GROUP - 1

Machine Learning Project Summary

TASK 2 –

Predict how many medals a country will win based on historical and current data.

Dataset –

Historical Olympic games.

Project Steps –

Outline used in this project are as follows:

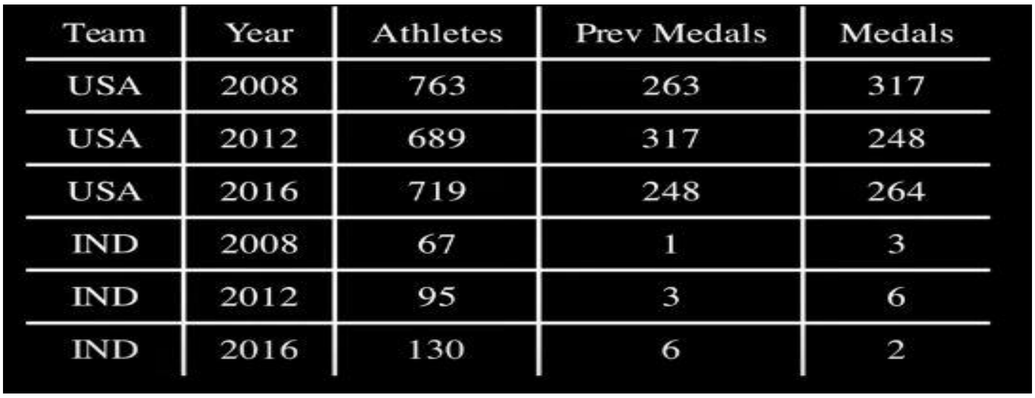
- ❖Form a hypothesis.
- ❖Find and explore the data.
- ❖(If necessary) Reshape the data to predict your target.
- ❖Clean the data for ML.
- ❖Pick an error metric.
- ❖Split your data.
- ❖Train a model.

Form a hypothesis.

A hypothesis is a statement that we can prove or disprove using data.

Hypothesis in the given dataset is to predict how many medals a country will win in the Olympics.

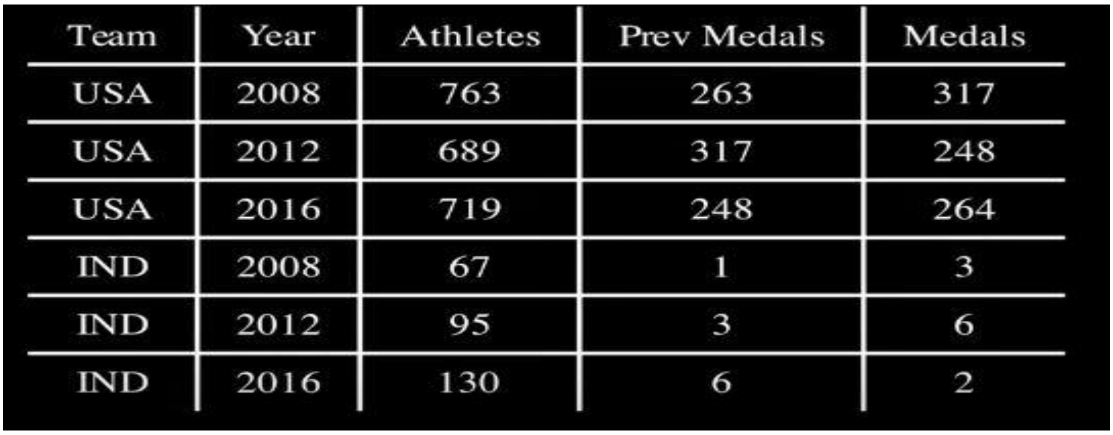
Find & explore the data.



Team	Year	Athletes	Prev Medals	Medals
USA	2008	763	263	317
USA	2012	689	317	248
USA	2016	719	248	264
IND	2008	67	1	3
IND	2012	95	3	6
IND	2016	130	6	2

We need to find the data to prove or disprove it. Above is the data from the summer Olympics dataset where each row is a single country in a Olympic game. Last column is the number of medals won by team in particular which is our prediction task.

