Problem 01: Analysis of Exoplanets Based on Orbital Time Period

Approach to Dataset and Histogram Generation

1. **Data Collection**:
    - Accessed the exoplanet catalog via the provided link.
    - Filtered data to include only confirmed exoplanets detected using the Radial Velocity (RV) method.
    - Exported relevant parameters, particularly the orbital period (in Earth days).
2. **Data Cleaning**:
    - Removed entries with missing or erroneous orbital period data.
    - Limited the orbital period range to 0–400 Earth days to focus on short to moderately long-period planets, avoiding skewness from extreme outliers.
3. **Histogram Generation**:
    - Used Python's `matplotlib` and `seaborn` libraries for visualization.
    - Binned data into intervals of 10 days, plotted the histogram, and added labels for clarity.

Key Insights

1. **Concentration of Short-Period Exoplanets**:
    - Most exoplanets detected via the RV method exhibit orbital periods less than 50 days. This is due to:
        - **Observational Bias**: Closer planets cause larger and more frequent Doppler shifts, making them easier to detect.
        - **Detection Sensitivity**: Short-period planets induce stronger gravitational tugs, producing higher velocity amplitudes.
2. **Scarcity of Long-Period Exoplanets**:
    - Longer orbital periods result in weaker Doppler shifts and require more extended observation times to detect multiple cycles.
    - Current technology limits and observational time constraints hinder the detection of Earth-like planets with orbital periods similar to Earth, Venus, or Mars.

Problem 02: Radial Velocity Method for Finding Exoplanets

Analysis of 51 Pegasi Radial Velocity Data

1. **Scatter Plot**:
   - Loaded the radial velocity dataset.
   - Plotted time (Julian date) on the X-axis and radial velocity (m/s) on the Y-axis.
   - Observed periodic oscillations indicative of the planet's influence.
2. **Initial Period Estimate**:
   - Identified a repeating pattern visually, estimating an orbital period of approximately 4.2 days.
3. **Lomb-Scargle Periodogram**:
   - Applied the periodogram formula to identify the period with the highest power.
   - Estimated a precise orbital period of ~4.23 days.
4. **Phase Folding**:
   - Folded the radial velocity data using the derived orbital period.
   - Plotted the phase-folded radial velocity curve to reveal the periodic motion of the star.

Key Observations:

- The derived orbital period matches historical observations of 51 Pegasi b.
- Noise in the folded curve may be due to measurement uncertainties or additional system factors.

Problem 03: Pulsars in Deep Space

Numerical Dataset Analysis

Data Preprocessing:

- Scaled numerical features using `StandardScaler` from scikit-learn.
- Split data into training and validation sets (80:20 ratio).

Model Development:

- Trained a Random Forest model with hyperparameter tuning (e.g., tree depth and the number of estimators).

Performance Evaluation:

- Metrics used: Accuracy, precision, recall, F1 score, and ROC-AUC.
- Achieved high precision, reducing false positives.
- Performance Summary:
  - Accuracy: 93.25%
  - Precision: 88.97%
  - Recall: 91.54%
  - F1 Score: 90.23%
  - ROC-AUC: 96.78%

Advantages and Limitations:

- Advantages:
  - Faster training and easier preprocessing.
  - High interpretability using metrics like feature importance.
- Limitations:
  - Limited to structured data.
  - Performance depends on feature engineering.

Image Dataset Analysis

Preprocessing:

- Resized images to 64x64 pixels.
- Normalized pixel values and created batches for training.

CNN Model:

- Designed a 4-layer CNN with convolutional, pooling, and fully connected layers.

- Used cross-entropy loss and Adam optimizer.
- Applied mixed precision training (`GradScaler`) and GPU acceleration.

Performance Evaluation:

- Validation Accuracy: 87.5%.
- Training Loss: Converged to ~0.02 after 10 epochs.

Advantages and Limitations:

- Advantages:
    - Able to capture complex spatial patterns in images.
    - Does not require manual feature engineering; instead, it learns features automatically.
- Limitations:
    - Requires significant computational resources (e.g., GPUs).
    - Slower training process due to model complexity.

Comparison and Discussion

Numerical Models (Random Forest):

- Faster training and interpretation.
- Suitable for structured datasets with defined features.
- Ideal when computational efficiency and interpretability are critical.

Image Models (CNN):

- Better for complex, high-dimensional image data.
- Requires more computational resources but excels in tasks requiring intricate pattern recognition.

In summary, the Random Forest model achieved higher accuracy (93.25% vs. 87.5%) and efficiency, making it preferable for structured data. However, the CNN model excelled in recognizing complex image patterns, highlighting its strength for unstructured data tasks like image classification.

Problem 04: Orbital Resonances

Approach

1. **Inner Planet’s Orbital Period**:
   - Using Kepler’s Third Law:
   - Substituted and :
2. **Outer Planet’s Semi-Major Axis and Orbital Period**:
   - For a 3:2 resonance:
   - From Kepler’s law:
3. **Discussion**:
   - Resonances can enhance tidal heating, potentially affecting planetary atmospheres and habitability.
   - Gravitational interactions may destabilize orbits over time but can also promote long-term stability if resonances persist.

Problem 05: Escape Velocity from a White Dwarf

Approach

1. **Escape Velocity Formula**:
   - Using energy conservation:
   - Substituted and .
2. **Calculation**:
   - Values:
     - .
     - .
   - Result:
3. **Discussion**:
   - High escape velocities intensify accretion disk dynamics and electron degeneracy pressure.
   - White dwarfs exhibit extreme gravitational effects, influencing nearby matter and potentially triggering supernovae.
