2023 Group 2 Project 4

Stars Classification

Welcome To Our Presentation

Cheila, Grace, Helen, Jacob,
Kassem & Rami

2023

Let's get to know about stars with the help of machine learning

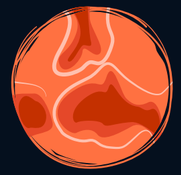
2023

Introduction

xxxxxxxxxx

Earth

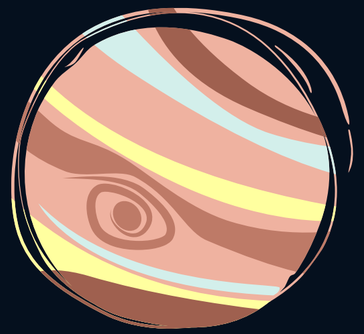
Mars



Mercury

Venus

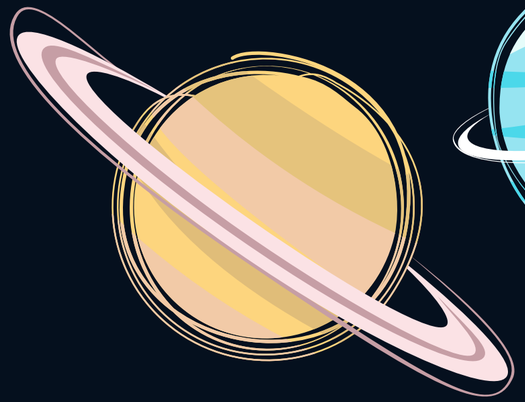
Uranus



Pluto

Neptune

Saturnus



Jupiter

2023

AIM!

The aim of this study is to utilise the Morgan-Keenan (MK) classification system, which incorporates the HR classification system, to categorise stars by their chromaticity and size using spectral data. The study will focus on categorising stars into the main Spectral Types using the Absolute Magnitude and B-V Color Index within a specific dataset.

2023

Dataset & preprocessing

.....

