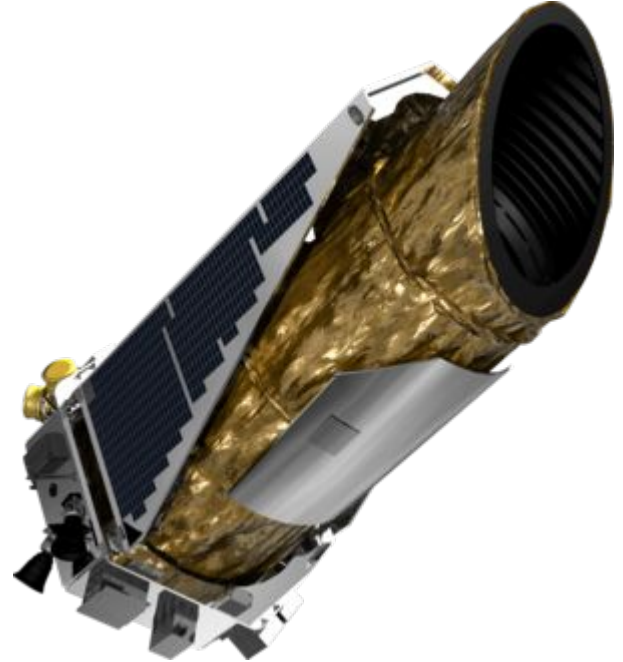


Classifying Observations from the Kepler Space Telescope

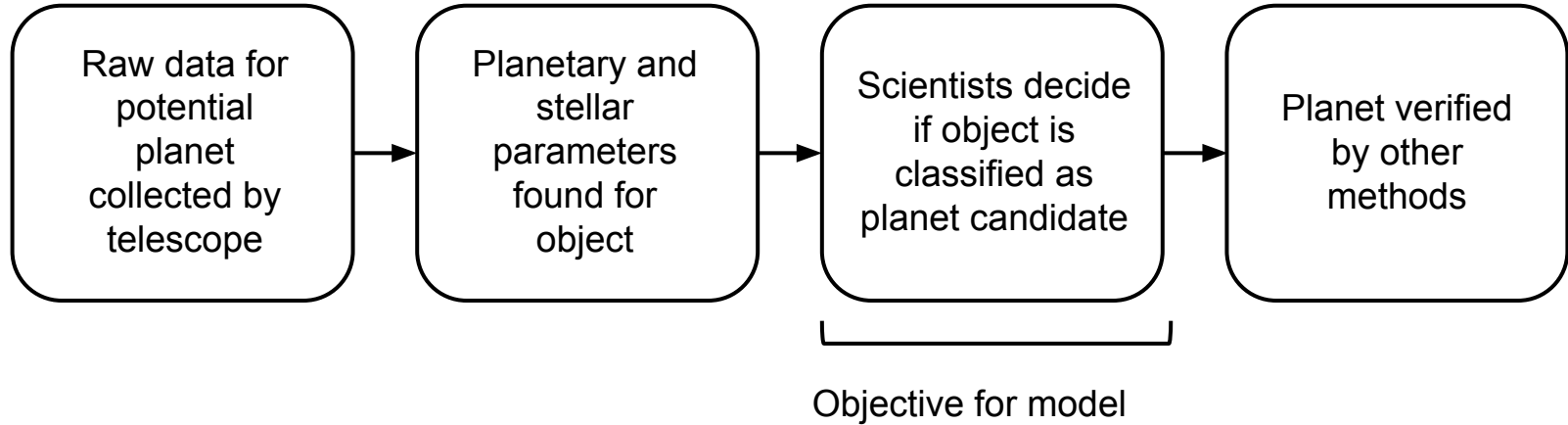
The background of the slide is a deep space image. It features a large, vibrant blue and white Earth in the center-right, showing swirling cloud patterns and the curvature of the planet. The rest of the background is a dark, star-filled space with faint blue nebulae and numerous small, distant stars.

