|  |  |  |
| --- | --- | --- |
| **Data Science Project Training Report**  **on**    **Machine Learning Domain Projects for Regression,**  **Classification and Clustering using Various Datasets**    **BACHELOR OF TECHNOLOGY**      **Session 2021-22**    **In**    **Computer Science**      **By**  **Kartik Saroha**  **2000320120095**  **Dev Tayal**  **2000320120070**  **Navam Sharma**  **2000320120113**    **PROF MS SAPNA JAIN**      **DEPARTMENT OF COMPUTER SCIENCE**  **ABES ENGINEERING COLLEGE, GHAZIABAD** | | |
|  |  |  |
| **AFFILIATED TO**  **DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW**  **(Formerly UPTU)** | | |

**Student’s Declaration**

I / We hereby declare that the work being presented in this report entitled **“GENDER CLASSIFICATION”** is an authentic record of my / our own work carried out under the supervision of Prof Ms. **SAPNA JAIN MA’AM, Professor, Computer Science.**

**Date: 26/06/2022**

**(Name: Kartik Saroha)**  **(Name: Dev Tayal)**  **(Name: Navam Sharma)**

**(Roll No.2000320120095) (Roll No.2000320120070) (Roll No.2000320120113)**

**Department: CS**

This is to certify that the above statement made by the candidate(s) is correct to the best of my knowledge.

**Signature of HOD Signature of Teacher**

**Prof.(Dr.) Pankaj Kumar Sharma Ms. Sapna Jain**

**Computer Science**   **Professor**

**Computer Science**

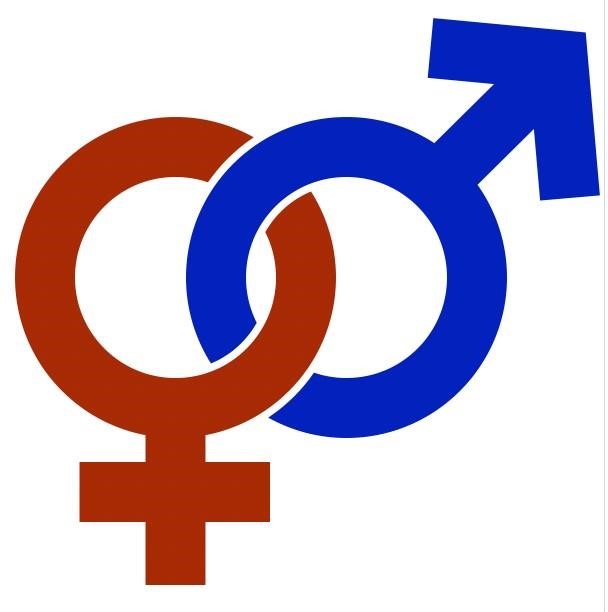
**Date:..........................**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Contents** | **Page No.** |
|  | Student’s Declaration |  |
| 1 | Introduction | 1 |
| 2 | Motivation | 1 |
| 3 | Project Explanation | 2-6 |
| 4 | Conclusion | 6 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Introduction

* This dataset consists of three columns which is Height, Weight and Gender.
* In this project we are going to classify person's gender based on his/her height and weight.
* Our goal is to train a model with the given data to classify gender. • Let's look more into the data



# Motivation

* We have done this project to stay motivated on Data Science.
* In this project, We have learnt about classification kind of problems.
* We done this project to learn, how the classification works.

# Project Explanation

## Data Collection

•The dataset is collected from kaggle.

* The dataset consists of a CSV file name weight-height.csv. Let's load the collected data.

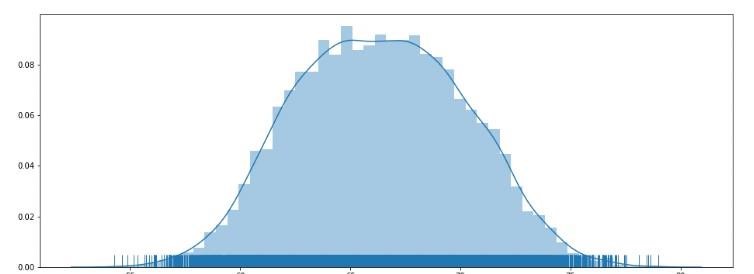


* The dataset consists of 3 features i.e., Height, Weight and Gender.
* The dataset consists of 10,000 samples of different persons. Feature

