Artificial Intelligence for Real World Application

16: Time Series Analysis

Introduction

- Every company in this world faces certain challenges and risks such as high competition , failure of technology, labor unrest, inflation, recession, and change in government laws. Thus, we can say that every business operates under risk and uncertainty. That's why forecast is necessary to lessen the adverse effect of the risks and to tell us in advance of any incoming dangers.
- Forecasting is a technique that uses historical data as inputs to make informed estimates that are predictive in determining the direction of future trends. Businesses utilize forecasting to determine how to allocate their budgets or plan for anticipated expenses for an upcoming period of time.

Time Series Analysis

- A time series is a series of measurements on the same variable collected over time. These measurements are made at regular time intervals. Thus it is a sequence of discrete-time data.
- Intervals of the Time Series Data
  1. Yearly :- GDP , Macro-economic series
  2. Quarterly :- Revenue of a company.
  3. Monthly:- Sales, Expenditure, salary
  4. Weekly:- Demand , Price of Petrol and diesel
  5. Daily:- Closing price of stock, Sensex value, daily transaction of ATM machine
  6. Hourly:- AQI
- Time series analysis can be useful to see how a given asset, security, or economic variable changes over time. It can also be used to examine how the changes associated with the chosen data point compare to shifts in other variables over the same time period.

Time Series Analysis

- Definition:
- ***A time series is nothing but a sequence of various data points that occurred in a successive order for a given period of time.***
- ***Objectives:***
- To understand how time series works, what factors are affecting a certain variable(s) at different points of time.
- Time series analysis will provide the consequences and insights of features of the given dataset that changes over time.
- Supporting to derive the predicting the future values of the time series variable.
- Assumptions: There is one and the only assumption that is “stationary”, which means that the origin of time, does not affect the properties of the process under the statistical factor.

How to analyse Time Series?

1. Collecting the data and cleaning it
2. Preparing Visualization with respect to time vs key feature
3. Observing the stationarity of the series
4. Developing charts to understand its nature.
5. Model building – AR, MA, ARMA and ARIMA
6. Extracting insights from prediction

Significance of Time Series

- TSA is the backbone for prediction and forecasting analysis, specific to the time-based problem statements.
  - Analyzing the historical dataset and its patterns
  - Understanding and matching the current situation with patterns derived from the previous stage.
  - Understanding the factor or factors influencing certain variable(s) in different periods.
- With help of “Time Series” we can prepare numerous time-based analyses and results.
  - Forecasting
  - Segmentation
  - Classification
  - Descriptive analysis`
  - Intervention analysis

Components of Time Series

- **Trend**: In which there is no fixed interval and any divergence within the given dataset is a continuous timeline. The trend would be Negative or Positive or Null Trend
- **Seasonality**: In which regular or fixed interval shifts within the dataset in a continuous timeline. Would be bell curve or saw tooth
- **Cyclical**: In which there is no fixed interval, uncertainty in movement and its pattern
- **Irregularity (Noise)**: Unexpected situations/events/scenarios and spikes in a short time span.

Components of Time Series

	Trend	Seasonality	Cyclical	Irregularity
Time	Fixed Time Interval	Fixed Time Interval	Not Fixed Time Interval	Not Fixed Time Interval
Duration	Long and Short Term	Short Term	Long and Short Term	Regular/Irregular
Visualization				
Nature - I	Gradual	Swings between Up or Down	Repeating Up and Down	Errored or High Fluctuation
Nature – II	Upward/Down Trend	Pattern repeatable	No fixed period	Short and Not repeatable
Prediction Capability	Predictable	Predictable	Challenging	Challenging
Market Model				Highly random/Unforeseen Events – along with white noise.

Limitations of Time Series Analysis

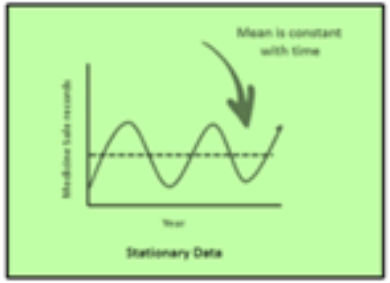
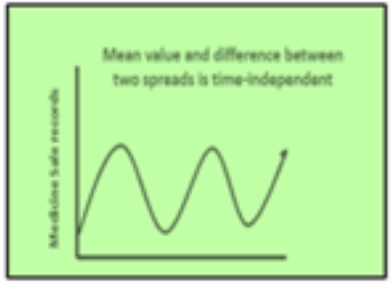
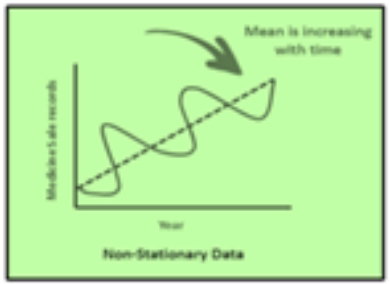
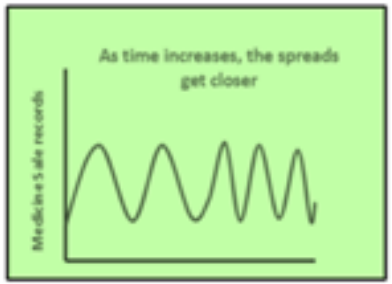
- Time series has the below-mentioned limitations, we have to take care of those during our analysis,
  - Similar to other models, the missing values are not supported by TSA
  - The data points must be linear in their relationship.
  - Data transformations are mandatory, so a little expensive.
  - Models mostly work on Uni-variate data.

