

Time Series Forecasting (Trend Forecasting)

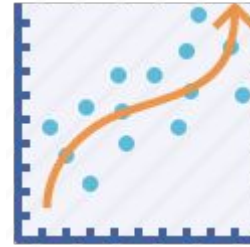




ข้อมูล time series



พัฒนาการของ time series forecasting

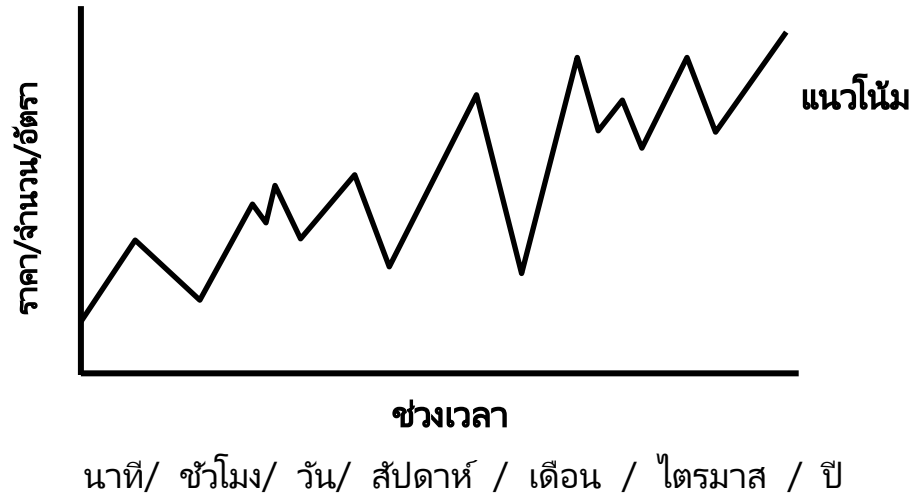


ขั้นตอนการทำ time series forecasting

ข้อมูล time series คืออะไร



Time series data = ข้อมูลที่เปลี่ยนแปลงไปตามเวลา

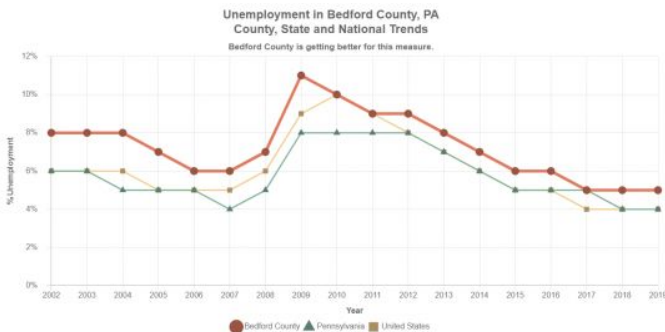


ตัวอย่าง time series data

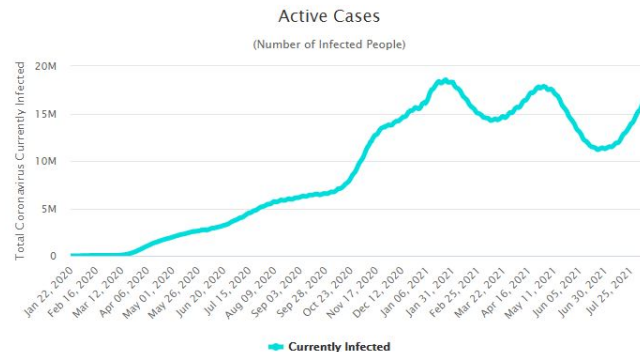
กราฟราคาหุ้น หรือ ดัชนี



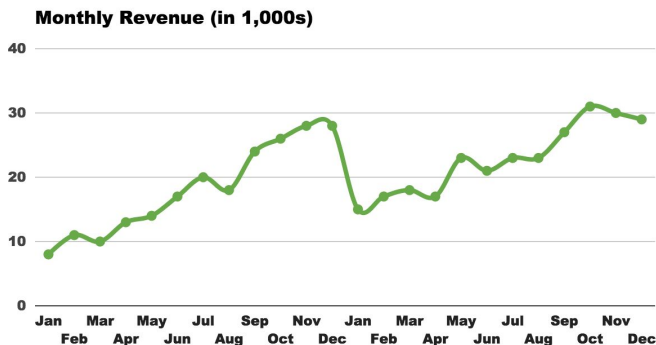
กราฟแนวโน้มแสดงอัตราการว่างงาน



กราฟจำนวนผู้ติดเชื้อ COVID-19



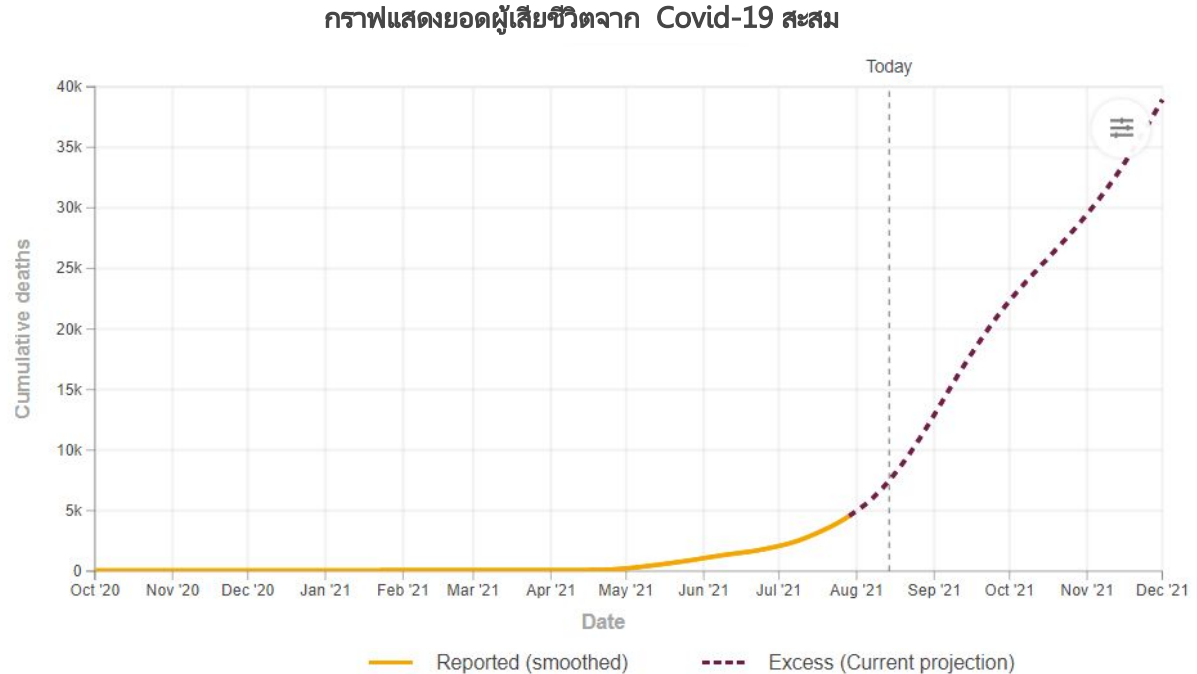
กราฟแนวโน้มรายได้ หรือค่าใช้จ่าย



ตัวอย่างการใช้ time series forecasting

Time series or trend forecasting to predict number of deaths from Covid-19

Reported deaths are the number of deaths officially reported as COVID-19. **Excess deaths** are the number of deaths estimated as attributed to COVID-19, including unreported deaths.

