement des paramètres de la fenêtre d'attention.
- Limites et problèmes récurrents :
 - Sensibilité aux grandes séquences.
- Initialisation & Hyperparamètres :
 - Optimisation des couches d'attention et de leur configuration.

4. Modélisation avec Transformers *(indépendante du nombre dans le groupe)*

- Proposition d'architecture :
 - Architecture Encoder-Decoder (exemple : BERT ou GPT).
 - Définir le nombre de têtes d'attention, de couches, et autres hyperparamètres.
- Bonnes pratiques :
 - Optimisation des paramètres pour des séquences longues.
 - Utilisation de préentraînements pour accélérer l'apprentissage.
- Limites et problèmes récurrents :
 - Ressources nécessaires pour l'entraînement.
 - Difficulté d'interprétation des résultats.
- Initialisation & Hyperparamètres :
 - Réglage des couches d'attention et du Learning rate Scheduler.

5. Perspectives et Prolongements

- Analyse des limites des approches utilisées.

- Propositions pour intégrer des modèles préentraînés (ex : GPT, BERT).
- Exploration des Transformers hybrides pour séries temporelles.

- **Dates de présentation : la troisième semaine de février**
- **Livrables : code sources + exécution sous forme HTML**
- Présentations power points**
- Copie sur google classroom + copie par email à : a.naji@ump.ac.ma**
- Date limite de dépôt de travaux : Lundi 24 février
- Voici le lien où il faut définir les éléments de groupe

https://docs.google.com/spreadsheets/d/1ZqdFRhQuumhiEyomk_Q56vxE-8b6VwMInwyMmrT2g4o/edit?usp=sharing

Bon courage !