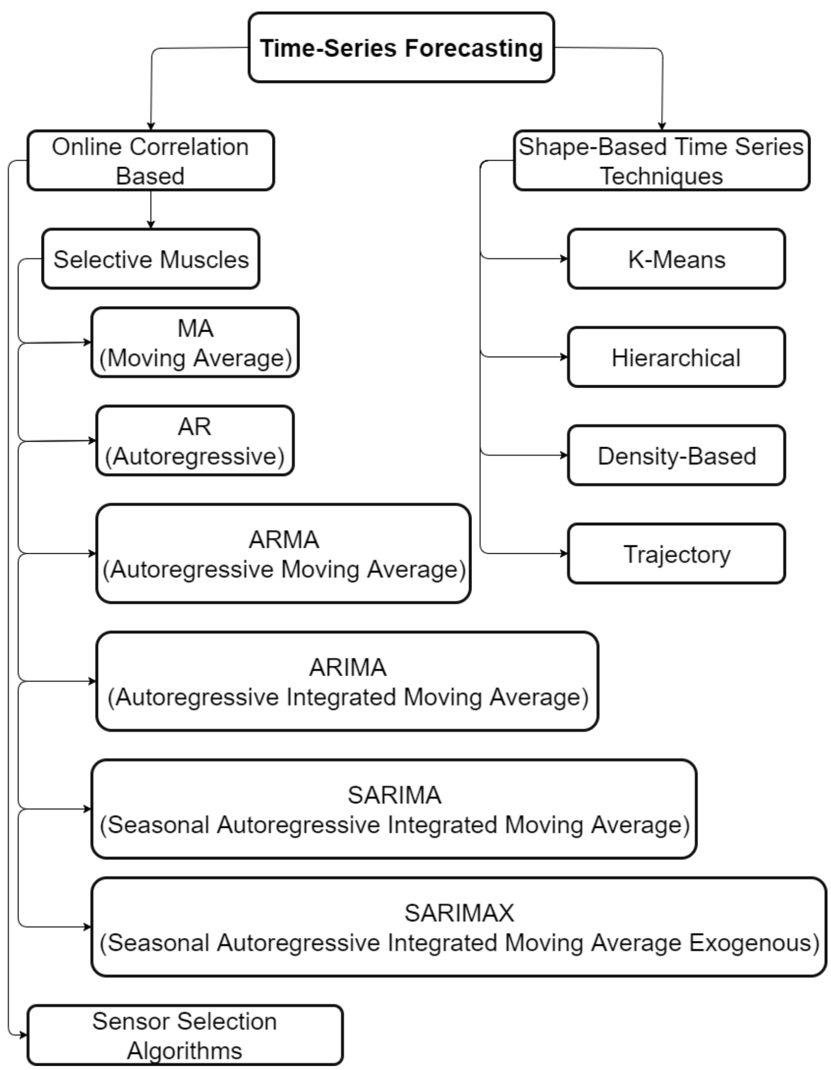
What is Time Series Forecasting?

عبارة عن مفهوم يمثل عملية مراقبة أو ملاحظة التشابه بين القيم السابقة والحالية لعنصر معين ضمن مجموعة من البيانات، بهدف التنبؤ بالقيمة المستقبلية لهذا العنصر.

تعتمد على مراقبة التشابه بسبب وجود ترابط تلقائي **Autocorrelation** بالتأكيد ضمن السلسلة المُعالجة.

على سبيل المثال، بمعرفة سعر منتج معين لليوم الحالي، يمكن التنبؤ (توقع) بسعرهُ غداً بشكل تقريبي.

