

# Searching For Exoplanets

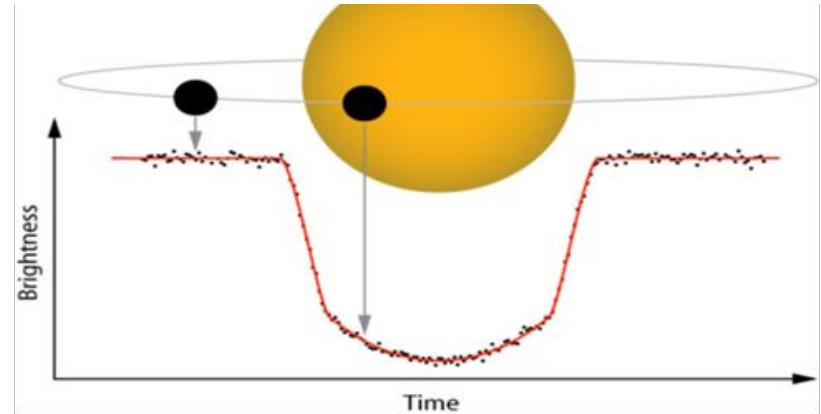
By: Aryaa, Elizabeth, Kaira, Paco, Sofie, Sourish, and Abhi

# Introduction: What are Exoplanets? - Kaira Rawert

- Exoplanets are planets orbiting a star other than our Sun.
- Exoplanets were first discovered in 1992.
- They vary in size and orbital period length.
- In this project we aimed to use machine learning and AI to search for exoplanets using transit photometry.

# Introduction: What are Exoplanets? - Kaira Rawert

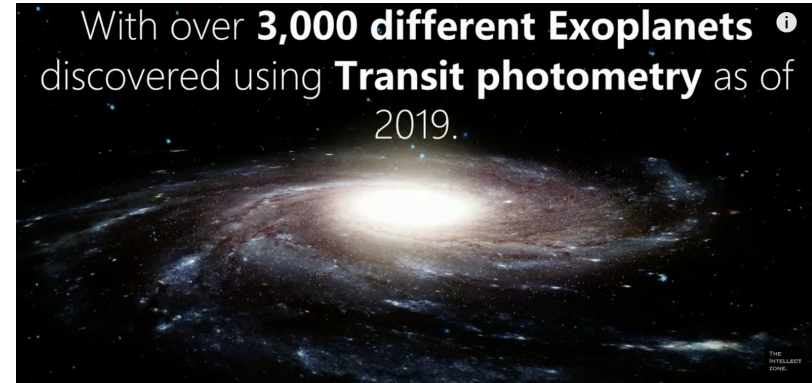
- Transit photometry: indirectly detecting exoplanets based on their effect on star brightness
- Exoplanet detection and confirmation can lead to further observations of the object.



# Task - Aryaa Mutha

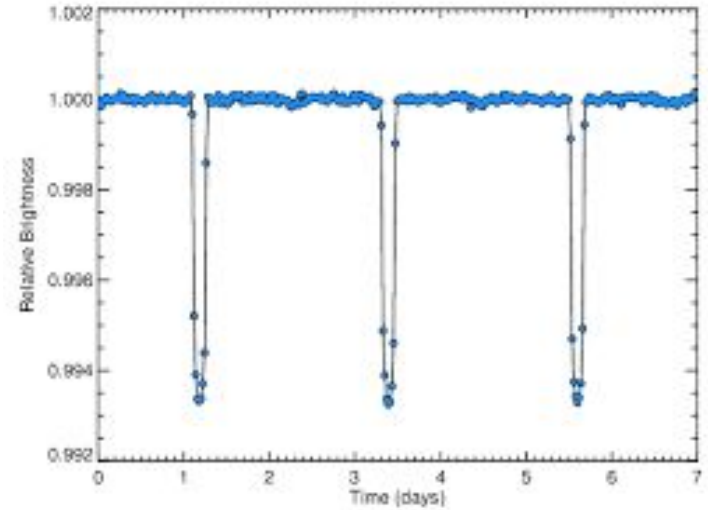
Using the power of AI to detect exoplanets in space

1. AI should be able to classify a star that has or does not have an exoplanet orbiting it
2. Using transit photometry data
3. Models used:
  - a. KNN (k-nearest-model)
  - b. Logistic Regression
  - c. Decision Trees
  - d. Convolution Neural Network (CNN)



# Data - Aryaa Mutha

1. Training Data:
  - a. Labeled data fed into AI which based the classifications
  - b. Labeled graphs from NASA
2. Testing Data
  - a. Unclassified data: AI would classify based from the training data
  - b. Unlabeled graphs: Classified as a star with or without an exoplanet

