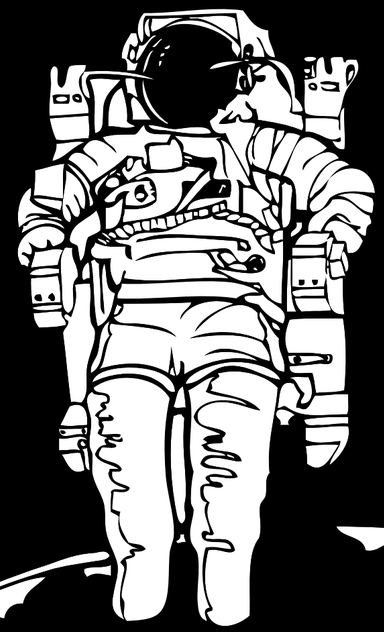
Dobby's Cult presents

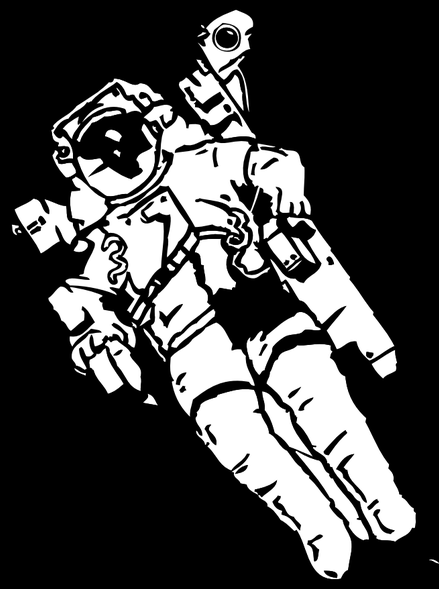
# A WORLD AWAY: HUNTING FOR EXOPLANETS WITH AI

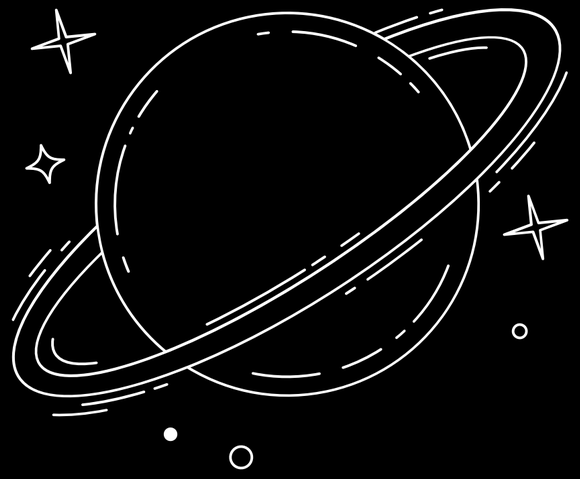


NASA Space App Challenge 2025

# OUR APPROACH

- **Data Foundation:** Utilized the Kepler Object of Interest (KOI) dataset provided by NASA.
- **Preprocessing & Feature Engineering:** Implemented a rigorous data cleaning pipeline and engineered new features like transit_shape to enhance predictive power.
- **Model Exploration:** Systematically evaluated a suite of powerful ensemble models, including RandomForest and LightGBM.
- **Intensive Optimization**: Employed RandomizedSearchCV for extensive hyperparameter tuning, with AI tools providing slight assistance in accelerating our coding workflow, which ultimately identified koi_score as the most influential feature.





# INITIAL CHALLENGES

**Significant Class Imbalance:**

The dataset was heavily skewed towards FALSE POSITIVE signals, risking a biased model. We addressed this with class weighting techniques during training

