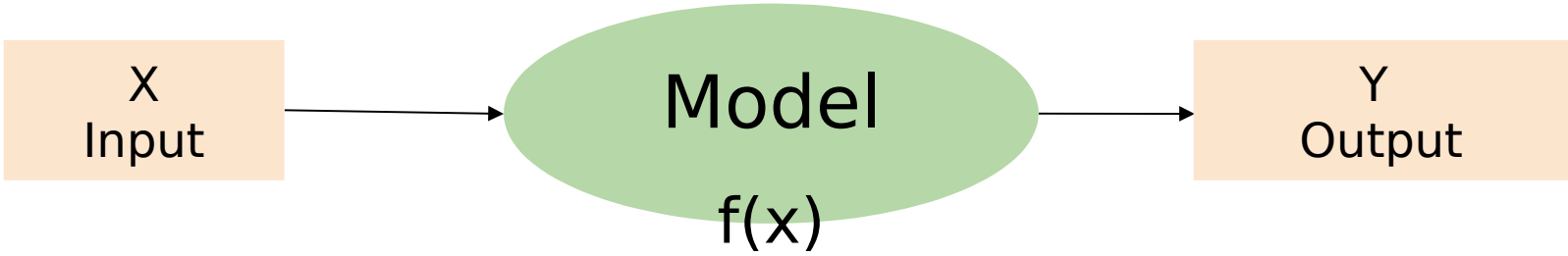
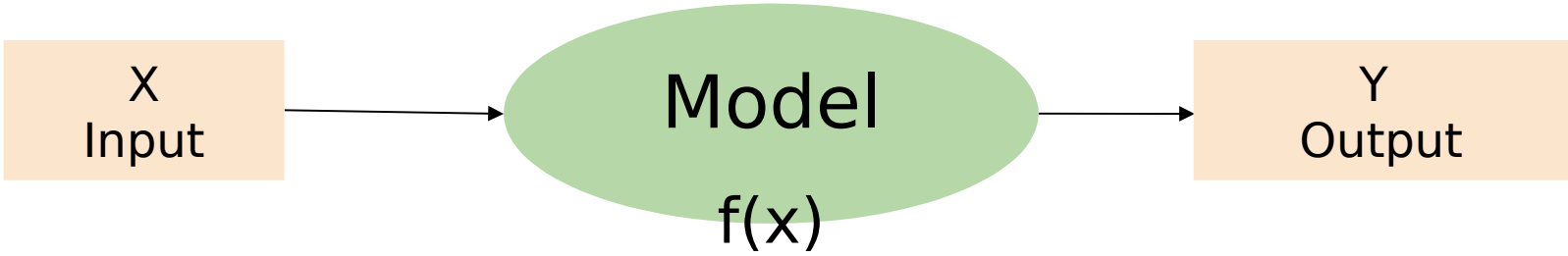
Introduction to Feature Engineering

Machine Learning Pipeline



Analytics Vidhya
Learn everything about analytics

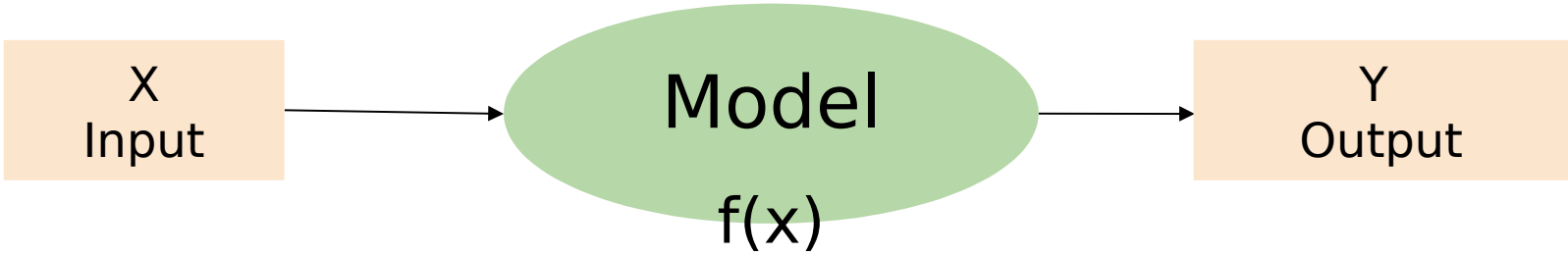
Machine Learning Pipeline



- Basic Machine Learning Models
  - kNN
  - Linear Regression
  - Logistic Regression
  - Decision Tree

Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline

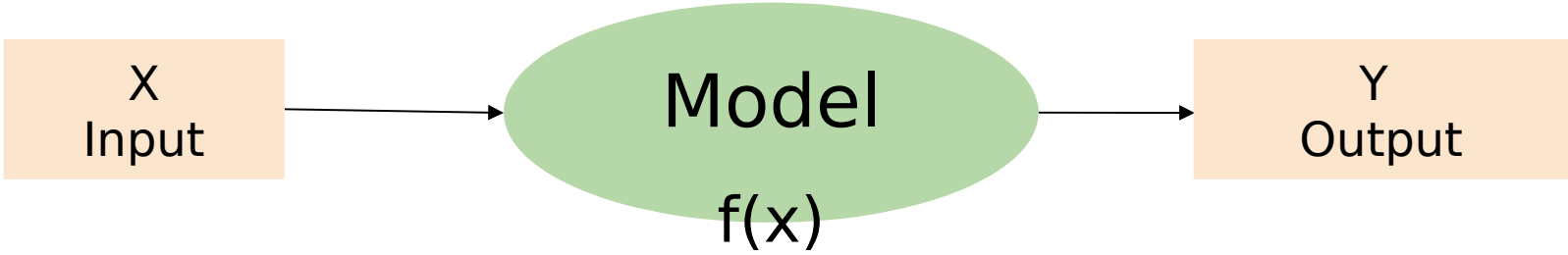


- Basic Machine Learning Models
  - kNN
  - Linear Regression
  - Logistic Regression
  - Decision Tree

$$y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+.....+\beta_nX_n$$

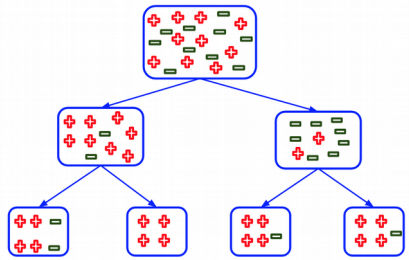
Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline



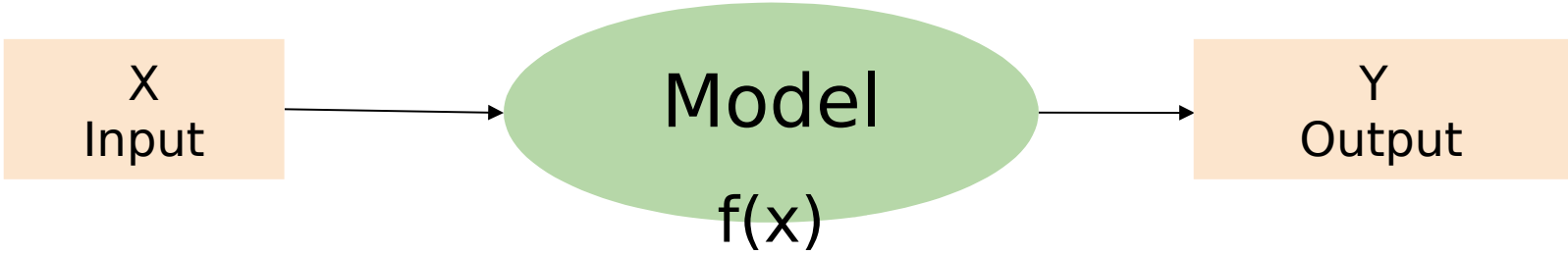
- Basic Machine Learning Models
  - kNN
  - Linear Regression
  - Logistic Regression
  - Decision Tree

$$y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+.....+\beta_nX_n$$



Analytics Vidhya
Learn everything about analytics

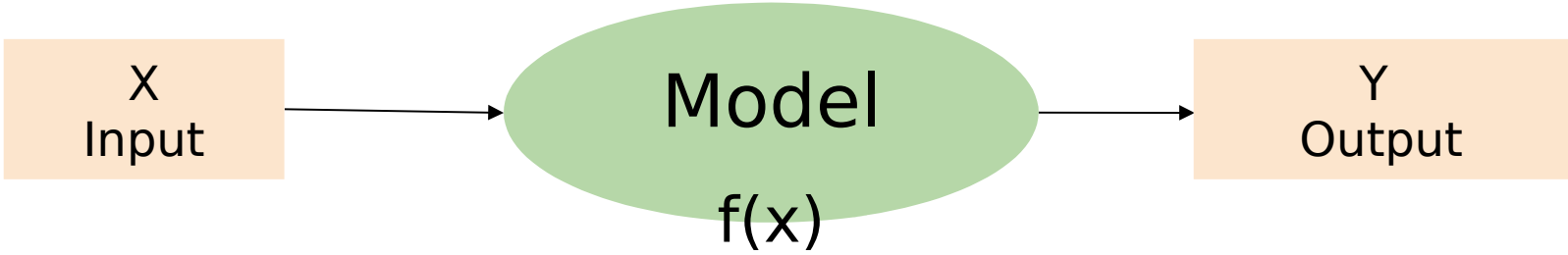
Machine Learning Pipeline



- Basic Machine Learning Models
  - kNN
  - Linear Regression
  - Logistic Regression
  - Decision Tree
- Advanced Machine Learning Models
  - Stacking
  - Blending
  - Bagging
  - Boosting

