Time Series Frameworks

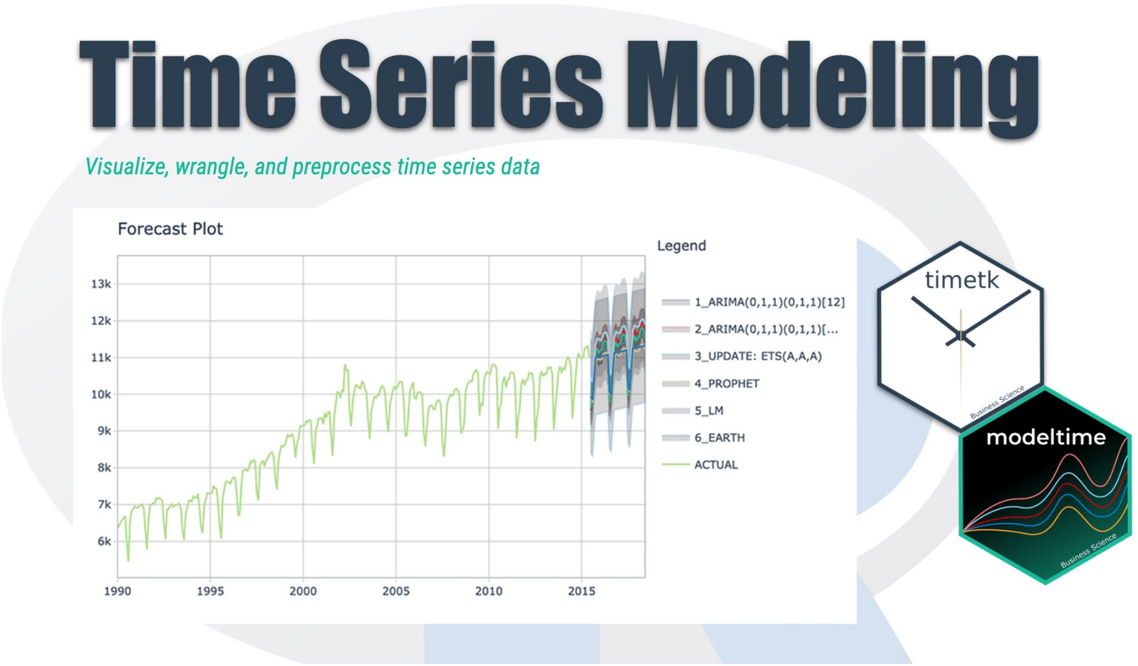
Fbprophet & Cesium

Agenda

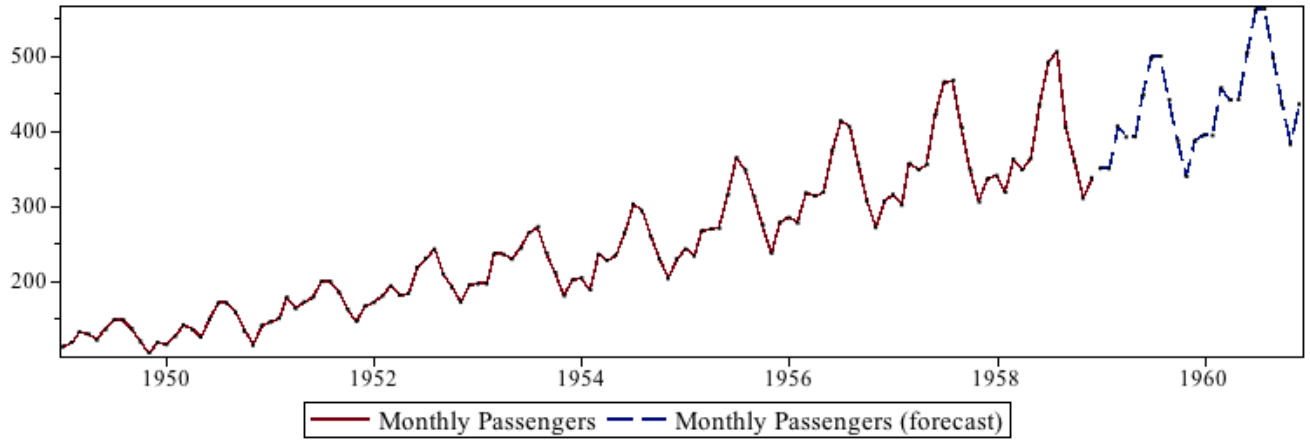
- Introduction to Time Series
- Predictions Time Series
- Frameworks
- FbProphet
- Cesium

Time Series Introduction

- We're going to discover how it differs from other types of data you've previously encountered and why
- Time series is a sequence of information which attaches a time period to each value
- The value can be pretty much anything measurable.
- It depends on time in some way, like prices, humidity or number of people.
- As long as the values we record are unambiguous, any medium could be measured with Time series.
- There aren't any limitations regarding the total time span of our Time series.
- It could be a minute, a day, a month or even a century.
- All we need is a starting and an ending point.



