Stars Classifacation

Outline :

Introduction

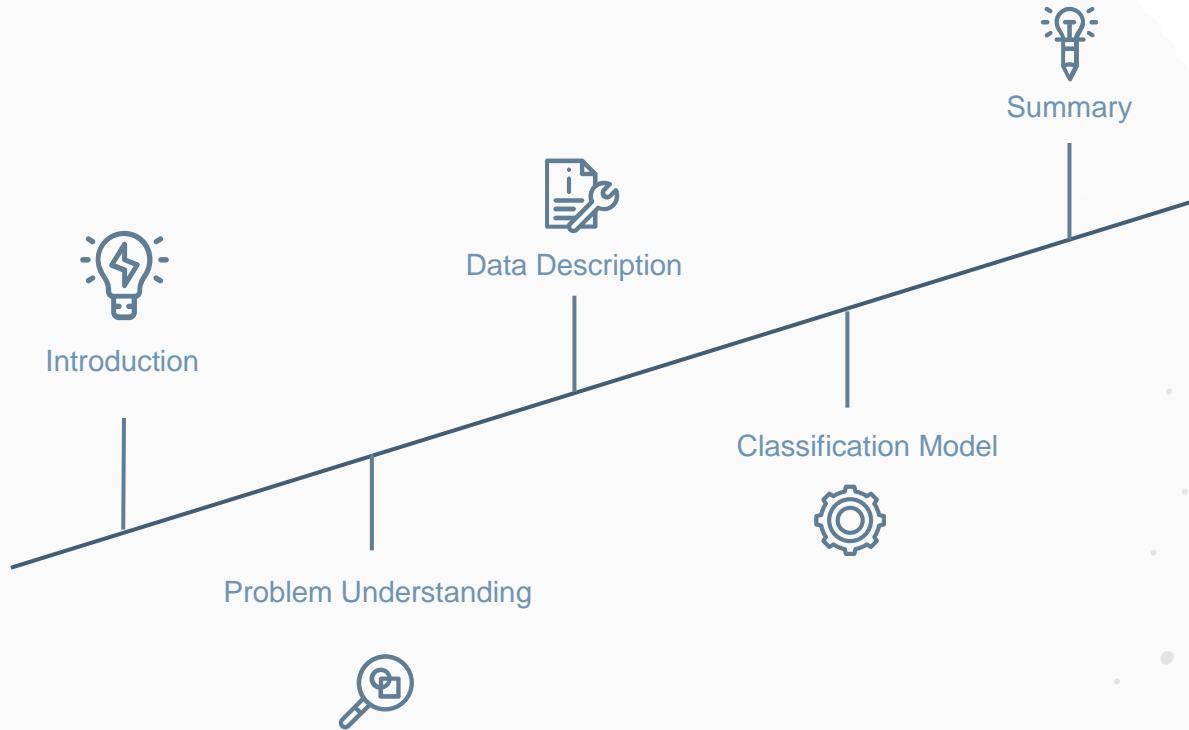Problem Understanding

Data Description

Classification Model

Summary

Introduction :

A star is an astronomical object consisting of a luminous spheroid of plasma held together by its own gravity. The nearest star to Earth is the Sun. Many other stars are visible to the naked eye at night, but due to their immense distance from Earth they appear as fixed points of light in the sky. The most prominent stars are grouped into constellations and asterisms, and many of the brightest stars have proper names. The observable universe contains an estimated 1022 to 1024 stars, but most are invisible to the naked eye from Earth.

Problem Statement :

We want to determine the types of stars according to our data and to know whether the star is a giant or a dwarf for the purpose of studying the stars and galaxies surrounding us.

Data Description :

Our data set include many thousand of stars that have been discovered by scientist.

The data set is provided in .csv format, contains information like visual Apparent Magnitude of the Star(Vmag), Distance Between the Star and the Earth (Plx), spectral type, B-V color index

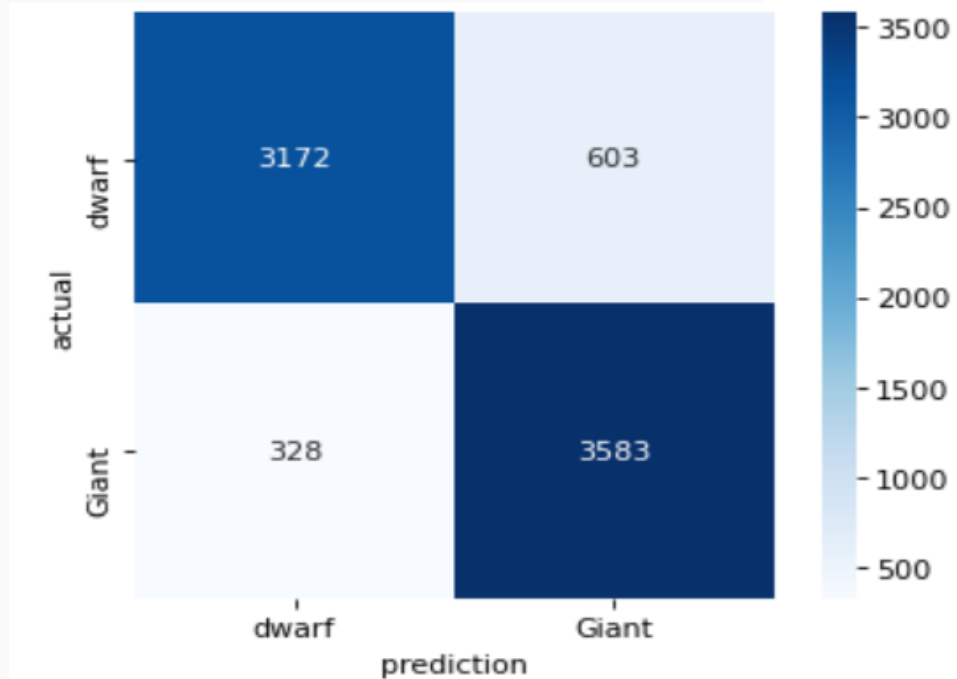Our data set contains 39,552 of rows and 7 columns of information of stars.