Analytics Vidhya
Learn everything about analytics

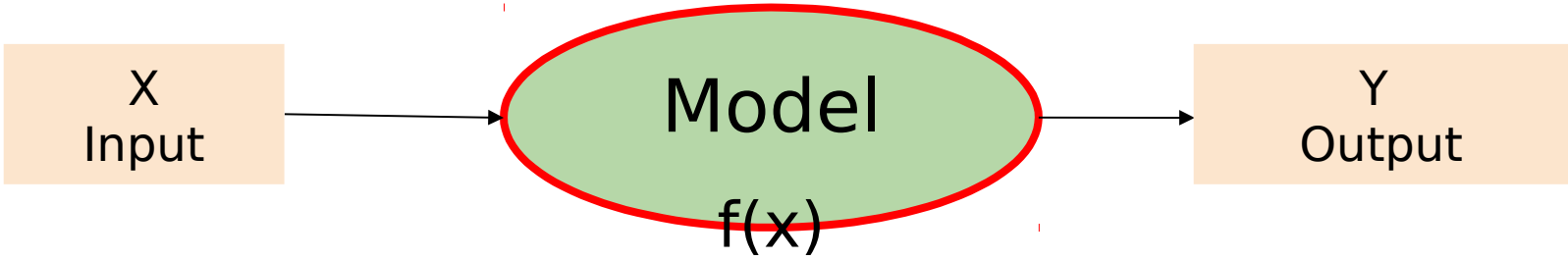
Machine Learning Pipeline



- Basic Machine Learning Models
  - kNN
  - Linear Regression
  - Logistic Regression
  - Decision Tree
- Advanced Machine Learning Models
  - Stacking
  - Blending
  - Bagging
  - Boosting
- Hyperparameter Tuning

Analytics Vidhya
Learn everything about analytics

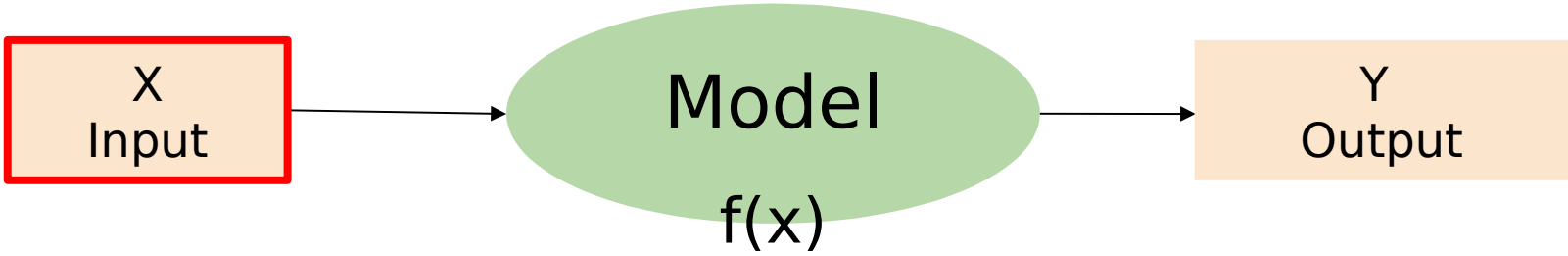
Machine Learning Pipeline



- Basic Machine Learning Models
  - kNN
  - Linear Regression
  - Logistic Regression
  - Decision Tree
- Advanced Machine Learning Models
  - Stacking
  - Blending
  - Bagging
  - Boosting
- Hyperparameter Tuning

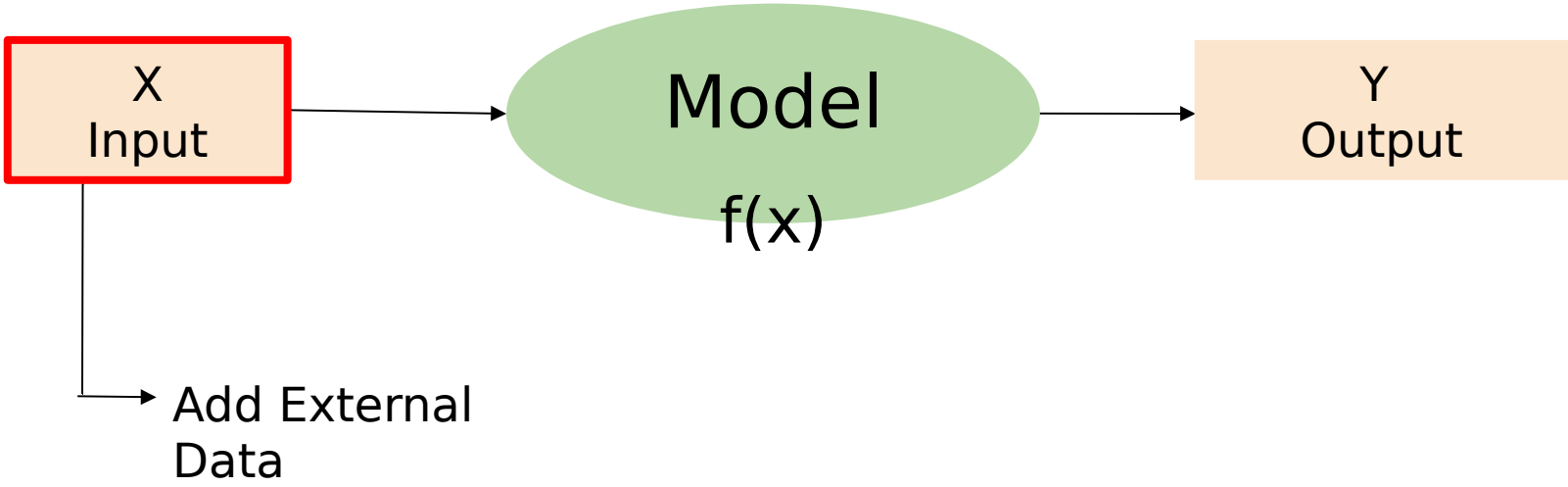
Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline



