Project Proposal Justification Form

1. Project Title

Developing a Hybrid Approach Combining QhX and Machine Learning for Quasar Light Curves

2. Student Name

Thara Rubi Caba Pineda

3. Number of Hours Required

Task	**Estimated Hours**
Literature Review	8 hours
Statistical Analysis Design	10 hours
Data Collection and Preparation	8 hours
Implementation of Improvements	14 hours
Validation and Testing	10 hours
Report Writing and Final Presentation	10 hours
Total	60 hours

4. Literature Base

- **Quasar Harmonic Explorer (QhX) Documentation**: Understanding the current implementation of QhX for periodicity detection.

- **Machine Learning Time-Series Techniques**: Focus on recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, and possibly transformers for sequential data.
- **Astrostatistical Methods**: Papers and textbooks on Bayesian statistics, Fourier analysis, and wavelet-based techniques in astronomical data analysis.
- **Key Papers**:
  - "Oscillatory Patterns in the Light Curves of Active Galactic Nuclei" (Kovačević et al., 2018).
  - "Neural Networks for Time-Series Analysis" (Graves et al., 2016).

5. Overview of Suggested Improvements

Scientific Problem: Quasar light curves exhibit complex multiperiodic behavior, but traditional methods used in QhX (like Fourier transforms and wavelet analysis) can struggle with noisy, irregularly sampled, or incomplete data. Machine learning, particularly neural networks, has the potential to enhance pattern recognition in these challenging datasets by learning temporal dependencies and reducing noise interference.

Proposed Solution: This project proposes to integrate QhX with machine learning techniques to form a **hybrid approach**. The workflow would involve QhX's feature extraction methods (such as wavelet and Fourier analysis) followed by a **machine learning layer** (e.g., an LSTM model) to further improve periodicity detection. The machine learning model will be trained on quasar light curves to detect additional signals missed by QhX alone, and handle noisy data more effectively.

Relation to Existing Pipelines:

- **Current Limitations of QhX**: While QhX is effective for clean, well-sampled data, it struggles when the data are noisy or when complex periodicities are present.
- **Improved Hybrid Approach**: By integrating machine learning with QhX, the periodicity detection process will become more flexible and robust, allowing it to handle more challenging data conditions and potentially discover subtle patterns.

6. Statistical Analysis Approach

Hybrid Model Design:

- **Step 1: Feature Extraction with QhX**: The existing QhX method will be used to extract key features from the quasar light curves, such as candidate periodicities and noise characteristics.
- **Step 2: Machine Learning Layer**: These features will be passed into a machine learning model—specifically an **LSTM (Long Short-Term Memory)** network or a **Convolutional Neural Network (CNN)** designed for time-series analysis. The ML model will learn to recognize patterns and refine the periodicity detection, especially in cases where QhX struggles with noise or irregular sampling.
    - **Training the Model**: The model will be trained on a set of quasar light curves from the SDSS dataset. The goal is to improve accuracy in detecting true periodicities and reduce false positives due to noise.
    - **Evaluation**: The hybrid model will be compared to both the standalone QhX method and a purely machine learning approach. Performance will be measured by precision, recall, and computational efficiency.

Key ML Techniques:

- **LSTM Networks**: To capture long-term dependencies in time-series data and predict periodic signals.
- **CNNs for Time-Series**: To detect localized patterns and features within the light curves.
- **Transfer Learning**: Pre-train the machine learning model on synthetic datasets or other time-series data, then fine-tune it for quasar light curves.

7. Specify data that will be used for the project

The **SDSS Quasar Dataset** will serve as the primary source of quasar light curves for this project. This dataset contains a large volume of quasar variability data, which is ideal for training and testing both the QhX and machine learning models.

Additional Data Sources:

- **ZTF (Zwicky Transient Facility)** for cross-validation with light curves from a different survey.
- **GAIA Data** for supplementary astrometric data if needed for contextual analysis.

8. Potential Outcome and Impact

Expected Outcomes:

- **Improved Periodicity Detection**: The hybrid model is expected to outperform QhX alone in detecting complex periodic signals, especially in noisy and irregularly sampled data.
- **Machine Learning's Role**: The ML component will reduce the impact of noise and enhance QhX's sensitivity to subtle periodicities, making the detection process more reliable.
- **Impact**: This project could lay the foundation for a new generation of astrostatistical tools, combining traditional methods with machine learning to improve performance across a range of time-series astronomical data.

9. Plan for Presentation at the Mathematical Symposium or Artificial Intelligence

I plan to present this project at the **Mathematical Symposium** in November.

10. Project Progress and Reporting

Progress Reports:

- Reports will be submitted **every two weeks** to track the following:
    - Hours worked.
    - Specific tasks completed.
    - Challenges encountered and how they were addressed.