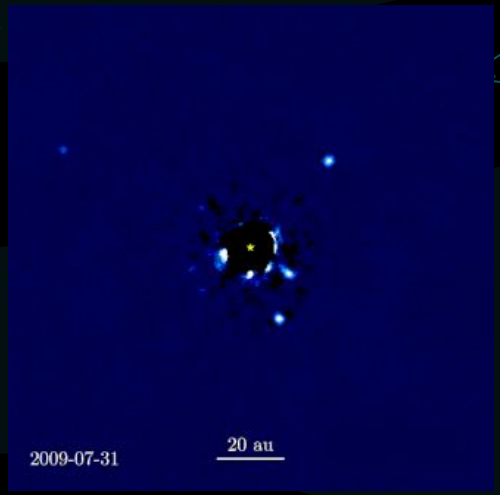
# Planet Hunters

Tanisha Patil, Josephine Bhadran, Arjun Kashyap, Tia Rojas, Ekin Akman & Ashish Kashyap

# What are Exoplanets?

- An exoplanet is any planet that is not within our solar system (do not orbit the sun)
- The search for exoplanets is important because they are the “prime target in the search for life beyond Earth” (planetary)
- What could make exoplanets difficult to detect? What is blocking the exoplanet from being seen?



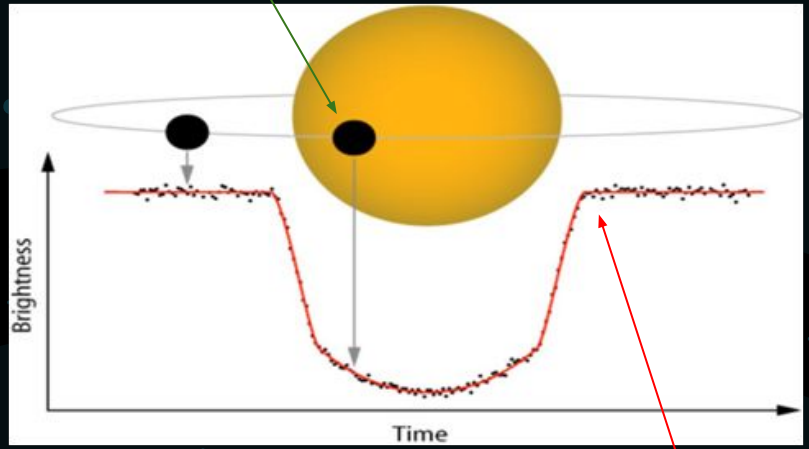

This is star HR 8799. We can roughly tell where the exoplanets are but unfortunately this is not always the case. So what way can we get more precise data about exoplanets?

Exoplanets are incredibly hard to detect with telescopes because they are ‘hidden’ by the light of the star they orbit. They can also be hidden by ISM (interstellar medium or basically ‘stuff in space’) like clouds of gas. Also because they are very far away.

# Finding Exoplanets: Transit Photometry

- Transit method
- Exoplanets are too dark and far away to be directly imaged
- Detect a planet's presence by observing a star's brightness **(Flux: A star's changing brightness)** over time
- Flashlight example

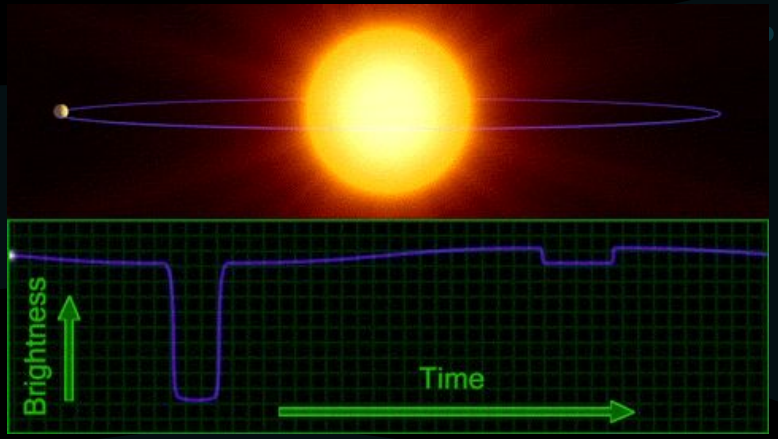
Every time an exoplanet's orbit crosses its stars path, it creates a shadow that decreases the stars overall **flux**