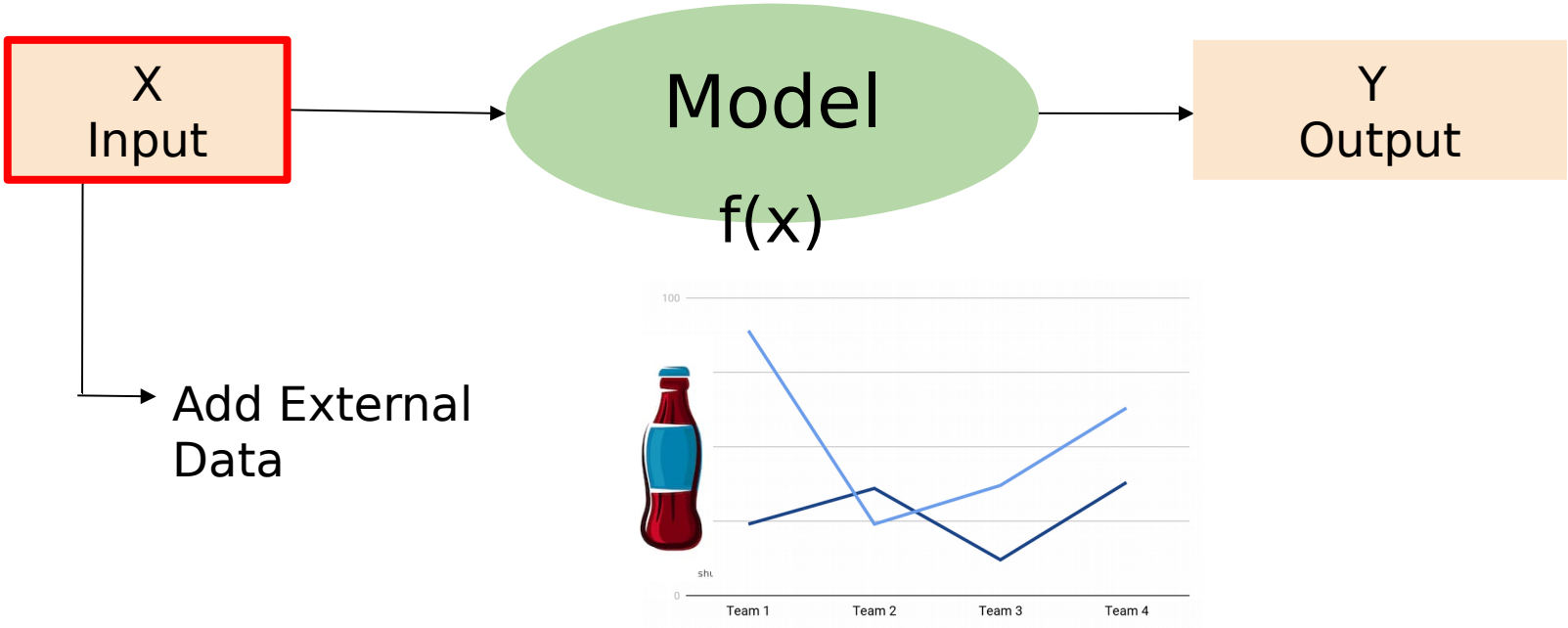
Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline



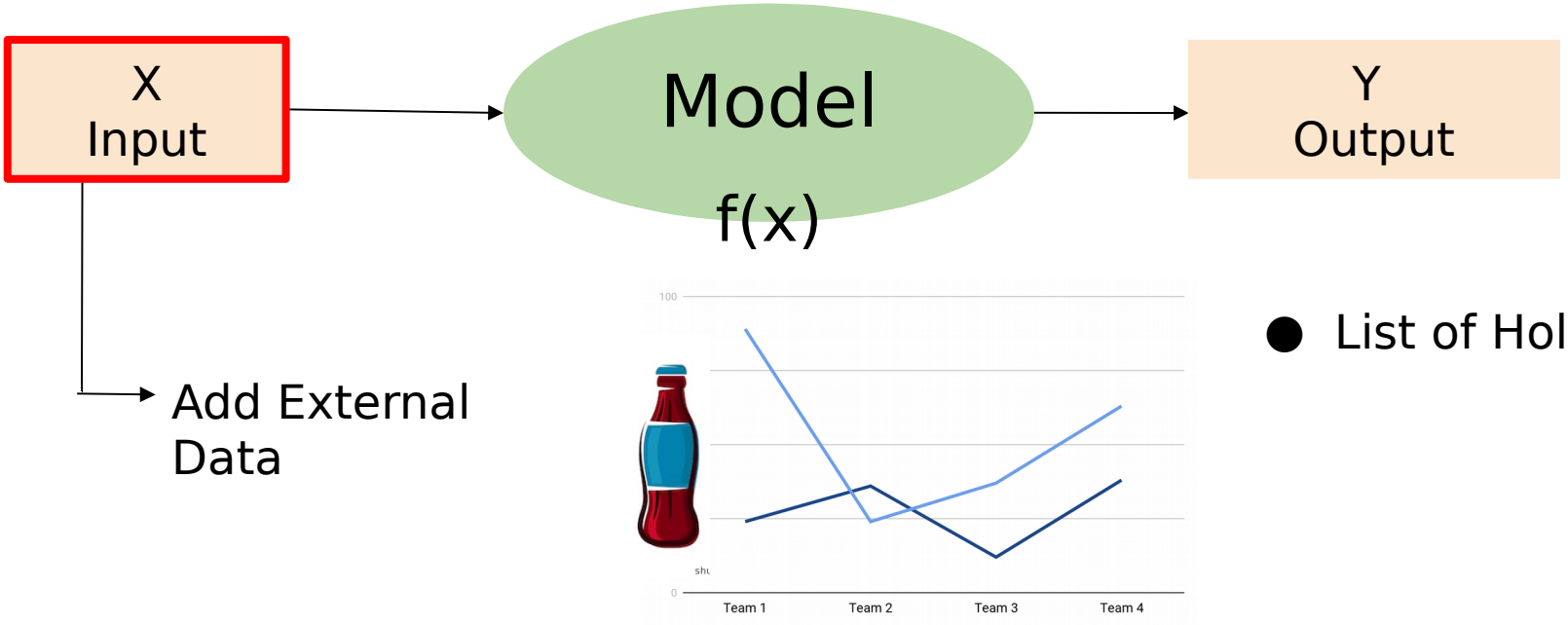
Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline



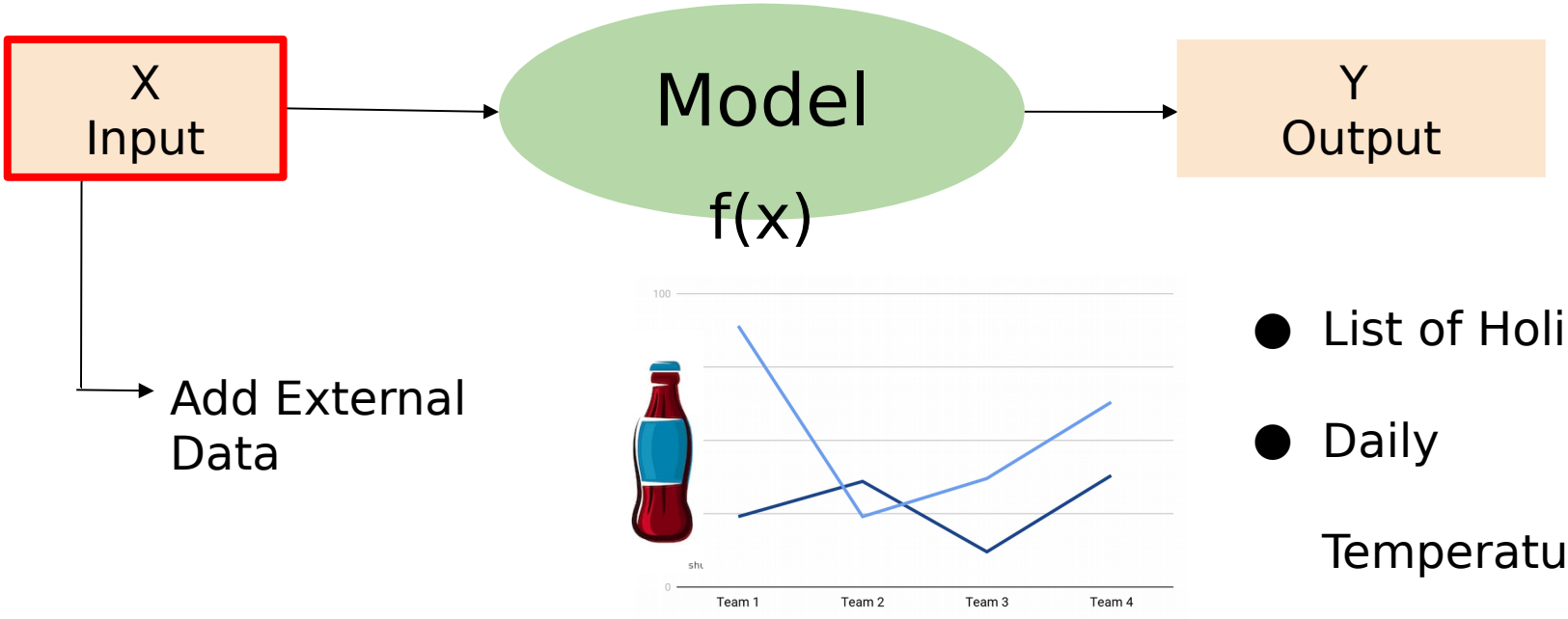
Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline



Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline

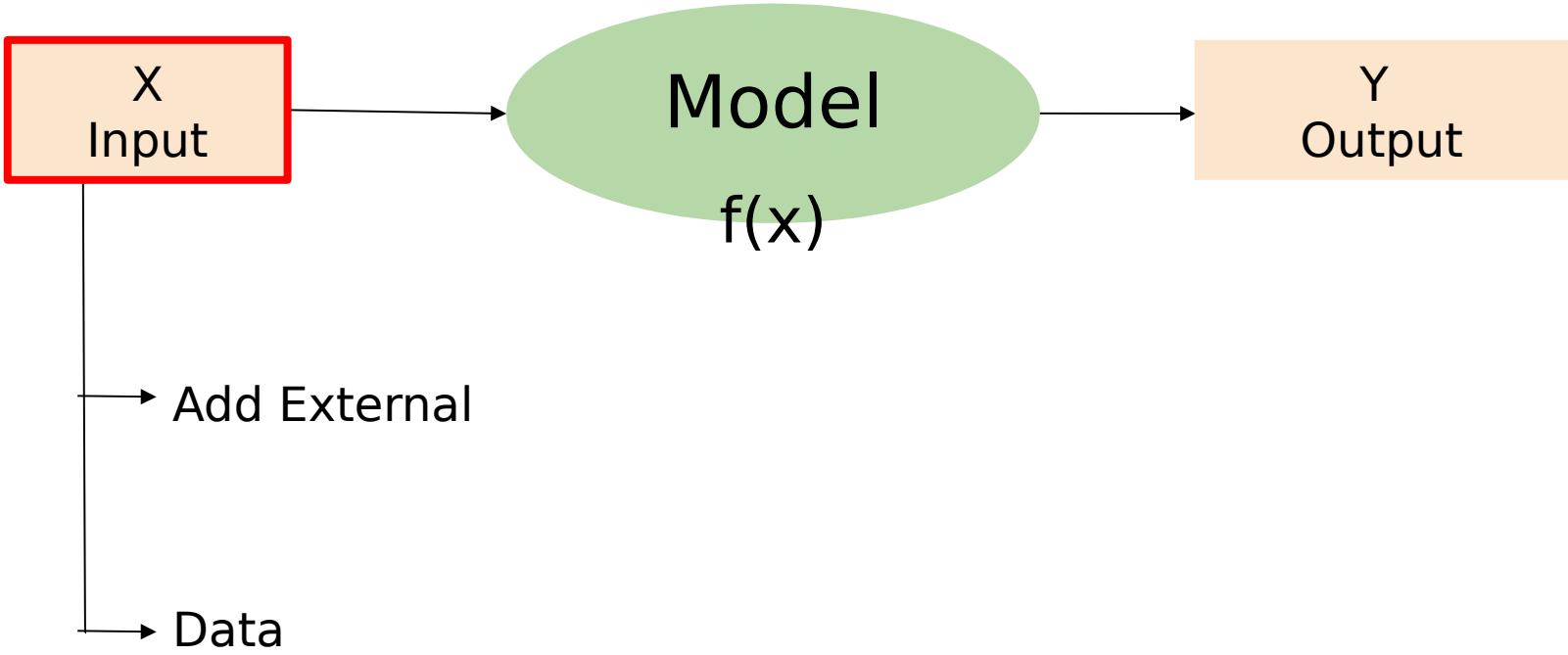


- List of Holidays