**Complex Signal Patterns:**

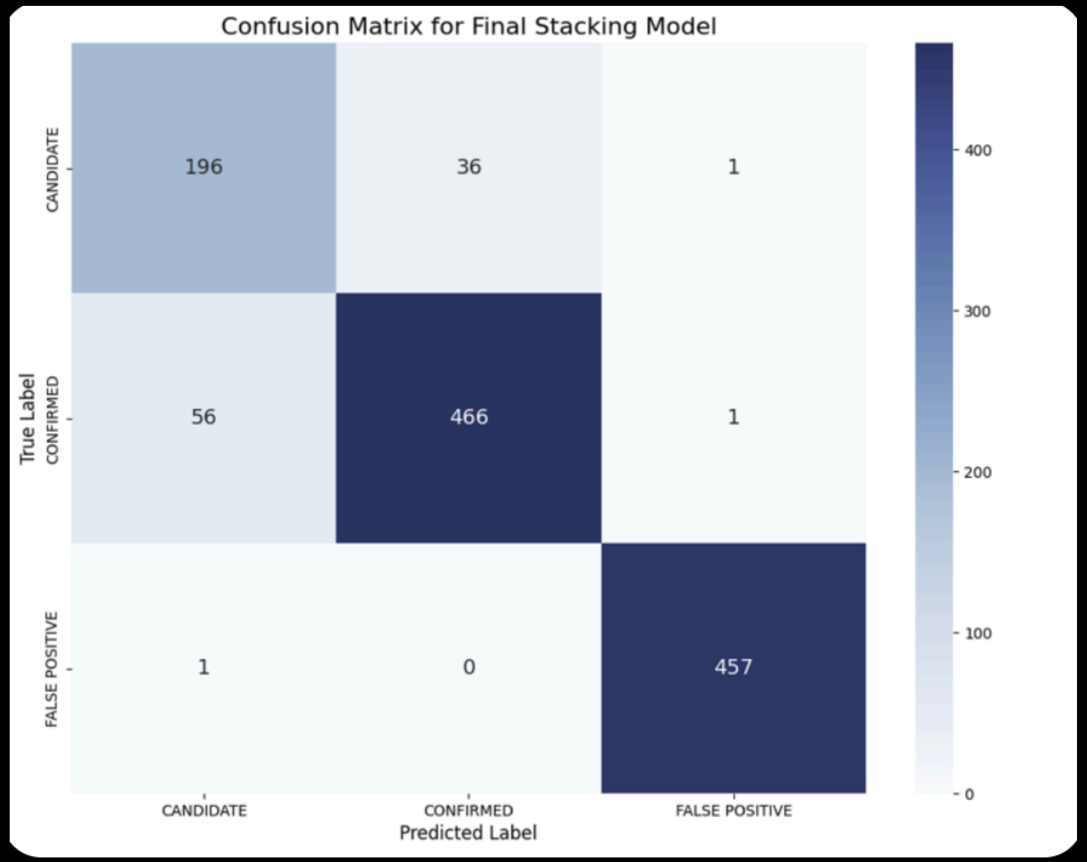
The initial features were insufficient. We performed feature engineering to better capture the nuanced signatures of a true planetary transit.

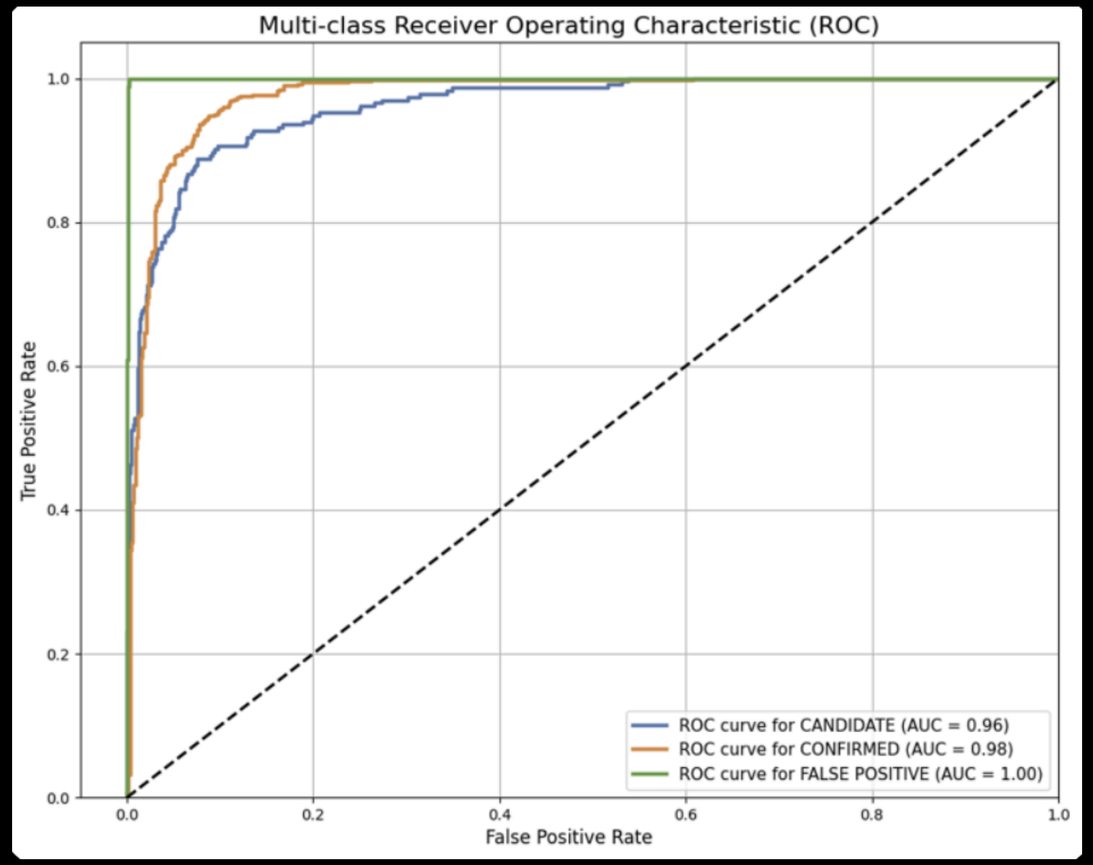
**Optimal Model Selection**:

With many high-performing models, identifying the best approach required a direct, systematic comparison of multiple advanced algorithms.

# OUR FINAL SOLUTION

- Our final solution is a Stacking Classifier, an advanced ensemble that integrates multiple fine-tuned models to achieve superior accuracy and robust generalization.

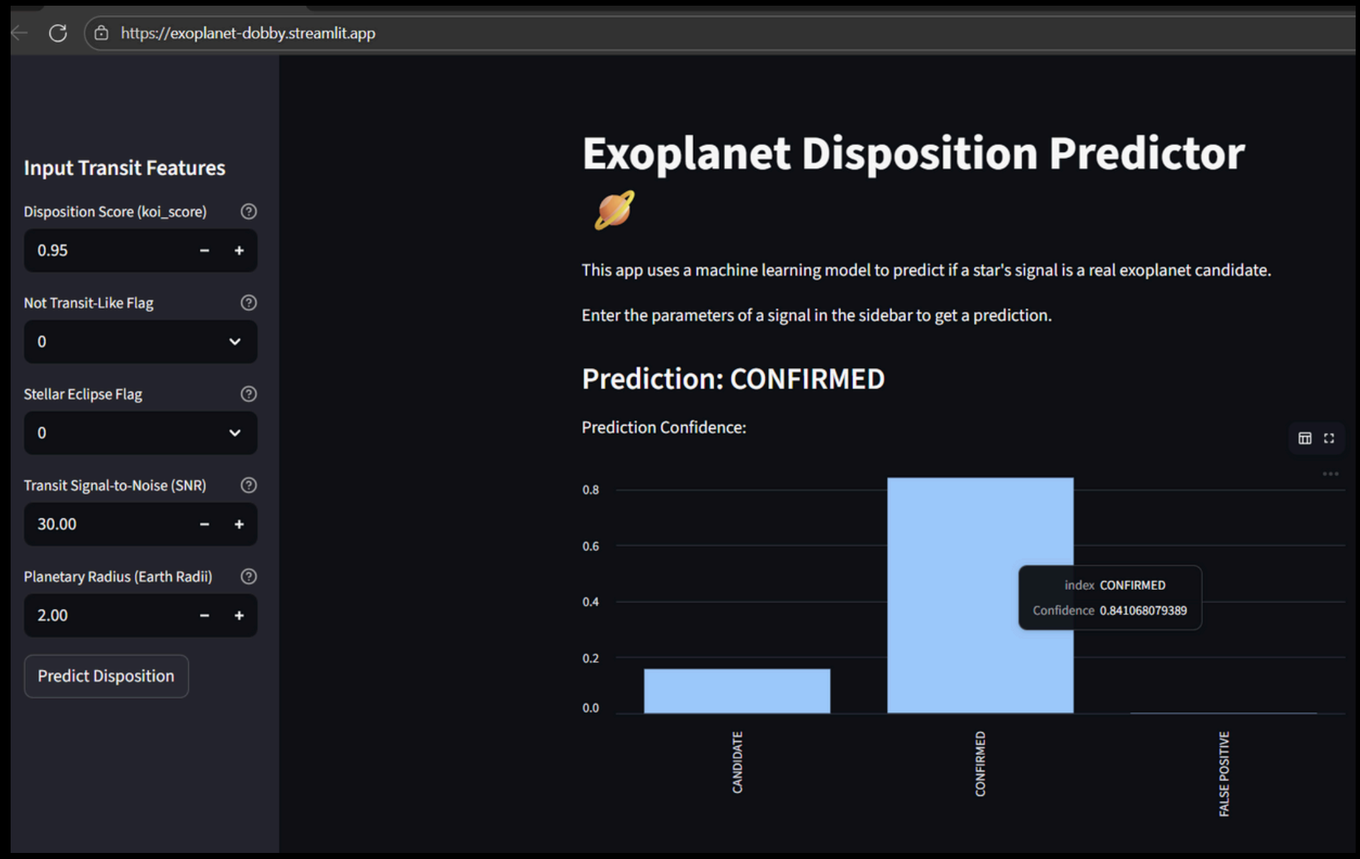


# USER INTERFACE

To make our model accessible, we developed an interactive web application for on-demand classification.

- **Input Data**: The user enters key observational data for a signal (e.g., koi_score, SNR) into a simple web interface.
- **Initiate Prediction**: The user clicks the 'Predict Disposition' button.
- **Receive Instant Results**: The application displays a clear, readable classification: CANDIDATE, CONFIRMED, or FALSE POSITIVE.