يمثل المخطط التالي، ملخص لأنواع والنماذج (Models) التي تندرج تحت مفهوم Time-Series forecasting:



في هذهِ الوظيفة قمنا بالاعتماد على عدة نماذج تعكس الارتباطات بين قيم البيانات المدروسة،

فسوف نتحدث عن النماذج من نوع **Selective Muscles**، ونوضح طريقة عمل كل منها:

1. AR (Autoregressive):

- نموذج خطي يعتمد على قيم الفترة الزمنية السابقة من اجل التنبؤ بالقيم المستقبلية، بحيث تمثل الحالية مجموع القيم السابقة مضروبة بعامل رقمي معين (أوزان).
- يرمز لهُ AR(P)، حيث P هي درجة النموذج، وتعبر عن عدد قيم lags التي نريد تضمينها بهذا النموذج.
- لنفرض أن P = 1، فتكون المعادلة الممثلة للنموذج AR(1) هي كالتالي:

$$X_t = C + \emptyset_1 X_{t-1} + \varepsilon_t$$

بحيث:

X_{t-1}: هي قيمة X في الوقت السابق (lag)، أي لنفرض أن t هي اليوم (الوقت الحالي)، وبفرض t-1 هي الأسبوع الماضي، فتكون X_{t-1} هي القيمة المسجلة لX في الأسبوع الماضي.

C: ثابت أساسي.

$\emptyset_1$: قيمة عددية، يمكن التعبير عنها بأنها جزء من القيمة السابقة التي لاتزال في المستقبل، يجب ان تكون هذهِ القيمة ضمن المجال [1+,1-]، وذلك لأنهُ اذا كانت القيمة المطلقة لها فرضاً أكبر من الواحد فسوف يحدث تزايد تدريجي لها الى اللانهاية.

ε_t: الرواسب (الأخطاء)، وهي القيمة الوحيدة ضمن المعادلة التي نسعى الى تقليلها، بحيث تمثل الفرق بين القيمة المتنبئ بها في الزمن t والقيمة الحقيقية: $(\varepsilon_t - y_t - \hat{y}_t)$، عادة لا يمكن التنبؤ باختلافات هذهِ القيم، لأنهُ لا يوجد نمط معين يمكن التقاطهُ.

- المعادلة العامة:

$$X_t = C + \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \ldots\ldots + \emptyset_n X_{t-n} + \varepsilon_t$$

2. ARMA (Autoregressive Moving Average):

- التنبؤ بالاعتماد على البيانات في الزمن الماضي قد لا يكفي في بعض الأحيان، مثلاً عندما تقع أحداث غير متوقعة مثل الكوارث الطبيعية أو الازمات المالية، بهذهِ الحالات يوجد تحول مفاجئ لا يمكن التوقع بهِ، لذلك نحتاج الى نماذج تستخدم البيانات السابقة كأساس للتنبؤ، ولكن يمكنها أيضاً التكيف بسرعة مع الاحداث غير المتوقعة.
- بالحديث عن هذا النموذج الذي يأخذ بعين الاعتبار القيم السابقة بالإضافة الى الأخطاء السابقة الناتجة عن التنبؤات المستقبلية التي قام بها في الوقت السابق.
- هذا النموذج عبارة عن دمج بين نموذجين بسيطين (AR و MA).
- يرمز لهُ ARMA(P,Q)، بحيث P درجة النموذج AR والتي تمثل عدد وما التي نريد تضمينها، Q هي درجة النموذج MA والتي تمثل عدد الأخطاء السابقة التي نريد تضمينها.
- لنفرض P = 2 و Q = 3، فتكون المعادلة المعبرة عن النموذج ARMA(2,3) هي كالتالي:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \varepsilon_t$$

بحيث:

y_t, y_{t-1}, y_{t-2}: القيم في الفترات السابقة والفترة الحالية على التوالي.

$\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}$: قيم الأخطاء في الفترات السابقة والحالية، بحيث يتم الاعتماد على خطأ الفترات الأخيرة بهدف تصحيح التوقعات من خلال معرفة مدى البعد عن الهدف، بحيث يمكن التنبؤ بشكل افضل من المرات المقبلة.

ϕ, θ: أوزان تتعلق بالنموذج MA و AR على التوالي، يجب أن يكونا ضمن المجال [-1,+1].

C: ثابت أساسي، يكون عادة C = 0.