- Daily

  Temperature

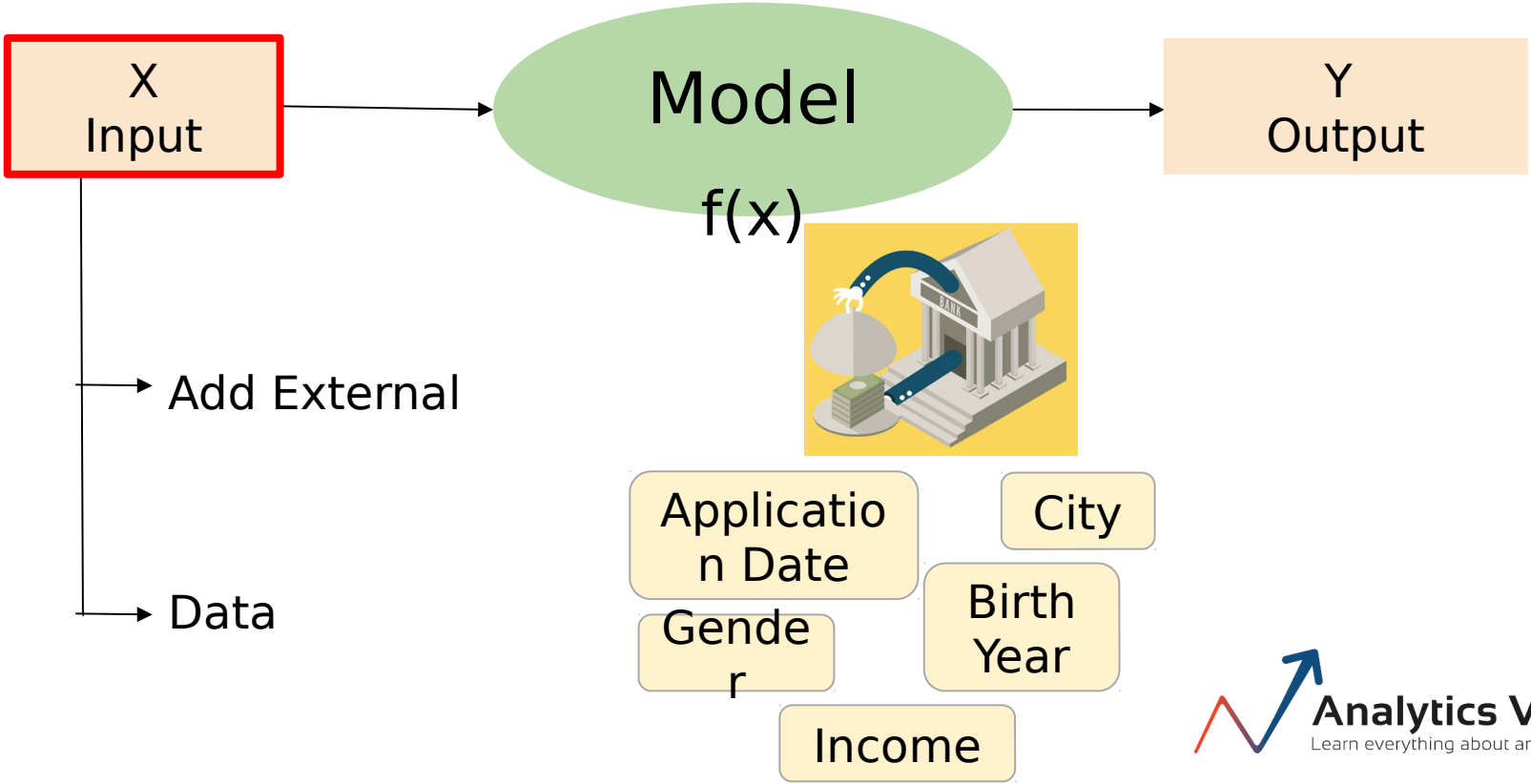
Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline

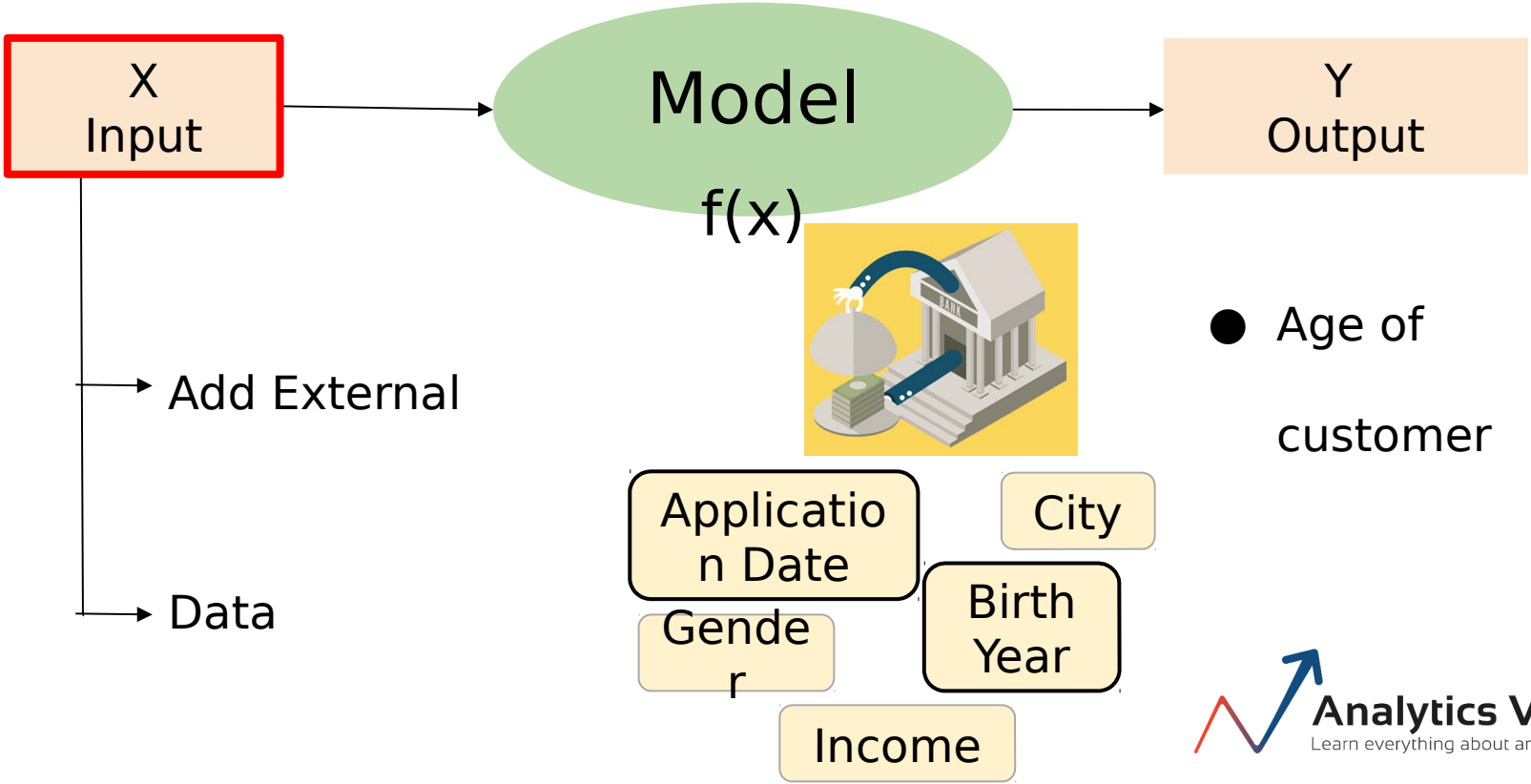


Analytics Vidhya
Learn everything about analytics

Machine Learning Pipeline



Machine Learning Pipeline



Machine Learning Pipeline

X
Input

Model

f(x)

Y
Output

Add External

Data

Applicatio n Date

City

Birth Year

Gende r

Income

- Age of customer
- Category of Income

Analytics Vidhya
Learn everything about analytics

Feature Engineering

Analytics Vidhya
Learn everything about analytics

Feature Engineering

Extracting more information/signal from the existing data

Analytics Vidhya
Learn everything about analytics

Feature Engineering

Extracting more information from the existing data

Titanic Survival Prediction

Id	Pclass	Name	Sex	Age	SibSp	Parch	Fare	Embarked	Survived
1	3	Braund, Mr. Owen Harris	male	22	1	0	7.25	S	0
2	1	Cumings, Mrs.John Bradley	female	38	1	0	71.2833	C	1
3	3	Heikkinen, Miss. Laina	female	26	0	0	7.925	S	1
4	1	Futrelle, Mrs. Jacques	female	35	1	0	53.1	S	1
5	3	Allen, Mr. William Henry	male	35	0	0	8.05	S	0
6	3	Moran, Mr. James	male		0	0	8.4583	Q	0
7	1	McCarthy, Mr. Timothy J	male	54	0	0	51.8625	S	0
8	3	Palsson, Master. Gosta Leo	male	2	3	1	21.075	S	0
9	3	Johnson, Mrs. Oscar W	female	27	0	2	11.1333	S	1
10	2	Nasser, Mrs. Nicholas	female	14	1	0	30.0708	C	1
11	3	Sandstrom, Miss. Marguerite	female	4	1	1	16.7	S	1
12	1	Bonnell, Miss. Elizabeth	female	58	0	0	26.55	S	1
13	3	Saundercock, Mr. William H	male	20	0	0	8.05	S	0
14	3	Andersson, Mr.Anders Johar	male	39	1	5	31.275	S	0

Big Mart Sales Prediction

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Size	Establish Year	Item Outlet Sales
FDA15	9.3	Low Fat	0.0160473	Dairy	249.8092	Medium	1999	3735.138
DRC01	5.92	Regular	0.0192782	Soft Drinks	48.2692	Medium	2009	443.4228
FDN15	17.5	Low Fat	0.01676	Meat	141.618	Medium	1999	2097.27
NCD19	8.93	Low Fat	0	Household	53.8614	High	1987	994.7052
FDP36	10.395	Regular	0	Baking Goods	51.4008	Medium	2009	556.6088
FDO10	13.65	Regular	0.012741	Snack Foods	57.6588	High	1987	343.5528
FDP10		Low Fat	0.1274698	Snack Foods	107.7622	Medium	1985	4022.7636
FDH17	16.2	Regular	0.0166871	Frozen Foods	96.9726		2002	1076.5986
FDU28	19.2	Regular	0.094449	Frozen Foods	187.8214		2007	4710.535
FDA03	18.5	Regular	0.0454637	Dairy	144.1102	Small	1997	2187.153
FDS46	17.6	Regular	0.0472573	Snack Foods	119.6782	Small	1997	2145.2076
FDP49	9	Regular	0.0690889	Breakfast	56.3614	Small	1997	1547.3192