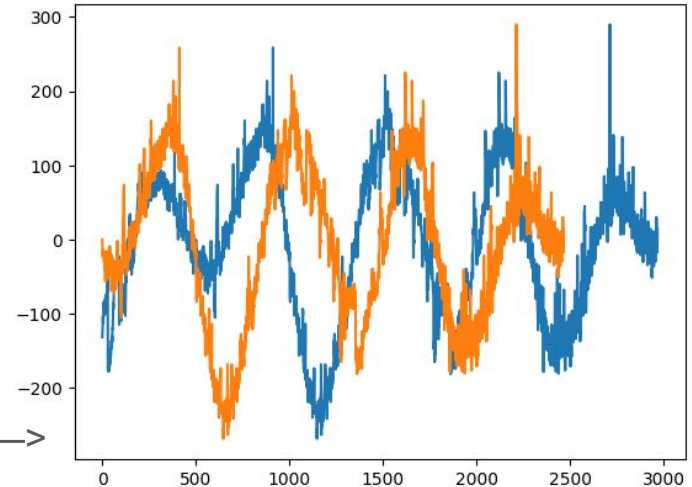


# Manual Efforts: Folding - Elizabeth Zhong

To manually determine if something is an exoplanet, we can use something called folding.





We do this by plotting all the periods on top of each other to see if there's a consistent trend.

We want to see if the dips occur roughly at the same places.



the two are overlaid —>

# Confusion Matrix Explanation - Performance Assessment

		Actual Values	
		1	0
Predicted Values	1	<b>TRUE POSITIVE</b> 	<b>FALSE POSITIVE</b>  <b>TYPE 1 ERROR</b>
	0	<b>FALSE NEGATIVE</b>  <b>TYPE 2 ERROR</b>	<b>TRUE NEGATIVE</b> 

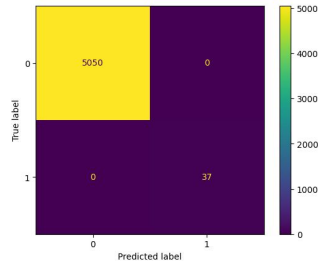
- Used to assess the accuracy of your model.
  - If the actual value = predicted, it is accurate.
  - If the actual value  $\neq$  predicted, it is inaccurate
- Using accuracy, precision, & recall
  - Accuracy = overall 'correctness' of the model
  - Precision = accuracy of positive predictions
  - Recall = ability of model to find all positive cases.
- Pros & Cons
  - Determine model's accuracy
  - Not good with multiclass
  - Can be misleading

# First Attempt Using ML Models - Paco Javier

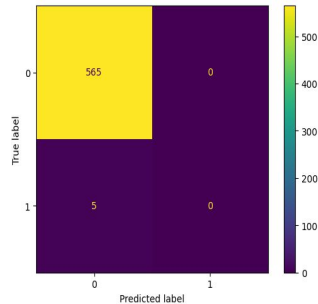
## KNN, Logistic Regression, and Decision Trees

In order to identify which stars have exoplanets, we can try using different ML models.

### KNN

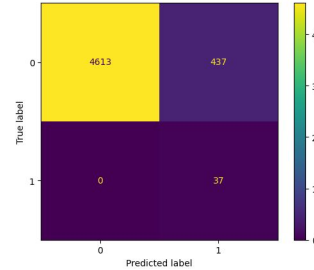


On the training data, the model has a 100% accuracy because this is what the model is trained on.

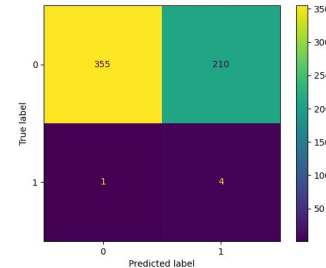


On the test data, it always predicts an exoplanet because the ratio of exoplanet to non-exoplanet is so high, and it can be 99% accurate if it just guesses non-exoplanet

### Logistic Regression



The Logistic Regression model specializes in classify different objects into categories, but the total accuracy of the training set is much lower than KNN

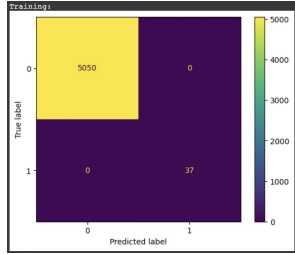


The test set accuracy is also much lower, but this model doesn't always choose to predict exoplanet at the cost of a lot of false positives.

