

# Prédiction du Taux de Chômage aux États-Unis

Analyse Exploratoire et Modélisation Temporelle

Youness Boumlik

Encadré par Sara Baghdadi

# Plan de la Présentation

Introduction

Objectifs de l'étude

Données et Prétraitement

Méthodologie de Modélisation

Régression Polynomiale

Modèle Prophet

Modèle ARIMA

Conclusion

- Le taux de chômage est un indicateur économique clé.
- Il reflète la santé du marché du travail et l'état général de l'économie.
- Sa prédiction est cruciale pour la planification politique, économique et pour les entreprises.
- **Objectif principal de cette étude** : Explorer des méthodes de prédiction du taux de chômage américain à partir de données historiques.

# Objectifs de l'étude

- Analyser l'évolution historique du taux de chômage aux États-Unis.
- Identifier les tendances, saisonnalités et cycles potentiels.
- Mettre en œuvre et évaluer différents modèles de prédiction :
  - Régression Polynomiale
  - Modèle Prophet
  - Modèle ARIMA/SARIMA
- Comparer les performances des modèles et discuter de leurs applicabilités.

- **Source** : Fichier 'UNRATE.csv' (série temporelle du taux de chômage mensuel aux USA).
- **Période** : De Janvier 1948 à Janvier 2025 (prévisions incluses dans les données initiales jusqu'à cette date).
- **Variables** :
  - 'observation\_date' : Date de l'observation (convertie en datetime).
  - 'UNRATE' : Taux de chômage (en pourcentage, type float).
- **Nombre d'observations** : 925 entrées.

- Chargement des données avec Pandas.
- Conversion de la colonne 'observation\_date' au format datetime.
- Vérification des doublons (aucun doublon majeur n'a été identifié et supprimé).
- Création d'une variable numérique 'date\_num' (nombre de jours depuis la première observation) pour la régression polynomiale.
- Mise en index de la date pour l'analyse des séries temporelles.

# Visualisation Initiale des Données

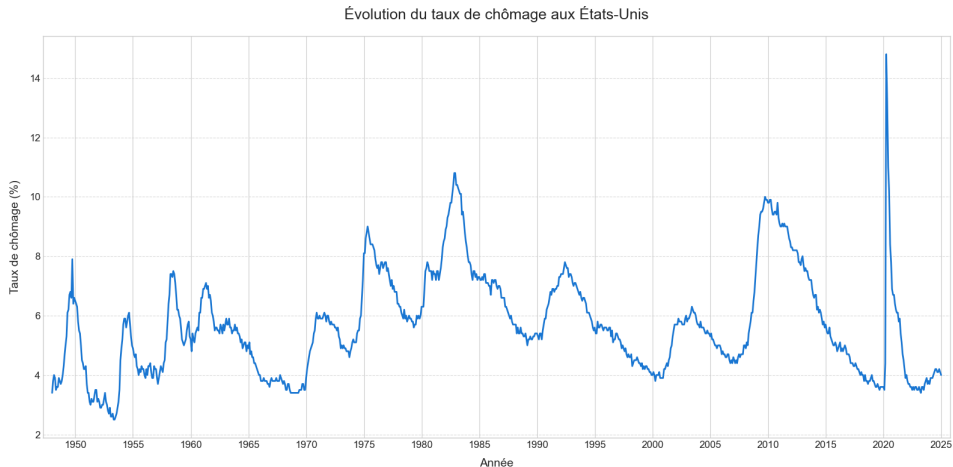
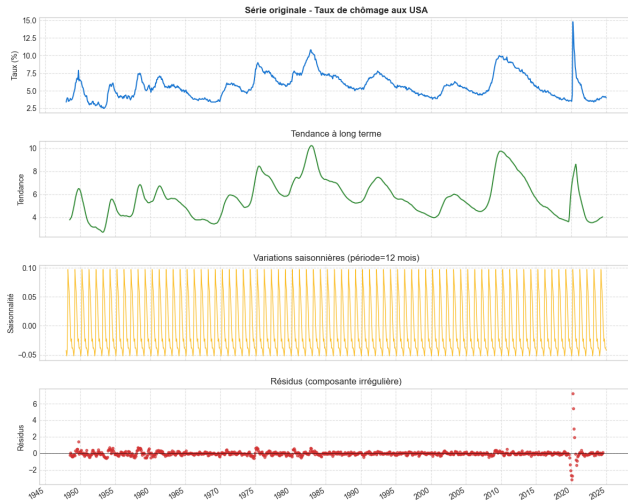


Figure – Évolution du taux de chômage aux États-Unis (1948-2025).