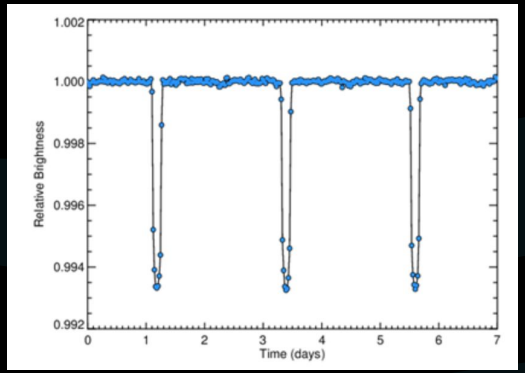


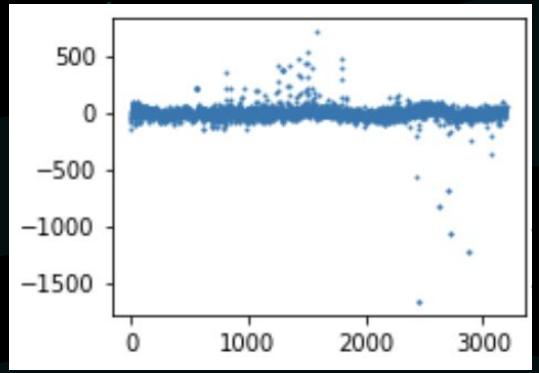

Plotting a star's brightness overtime is called a light curve.

# Finding Exoplanets: Transit Photometry

*It is important that the dips we observe are regular. If there are not regular (repeating at exact intervals) dips in the light curve, it could just be a cloud of gas passing through, decreasing the brightness of the star temporarily. We **KNOW** the object is an exoplanet only when we can detect a regular, repeating orbit.*



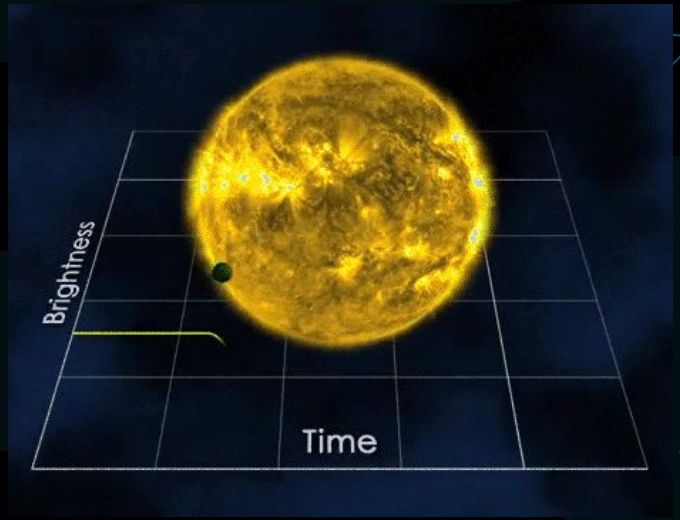



Now, knowing that exoplanets have regular, repeating light curves, which of the light curves represent an exoplanet and which does not?

