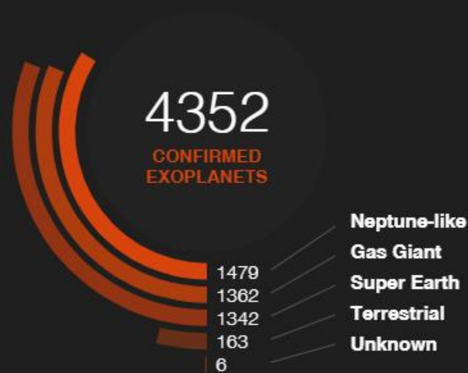
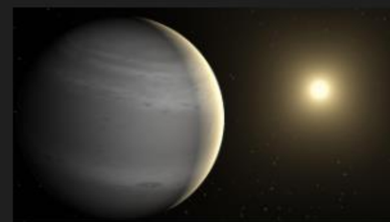
Planet Types



New Discovery

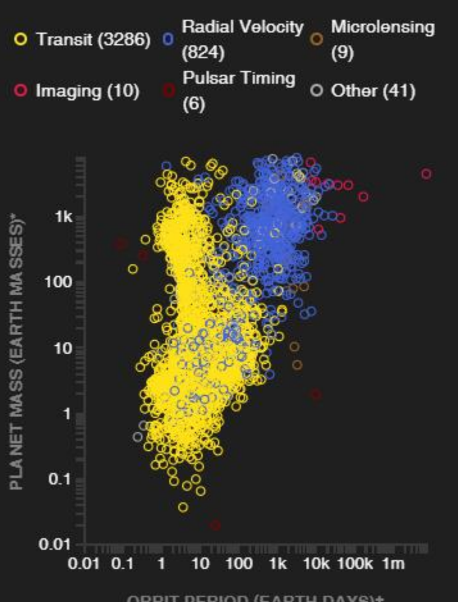


PLANET NAME	PLANET TYPE
TYC 0434-04538-1 b	Gas Giant
DISCOVERY DATE	**DETECTION METHOD**
2021	Radial Velocity

› More about this planet

Exoplanet Census

Display limited to planets with both measured or estimated orbital period and mass



YEAR 2021 | **DISCOVERIES‡ 4352**

1989 — 2021

k=thousand,m=million
*Masses and orbital periods are estimated for some planets based on other parameters
†Orbit period is equal to one trip around the star
‡Does not include discoveries where mass or orbit period is unknown or mass in Jupiters is > 25

By Method

- 76.2% Transit
- 19.0% Radial Velocity
- 2.4% Microlensing
- 1.2% Imaging

0.48% Transit Timing Variations, 0.37% Eclipse Timing Variations, 0.16% Pulsar Timing, 0.14% Orbital Brightness Modulation, 0.05% Pulsation Timing Variations, 0.02% Disk Kinematics, 0.02% Astrometry