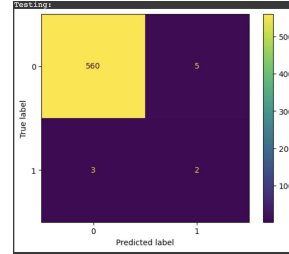


# First Attempt Using ML Models - Paco Javier

## Decision Tree



The Decision Tree model trained its “decisions” on the training model and thus, has a 100% accuracy just like the KNN model



Unlike the KNN and Logistic Regression model, the Decision Tree doesn't just guess exoplanet and doesn't have a lot of false positives.

# Data Augmentation - Normalization - Elizabeth Zhong

Because our range of flux values vary very wildly, we can use normalization to make our data more concise and easier.

In this case, we scaled our data so that all the points would lie between 0 and 1.

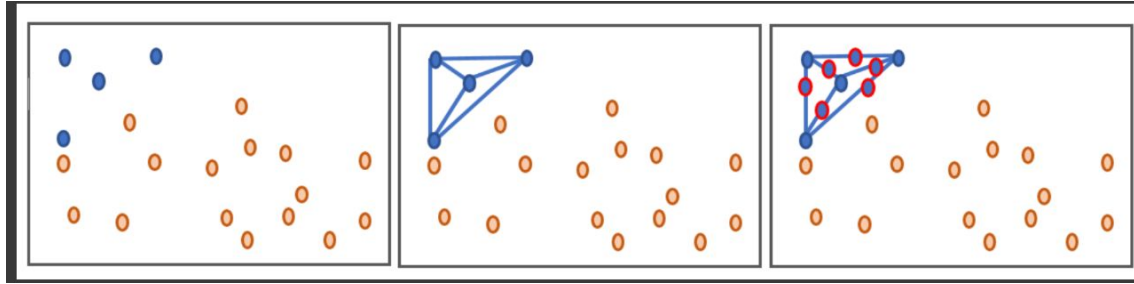
For example, if we were to normalize this list of values...

[0, 1, 2, 3, 4, 5]

The normalized list would be:

[0, 0.2, 0.4, 0.6, 0.8, 1]

# Data Augmentation - SMOTE - Abhi

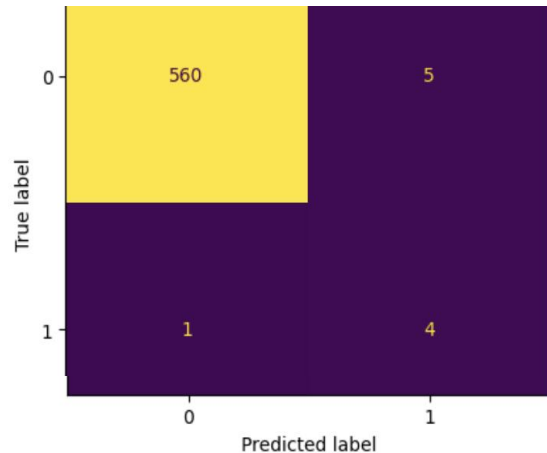


THE BLUE POINTS ARE CONFIRMED EXOPLANETS IN OUR HYPOTHETICAL SCENARIO. WHAT SMOTE DOES, IS THAT IT CONNECTS THE DOTS AND PLACES NEW SYNTHETIC DOTS AT ABOUT  $\frac{2}{3}$  OF THE DISTANCE FROM POINT TO POINT. THIS NEW ADDITION ASSISTS THE AI TO HELP US AS MUCH AS POSSIBLE TO HELP US MAKE MORE OBSERVATIONS.

# Second Attempt Using ML Models (With Data Augmentation) - Sourish Rikkala

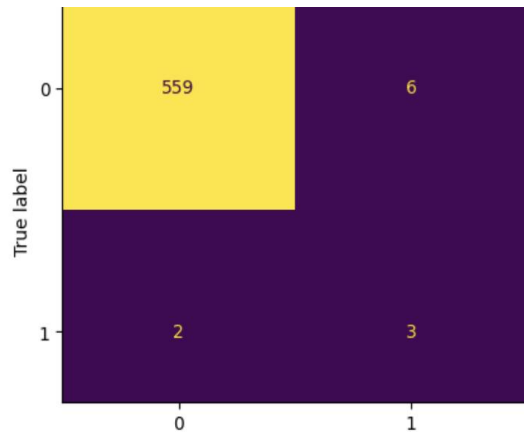
## Logistic Regression

- This model correctly predicted 99% of the non-exoplanets, which is better than the non-augmented model's 62%
- It correctly identified 4 of the 5 exoplanets, which is the same as the model without augmentation



## Decision Trees

- This model with augmented data correctly identified the same amount of non-exoplanets as the one without augmentation
- It identified 3 of the 5 exoplanets, which is better than the non-augmented model's 1 out of 5



# Neural Networks and CNNs- Abhi

What are Neural Networks and CNNs (Convolutional neural networks)?