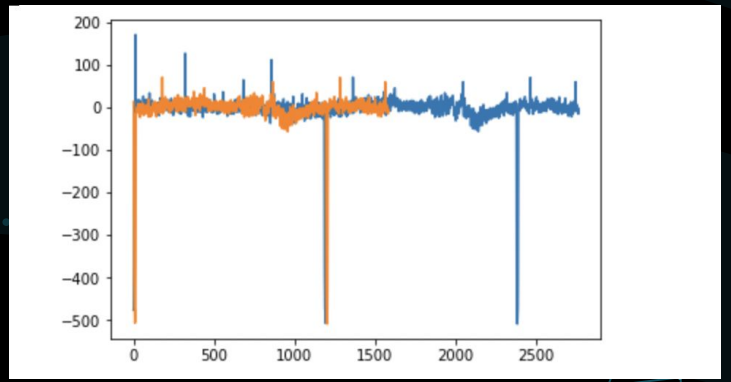
# What is a light curve?

- A light curve is a graph of light intensity of a celestial object as a function of time.

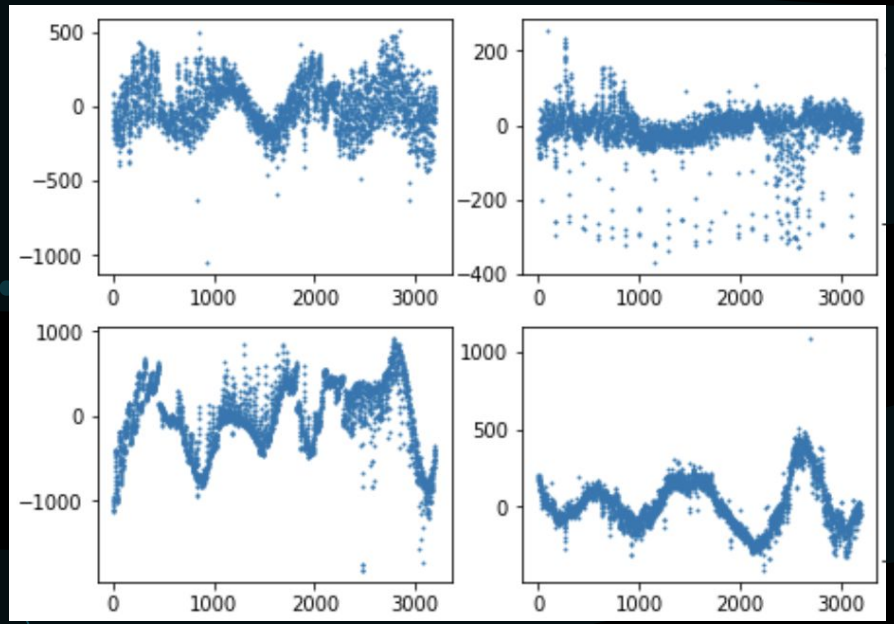


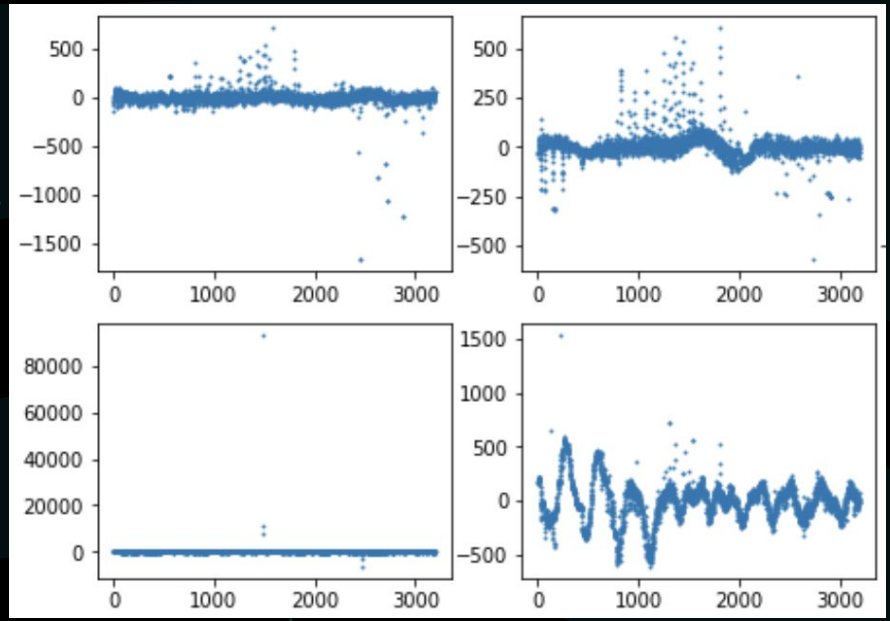

# What is light curve folding?

- Folding method is used to determine if there is a consistent trend between light curves of exoplanets by plotting all the periods on top of each other.