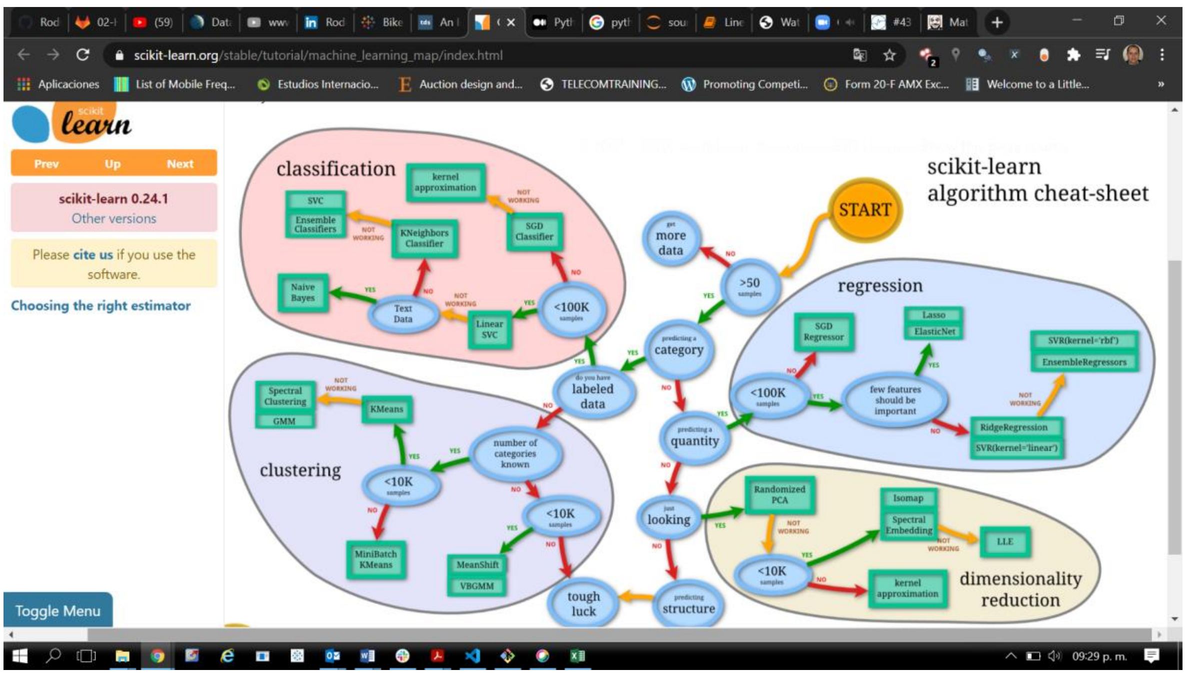
› More about planet-hunting methods

The Kepler Space Observatory is a NASA-build satellite that was launched in 2009. The telescope is dedicated to searching for exoplanets in star systems besides our own, with the ultimate goal of possibly finding other habitable planets besides our own. The original mission ended in 2013 due to mechanical failures, but the telescope has nevertheless been functional since 2014 on a "K2" extended mission.

Kepler had verified 1284 new exoplanets as of May 2016. As of October 2017 there are over 3000 confirmed exoplanets total (using all detection methods, including ground-based ones). The telescope is still active and continues to collect new data on its extended mission.

The main objective of the machine learning model is classify candidate exoplanets from the raw dataset. In other words, it aims to predict a category based on the following labels or strings: Confirm, false positve or candidate.

Having in mind that there is a preexistent classificaction, we should use models for classification: SVC, Kernel aproximation, SGD classifier, such as the following schema:



Dictonary:

Variable	Description
KOI	Kepler Objects of Interest
KOI_disposition	The category of this KOI from the Exoplanet Archive. Current values are CANDIDATE, FALSE POSITIVE, NOT DISPOSITIONED or CONFIRMED. All KOIs marked as CONFIRMED are also listed in the Exoplanet Archive Confirmed Planet table. Designations of CANDIDATE, FALSE POSITIVE, and NOT DISPOSITIONED are taken from the Disposition Using Kepler Data.
koi_fpflag_nt	A KOI whose light curve is not consistent with that of a transiting planet. This includes, but is not limited to, instrumental artifacts, non-eclipsing variable stars, and spurious (very low SNR) detections.

koi_fpflag_ss	A KOI that is observed to have a significant secondary event, transit shape, or out-of-eclipse variability, which indicates that the transit-like event is most likely caused by an eclipsing binary. However, self-luminous, hot Jupiters with a visible secondary eclipse will also have this flag set, but with a disposition of PC.
koi_fpflag_co	The source of the signal is from a nearby star, as inferred by measuring the centroid location of the image both in and out of transit, or by the strength of the transit signal in the target's outer (halo) pixels as compared to the transit signal from the pixels in the optimal (or core) aperture.
koi_fpflag_ec	The KOI shares the same period and epoch as another object and is judged to be the result of flux contamination in the aperture or electronic crosstalk.
koi_period	The interval between consecutive planetary transits.
koi_period_err1	
koi_period_err2	
koi_time0bk	The time corresponding to the center of the first detected transit in Barycentric Julian Day (BJD) minus a constant offset of 2,454,833.0 days. The offset corresponds to 12:00 on Jan 1, 2009 UTC.
koi_time0bk_err1	
koi_time0bk_err2	
koi_impact	The sky-projected distance between the center of the stellar disc and the center of the planet disc at conjunction, normalized by the stellar radius.
koi_impact_err1	
koi_impact_err2	
koi_duration	The duration of the observed transits. Duration is measured from first contact between the planet and star until last contact. Contact times are typically computed from a best-fit model produced by a Mandel-Agol (2002) model fit to a multi-quarter Kepler light curve, assuming a linear orbital ephemeris.
koi_duration_err1	

```
df = pd.read_csv("exoplanet_data.csv")

#Drop the null columns where all values are null

df = df.dropna(axis='columns', how='all')

# Drop the null rows

df = df.dropna()

df
```

Procedure standard:

1. We create the model using sklearn

from sklearn.linear_model import LinearRegression

model = LinearRegression

model

2. We fit the model to our data using the fin method (runing the model)

    model.fit(X,y)

    <is an object our model is now train>

3. We can predict

    y_predicted = model.predict(X_test)