## Engineering

### 1. Height

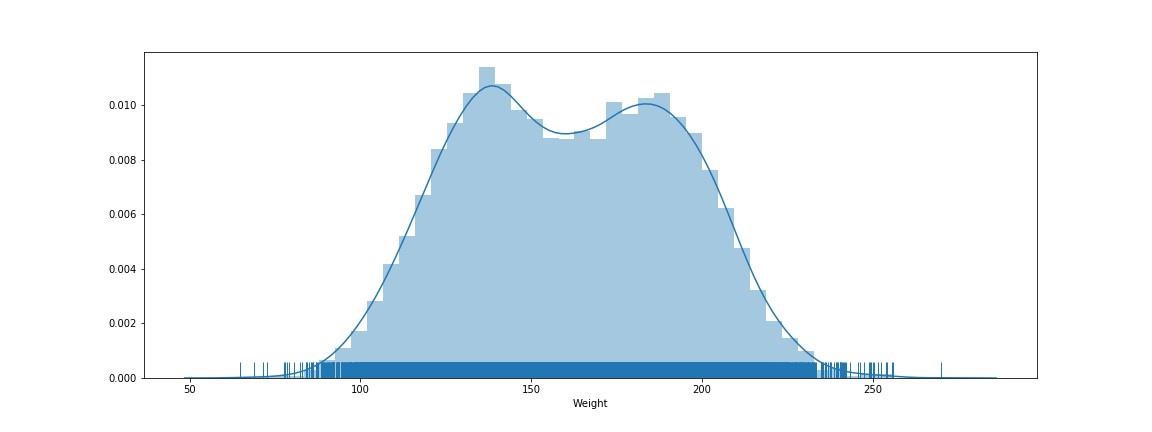
* The heights are mentioned in inches.
* There are no null values in the Height column.
* Let's see the distribution of height data.



* Here we get a well distributed data and a bell shaped curve shows that it is apt for analysis.

### 2. Weight

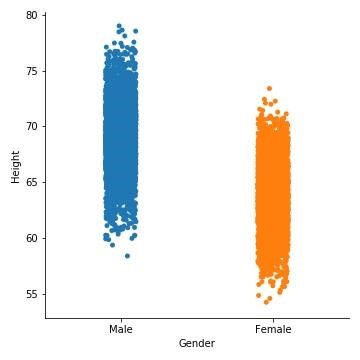
* The weights are mentioned in pounds.
* There are no null values in the Weight column.
* Let's see the distribution of weight data.

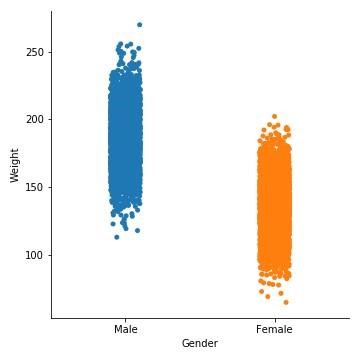


* Here we get a well distributed data and a bell shaped curve shows that it is apt for analysis.

### 3. Gender

* Gender is a **nominal** kind of **qualitative data**, because there is no numerical relation between different genders.
* Let's see the relation of Gender with height and weight.