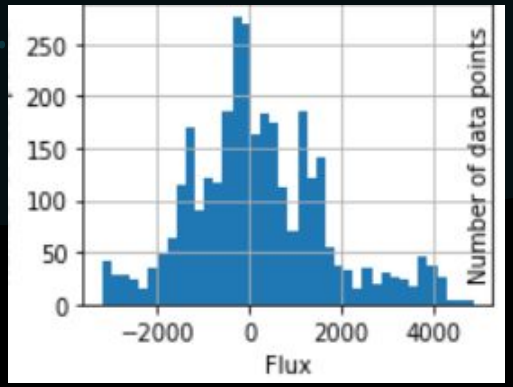


# What do flux curves for exoplanets look like?

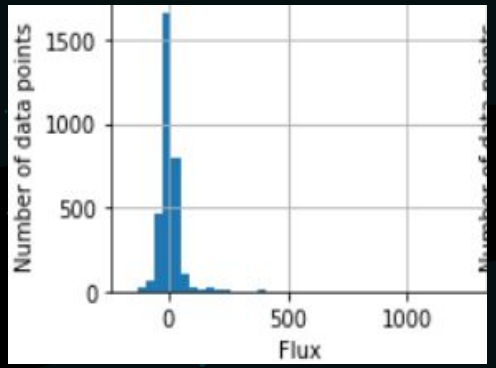


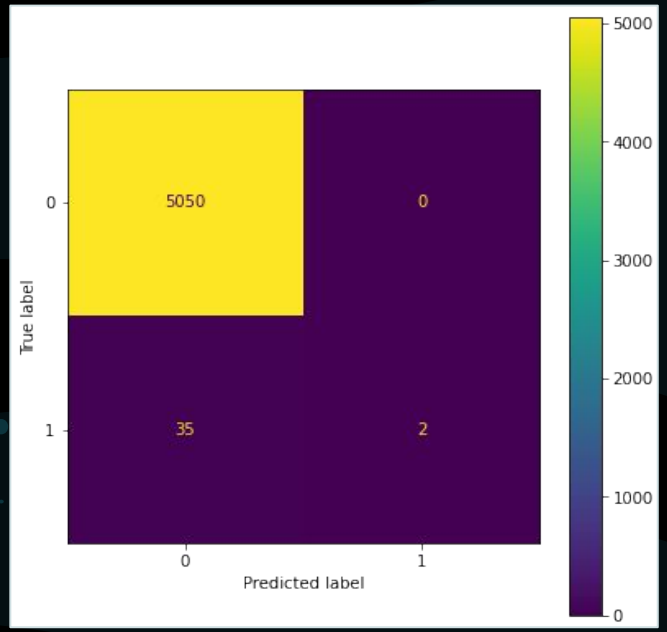
# Challenges With Our Data




Exoplanet Histogram




Non-exoplanet Histogram



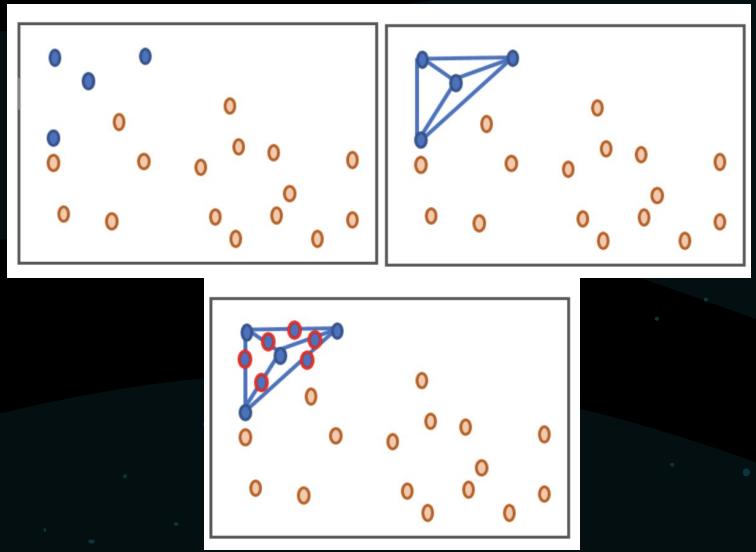

Using a KNN model with an imbalance of data, we have an extreme bias towards the variable with more data points (non-exoplanets)