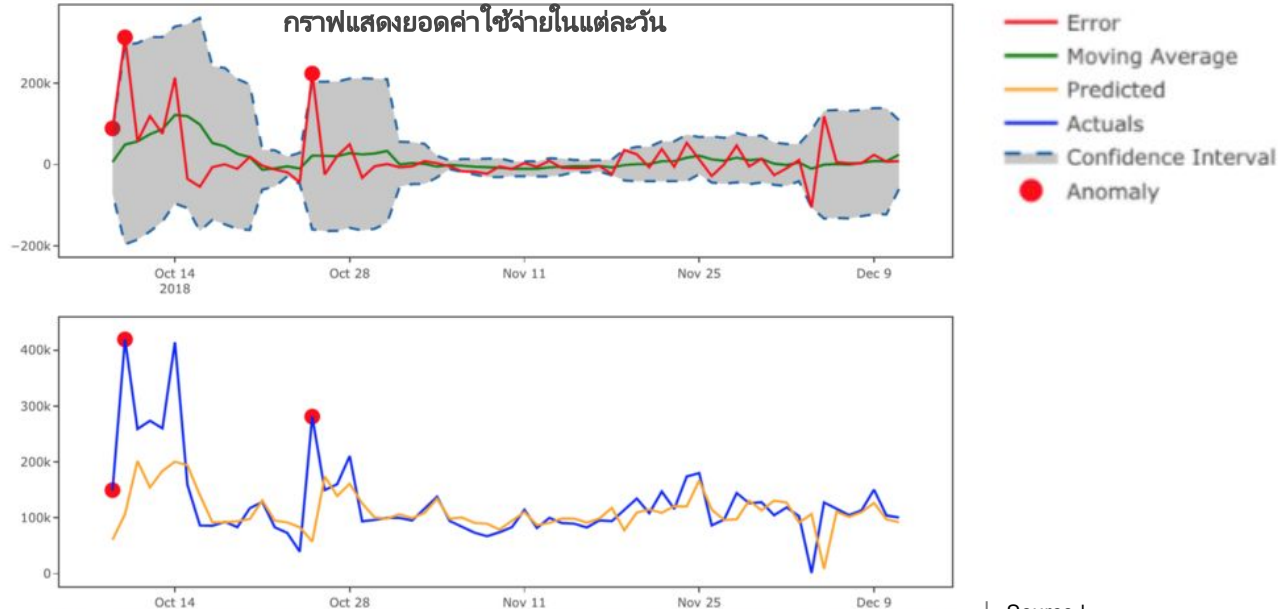


Source | <https://covid19.healthdata.org/thailand?view=cumulative-deaths&tab=trend>

ตัวอย่างการใช้ time series forecasting

Anomaly detection with time series forecasting

ไม่ใช่เพียงแค่การทำนายยอดขาย หรือราคาหุ้น time series forecasting ยังสามารถนำมาประยุกต์เพื่อหารายการผิดปกติ (anomaly detection)



Date	Actual Values	Predicted	% Difference	Severity (0-3)
2018-11-23	113732	126306.074	-9.333%	0
2018-11-22	147069	108739.168	26.062%	0
2018-11-21	108126	114978.294	-6.337%	0
2018-11-20	134514	109228.065	18.798%	0
2018-11-19	113630	77791.011	31.54%	2
2018-11-18	94102	117629.725	-25.002%	1
2018-11-17	95410	98640.56	-3.386%	0

Source | <https://towardsdatascience.com/anomaly-detection-with-time-series-forecasting-c34c6d04b24a>

ตัวอย่างการใช้ time series forecasting



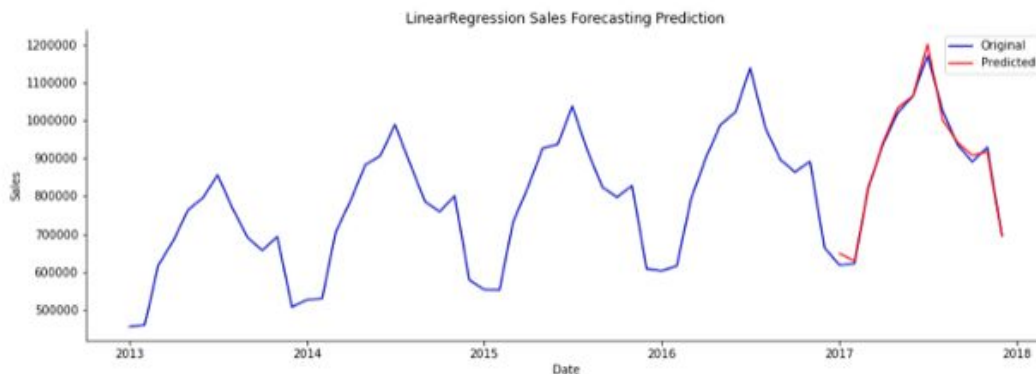
ML trend forecasting for sale or revenue