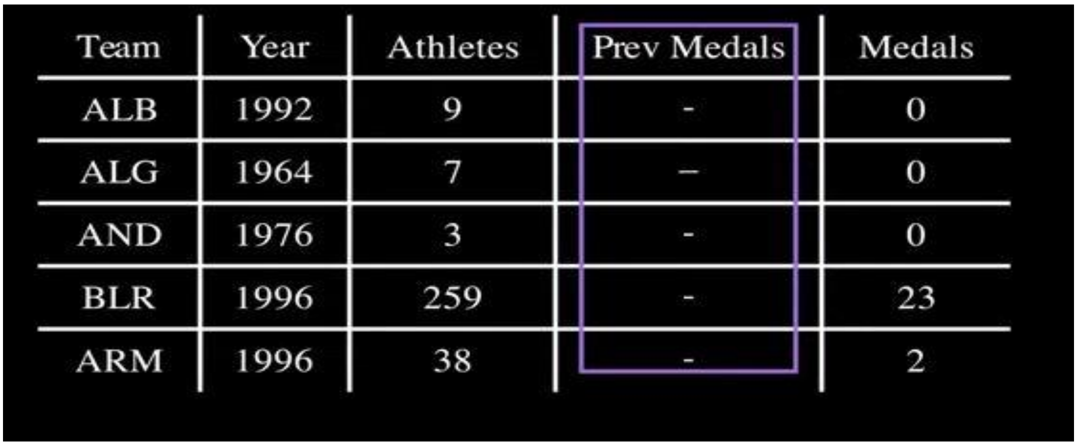
Reshape the Data



Team	Year	Athletes	Prev Medals	Medals
USA	2008	763	263	317
USA	2012	689	317	248
USA	2016	719	248	264
IND	2008	67	1	3
IND	2012	95	3	6
IND	2016	130	6	2

Once we have data, we need to reshape it to make machine learning predictions possible. But in this case, we do not need to do reshaping as our data is already in the form where it is easy to pull data from a single row.

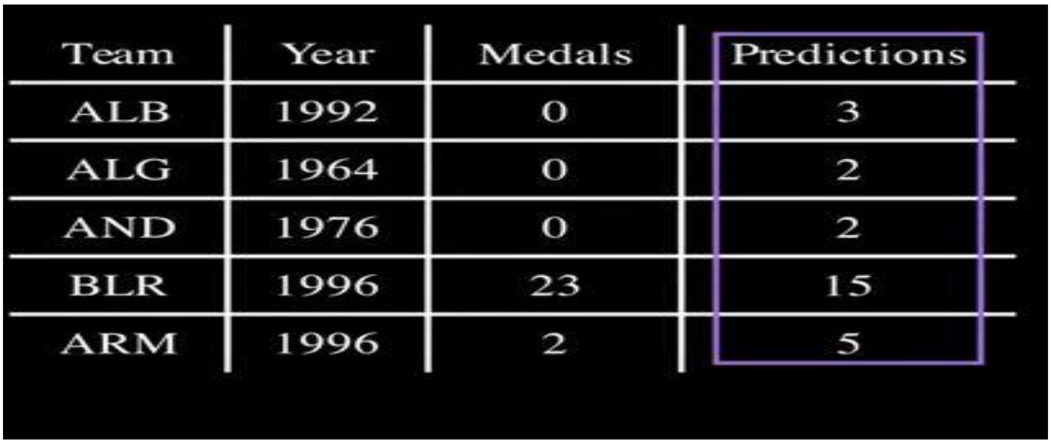
Clean the data



Team	Year	Athletes	Prev Medals	Medals
ALB	1992	9	-	0
ALG	1964	7	–	0
AND	1976	3	-	0
BLR	1996	259	-	23
ARM	1996	38	-	2

Cleaning the data involves making sure that the data is ready for machine learning. In this case, Prev Medals column contains missing values which need to be removed for proper functioning of machine learning algorithm.