4. R2

    model.score(X,y)

5. Performance Metric

    from sklearn.metrics import f1 score
    f1_score(y_test, y_predicted)

    Mean squared Error

    from sklearn.metrics import mean_square_error, r2_score

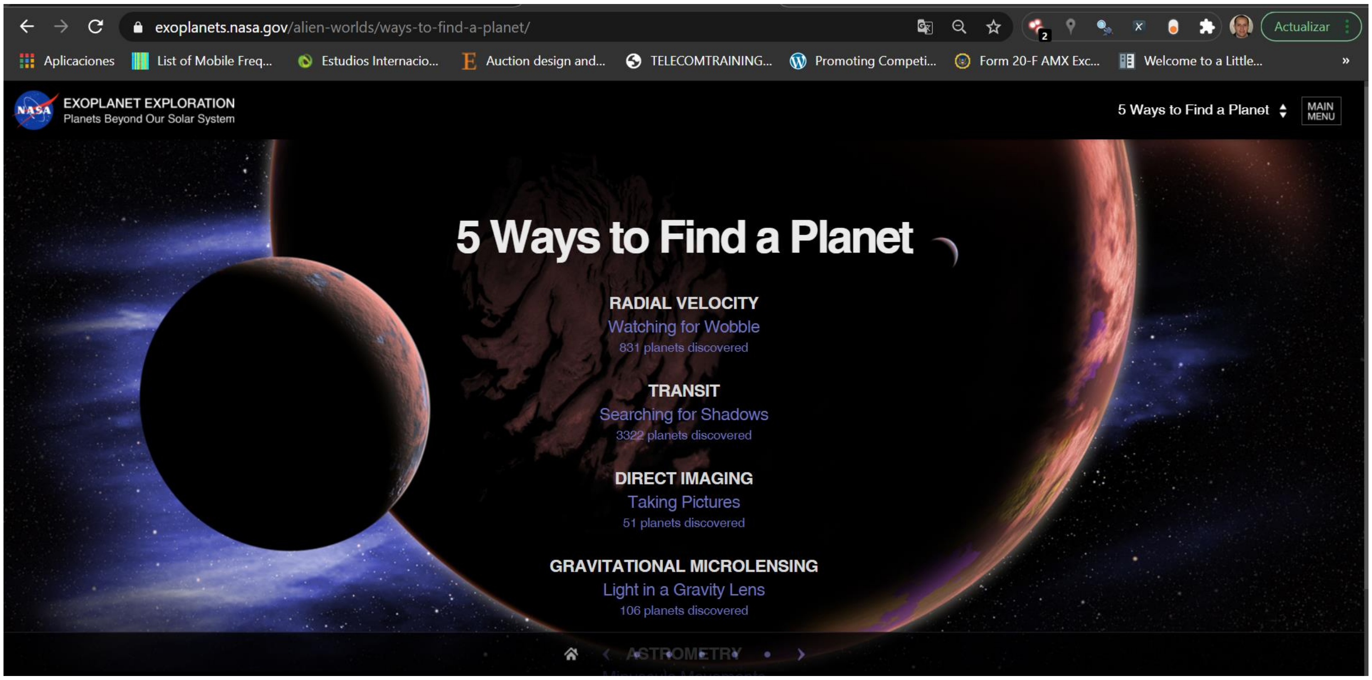
    predicted = model.predict(X_test)

