**Model Card: Random Forest for Solar Flare Prediction**

**Model Details**

* **Model Type**: Random Forest Classifier
* **Framework**: Scikit-learn
* **Hyperparameters**: Tuned using Optuna
  + n\_estimators: Number of trees in the forest
  + max\_depth: Maximum depth of the tree
  + min\_samples\_split: Minimum number of samples required to split an internal node
  + min\_samples\_leaf: Minimum number of samples required to be at a leaf node
  + max\_features: Number of features to consider when looking for the best split

**Intended Use**

* **Primary Use Case**: Predicting the occurrence of solar flares based on various input features.
* **Target Users**: Researchers, meteorologists, and space weather forecasters.
* **Input Data**: Features related to solar activity (e.g., magnetic field measurements).
* **Output**: Binary value indicating the likelihood of a solar flare occurring within 24 hours.

**Performance Metrics**

* **Evaluation Metrics**:
  + Accuracy
  + Precision
  + Recall
  + F1-Score
* **Performance on Test Set**:
  + Accuracy: 0.9073
  + Precision: 0.9073
  + Recall: 0.9073
  + F1-Score: 0.9073

**Feature Importances**

The following features were identified as the most important by the Random Forest model:

| **Feature** | **Description** | **Importance** |
| --- | --- | --- |
| R\_VALUE | Complexity of the magnetic field configuration | 0.152 |
| TOTUSJH | Total unsigned magnetic flux in the horizontal(H) | 0.131 |
| TOTUSJZ | Total unsigned magnetic flux in the vertical(Z) | 0.084 |
| USFLUX | Unsigned magnetic flux | 0.083 |
| ABSNJZH | Absolute value of the net magnetic flux in the vertical component. | 0.690 |

**Considerations**

* **Data Quality**: Ensure that the input data is clean and preprocessed correctly. Missing values and outliers should be handled appropriately.
* **Model Limitations**: The model's performance is dependent on the quality and relevance of the input features. It may not generalise well to unseen data if the training data is not representative.
* **Ethical Considerations**: The predictions made by the model should be used as one of several inputs in decision-making processes. Users should be aware of the model's limitations and potential biases.

**Training and Evaluation**

* **Training Data**: The model was trained on a dataset containing features related to solar activity.
* **Test Data**: The model was evaluated on a separate test set to ensure its performance generalises to unseen data.
* **Training Procedure**: The model was trained using the Random Forest algorithm with hyperparameters tuned using Optuna.

**Usage**

To use the model, follow these steps:

1. **Preprocess the Input Data**: Scale the features using a standard scaler.
2. **Load the Model**: Load the trained Random Forest model.
3. **Make Predictions**: Use the model to predict the likelihood of different types of solar flares.