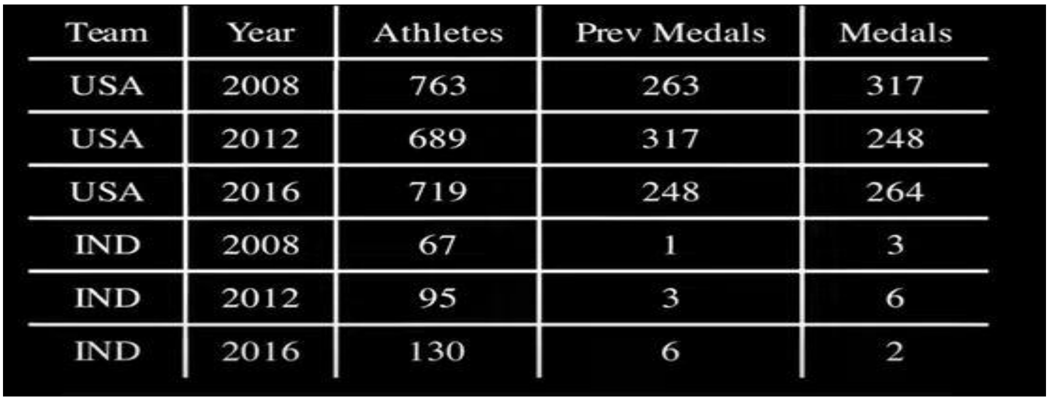
Error Metric



Team	Year	Medals	Predictions
ALB	1992	0	3
ALG	1964	0	2
AND	1976	0	2
BLR	1996	23	15
ARM	1996	2	5

Error Metric is used to evaluate the performance of our machine learning model. In this case, prediction can be done on the basis of how many medals we think a country should have earned in a given Olympics.

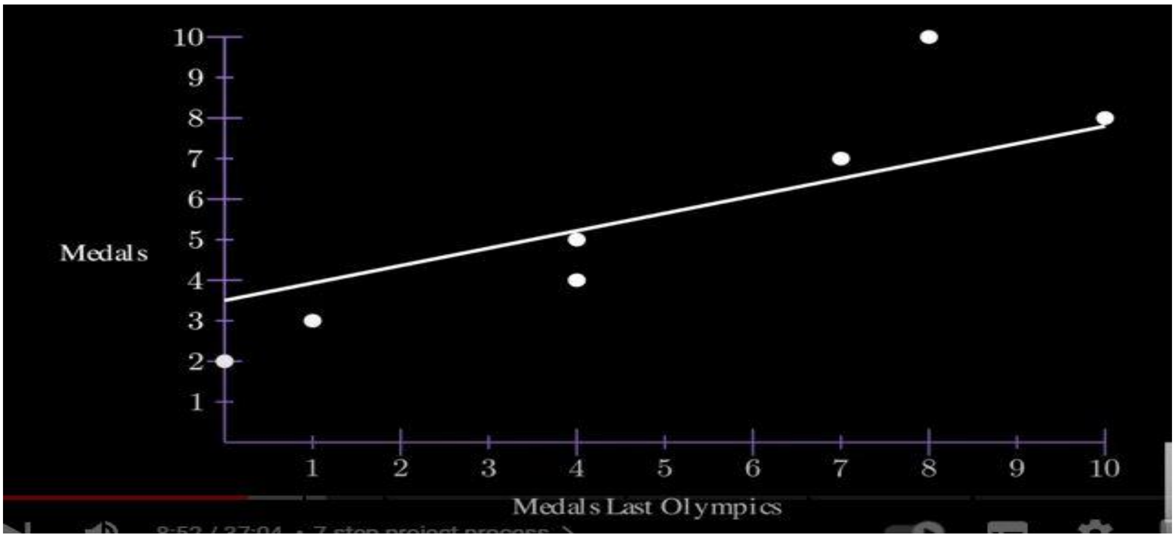
Split the data



Team	Year	Athletes	Prev Medals	Medals
USA	2008	763	263	317
USA	2012	689	317	248
USA	2016	719	248	264
IND	2008	67	1	3
IND	2012	95	3	6
IND	2016	130	6	2

It is important to split the data because one part of the data is needed to train the model known as train data and another is needed to make predictions known as test data.

Train a Model



Use linear regression model with an equation(Y=ax+B) to train the model. Linear Regression model has the ability to provide the probabilities and classify new data using continuous and discrete datasets.

▪Conclusion

This project shows the implementation of the Olympics Data prediction model. This gives an insight into how to analyze a given raw data and convert that into useful features by removing unwanted features.