- **Review Confidence**: A bar chart shows the model's confidence probability for each of the three possible classes.

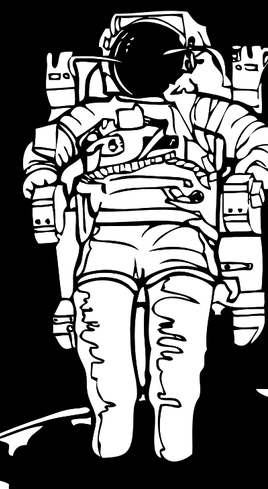


# FUTURE IMPROVEMENTS

- **Incorporate New Data**: Enhance model robustness by training on more diverse datasets from missions like NASA's TESS.

- **Explore Deep Learning**: Investigate Convolutional Neural Networks (CNNs) to analyze raw transit light curves directly and uncover more complex patterns.

- **Enhance Application Functionality:** Expand the web app to support batch processing, provide model interpretability results (e.g., SHAP plots), and integrate with live astronomical databases.

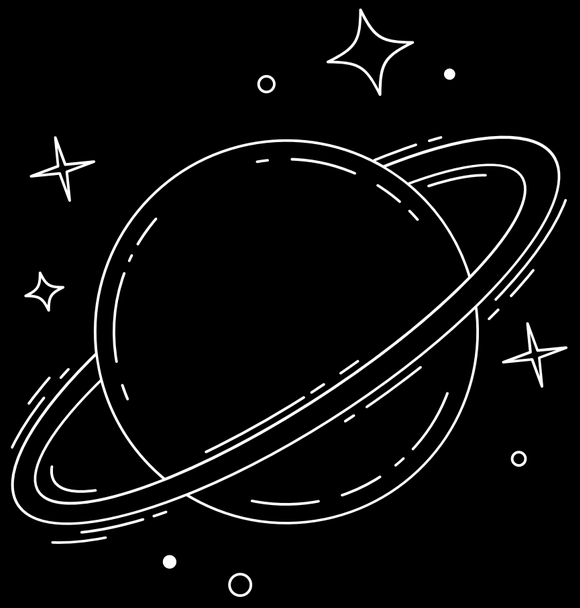


# MEET THE CULT

Sharad Yadav

Smit Ashvinbhai Shingala

Parigna Bhavik Rathod



# THANK YOU !