ตัวอย่างข้อมูลยอดขายรายเดือน

	date	sales
0	2013-01-01	454904
1	2013-02-01	459417
2	2013-03-01	617382
3	2013-04-01	682274
4	2013-05-01	763242

Total Monthly Sales

กราฟที่ predict เทียบกับข้อมูลจริง



ความแตกต่างระหว่าง time series และ non-time series



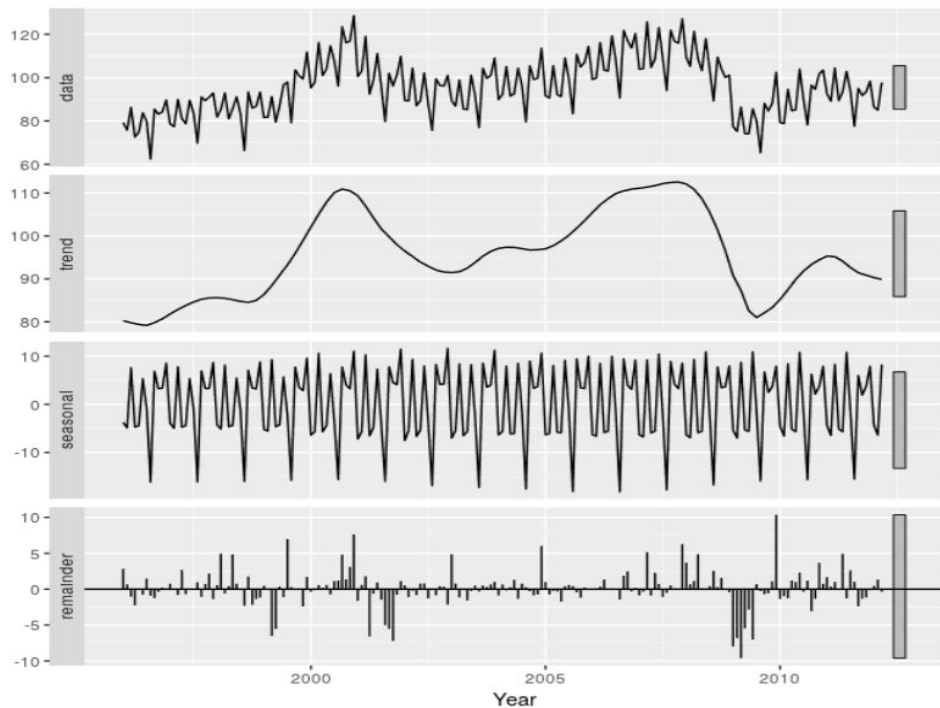
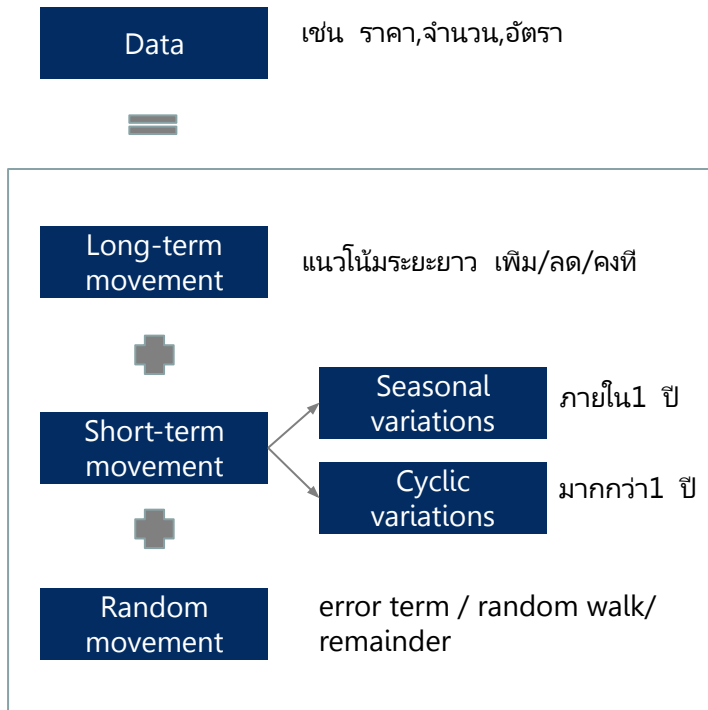
Non-time series

- จำนวนผู้พักอาศัยในคอนโด A เทียบกับจำนวนผู้พักอาศัยในคอนโด B
- จำนวนนักเรียนที่เข้าเรียนคณะ Data Science มหาวิทยาลัย Botnoi
- ราคาทองคำแท่ง น้ำหนัก 10 กิโลกรัม
- จำนวนผู้ป่วย COVID ทั่วโลก ณ เดือน ธันวาคม 2020

Time Series

- จำนวนผู้พักอาศัยในคอนโด A **ระหว่างปี 2017 - 2020**
- จำนวนนักเรียนที่จบการศึกษาในแต่ละปี **ตั้งแต่ปี 2010 ถึงปี 2020**
- ราคาทองคำ**ย้อนหลัง 6 เดือน**
- จำนวนผู้ติดเชื้อ COVID ทั่วโลก**ในแต่ละวัน ตั้งแต่เดือน มกราคม ถึงเดือน ธันวาคม**