# Solutions: Data Augmentation

**Normalization**

- Eliminates the issue of data having a high range of values
- Scaling all data points so they lie between 0 and 1

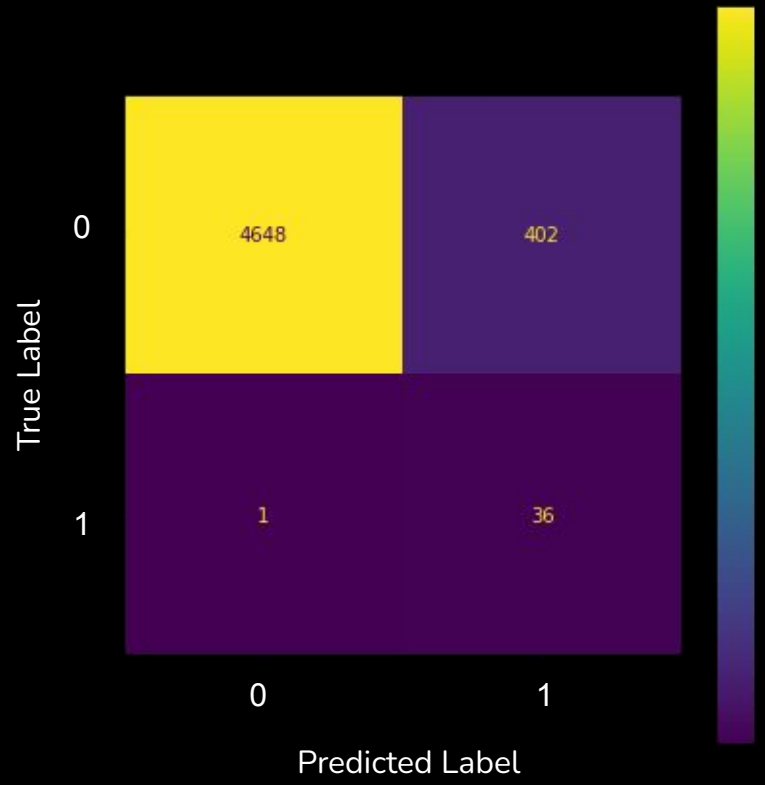
Example:
Before normalization [0, 1, 2, 3, 4, 5]
After normalization [0, 0.2, 0.4, 0.6, 0.8, 1]



SMOTE

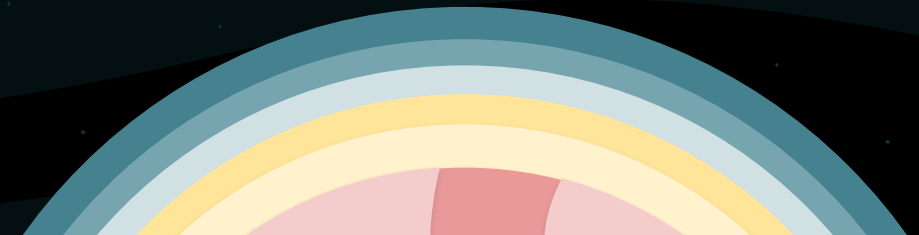
- Synthesizes new data based on existing data to balance the dataset
- This way the model does not always pick the majority class