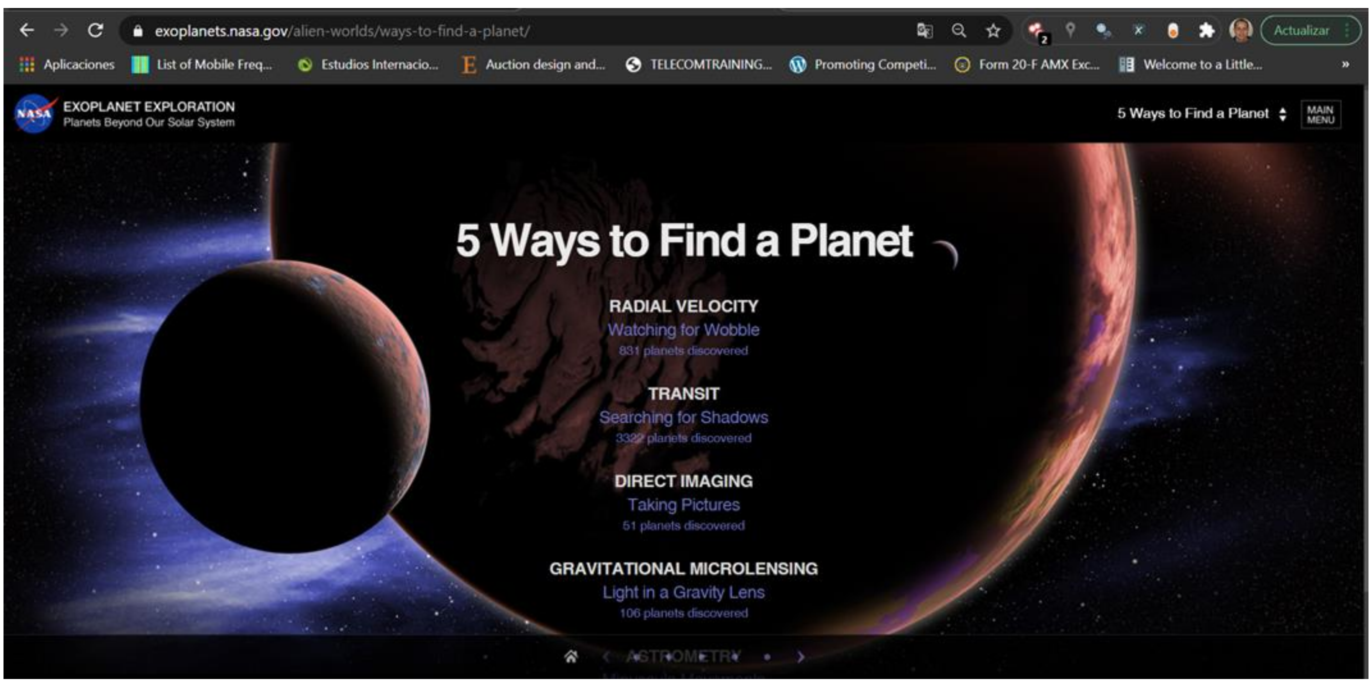
    MSE = mean_square_error(y, predicted)

    R2 = r2_socore(y, predicted)

selected_features = df[['koi_fpflag_nt', 'koi_fpflag_ss', 'koi_fpflag_co', 'koi_fpflag_ec', 'koi_period', 'koi_time0bk', 'koi_impact', 'koi_duration']]

selected_features





Sources

Label Encoding sklearn.preprocessing:

https://scikit-learn.org/stable/modules/preprocessing_targets.html#preprocessing-targets

Approach

Supervised learning (classification)

This aims to predict a category, with a preexistence labeled as CANDIDATE, CONFIRMED and FALSE POSITIVE.

Comparing models

Logistic Regression			
Training data score	0.7451840549303834		
Testing data score	0.7185354691075515		
Outputs	CONFIRMED	CANDIDATE	FALSE POSITIVE
Precision	0.46	0.98	0.00
F1 score	0.63	0.99	0.00
recall	0.99	1	0.00
Hyperparameter Tuning	0.79		
Accuracy	0.72 very low		

Random Forest			
Training data score	1		
Testing data score	0.8180778032036613		
Outputs	CONFIRMED	CANDIDATE	FALSE POSITIVE
Precision	0.64	0.99	0.62
F1 score	0.59	0.99	0.66
recall	0.55	1	0.69
Hyperparameter Tuning	0.8159		
Accuracy	0.82		

K Nearest Neighbor			
Training data score	**0.86382**		
Testing data score	**0.81522**		
Outputs	CONFIRMED	CANDIDATE	FALSE POSITIVE
Precision	0.62	0.99	0.62
F1 score	0.59	0.99	0.64
recall	0.56	1	0.66
Accuracy	0.82		

Support Vector Model			
Training data score	**0.7770360480640854**		
Testing data score	**0.7877574370709383**		
Outputs	CONFIRMED	CANDIDATE	FALSE POSITIVE
Precision	0.83	0.54	0.99
F1 score	0.22	0.69	0.99
recall	0.13	0.96	1.00
Hyperparameter Tuning	0.7806584242353678		

Accuracy	0.79

Deep Learning – Neural Network			
Training data score	**0.7770360480640854**		
Testing data score	**0.7877574370709383**		
Outputs	CONFIRMED	CANDIDATE	FALSE POSITIVE
Precision	0.83	0.54	0.99
F1 score	0.22	0.69	0.99
recall	0.13	0.96	1.00
Hyperparameter Tuning	0.7806584242353678		
Accuracy	0.8278		

		Predicted values		
Observed values		CONFIRMED	FALSE POSITIVE	CANDIDATE
	CONFIRMED			
	FALSE POSITIVE			
	CANDIDATE			

Accuracy defined as TP+TN/Total observations is higher in