A neural network is a replication of the human brain that uses nodes as neurons. They come to a conclusion using logic rather than a pre written program. Convolutional neural networks are networks that specialize in image recognition and classifications.

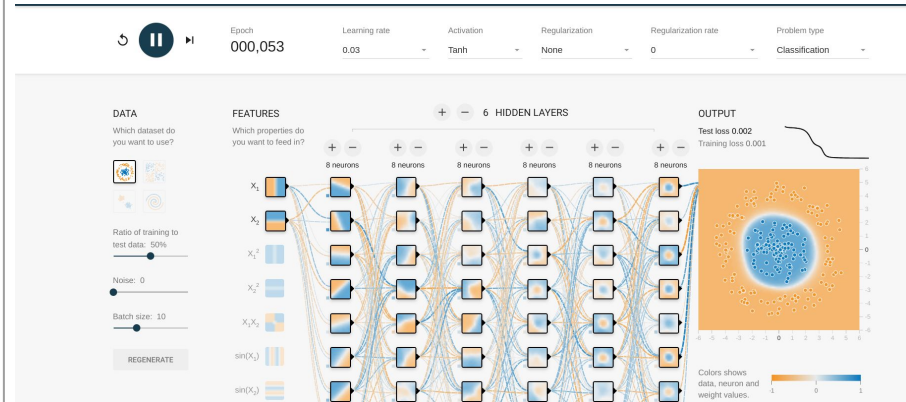
Where are these two useful?

## Neural Networks

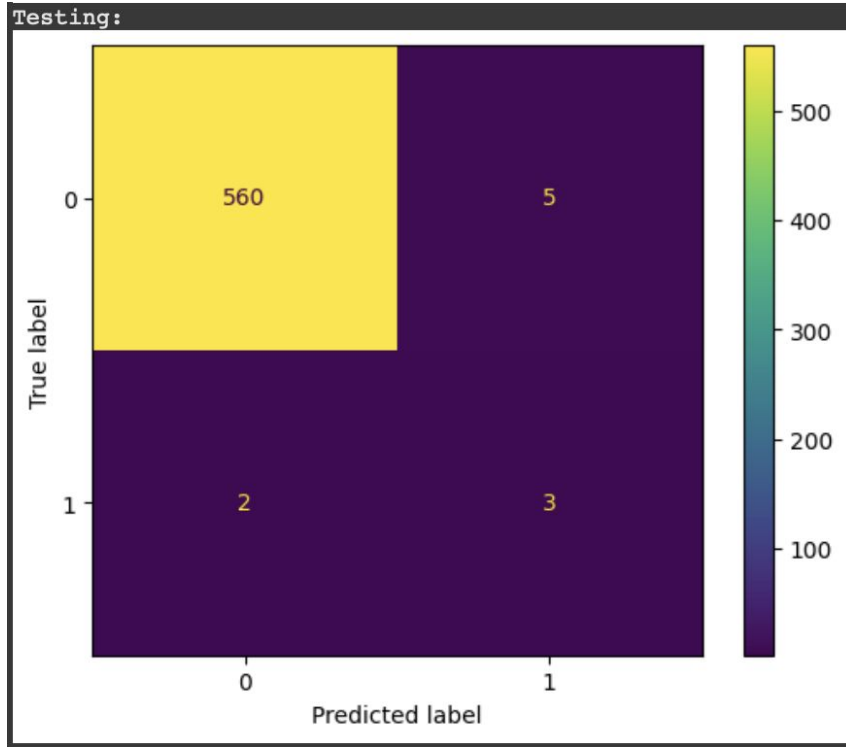
- Medical diagnosis
- Targeted Marketing
- Financial Predictions
- Self-driving cars
- Chat AI (Chat GPT, Gemini, Website Landing Page Helpers, etc)

## Convolutional Neural Networks

- Facial Recognition
- Image Classification
- Object Recognition
- Recommender Systems
- Health Risk Assessment
- Autonomous Systems



# How it applied to our project (NN and CNN continued)



AS WE CAN SEE, THE AI USED THE FLUX DATA TO PREDICT THE CLASSIFICATION TYPE OF THE DATA POINTS.

# Conclusion - Kaira Rawert

- In this project, we have learned how to train and test various AI models to detect exoplanets using transit photometry data.
- Models were assessed using a confusion matrix.
- Accuracy varied among methods.
- As we increase the complexity of our models(normalization, augmentation, hidden layers in CNNs), the prediction accuracy increases.