**Imperial College – Professional Certificate in Machine Learning – Capstone Project- Model Card**

**MODEL DETAILS**

A lightweight model to classify stars based on spectral characteristics from photometric telescope systems.

**MODEL SPECIFICATIONS**

Decision Tree with tree depth 9.

**MODEL INPUTS**

Following inputs from instruments on the telescopes photometric system:

* Ultraviolet filter
* Green filter
* Red filter
* Near Infrared filter
* Infrared filter

It also uses the calculated redshift value from change in wavelength in prediction.

**MODEL OUTPUTS**

One of three classifications

* Star
* Quasar
* Galaxy

**APPLICATIONS**

Classification of celestial objects from spectral data from photometric telescopes.

**USERS**

Digital sky surveys that require automated classification from large volumes of data captured.

**LIMITATIONS**

Model is trained on cleaned data sets from Sloan digital sky survey. Model is only expected to perform well when input data is similarly well cleaned.

**TRADE OFF**

Simplistic model, we prefer interpretability and speed of execution.

**DATASET**

Training data is from the Sloan Digital Sky Survey release number RD17. Details on the imaging camera can be found here: <https://www.sdss4.org/faceplat/>.

80% of data was used for training

20% of data was used for testing

**MODEL TRAINIGN DETAILS**

Model trained of various tree depths, with tree depth 9 selected with optimal classification power.

Depth: 1 Train Score: 0.8054350679383492 Validation Score: 0.80455

Depth: 2 Train Score: 0.9475118438980488 Validation Score: 0.9465

Depth: 3 Train Score: 0.9516118951486894 Validation Score: 0.95055

Depth: 4 Train Score: 0.96582457280716 Validation Score: 0.9648

Depth: 5 Train Score: 0.9658995737446718 Validation Score: 0.9646

Depth: 6 Train Score: 0.9701121264015801 Validation Score: 0.9666

Depth: 7 Train Score: 0.9764747059338241 Validation Score: 0.9729

Depth: 8 Train Score: 0.9777872223402793 Validation Score: 0.9741

Depth: 9 Train Score: 0.9804872560907011 Validation Score: 0.9751

Depth: 10 Train Score: 0.9822247778097226 Validation Score: 0.97265

Depth: 11 Train Score: 0.9851498143726797 Validation Score: 0.97225

Depth: 12 Train Score: 0.9875248440605507 Validation Score: 0.9719

Depth: 13 Train Score: 0.989874873435918 Validation Score: 0.9713

Depth: 14 Train Score: 0.9921999024987812 Validation Score: 0.97035

Depth: 15 Train Score: 0.9940374254678184 Validation Score: 0.96905

Depth: 16 Train Score: 0.9953249415617695 Validation Score: 0.96875

Depth: 17 Train Score: 0.9963249540619258 Validation Score: 0.96785

Depth: 18 Train Score: 0.9972499656245704 Validation Score: 0.9671

Depth: 19 Train Score: 0.997862473280916 Validation Score: 0.96715

Depth: 20 Train Score: 0.9983124789059863 Validation Score: 0.96745

Best Tree Depth 9

Tree Model Training time in seconds: 9.807419061660767

**MODEL ACCURACY**

97.5% classification accuracy Worst classification on quasars, with 92% accuracy.

CONFUSION MATRIX

A chart of different colors

Description automatically generated with medium confidence

FEATURE IMPORTANCES

TEST SET

A graph with blue squares

Description automatically generated

PERMUTATION IMPORTANCES

A graph of a train set

Description automatically generated

A graph with lines and numbers

Description automatically generated

Below is the correlation matrix for the data:

A table with numbers and symbols

Description automatically generated with medium confidence

You can u,g and r highly correlated, hence their importance might be low here.

Green = Galaxy

Blue = Star

Yellow = Quasar

A diagram of a number of dots

Description automatically generated

A diagram of a number of dots

Description automatically generated