องค์ประกอบของข้อมูล time series

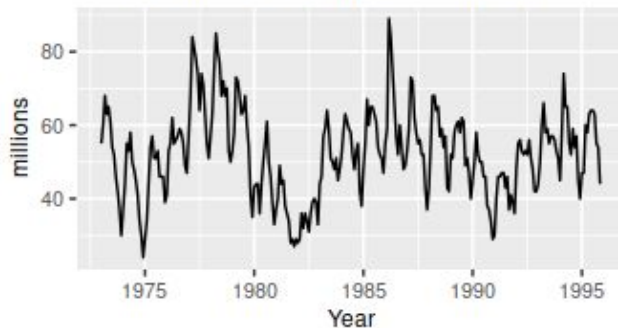


Reference : <https://blog.datath.com/time-series-forecasting/>

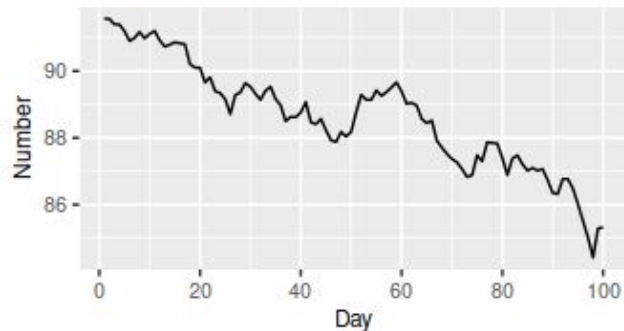
ตัวอย่างการวิเคราะห์องค์ประกอบของ time series

Trend —
Cycle ✓
Seasonal ✓

Sales of new one-family houses, USA



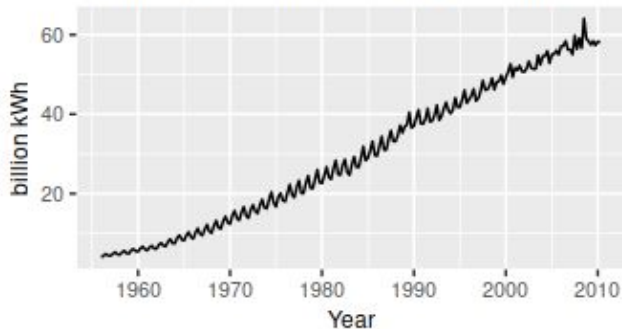
US treasury bill contracts



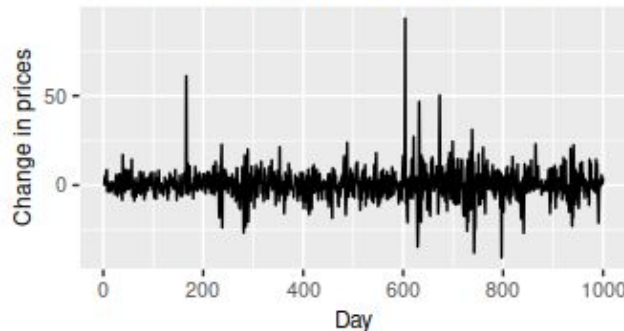
Trend ↓
Cycle ✗
Seasonal ✗

Trend ↑
Cycle ✗
Seasonal ✓

Australian quarterly electricity production



Google daily changes in closing stock price



Trend —
Cycle ✗
Seasonal ✗

Traditional vs current time series forecasting

ความแตกต่างของการทำนายยอดขายระหว่าง แบบดั้งเดิม และปัจจุบัน

		1	2
เดือน	ยอดขายจริง	ทำนายจาก ยอดขาย ย้อนหลังเฉลี่ย 3 วัน	ยอดขายทำนายด้วย linear regression
1	1000		
2	1100		
3	1050		
4	1250		
5	1280		
6	1190		
7	1230		
8	1200		
9	1240		
10	1150		
11		1197	1267
12		1196	1284
13		1181	1302

1 วิธีทำนายยอดขายแบบดั้งเดิม

เช่น การทำนายยอดขายด้วยการใช้ราคา
ย้อนหลังเฉลี่ย 3 วัน

$$\text{ราคาวันที่ 11} = (1200 + 1240 + 1150) / 3 = 1197$$

2 วิธีทำนายยอดขายแบบปัจจุบัน

เช่น การทำนายยอดขายด้วยการใช้
linear regression

$$\text{ราคาวันที่ 11} = 17.758 (11) + 1071.3$$

เนื่องจากวิธีการคำนวณต่างกัน ผลลัพธ์จึงต่างกัน

วิธีการทำนาย

Naive approach

Simple average

Weight moving average

Single exponential smoothing

Holt's linear trend method

Holt's Winter seasonal method

ARIMA*

ทำนายราคาหุ้นโดยใช้

ราคาเมื่อวาน

ค่าเฉลี่ยของช่วงเวลาในอดีต

ค่าเฉลี่ยของช่วงเวลาในอดีต ถ่วงน้ำหนักด้วยเวลา (ยิ่งใกล้ยิ่งให้น้ำหนักเยอะ)

ค่าเฉลี่ยของช่วงเวลาในอดีต ถ่วงน้ำหนักด้วยเวลา (ยิ่งใกล้ยิ่งให้น้ำหนักเยอะมากๆ)

ใช้ single exponential smoothing ผสม trend

ใช้ single exponential smoothing ผสม trend และ seasonal

มองว่าราคาในแต่ละวันมีความสัมพันธ์กัน (correlation)