# Confusion Matrix



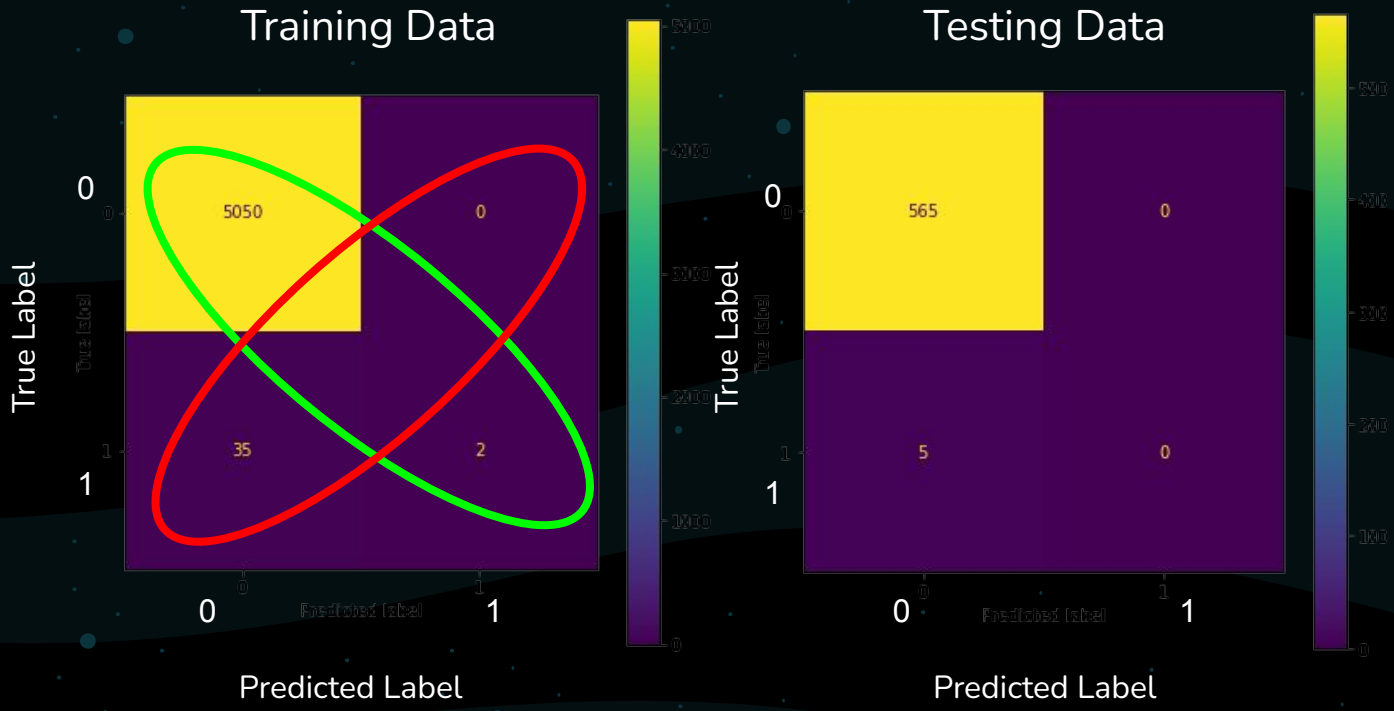

- Tells us how well the model performed
- Shows us correct and incorrect guesses
- When the predicted label matches the true label, we know the guess was correct!