01

Dataset – Kaggle, Raw File, Clean, Final Dataset

02

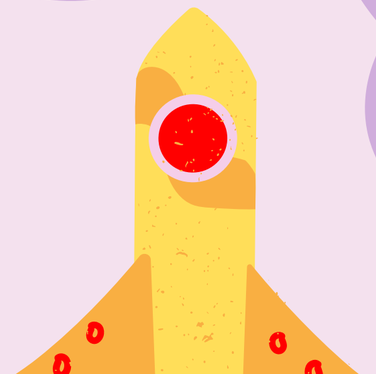
ETL – Extract, Transform & Load

03

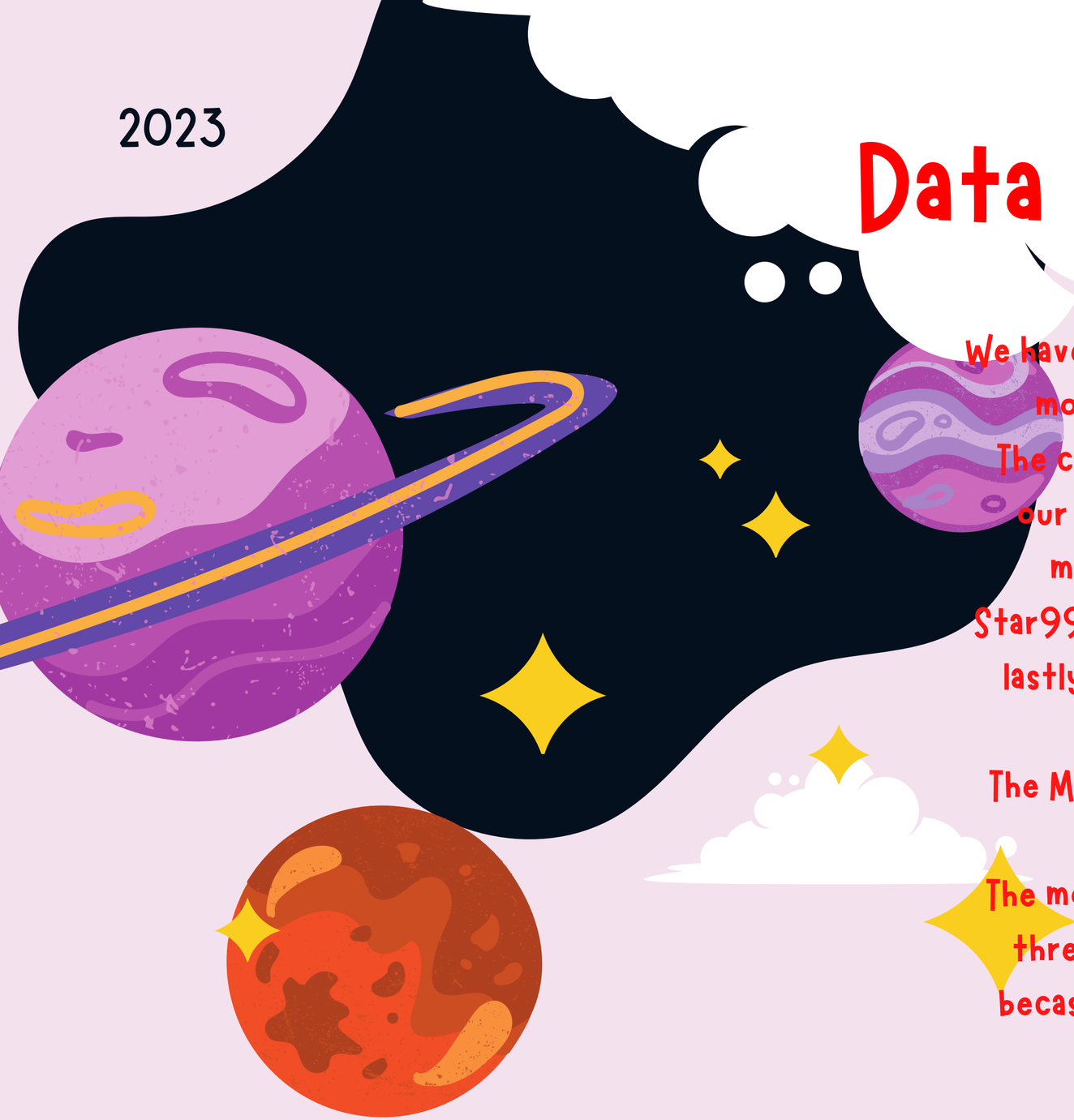
Pre Processing

MACHINE LEARNING!

we used unsupervised machine learning because.....



2023



Data Model Implementation

We have produced a Python script that initializes, trains, and evaluates a model based on our cleaned up dataset which is the Final_Stars.csv. The cleaned, normalized, and standardized data was achieved as part of our pre-processing work at the begining of the project and prior to modeling. This is evidenced under the resources files starting with Star9999_raw then further formating within the Clean_stars.csv and lastly leading to our final Final_Stars.csv which is our fully formatted and cleaned dataset file.