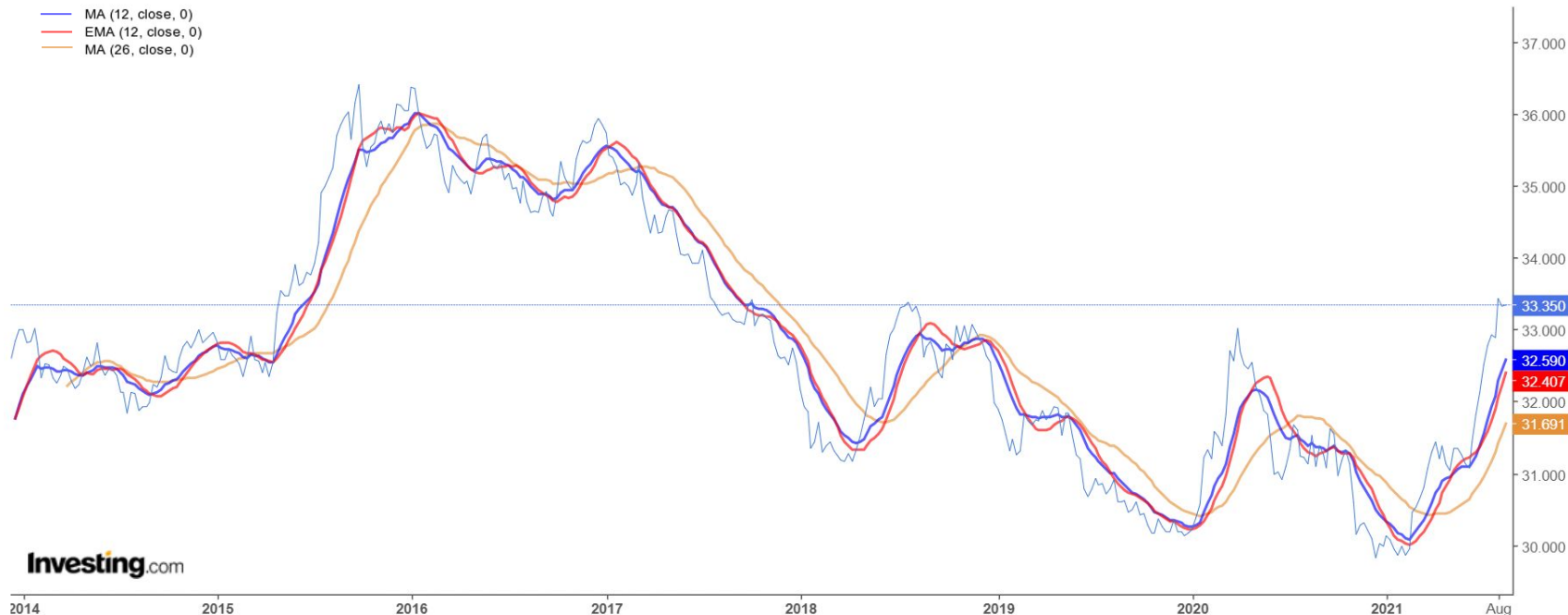
*ARIMA = Autoregressive integrated moving average

ตัวอย่างของเส้นค่าเฉลี่ยที่คำนวณจากราคาย้อนหลัง

Published on Investing.com, 21/Aug/2021 - 9:19:36 GMT, Powered by TradingView.

USD/THB, W

MA (12, close, 0)
EMA (12, close, 0)
MA (26, close, 0)

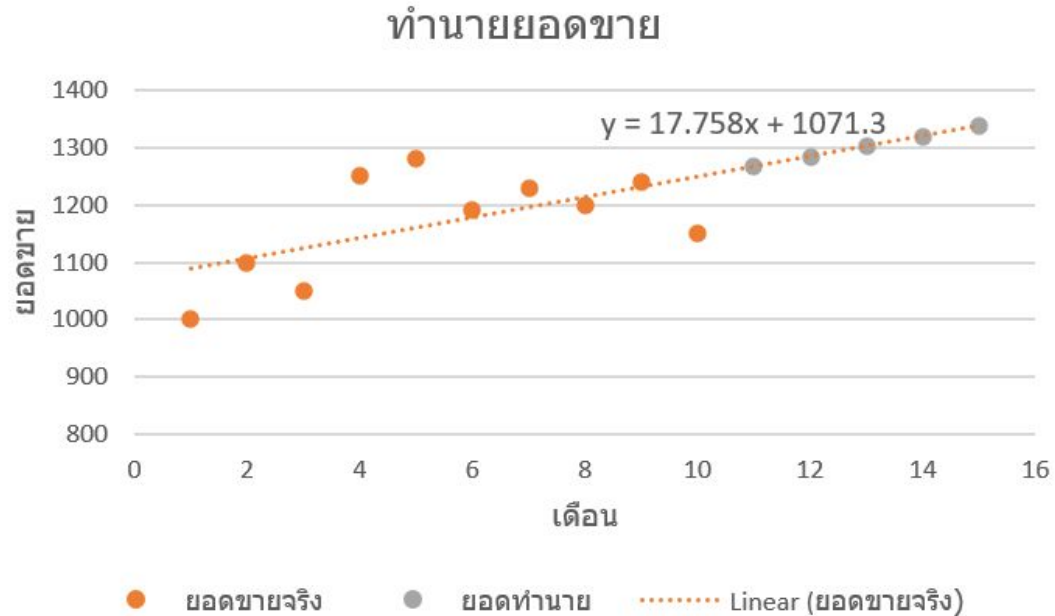


Investing.com

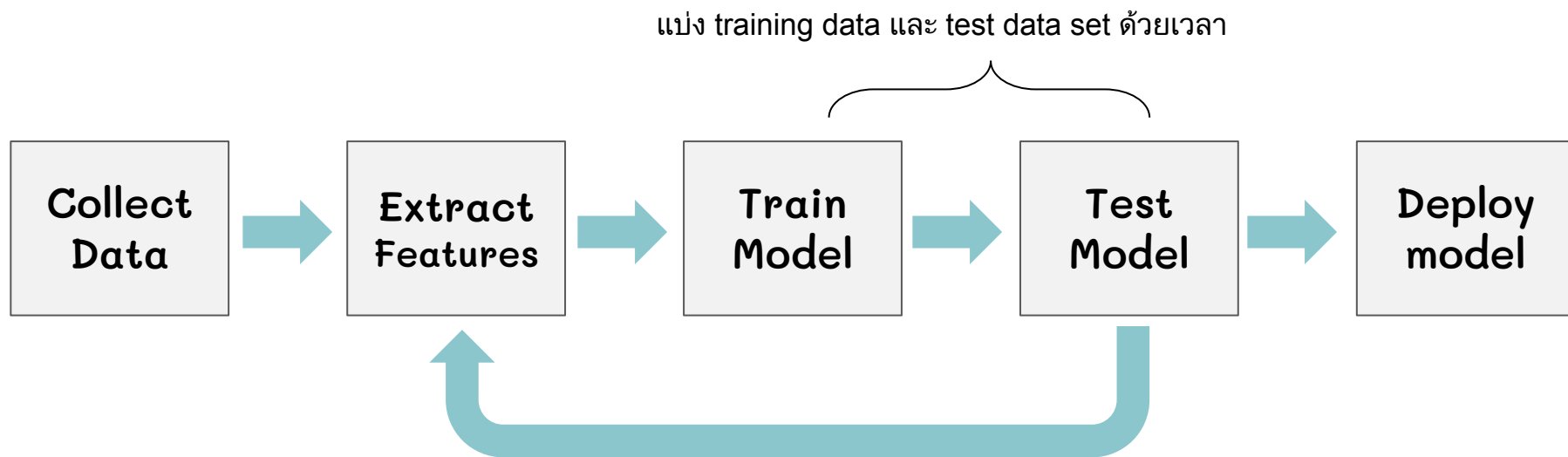
ตัวอย่างการทำนายยอดขายโดยใช้ linear regression