- ليس من الضروري أن تتساوى قيمتي P و Q كما لاحظنا في المثال السابق.
- ملاحظة:

ARMA(P,0) = AR(P)

ARMA(0,Q) = MA(Q)

3. ARIMA (Autoregressive Integrated Moving Average):

- توسيع لنموذج ARMA السابق، بهدف التعامل مع البيانات الغير مستقرة non-Stationary.
- يتم تحديدهُ من خلال order(p,d,q) 3، بحيث p تتلق بالنموذج AR، q تتعلق بالنموذج MA، d هو معامل التكامل، بحيث يعبر عن عدد المرات التي نحتاجها لدمج السلسلة الزمنية بهدف ضمان استقرارها Stationarity.
- بالنسبة لمكونات AR ضمن هذا النموذج، يتم التعامل مع الفرق بين قيم السلسلة الزمنية $\triangle P$ بدلاً من القيم بحد ذاتها، بالتالي يتم التكامل d مرة لبناء سلسلة زمنية جديدة مستقرة، ومن ثم يتم تدريب نموذج ARIMA على هذهِ السلسلة الجديدة.
- لنفرض أن $p = 1, d = 1, q = 1$، فتكون المعادلة المعبرة عن نموذج ARIMA(1,1,1) هي:

$$\triangle P_t = C + \phi_1 \triangle P_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

بحيث:

$\triangle P_t = P_{t-1} - P_t,\ \triangle P_{t-1} = P_{t-2} - P_{t-1}$، وهكذا......

- ملاحظة:

ARIMA(p,0,q) = ARMA(p,q)

ARIMA(p,0,0) = AR(p)

ARIMA(0,0,q) = MA(q)

4. SARIMAX (Seasonal Autoregressive Integrated Moving Average Exogenous)

- كل النماذج السابقة التي تم الحديث عنها يكون لها مكافئ موسمي Seasonal.
- يعني مصطلح "موسمي" Seasonal أنهُ ضمن سلسلة زمنية معينة، إذا وجدت أنماط ضمنها غير متسقة، ولكنها تظهر بشكل دوري، فهذهِ سلسلة زمنية موسمية.
- لذلك في حال كانت السلسلة الزمنية تحتوي على أنماط موسمية، لا يمكن لأي من النماذج التي ذكرناها سابقاً أن تتنبأ بشكل صحيح من اجل هذه السلسلة، مما يفسر حاجتنا لهذا النموذج.
- ففي مسألتنا هذهِ، تعتبر السلسلة الزمنية المعبرة عن إجمالي كمية المبيعات لكل متجر، سلسلة زمنية موسمية لأنهُ ممكن أن تختلف كمية المبيعات بحسب الموسم الوارد زمنياً.

- المبدأ الأساسي لعمل هذا النموذج هو إيجاد سلسلة زمنية أخرى تنتشر بالوقت المناسب (موسمياً) أفضل من السلسلة الزمنية الاصلية.
- مقارنتاً بالنموذج السابق ARIMA، يحتاج هذا النموذج الى 4 معاملات إضافية (بالإضافة الى order(p,d,q)، يوجد معاملات موسمية Seasonal_order(sp,sd,sq,s)، بحيث:
  - sp: Seasonal autoregressive order
  - sd: Seasonal Integrated order
  - sq: Seasonal Moving Average order
  - s: هو طول الدورة، على سبيل المثال: لدينا سلسلة زمنية يومية، وطول الدورة فيها s = 4 فرضاً، فسوف يظهر النمط الموسمي مرة واحدة كل 4 أيام متتالية.
- لنفرض الآن أنهُ لدينا order(1,0,1)، و Seasonal_order(2,0,1,4) ، فيكون شكل المعادلة المعبرة عن نموذج SARIMAX(1,0,1,2,0,1,4)، هو:

$$y_t = C + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \varphi_1 (y_{t-4} + \varphi_1 y_{t-5}) + \varphi_2 (y_{t-8} + \varphi_1 y_{t-9}) + \theta_1(\varepsilon_{t-4} + \theta_1 \varepsilon_{t-5} + \theta_2 \varepsilon_{t-6}) + \varepsilon_t$$

تم شرح جميع البارامترات في النماذج السابقة.

ملاحظة هامة:

في هذا الملف تم شرح جميع المفاهيم التي تم دراستها وفهمها.

شرح طريقة الحل سوف يتم تضمينهُ ضمن ال Notebook مع الحل، بحيث قبل كل خانة تنفيذية سوف يتم إضافة الشرح الخاص بها.