Exoplanets and the Kepler space telescope

- Exoplanets: planets orbiting stars outside the Solar System
- Kepler space telescope: NASA mission to find Earth-size exoplanets
 - Found exoplanets by looking at dimming of stars as planets pass in front
 - Found over 1000 verified planets, ~30 potentially habitable planets

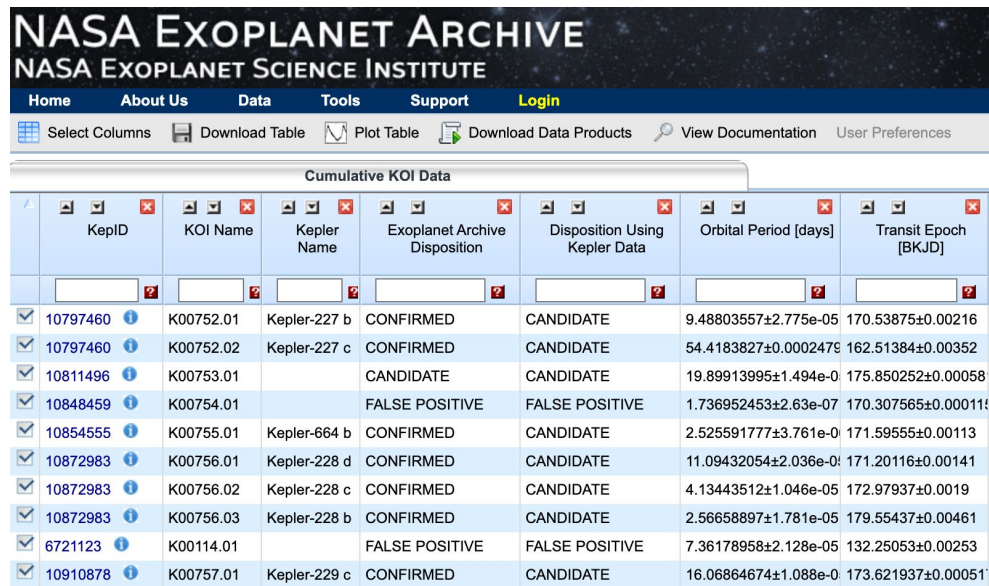


Objective: match human classification



Data source and features

- Data sourced from NASA exoplanet archive
- Columns used are fit parameters of light curves and the proportional size of the error bars on those fits
- Focusing on recall



The screenshot shows the NASA Exoplanet Archive website interface. At the top, the header reads "NASA EXOPLANET ARCHIVE" and "NASA EXOPLANET SCIENCE INSTITUTE". Below this is a navigation bar with links: Home, About Us, Data, Tools, Support, and Login. A secondary bar contains links for Select Columns, Download Table, Plot Table, Download Data Products, View Documentation, and User Preferences. The main content area displays a table titled "Cumulative KOI Data". The table has columns for KeplID, KOI Name, Kepler Name, Exoplanet Archive Disposition, Disposition Using Kepler Data, Orbital Period [days], and Transit Epoch [BKJD]. The table lists various exoplanets with their corresponding IDs, names, and dispositions.