สมมติว่าเรามียอดขาย 10 เดือนย้อนหลัง และต้องการทำนายยอดขายในอีก 5 เดือนข้างหน้า

เดือน	ยอดขายจริง	ยอดขายทำนาย
1	1000	
2	1100	
3	1050	
4	1250	
5	1280	
6	1190	
7	1230	
8	1200	
9	1240	
10	1150	
11		1267
12		1284
13		1302
14		1320
15		1338



ขั้นตอนการทำ time series forecasting ด้วย ML



เราจะใช้โมเดลตัวไหนในการทำนายดี ?



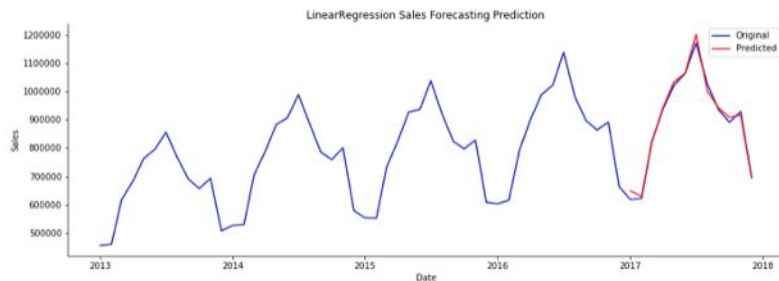
ML trend forecasting for sale or revenue

ตัวอย่างข้อมูลยอดขายรายเดือน

	date	sales
0	2013-01-01	454904
1	2013-02-01	459417
2	2013-03-01	617382
3	2013-04-01	682274
4	2013-05-01	763242

Total Monthly Sales

กราฟที่ predict เทียบกับข้อมูลจริง



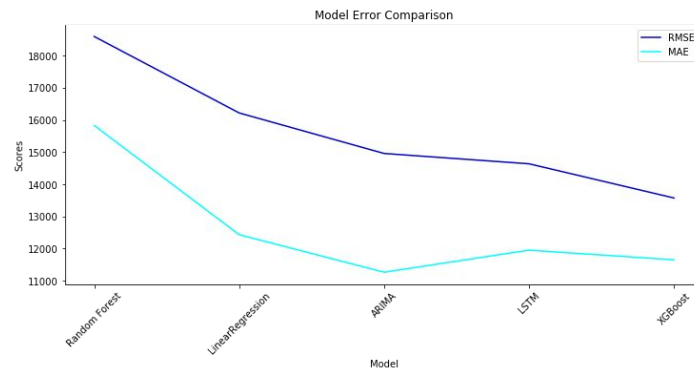
ผลเปรียบเทียบของโมเดล

	index	RMSE	MAE
0	Random Forest	18599.232966	15832.750000
1	LinearRegression	16221.040791	12433.000000
2	ARIMA	14959.893467	11265.335749
3	LSTM	14638.748350	11951.083333
4	XGBoost	13574.792632	11649.666667

MAE = Mean absolute error

RMSE = Root mean squared error

ค่ายิ่งน้อยยิ่งดี 👍



ภาคผนวก

Type of linear regression

1. Simple Linear Regression

$$\hat{y} = b_0 + b_1X_1$$

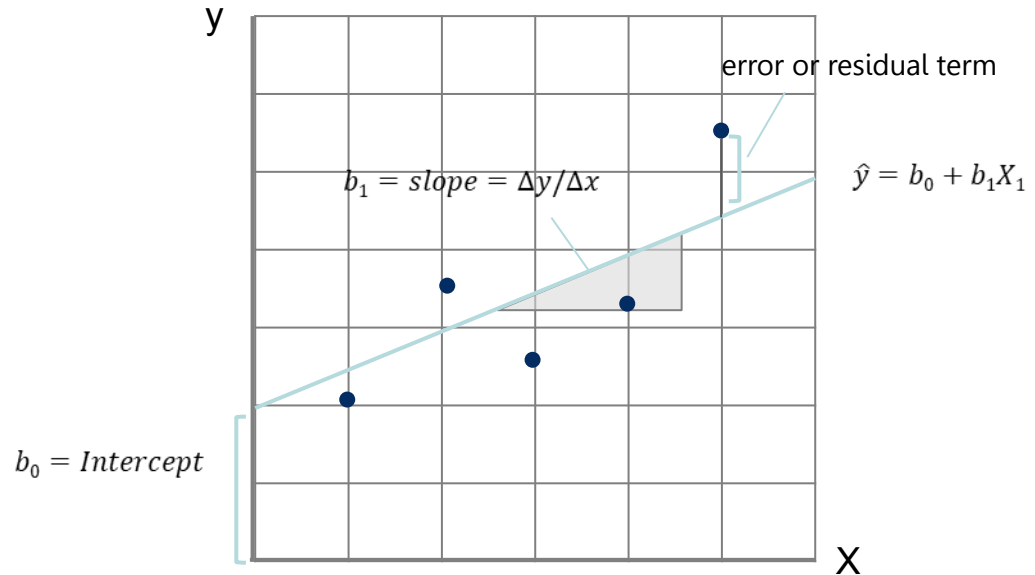
2. Multiple Linear Regression

$$\hat{y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Example of use Cases