Target histogram:

We found that our target class is balanced, so the imbalance will not affect much.
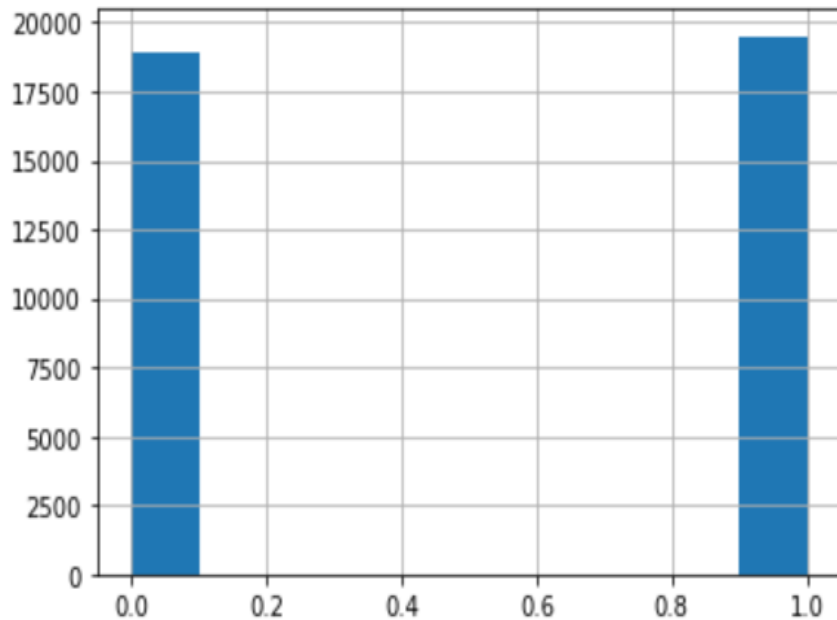
Confusion Matrix :

The graph shows the what is actual class and what model predicted.

Classification Model :

Model	F1	Recall	Precision
baseline	85%	88%	84%
KNN	87%	89%	86%
Logistic Regression (GS)	88%	91%	85%
Voting(Soft)	88%	91%	86%
XGBoost	88%	90%	87%

Classification Model :

Model	Train	Test
Baseline	86%	84%
KNN	88%	88%
Logistic Regression (GS)	88%	87%
Voting(soft)	88%	89%
XGBoost	89%	89%

Classification Model :

This graph represents
how our models do.
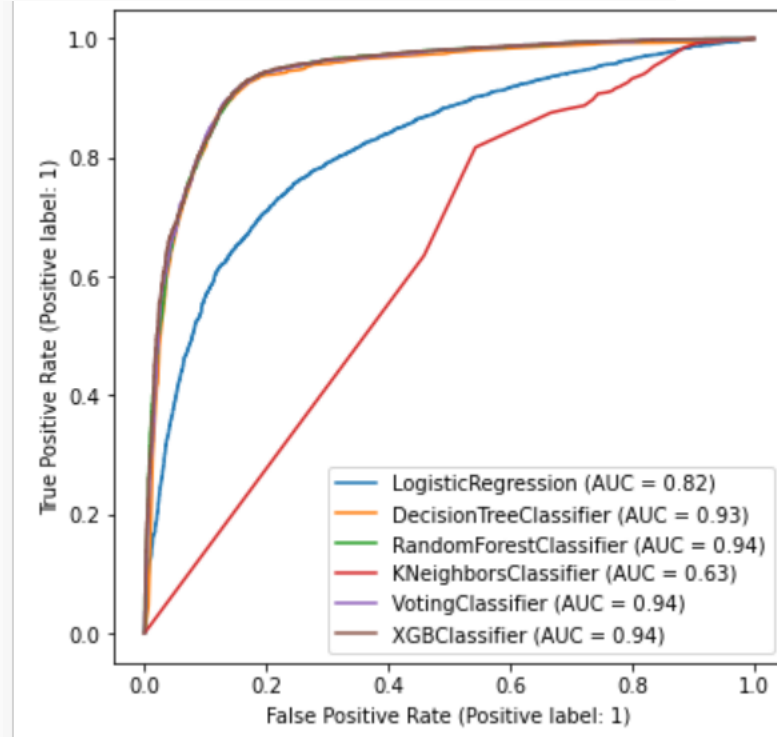
Summary

The results was very close to each other.

We find The best model is **XGBoost Model.**

So , our model ready to predict the stars.

THANKS !

Any questions?