

---

---

# ExoClassify

## Unveiling the Unseen Universe

*ExoClassify: Exoplanet Classification*

**Team Name:** Exo\_worlds

**Members:** Shreya S Suranagi

**Local Event:** Harohalli, (Karnataka, India)

**Challenge Chosen:** "A World Away: Hunting for Exoplanets with AI"

---

# CHALLENGE

- **What challenge does this project address?**

Current exoplanet classification relies heavily on manual analysis of complex Kepler data. I was inspired to use AI/ML to improve and accelerate classification using public NASA data, particularly Kepler data



# SOLUTION

## Project Summary:

I built a machine learning web app that predicts whether a celestial object is a confirmed exoplanet, a candidate, or a false positive — using Kepler mission data and SHAP explainability to improve transparency and trust in the AI's decisions and uses Ensemble algorithms

RandomForest Accuracy: 0.8447

Classification Report for RandomForest:

	precision	recall	f1-score	support
CANDIDATE	0.84	0.84	0.84	943
FALSE POSITIVE	0.85	0.85	0.85	970
accuracy			0.84	1913
macro avg	0.84	0.84	0.84	1913
weighted avg	0.84	0.84	0.84	1913

Confusion Matrix for RandomForest:

```
[[792 151]
 [146 824]]
```

AdaBoost Accuracy: 0.8129

Classification Report for AdaBoost:

	precision	recall	f1-score	support
CANDIDATE	0.79	0.85	0.82	943
FALSE POSITIVE	0.84	0.78	0.81	970
accuracy			0.81	1913
macro avg	0.81	0.81	0.81	1913
weighted avg	0.82	0.81	0.81	1913

Confusion Matrix for AdaBoost:

```
[[802 141]
 [217 753]]
```

GradientBoosting Accuracy: 0.8406

Classification Report for GradientBoosting:

	precision	recall	f1-score	support
CANDIDATE	0.83	0.85	0.84	943
FALSE POSITIVE	0.85	0.83	0.84	970
accuracy			0.84	1913
macro avg	0.84	0.84	0.84	1913
weighted avg	0.84	0.84	0.84	1913

Confusion Matrix for GradientBoosting:

```
[[804 139]
 [166 804]]
```

XGBoost Accuracy: 0.8453

Classification Report for XGBoost:

	precision	recall	f1-score	support
CANDIDATE	0.83	0.86	0.85	943
FALSE POSITIVE	0.86	0.83	0.85	970
accuracy			0.85	1913
macro avg	0.85	0.85	0.85	1913
weighted avg	0.85	0.85	0.85	1913