# Different methods we tried



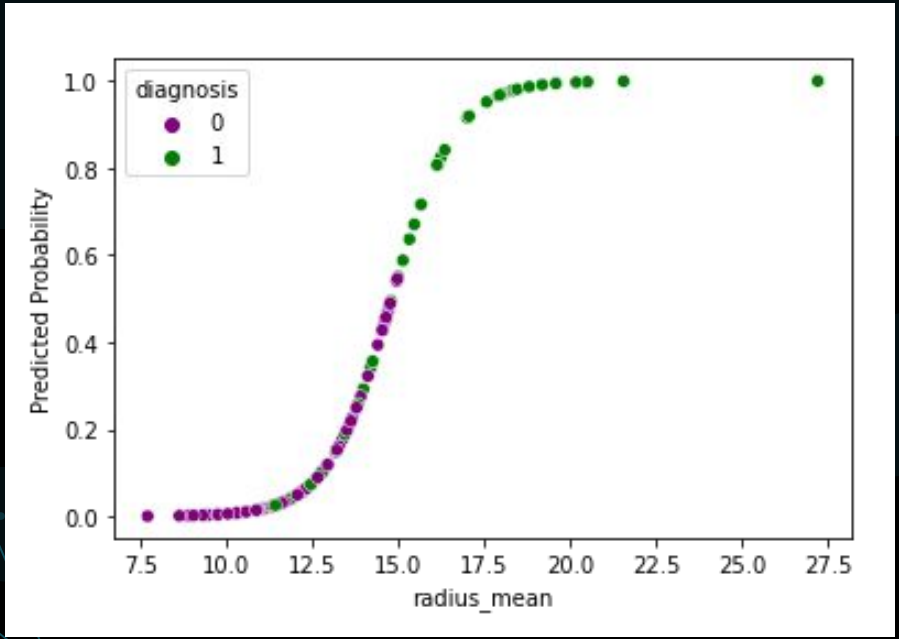
# Classification: KNN

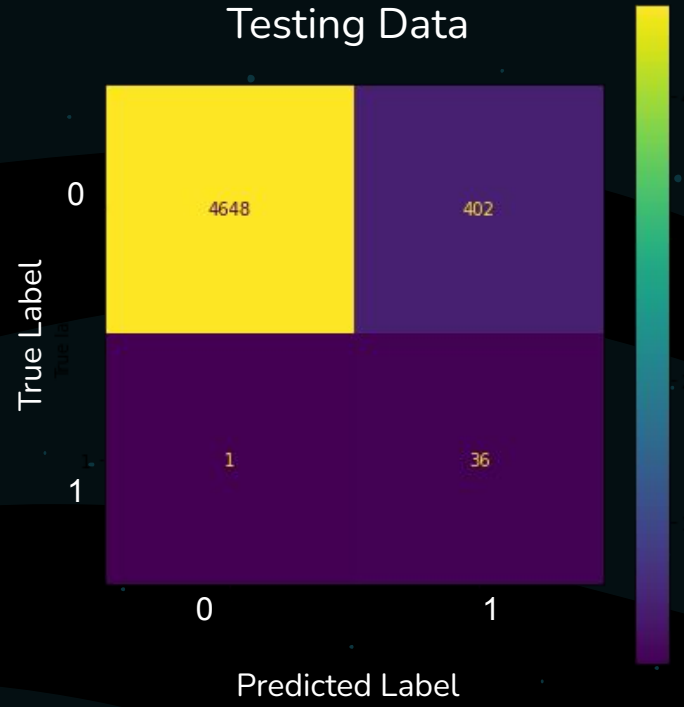
K-nearest neighbours uses proximity to classify the exoplanets vs. non-exoplanets since this method assumes that similar things exist in close proximity to each other.