Time Series Basics



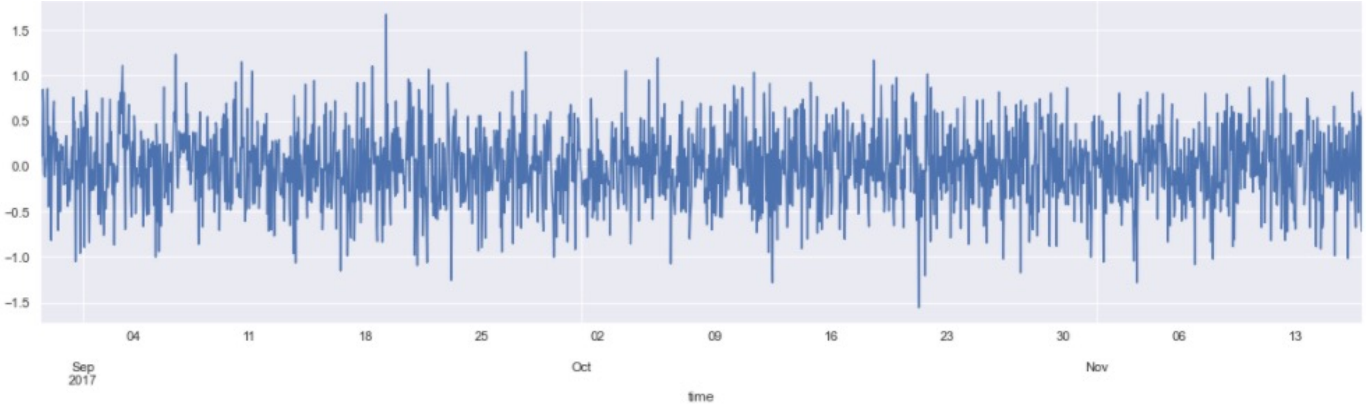
- Chronological Data
- Cannot be shuffled
- Each row indicate specific time record
- Train – Test split happens chronologically
- Data is analyzed univariately and multivariate
- Nature of the data represents if it can be predicted or not

time	IOT_Sensor_Reading	Error_Present	Sensor_2	Sensor_Value
2017-08-29 11:00:00	-0.015871	0.353986	-0.787655	0.008144
2017-08-29 12:00:00	-0.101576	0.353986	-0.787655	-0.029860
2017-08-29 13:00:00	-0.118241	0.353986	-0.787655	-0.021717
2017-08-29 14:00:00	-0.214262	0.353986	-0.787655	0.008144
2017-08-29 15:00:00	-0.249972	0.353986	-0.787655	-0.108583

Data Examining & Preprocessing

- Reading Data using Pandas – describe, head
- Check for null values
- Line plot for each feature
- Convert Date from String to Date time feature
- Setting the desired frequency
- Handling missing Values
- QQ plots

time	IOT_Sensor_Reading	Error_Present	Sensor_2	Sensor_Value
2017-08-29 11:00:00	-0.015871	0.353986	-0.787655	0.008144
2017-08-29 12:00:00	-0.101576	0.353986	-0.787655	-0.029860
2017-08-29 13:00:00	-0.118241	0.353986	-0.787655	-0.021717
2017-08-29 14:00:00	-0.214262	0.353986	-0.787655	0.008144
2017-08-29 15:00:00	-0.249972	0.353986	-0.787655	-0.108583



Predictions on Time Series

- Nature of Time Series

  1) Seasonality : Seasonality is a simple term that means while predicting a time series data there are some months in a particular domain where the output value is at a peak as compared to other months. for example if you observe the data of tours and travels companies of past 3 years then you can see that in November and December the distribution will be very high due to holiday season and festival season. So while forecasting time series data we need to capture this seasonality.

  2) Trend: The trend is also one of the important factors which describe that there is certainly increasing or decreasing trend time series, which actually means the value of organization or sales over a period of time and seasonality is increasing or decreasing.

  3) Unexpected Events: Unexpected events mean some dynamic changes occur in an organization, or in the market which cannot be captured. for example a current pandemic we are suffering from, and if you observe the Sensex or nifty chart there is a huge decrease in stock price which is an unexpected event that occurs in the surrounding.

- Time series are widely used for non-stationary data, like economic, weather, stock price, and retail sales

Predictions on Time Series

- Forecasting data using time-series analysis comprises the use of some significant model to forecast future conclusions on the basis of known past outcomes.
- An objective of time series analysis is to explore and understand patterns in changes over time where these patterns signifies the components of a time series including trends, cycles, and irregular movements.
- When such components reside in a time series, the data model must be considered for these patterns for generating accurate forecasts, such as future sales, GDP, and global temperatures.