# Décomposition de la Série Temporelle

## Décomposition saisonnière du taux de chômage aux États-Unis



La décomposition saisonnière additive sépare une série temporelle en :

- Tendance: variation à long terme du taux de chômage
- Saisonnalité: variations régulières (cycles de 12 mois)
- Résidus: fluctuations irrégulières/aléatoires

## Décomposition saisonnière :

- **Tendance** : Variation à long terme  
(amplitude de 7.54)
- **Saisonnalité** : Variations régulières sur 12 mois  
(amplitude de 0.149)
- **Résidus** : Fluctuations irrégulières  
(écart-type de 0.423)



# Approche de Modélisation

Nous allons explorer trois approches pour la prédiction du taux de chômage :

## 1. Régression Polynomiale :

- Modèle simple pour capturer des tendances non linéaires.
- Utile comme baseline et pour comprendre la courbure générale des données.

## 2. Modèle Prophet :

- Développé par Facebook, adapté aux séries temporelles avec saisonnalités multiples et jours fériés.
- Robuste aux données manquantes et aux changements de tendance.

## 3. Modèle ARIMA/SARIMA :

- Approche statistique classique pour l'analyse et la prévision de séries temporelles.
- SARIMA étend ARIMA pour gérer explicitement la saisonnalité.

L'évaluation se fera à l'aide de métriques comme  $R^2$ , MAE, RMSE, MAPE.

# Régression Polynomiale : Principe et Formulation

La régression polynomiale est une extension de la régression linéaire simple. Elle permet de modéliser des relations non linéaires entre une variable indépendante  $x$  et une variable dépendante  $y$  à l'aide d'un polynôme de degré  $n$ .

## Principe :

- On ajuste une courbe polynomiale au lieu d'une droite.
- On génère de nouvelles variables explicatives :  $x^2, x^3, \dots, x^n$ .
- Cela permet de mieux capter les variations complexes des données.

## Formule (pour un polynôme de degré $n$ ) :

$$y = b_0 + b_1x + b_2x^2 + \dots + b_nx^n + \epsilon$$

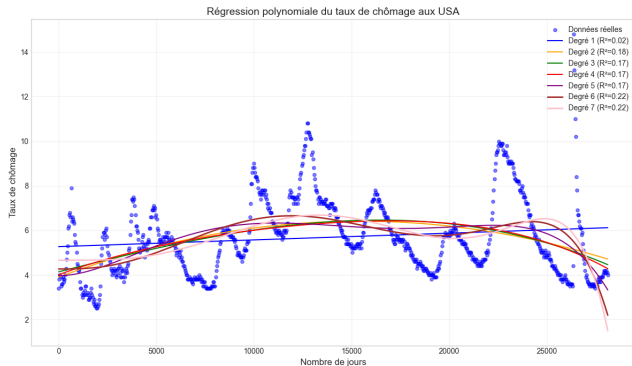
Où  $b_0, b_1, \dots, b_n$  sont les coefficients et  $\epsilon$  est le terme d'erreur.

## Avantages et Inconvénients :

- **Avantage** : Modélise les relations non linéaires.
- **Inconvénient** : Choix délicat du degré  $n$

# Régression Polynomiale : Résultats

- Test de différents degrés de polynômes (1 à 7).
- Le **polynôme de degré 6** a offert le meilleur score  $R^2$  sur l'ensemble de test, avec  $R^2 = 0.2179$ .
- Bien que capturant une certaine courbure, ce modèle a des limites pour prédire des dynamiques complexes de séries temporelles.



# Modèle Prophet : Fonctionnement

Prophet est un modèle additif où les non-linéarités sont ajustées avec des termes de saisonnalité annuelle, hebdomadaire, journalière et des effets de jours fériés. Il modélise la série temporelle  $y(t)$  comme :

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

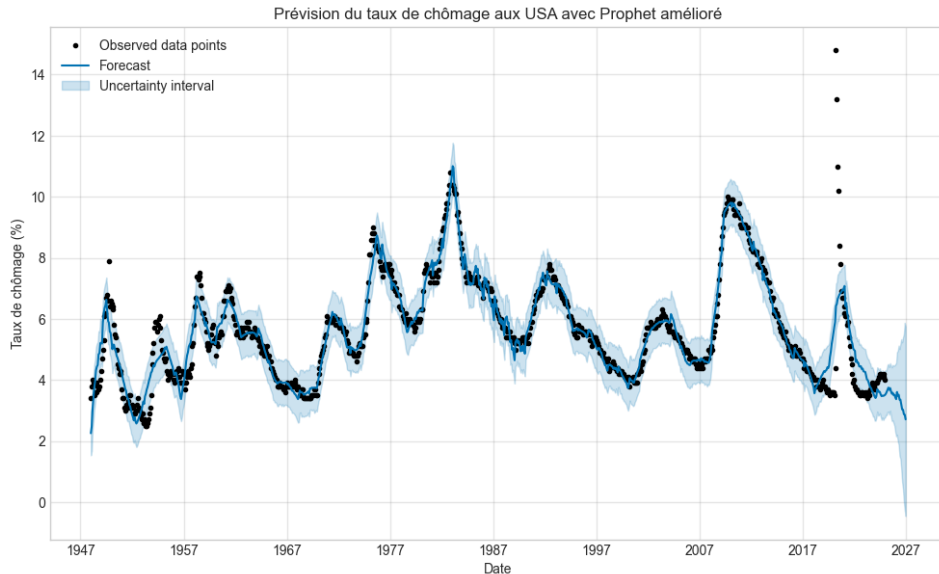
Où :