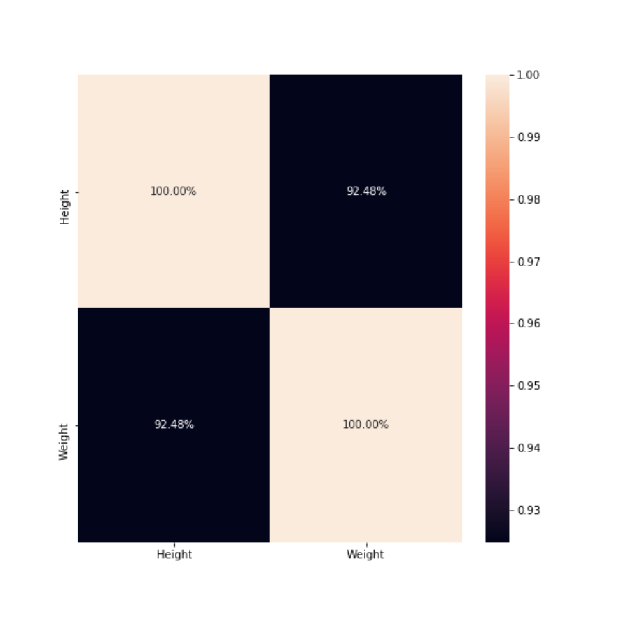


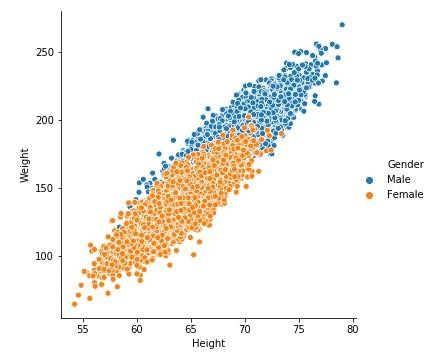
## Data pre processing

The data which we collect from kaggle is already well pre-processed. So no need of pre-processing.

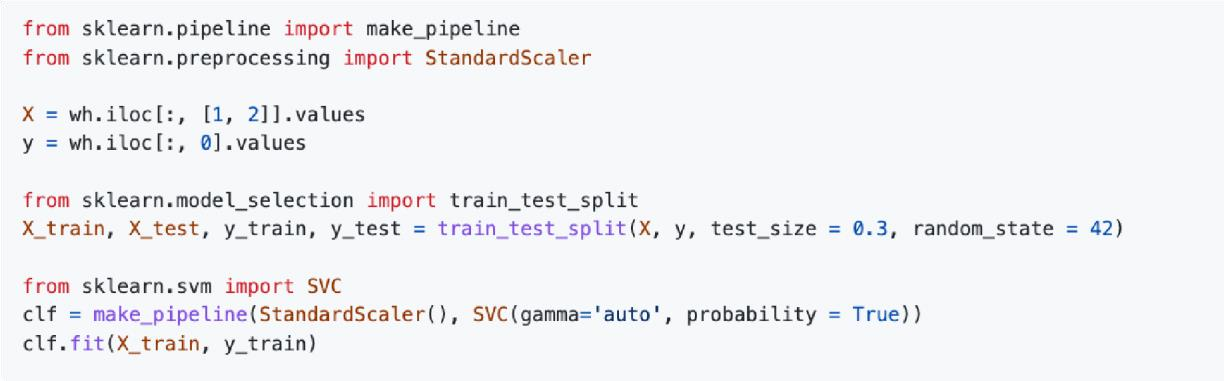
## Feature Observations

Let's see the correlation of features with a heatmap.



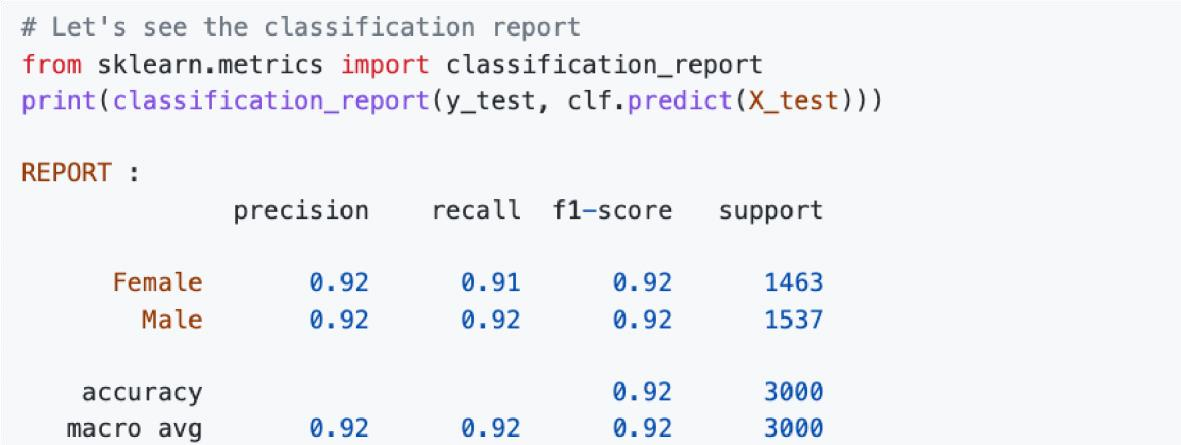


## Building Machine Learning Model





## Model Performance



# Conclusion

* Finally we finished our gender classification project.
* We have built a Support Vector Machine Classifier which performs well with given features and having the Training accuracy of 91.47% and Testing accuracy of 91.90%.