Data Types of Time Series Analysis

- There are two major types:
- Stationary
- Non- Stationary

- **Stationary**: A dataset should follow the below thumb rules, without having Trend, Seasonality, Cyclical, and Irregularity component of time series
- The MEAN value of them should be completely constant in the data during the analysis
- The VARIANCE should be constant with respect to the time-frame
- The COVARIANCE measures the relationship between two variables.
- **Non- Stationar**y: This is just the opposite of Stationary.

Data Types of Time Series Analysis

	MEAN	Variance	Covariance
Stationary			
Non-Stationary			

Methods to check Stationarity

- During the TSA model preparation workflow, we must access if the given dataset is Stationary or NOT.
- **Statistical Test:** There are two tests available to test if the dataset is Stationary or NOT.
  - **Augmented Dickey-Fuller (ADF)** Test or Unit Root Test: The ADF test is the most popular statistical test and with the following assumptions.
    - Null Hypothesis (H0): Series is non-stationary
    - Alternate Hypothesis (HA): Series is stationary
      - p-value >0.05 Fail to reject (H0)
      - p-value <= 0.05 Accept (H1)
  - **Kwiatkowski–Phillips–Schmidt–Shin (KPSS):** these tests are used for testing a NULL Hypothesis (HO), that will perceive the time-series, as stationary around a deterministic trend against the alternative of a unit root. Since TSA looking for Stationary Data for its further analysis, we have to make sure that the dataset should be stationary.

Converting Non-stationary into stationary

- There are three major methods available for this conversion.
- **Detrending:** It involves removing the trend effects from the given dataset and showing only the differences in values from the trend. it always allows the cyclical patterns to be identified.
- **Differencing:** This is a simple transformation of the series into a new time series, which we use to remove the series dependence on time and stabilize the mean of the time series, so trend and seasonality are reduced during this transformation.

$$Yt = Yt - Yt\text{-}1$$

Yt=Value with time

- **Transformation:** This includes three different methods they are Power Transform, Square Root, and Log Transfer., most commonly used one is Log Transfer.

Converting Non-stationary into stationary

- The commonly used time series method is Moving Average. This method is slick with random short-term variations. Relatively associated with the components of time series.
- **The Moving Average (MA) (Or) Rolling Mean:** In which MA has calculated by taking averaging data of the time-series, within k periods.
- Let's see the types of moving averages:
  - Simple Moving Average (SMA),
  - Cumulative Moving Average (CMA)
  - Exponential Moving Average (EMA)

Converting Non-stationary into stationary

- **Simple Moving Average (SMA)**
- The SMA is the unweighted mean of the previous M or N points. The selection of sliding window data points, depending on the amount of smoothing is preferred since increasing the value of M or N, improves the smoothing at the expense of accuracy.

- **Cumulative Moving Average (CMA)**
- The CMA is the unweighted mean of past values, till the current time.

- **Exponential Moving Average (EMA)**
- EMA is mainly used to identify trends and to filter out noise. The weight of elements is decreased gradually over time. This means It gives weight to recent data points, not historical ones. Compared with SMA, the EMA is faster to change and more sensitive.

Converting Non-stationary into stationary

- **Simple Moving Average (SMA)**
- The SMA is the unweighted mean of the previous M or N points. The selection of sliding window data points, depending on the amount of smoothing is preferred since increasing the value of M or N, improves the smoothing at the expense of accuracy.

- **Cumulative Moving Average (CMA)**
- The CMA is the unweighted mean of past values, till the current time.

- **Exponential Moving Average (EMA)**
- EMA is mainly used to identify trends and to filter out noise. The weight of elements is decreased gradually over time. This means It gives weight to recent data points, not historical ones. Compared with SMA, the EMA is faster to change and more sensitive.

Decomposition of Time Series

- **Additive Model**
- Observation = Trend + Seasonality + Error
- If Seasonality is constant

- **Multiplicative Model**
- Observation = Trend * Seasonality * Error
- If Seasonality is variant

Decomposition of Time Series

- **Interpretation - Forecasting**
- Sales = Trend + Seasonality + Error
  - Trend → Business growth
  - Seasonality → weather
  - Error → theft / calamity

References

- Introduction to Time Series Analysis in Machine learning | Analytics Steps
- Time Series (oracle.com)
- Time Series Analysis in Python: An Introduction | by Will Koehrsen | Towards Data Science
- Time Series Analysis for Machine Learning | by Mauro Di Pietro | Towards Data Science

THANK YOU