- $g(t)$  : fonction de **tendance** (trend), modélise les changements non périodiques.
  - Peut être linéaire ou logistique.
  - Gère les points de changement (changepoints).
- $s(t)$  : fonction de **saisonnalité** (seasonality), modélise les changements périodiques (ex : annuel, mensuel).
  - Approximée par des séries de Fourier.
  - Permet des saisonnalités multiples.
- $h(t)$  : fonction des **effets de jours fériés** (holidays) et événements spéciaux.
- $\epsilon_t$  : terme d'erreur, supposé normalement distribué.

## Performances sur les données historiques :

- Erreur absolue moyenne (MAE) : 0.323
- Erreur quadratique moyenne (RMSE) : 0.581
- Erreur moyenne absolue en pourcentage (MAPE) : 6.276%

# Modèle Prophet : Visualisation des Prévisions



# Modèle Prophet : Composantes

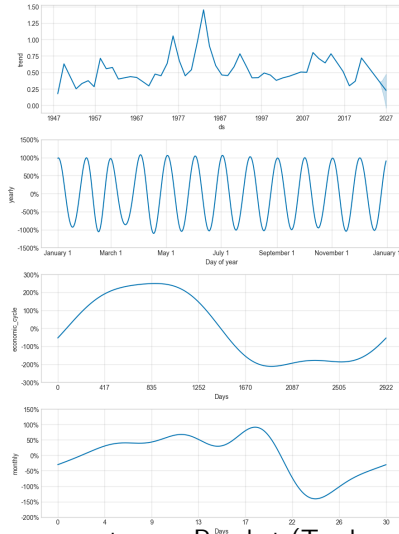


Figure – Décomposition des composantes par Prophet (Tendance, Saisonnalités).

# Modèle ARIMA : Fonctionnement

**ARIMA** (AutoRégressif Intégré à Moyenne Mobile) est un modèle utilisé pour prédire des séries temporelles.

- **AR (AutoRégressif,  $p$ )** : dépend des valeurs passées.
- **I (Intégré,  $d$ )** : rend la série stable (stationnaire).
- **MA (Moyenne Mobile,  $q$ )** : utilise les erreurs précédentes.



# Modèle ARIMA : Fonctionnement

**ARIMA** (AutoRégressif Intégré à Moyenne Mobile) est un modèle utilisé pour prédire des séries temporelles.

- **AR (AutoRégressif,  $p$ )** : dépend des valeurs passées.
- **I (Intégré,  $d$ )** : rend la série stable (stationnaire).
- **MA (Moyenne Mobile,  $q$ )** : utilise les erreurs précédentes.

**Méthode utilisée :**

- On teste si la série est stable avec un test statistique (ADF).
- Puis, on choisit automatiquement les bons paramètres ( $p$ ,  $d$ ,  $q$ ) avec Python grâce à la fonction `auto_arima`.

# Modèle ARIMA : Application et Résultats

## Étapes et Paramètres :

- **Test de stationnarité (ADF)** : La série est stationnaire (p-value = 0.0019).
- **Sélection automatique des paramètres ('auto\_arima')** :
  - Modèle retenu : ARIMA(1,1,1)
  - Le 'auto\_arima' a identifié 'd=1' (une différenciation).

## Étapes et Paramètres :

- **Test de stationnarité (ADF)** : La série est stationnaire ( $p\text{-value} = 0.0019$ ).
- **Sélection automatique des paramètres ('auto\_arima')** :
  - Modèle retenu : ARIMA(1,1,1)
  - Le 'auto\_arima' a identifié 'd=1' (une différenciation).