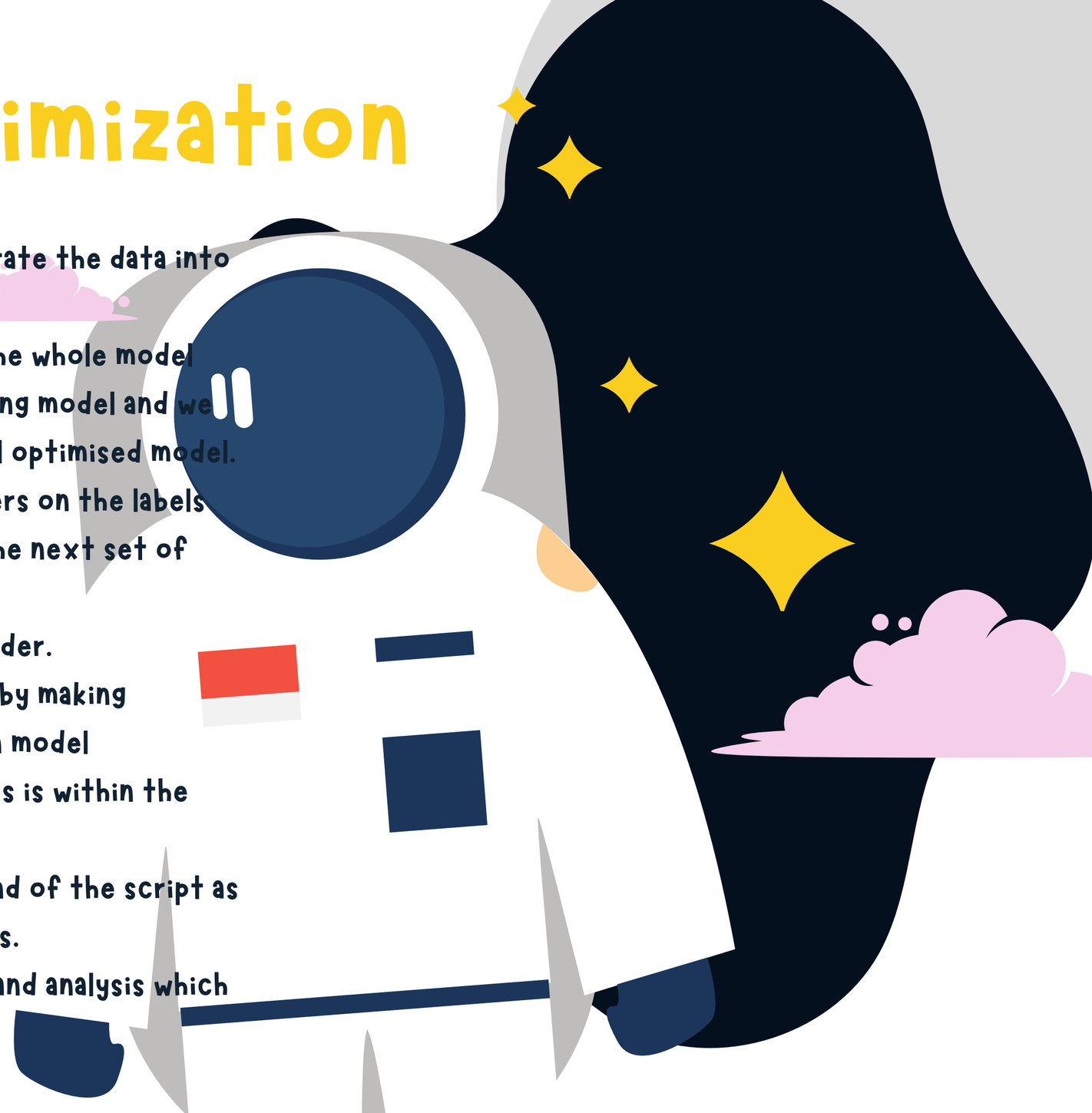
The Model utilizes data retrieved from SQL which is visable iwithin out Jypeter Notebook script.

The model demonstrates meaningful predictive power of over the 75% threshold set within the rubric at 75.10% classification accuracy. It becasue of the for loop in the pre processing that cleaned the target class to give us a higher classification accuracy.

2023

Data Model Optimization

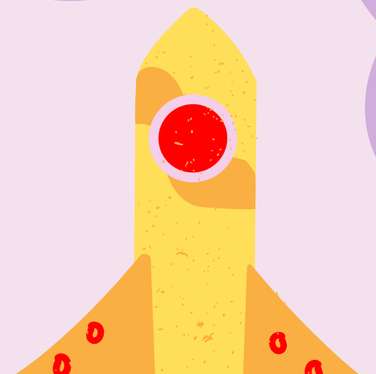
- When we are training and testing the model we had to separate the data into two models and data sets.
- So that we were not cheating the models as if we tested the whole model and dataset we've would seen what we want from the training model and we would be testing the same in the test set model in the final optimised model.
- So by splitting the datasets and looping through the numbers on the labels we would be training on one set of data and then test on the next set of data after making our predictions.
- All this work is saved within our repo in the two-target folder.
- Then we ran the model optimization and evaluation process by making iterative changes to the model and the resulting changes in model performance is documented at slightly higher to 75.6% this is within the Python script itself.
- Overall model performance is printed or displayed at the end of the script as mentioned is at 75.6% post the model optimisation process.
- Within the two-target folder is our final optimised model and analysis which contains an accuracy of 89.2%.



2023

VISUALISATIONS & ANALYSIS

xxxxxxxxx



2023

Visualisations & Analysis

Matplotlib

Write a topic or a highlight of the presentation here

Seaborne

Write a topic or a highlight of the presentation here

Matplotlib

Write a topic or a highlight of the presentation here

Tableau

Stellar Classifications - Rahmi

Visual & Absolute Magnitude - Grace

Thank You

for your attention &
any questions

The lesson chart with the best achievements

With Canva Presentations, you can collaborate in real-time and feel like you're in the same room as your teammates or co-presenters.