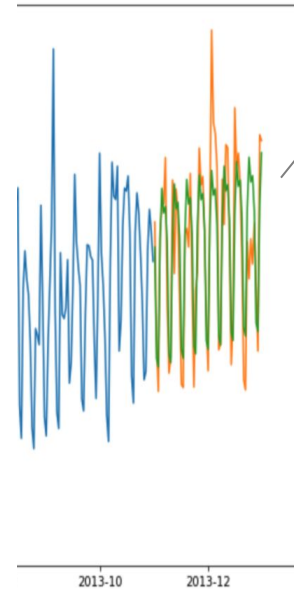
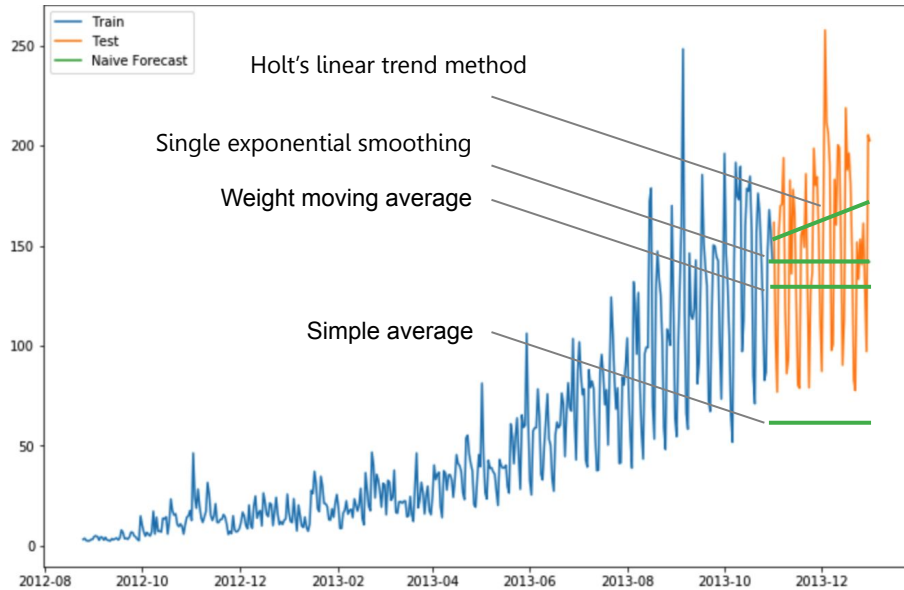
- Predict future stock price
- Predict future sales or expense

Visualizing terms

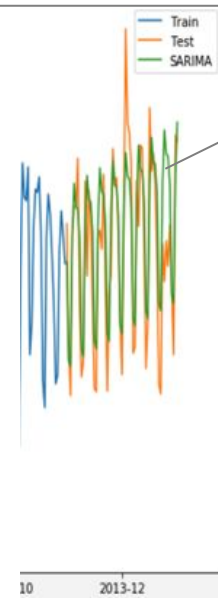


พัฒนาการของ traditional trend forecasting

เมื่อนำค่าที่พยากรณ์ในแต่ละ model มาเทียบกันจะพบว่ามีความแตกต่างกัน



Holt's Winter
seasonal
method

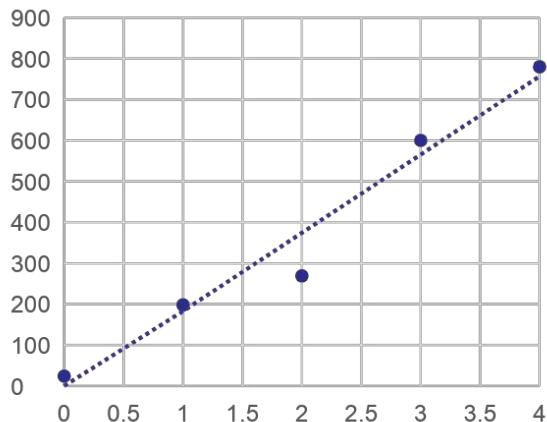


ARIMA

ตัวอย่างการคำนวณโดยใช้ linear regression

สมมติว่าเรามีราคาและพื้นที่ของห้อง

Area	Actual price	Predict Price	Error term
0	24	(8.6)	32.6
1	198	182.8	15.2
2	269	374.2	(105.2)
3	600	565.6	34.4
4	780	757	23



ผลการคำนวณ regression โดยใช้ excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.981328411
R Square	0.963005451
Adjusted R Square	0.950673935
Standard Error	68.49136199
Observations	5

ANOVA

	df	SS	MS	F	Significance F
Regression	1	366339.6	366339.6	78.09302788	0.003054117
Residual	3	14073.2	4691.066667		
Total	4	380412.8			

	Coefficients	Standard Error	t Stat	P-value
Intercept	-8.6	53.05318087	-0.162101496	0.88152873
Area	191.4	21.65887039	8.837025964	0.003054117

Linear regression (supervised learning method)