## Performances sur les données historiques :

- Erreur absolue moyenne (MAE) : 0.167
- Erreur quadratique moyenne (RMSE) : 0.415
- Erreur moyenne absolue en pourcentage (MAPE) : 2.89%
- Coefficient de détermination ( $R^2$ ) : 0.941

# Modèle ARIMA : Ajustement et Prévisions

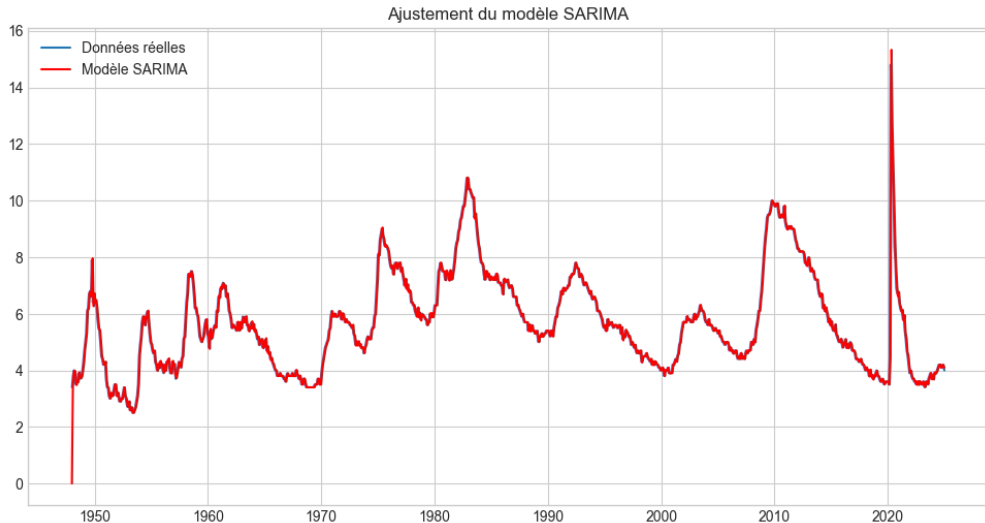


Figure – Ajustement du modèle ARIMA sur les données historiques.

# Comparaison des Prévisions : Prophet et ARIMA (jusqu'en 2027)

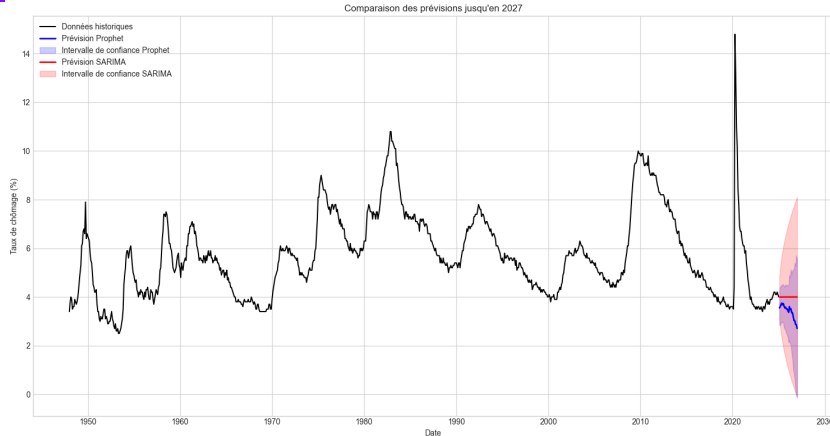


Figure – Comparaison des prévisions du taux de chômage avec Prophet et ARIMA, et leurs intervalles de confiance à 95% jusqu'en 2027.

- **Intervalles de confiance** : Celui de Prophet s'élargit plus rapidement, indiquant une plus grande incertitude à long terme comparé à ARIMA dans cet exemple.

# Comparaison des performances : Prophet vs ARIMA

Métrique	Prophet	ARIMA	Meilleur modèle
Erreur absolue moyenne (MAE)	0.322806	0.166880	ARIMA
Racine de l'erreur quadratique moyenne (RMSE)	0.581035	0.414739	ARIMA
Erreur absolue moyenne en pourcentage (MAPE)	6.275588	2.885327	ARIMA
Coefficient de détermination ( $R^2$ )	0.884278	0.940989	ARIMA

# Conclusion

- **Régression polynomiale** : Peu adaptée ( $R^2$  faible à 0.22).
- **Prophet** :
  - Bonne capture des tendances et cycles.
  - Modèle flexible et interprétable.
- **ARIMA** :
  - Meilleures performances globales (MAPE : 2.89%,  $R^2$  : 0.941).
  - Modèle ARIMA(1,1,1) simple et efficace.
- **Comparaison** : ARIMA est le plus précis. Prophet reste utile pour l'analyse des composantes.
- **Limites** : Prévisions sensibles aux événements imprévus. Modèles fondés sur les données passées.

**Merci de votre attention !**

Des questions ?