KeplID	KOI Name	Kepler Name	Exoplanet Archive Disposition	Disposition Using Kepler Data	Orbital Period [days]	Transit Epoch [BKJD]
10797460	K00752.01	Kepler-227 b	CONFIRMED	CANDIDATE	9.48803557±2.775e-05	170.53875±0.00216
10797460	K00752.02	Kepler-227 c	CONFIRMED	CANDIDATE	54.4183827±0.0002479	162.51384±0.00352
10811496	K00753.01		CANDIDATE	CANDIDATE	19.89913995±1.494e-0	175.850252±0.00058
10848459	K00754.01		FALSE POSITIVE	FALSE POSITIVE	1.736952453±2.63e-07	170.307565±0.00011
10854555	K00755.01	Kepler-664 b	CONFIRMED	CANDIDATE	2.525591777±3.761e-0	171.59555±0.00113
10872983	K00756.01	Kepler-228 d	CONFIRMED	CANDIDATE	11.09432054±2.036e-0	171.20116±0.00141
10872983	K00756.02	Kepler-228 c	CONFIRMED	CANDIDATE	4.13443512±1.046e-05	172.97937±0.0019
10872983	K00756.03	Kepler-228 b	CONFIRMED	CANDIDATE	2.56658897±1.781e-05	179.55437±0.00461
6721123	K00114.01		FALSE POSITIVE	FALSE POSITIVE	7.36178958±2.128e-05	132.25053±0.00253
10910878	K00757.01	Kepler-229 c	CONFIRMED	CANDIDATE	16.06864674±1.088e-0	173.621937±0.00051

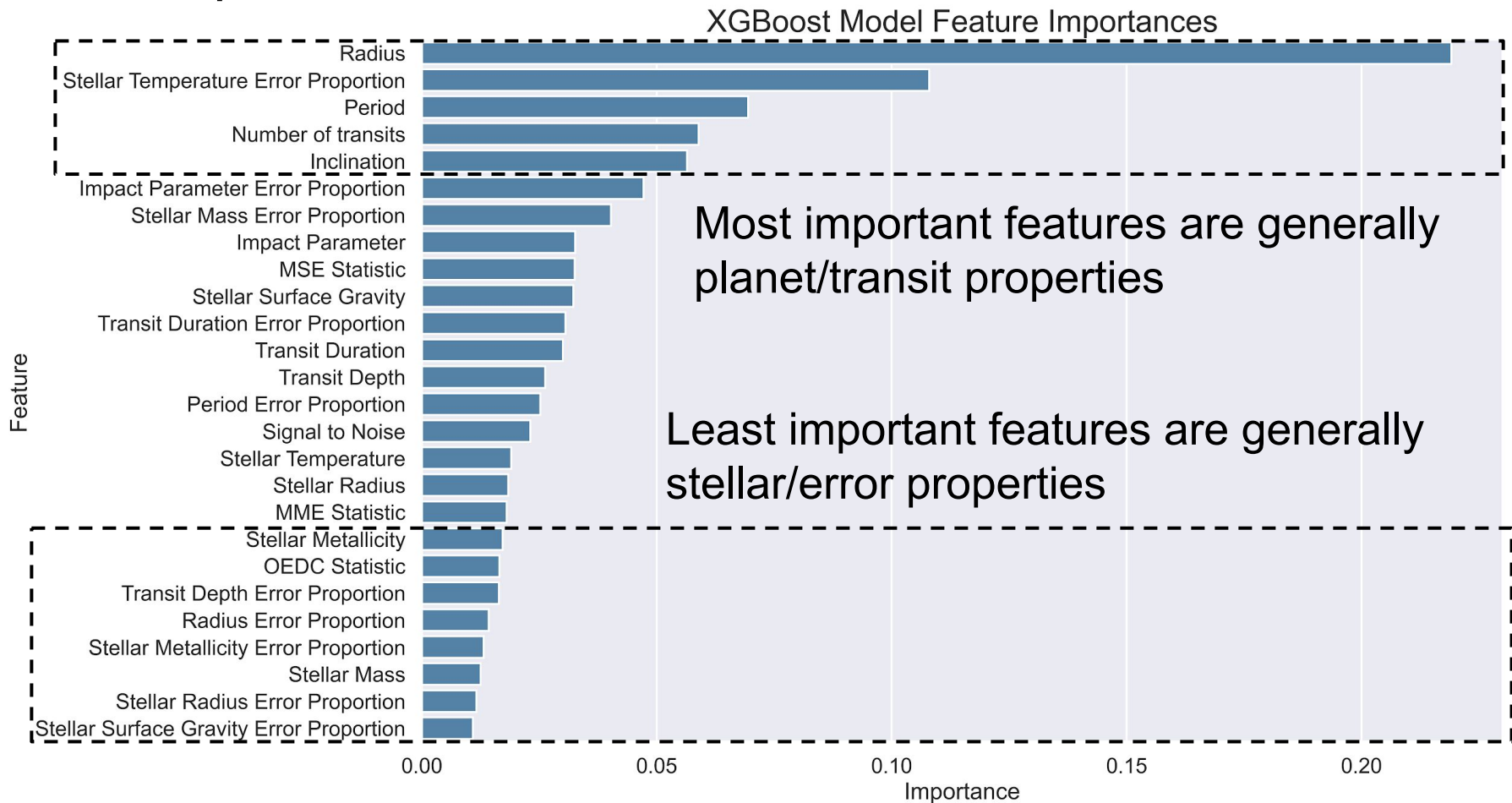
An XGBoost Classifier catches most Candidates