Regression algorithm

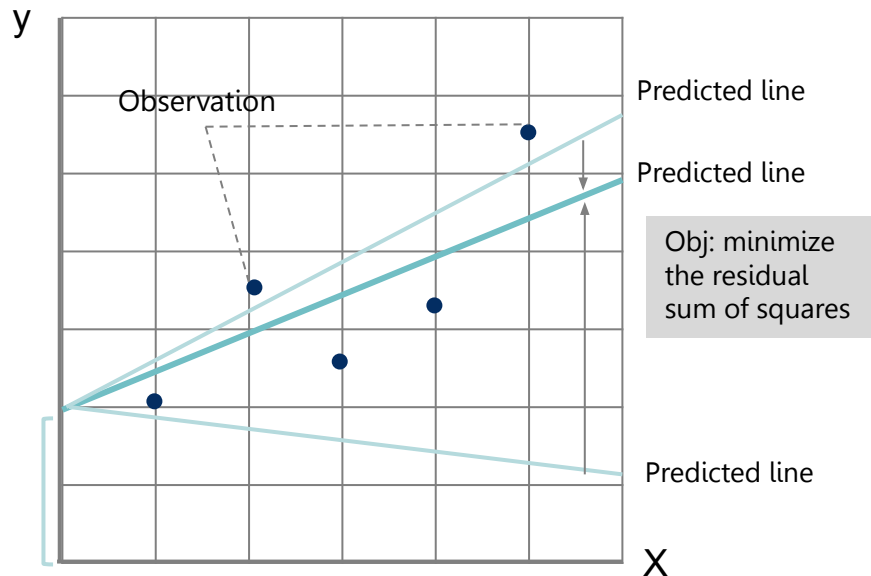
วัตถุประสงค์

ลดความห่างระหว่างข้อมูล กับเส้นที่ predict ไว้โดยรวมให้มากที่สุด (the residual sum of squares)

Programing package

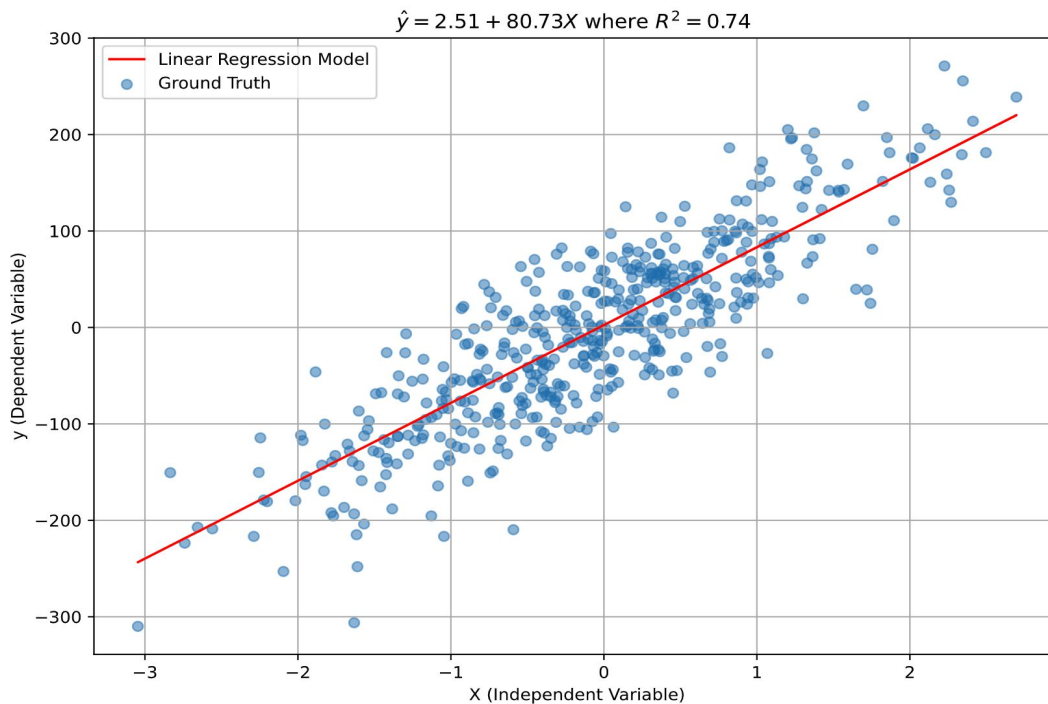
Programming Language: Python ([Link](#))

Package for Linear Regression:
Scikit-Learn ([Link](#)), Stats models ([Link](#))



Linear regression (supervised learning method)

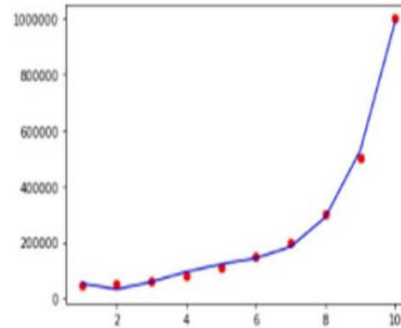
ตัวอย่างการใช้โมเดล linear regression ด้วย Python



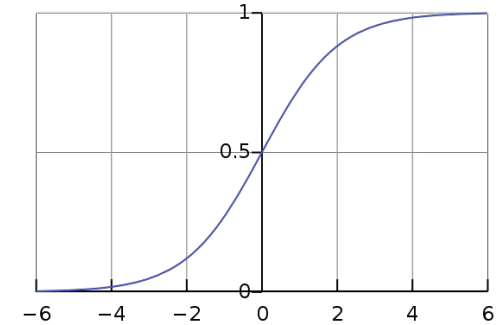
ตัวอย่าง regression models แบบอื่น ๆ

โมเดล regression ประเภทอื่น ๆ นอกจาก simple และ multiple linear regression

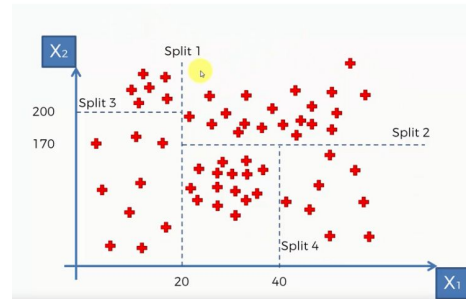
- Polynomial regression
- Logistic regression
- Ridge regression
- Lasso regression
- Bayesian linear regression
- Decision tree regression
- Random forest regression



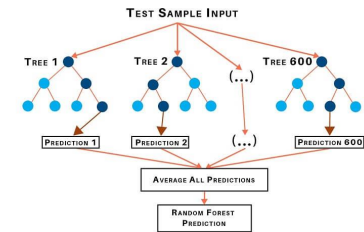
Polynomial regression



Logistic regression



Decision tree regression



Random forest regression

Source:
<https://favtutor.com/blogs/types-of-regression>

Thank you