ML Models For Time-Series Forecasting

- ARIMA Model
- Autoregressive
- ARCH / GARCH
- Deep Learning
- And many other models

Framework buildup

- The traditional development process of time series analysis is complex and time consuming. Framework development process better, faster and more efficient time series analysis.
- Classical time series forecasting techniques build on stats models which requires lots of effort to tune models and expect in data and industry.
- The person has to tune the parameters of the method with regards to the specific problem when a forecasting model doesn’t perform as expected.
- Tuning these methods requires a thorough understanding of how the underlying time series models work. It’s difficult for some organizations to handling those forecasting without data science teams
- And it might seem doesn’t profitable for an organization to have a bunch of expects on board if there is no need a build a complex forecasting platform or other services.

FbPhophet

- Facebook developed an open sourcing Prophet, a forecasting tool available in both Python and R.
- It provides intuitive parameters which are easy to tune. Even someone who lacks deep expertise in time-series forecasting models can use this to generate meaningful predictions for a variety of problems in business scenarios.
- Two main themes in the practice of creating a variety of business forecasts:
  - Completely automatic forecasting techniques can be brittle and they are often too inflexible to incorporate useful assumptions or heuristics.
  - Analysts who can product high quality forecasts are quite rare because forecasting is a specialized data science skill requiring substantial experience.

Highlights of Fbprophet

- Very fast, since it’s built in Stan, a programming language for statistical inference written in C++.
- An additive regression model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects:
  - A piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting changepoints from the data
  - A yearly seasonal component modeled using Fourier series
  - A weekly seasonal component using dummy variables
  - A user-provided list of important holidays.
- Robust to missing data and shifts in the trend, and typically handles outliers .
- Easy procedure to tweak and adjust forecast while adding domain knowledge or business insights.

The Prophet Forecasting Model

- The Prophet uses a decomposable time series model with three main model components: trend, seasonality, and holidays. They are combined in the following equation:

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

  - g(t): piecewise linear or logistic growth curve for modeling non-periodic changes in time series
  - s(t): periodic changes (e.g. weekly/yearly seasonality)
  - h(t): effects of holidays (user provided) with irregular schedules
  - εt: error term accounts for any unusual changes not accommodated by the model
- Using time as a regressor, Prophet is trying to fit several linear and non linear functions of time as components.
- Modelling seasonality as an additive component is the same approach taken by exponential smoothing in Holt-Winters technique .
- Prophet is framing the forecasting problem as a curve-fitting exercise rather than looking explicitly at the time based dependence of each observation within a time series.

ChangePoints

- Changepoints are moments in the data where the data shifts direction
- Using new COVID-19 cases as an example, it could be due to new cases beginning to fall after hitting a peak once a vaccine is introduced. Or it could be a sudden pick up of cases when a new strain is introduced into the population and so on.
- Prophet can automatically detect change points or you can set them yourself.
- You can also adjust the power the change points have in altering the growth function and the amount of data taken into account in automatic changepoint detection.

Changepoints Modifiers

- **changepoint_range:** the default setting for the range of data points considered when identifying changepoints is the first 80% of data in the time series. We can fix this by setting changepoint_range = 1 when instantiating the model which will incorporate 100% of the data. In other situations, it may be good to keep the changepoint range at 80% or lower to ensure that the model doesn’t overfit your data and can understand the last 20% on its own.
- The **changepoint_prior_scale** and the **n_changepoints** hyperparameters allow us to adjust this. By default, **changepoint_prior_scale** it is set to 0.05, increasing this value allows the automatic detection of more change points and decreases it allows for less.
- Alternatively, we can specify a number of changepoints to detect using **n_changepoints** or list the changepoints ourselves using changepoints. Be careful with this, as too many changepoints may cause overfitting.

Cesium

- Cesium is an open source library that allows users to extract features from raw time series data, build machine learning models from these features, as well as generate predictions for new data.
- The cesium library also powers computations within the cesium web interface, which allows similar time series analyses to be performed entirely within the browser.
- Building a functioning machine learning pipeline involves much more than choosing a mathematical model for your data.
- The goal of cesium is to simplify the analysis pipeline so that scientists can spend less time solving technical computing problems and more time answering scientific questions.

Cesium

- Cesium comes with a number of out-of-the-box feature engineering workflows, such as periodogram analysis, that transforms raw time series data to pull signal from the noise.
- By also streamlining the process of fitting models and studying relationships within datasets, cesium allows researchers to iterate rapidly and quickly answer new questions that arise out of previous lines of inquiry.
- Cesium is easily shareable and reproducible, so that an entire process of discovery can be shared with and reproduced by other researchers.
- Saved cesium workflows are meant to be production-ready, meaning that comprehensive machine learning can be applied not just to data in retrospect but to live, streaming data as well.