Test data confusion matrix

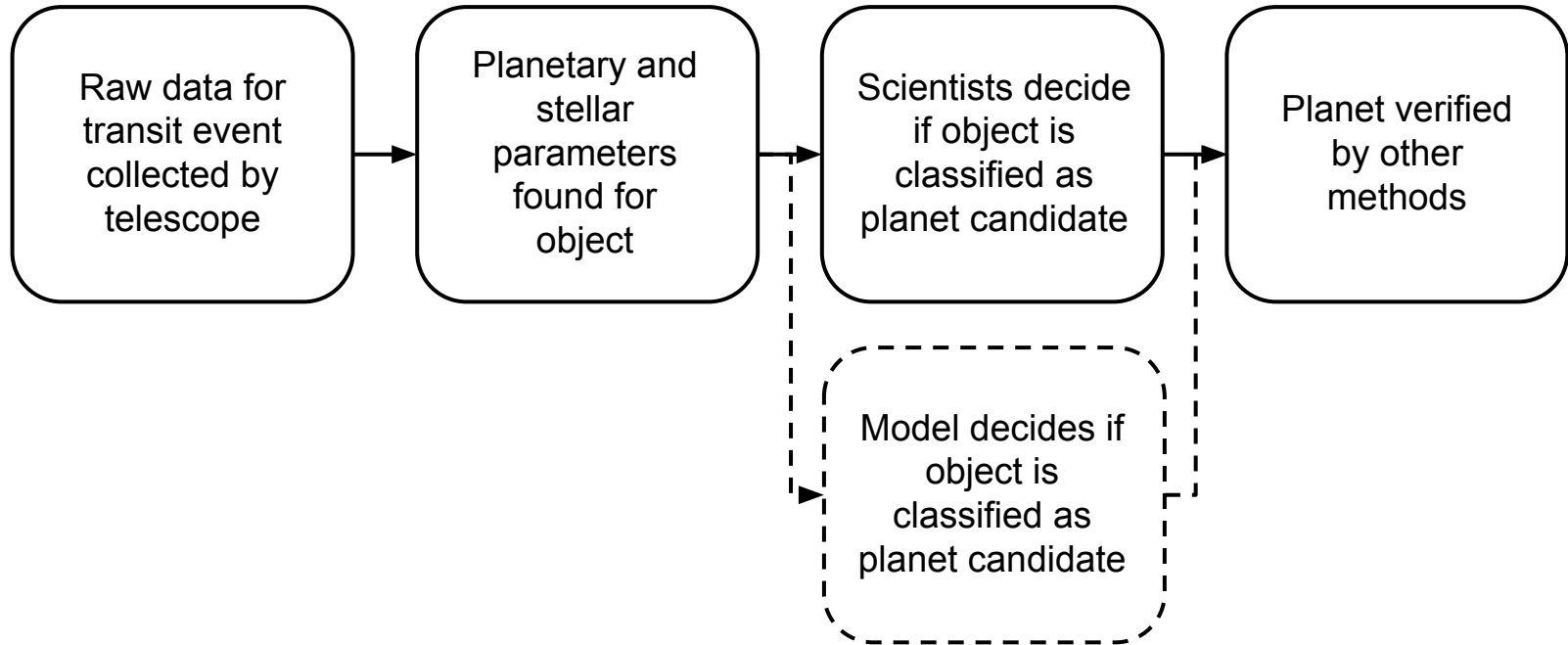
Actual	Candidate	Not Candidate
	Predicted	Predicted
Candidate	901 TP	91 FN
Not Candidate	123 FP	835 TN

