Exo-AI Explorer – Project Documentation

AI-Powered Exoplanet Classification Platform Using NASA Mission Data

✅ PROJECT OVERVIEW

Objective

Exo-AI Explorer automates the classification of exoplanet candidates using high-fidelity mission data from NASA's Kepler, K2, and TESS observatories. It replaces weeks of manual vetting with real-time AI inference.

Core Innovation

A custom-built ensemble machine learning system integrates four optimized classifiers and exposes predictions via a browser-based web interface.

Target Users

- NASA mission scientists & research teams
- Astronomy students and academic institutions
- Citizen scientists and discovery volunteers

✅ TECHNICAL ARCHITECTURE

Data Pipeline

Data Sources

- Kepler Confirmed Planets
- K2 Candidate Catalogues
- TESS Objects of Interest

Total Observations Processed: 206,065 entries

Robust Cleaning & Scaling

- Invalid/missing values replaced (fillna(0), ∞→0)
- Outlier resistance via RobustScaler
- Strict column validation before training

Feature Engineering (8 Scientific Predictors)

1. Planetary Radius
2. Orbital Period
3. Transit Duration
4. Transit Depth
5. Stellar Temperature
6. Equilibrium Temperature
7. Stellar Gravity
8. Confidence Score (detection reliability)

Each feature directly corresponds to entries used in real NASA catalog modeling.

✅ ENSEMBLE MACHINE LEARNING SYSTEM

Referencing scientific findings showing ensemble dominance (Malik et al., 2022; Priyadarshini & Puri, 2021), the final architecture uses:

Model Components

✅ **Random Forest (Primary)**

- 200 estimators
- max_depth=25
- min_samples_split=8
- min_samples_leaf=3
- class_weight="balanced"

✅ **Extra Trees Classifier**

- 150 estimators
- max_depth=20
- class_weight="balanced"

✅ **Gradient Boosting**

- 100 estimators
- max_depth=5
- learning_rate=0.1
- subsample=0.8

✅ **Conservative Random Forest**

- 100 estimators
- max_depth=15
- min_samples_split=12
- class_weight="balanced"

Voting Strategy

Weighted soft voting mechanism:

- RF Optimized → 1.4
- Extra Trees → 1.3
- Gradient Boosting → 1.2
- Conservative RF → 1.0

This structure improves interpretability and model stability across rare classes.

✅ **DATA BALANCING & BIAS MITIGATION**

Challenge

Overrepresentation of Hot Jupiter and underrepresentation of terrestrial planets.

Mitigation Measures

- **Stratified train-test split**
- **Class weighting** across all ensemble estimators
- **Synthetic oversampling** for rare planetary types
- **Balanced accuracy metrics** focus

✅ MODEL VALIDATION

Validation Strategy

- 3-Fold StratifiedKFold validation
- Manual iteration (not automated wrappers)
- Avoids compatibility issues with custom ensembles

Metrics

- **Balanced Accuracy**: 80%+
- **Per-Class Recall**:
  - False Positives → 85%
  - Planetary Candidates → 82%
  - Confirmed Exoplanets → 89%

A live confusion matrix and classification report are generated during training.

✅ WEB INTERFACE (GRADIO)

Real-Time Classification

- 7 interactive slider for key planetary parameters
- Confidence scoring built into outputs

User Experience

- Preloaded presets:
  - Earth-analog
  - Hot Jupiter
  - False Positive case
- Color-coded result visualization
- Immediate inference on CPU hardware

Deployment

- Hosted on Hugging Face Spaces

- Browser-only, no installation needed
- Public URL for global access

✅ PERFORMANCE & EXPORT SYSTEM

Model Persistence

- Saved using joblib with timestamp
- Exported as .pkl
- Supports reload & deployment

Metadata Recording

JSON export includes:

- Final balanced accuracy
- Cross-validation mean/std
- Feature names
- Class distribution
- Model weights

Quality Messaging

- ≥0.85 accuracy → "Production-ready"
- ≥0.75 → "Research-grade"
- <0.75 → "Eligible for retraining"

✅ RESEARCH FOUNDATION

Scientific grounding includes:

- **Ensembles outperform single classifiers**
  (Malik et al., 2022; Priyadarshini & Puri, 2021)
- **Feature-based prediction over deep learning**
  Favored for transparency and data limits

- **Transfer learning in astronomy**
  Inspires architecture adaptability for future missions

✅ INNOVATION CONTRIBUTIONS

1. **Accessibility**
   Publicly accessible AI for planetary science
2. **Efficiency**
   Reduces exoplanet vetting time from weeks to seconds
3. **Educational Impact**
   Practical learning tool for astrophysics students
4. **Scalability**
   Easily expandable for PLATO, Roman, and LUVOIR missions

✅ FUTURE ENHANCEMENTS

- Real-time ingestion from active TESS alerts
- 3D orbital and system visualizations
- Mobile app deployment
- Atmospheric composition as new feature class
- Integration with JWST follow-up data

Thank you for time to read this document for better understanding of the project.