Project Title: Olympics Medal Prediction

Team Information:

Name	Student ID
Sai Varshith Talluri	015952586
Teja Ramisetty	015906644
Hema Aishwarya Talluri	015948816
Prerana Uppalapati	015933138

Dataset Source: Kaggle

Project Idea:

The aim of our project is to understand and analyze the patterns and statistics of the Olympic Games. Here we focus on various attributes which contribute to the Olympic success. As a team we would like to work to produce a supervised learning model that can predict medal count for individual sports and countries.

Our project would aid governments and their athletic bodies to analyze their level and work on improving their representative's performance while enhancing chance at achieving a medal.

Description of the steps:

1) Identify suitable dataset, applying pre-processing concepts like handling null values, removing noise from the data and encoding columns into numeric values.
2) Dividing the data set into two groups train and test.
3) Train model on the best algorithm on basis of accuracy of the predicted output.
4) Visualization of the result.

Roles:

Name	Roles
Sai Varshith Talluri	Identifying dataset, Data transformation, Pattern evaluation.
Teja Ramisetty	Data cleaning, working on various models
Prerana Uppalapati	Identifying dataset, testing efficiency of the models.
Hema Aishwarya Talluri	Data cleaning, research on existing work, working on various models.

References:

Data Preprocessing:

I) **Big data preprocessing: methods and prospects:**
https://bdataanalytics.biomedcentral.com/articles/10.1186/s41044-016-0014-0

II) **Review of Data Preprocessing Techniques in Data Mining**
https://d1wqtxts1xzle7.cloudfront.net/54509277/4102-4107-with-cover-page-v2.pdf?Expires=1646623858&Signature=a~CGXTEXZ3JC2CMj3ndLnPE5jLjQ3RwD~9stjJaGYTxCFgrWbp8gLummGHS0cL7Z0eg1Im1yTlJ0FpR5zdR3eK8vckKNT-lNgYVac7zGlJRdoQuvv~5DdceS87pfcpiLde3KBHRAeyGDvNM7BnVunCRljZeESfjvnUZesuZc6~EaHWNJRAZYb1WCbeok8qbDSdDLU4N4GaOL0hof0S6CZt0zv82e7pvHy9rRLQHOfMDs1s9ClOFejBRr5MgfkPguFOdwYoUcxpf87wP2tqQlt3scLSXHbIQYd4gIyF6~p6BomDasY6b53CrHg9H0DkgU3dk2S8dAvhTUF0EKTR6ZBw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

III) **New data preprocessing trends based on ensemble of multiple preprocessing techniques:**

https://www.sciencedirect.com/science/article/pii/S0165993620302740

IV) **Data Preprocessing for Supervised Leaning:**

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.8413&rep=rep1&type=pdf

Regrssion Models:

I) **Validation techniques for logistic regression models**

https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780100805?casa_token=7WLlEI-wEdAAAAAA:l3qFR5EaBAesR06ZSRUHYBzl9FoSRpWLEo06dGLtYFmXA0ugd6avBKj46wkGAL4R2V-7WpPGowbt

I) **Evaluating the predictive performance of habitat models developed using logistic regression**

https://www.sciencedirect.com/science/article/pii/S0304380000003227?casa_token=fOVxoITMN0wAAAAA:f33qwytuSqF69xCxqYLL1kPmLn4G7Rx22HEdnQRzDr412uW5wvtKqC1uiFbOQIS2Iqe8W2ZY

II) **Developing prediction models for clinical use using logistic regression: an overview**

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6465431/

III) **Comparison of artificial neural network and logistic regression models for prediction of outcomes in trauma patients: A systematic review and meta-analysis**

https://www.sciencedirect.com/science/article/abs/pii/S0020138319300075?casa_token=Gym3YjEy_icAAAAA:QOpdBqQsWGTfVrQcsSVZDFDC9QXiVphuFCwhC6pRdDjjcRdyKytJda3pIF6gvpQ976cpCXLE