Recall: 0.91

Most important features



Model has several possible use cases in planet finding



Model has several possible use cases in planet finding

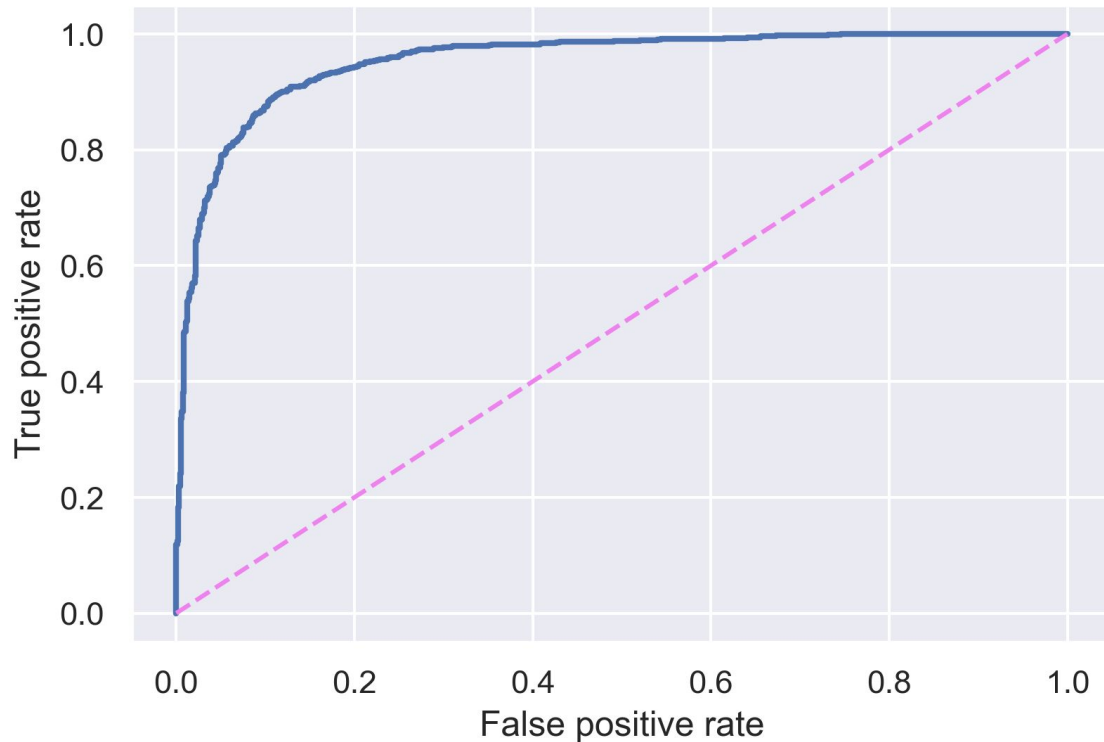
- Model could be used to prioritize human review or verification
- Feature importances could be used to look for unforeseen biases in observations or fitting methods

Conclusions and future directions

- ML models can classify exoplanet candidates with high recall
- Model can fit into several places in workflow, depending on what scientists want to optimize
- Future directions: Base model on raw light curve data, incorporate features that humans use to classify candidates

Thank you!

Other metrics



Accuracy: 0.89

Precision: 0.88

F1: 0.89

ROC AUC: 0.95