NASA Space Apps - BlackEye Exoplanet Detection System

Project Documentation

---

1. High-Level Project Summary

What We Developed

We developed **BlackEye**, a machine learning system that processes NASA's Kepler and TESS datasets to identify and characterize exoplanets using XGBoost classification and regression models.

How It Addresses the Challenge

BlackEye automates the analysis of light curve data from NASA's Kepler and TESS missions, reducing manual review time and improving classification accuracy.

Why It's Important

This system accelerates exoplanet discovery by providing automated analysis tools that can process NASA's datasets in real-time, making NASA data more accessible to the research community.

---

2. Project Details

What Exactly Does It Do?

BlackEye processes NASA's Kepler and TESS datasets to identify and characterize exoplanets using XGBoost to classify objects as CONFIRMED, CANDIDATE, FALSE_POSITIVE, or UNKNOWN while predicting scientific parameters.

How Does It Work?

The system provides a FastAPI backend for real-time analysis and web interface for visualization with XGBoost models for classification and parameter prediction.

What Benefits Does It Have?

- Accelerated discovery of exoplanet candidates
- Scientific accuracy in classification and parameter estimation

- Makes NASA data more accessible to researchers
- Reduces manual review time from days to minutes

What Do We Hope to Achieve?

- Contribute to exoplanet research and planetary system understanding
- Provide tools for astronomers and researchers worldwide
- Advance the field of automated exoplanet detection

Tools, Coding Languages, Hardware, and Software Used

Machine Learning: Python3, XGBoost, Scikit-learn, Pandas, NumPy, Matplotlib, Scipy **Backend:** FastAPI, Uvicorn, Docker **Frontend:** PHP, HTML5, CSS3, JavaScript **Development:** GitHub, Git

---

3. NASA Data

Specific NASA Data Sources Used

NASA Kepler Objects of Interest (KOI) Dataset - 9,564 objects from NASA's Kepler space telescope **NASA TESS Objects of Interest (TOI) Dataset** - 7,668 objects from NASA's TESS mission

How We Used NASA Data

1. **Merged Dataset**: Combined Kepler KOI and TESS TOI into unified dataset of 17,232 astronomical objects
2. **Feature Engineering**: Extracted 12 raw features plus 3 physics-based engineered features
3. **Machine Learning**: Multi-target learning with XGBoost for classification and regression
4. **Model Performance**: Achieved 74% classification accuracy across all classes

How NASA Data Inspired Our Project

The NASA data revealed the challenge of processing vast amounts of light curve data, which previously required manual analysis by expert astronomers. This inspired us to create an automated framework that reduces manual review time while improving classification accuracy.

---

4. Space Agency Partner & Other Data

NASA Datasets and Data Sources

- **NASA Kepler Objects of Interest (KOI) Dataset** - 9,564 objects from NASA's Kepler space telescope
- **NASA TESS Objects of Interest (TOI) Dataset** - 7,668 objects from NASA's TESS mission

Python Technologies and Libraries

- **XGBoost** - Core machine learning algorithm
- **Scikit-learn** - Data preprocessing and metrics
- **Pandas** - Data manipulation
- **NumPy** - Numerical operations
- **Matplotlib** - Data visualization
- **Scipy** - Scientific computing
- **FastAPI** - Python backend API server
- **Uvicorn** - ASGI server
- **Docker** - Containerized deployment

Web Technologies and Frontend

- **PHP** - Server-side web development
- **HTML5** - Web page structure
- **CSS3** - Styling and responsive design
- **JavaScript** - Interactive functionality
- **GitHub** - Version control
- **Git** - Distributed version control

---

5. Use of Artificial Intelligence (AI)

AI Tools Utilized in This Project

- **Cursor AI**: Used for thinking and planning during development phase
- **ChatGPT**: Utilized for brainstorming ideas and project structure planning