Confusion Matrix for XGBoost:

```
[[808 135]
 [161 809]]
```

Predicted class: FALSE POSITIVE

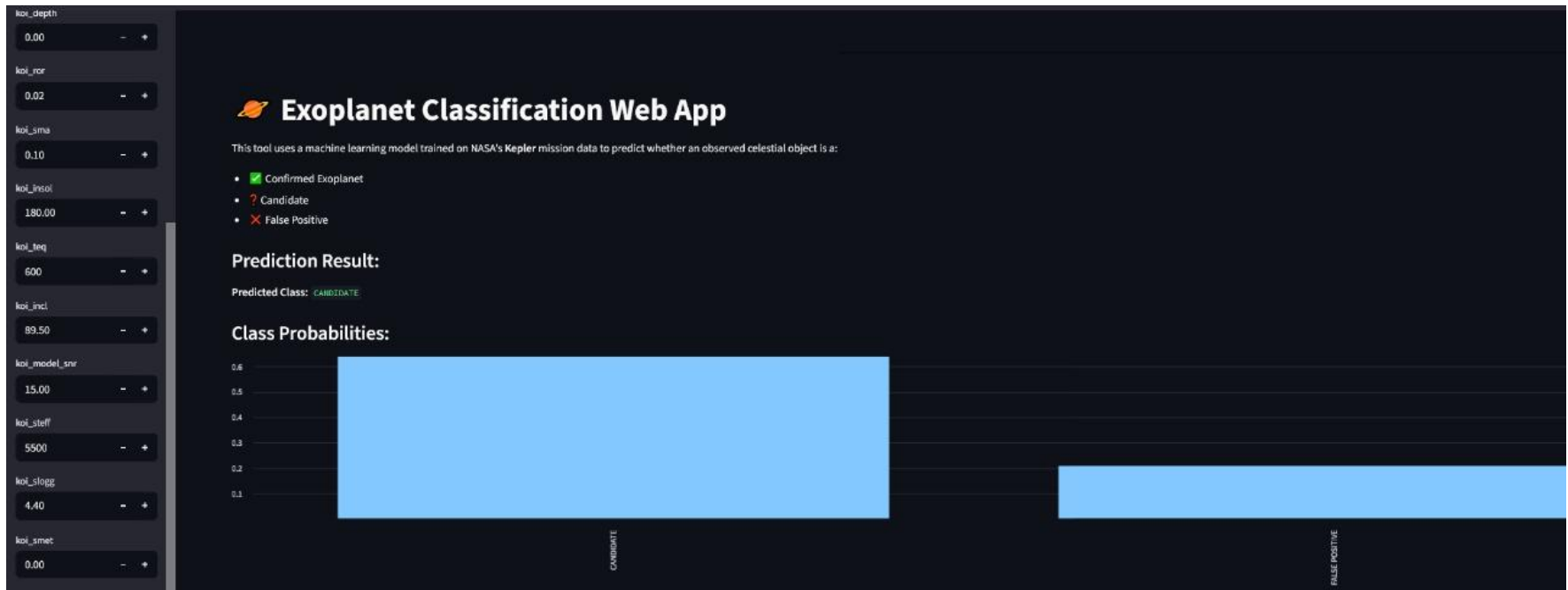
Class probabilities: [0.04603916 0.95396084]

Predicted class: CANDIDATE

Class probabilities: [0.79173326 0.2082667 ]

# HOW IT WORKS

Built using XGBoost + SHAP for predictions and explanations. Frontend built with Streamlit for easy use. CSV upload feature allows batch predictions.



# IMPACT AND FUTURE VISION

- **Looking Ahead:**
  - Improve accessibility for citizen scientists
  - Expand to other missions (e.g., TESS)
  - Enable anomaly detection for unknown planetary types
  - Help prioritize follow-up observations for promising candidates
- “Empowering planetary discovery, one prediction at a time.”



# USER EXPERIENCE



## Workflow:

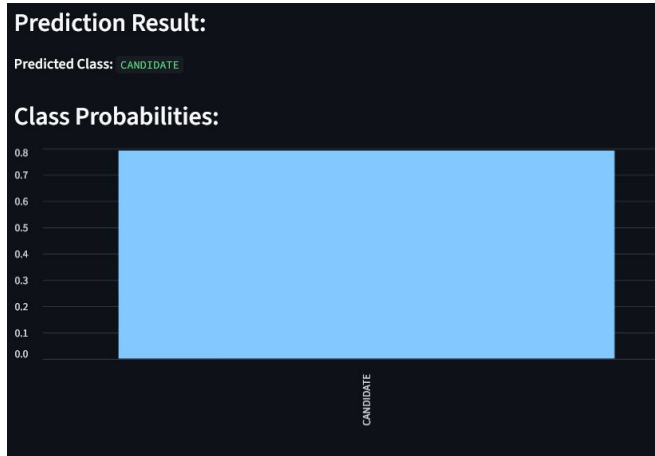
Input planet/star features manually or via CSV

Click "Predict"

View prediction + confidence

See SHAP feature importance

Download results as CSV



koi\_steff

5500 - +

koi\_slogg

4.40 - +

koi\_smet

0.00 - +

koi\_srad

1.00 - +

koi\_smass

1.00 - +

Predict from Input

## Batch Predictions via CSV Upload

Upload a CSV file with the same columns used by the model:

Upload your CSV



Drag and drop file here

Limit 200MB per file • CSV

# USE of NASA data and AI

- **NASA Data Used:**
- *Kepler Object of Interest (KOI)* catalog from the [NASA Exoplanet Archive](#)
- **Space Agency / Other Data:**
- Derived feature: `period_prad` ( $\text{period} \times \text{planet radius}$ )
- Data preprocessed with imputation and label encoding
- **AI Use:**
- Code for ensemble and SHAP
- Image generation in Slide 2