# Classification: Logistic Regression

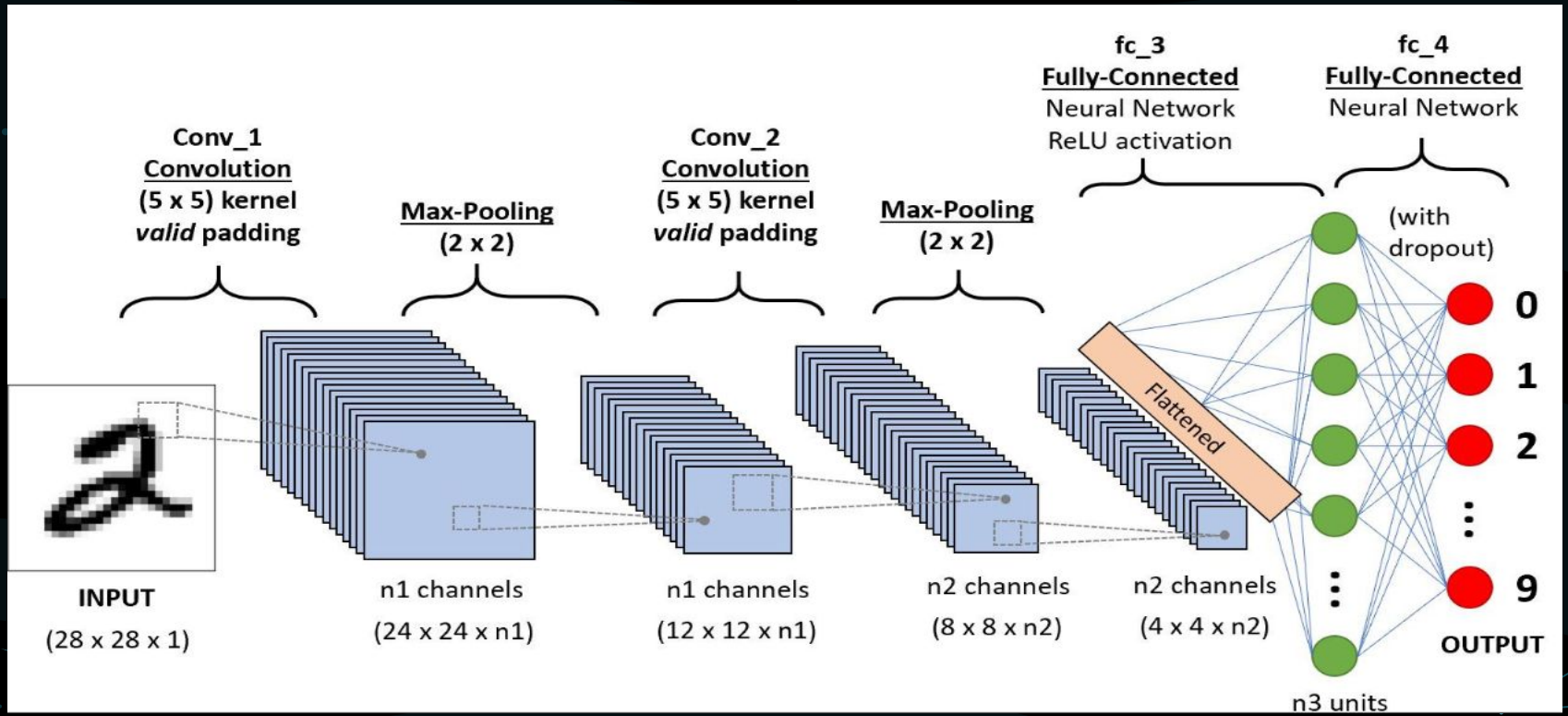
Estimates the probability of an event occurring based on a given dataset of independent variables.



# What Worked Best?

Convolutional Neural Network (CNN)



# How the model performed:

```
Epoch 20/20
111/111 [==============================] - 1s 6ms/step - loss: 0.0044 - accuracy: 0.9987 - val_loss: 0.0393 - val_accuracy: 0.9961
```