Illuminating the Stars: Predicting Stellar Luminosity with Explainable Machine Learning

Nikhil Kapoor
a1879608

September 3, 2024

Report submitted for **Data Science Research Project Part B** at the School of Mathematical Sciences, University of Adelaide

THE UNIVERSITY *of* ADELAIDE

Project Area: **SHAPLEY VALUES FOR EXPLAINING MACHINE LEARNING MODELS**
Project Supervisor: **Prof. DINO SEJDINOVIC**

Abstract

This study explores the integration of advanced machine learning techniques with astrophysical data analysis to predict stellar luminosity using the Gaia DR3 dataset. By employing various machine learning models, including Support Vector Regressor (SVR), Random Forest, Gradient Boosting, and others, we aimed to identify the best-performing model. The SVR model exhibited exceptional performance, achieving the lowest RMSE and MAE and the highest R^2 among all tested models. SHAP (SHapley Additive exPlanations) values were utilized to interpret the model predictions, consistently identifying log-transformed distance and photometric measurements as the most influential features. The SHAP Dependence plots and SHAP Value Uncertainty Analysis provided interpretable insights and bolstered confidence in the model's robustness and generalizability. The findings demonstrate the potential of combining machine learning with SHAP values to enhance the interpretability of stellar luminosity predictions, contributing to the advancement of computational astrophysics.

1 Introduction

The field of Astrophysics is entering a phase where advanced computational techniques broaden the scope of discovery and analysis. The Gaia mission's cataloguing of billions of stars presents a rich dataset for machine learning (ML) applications, that has the potential to revolutionize our understanding of the cosmos. Despite this potential, the complicated nature of astronomical data, with its high dimensionality and complexity, presents certain challenges, especially when decoding the 'black box' of ML predictions.

Central to astrophysical research is the measurement of absolute magnitude which is the intrinsic luminosity of celestial bodies and a foundation for understanding stellar classification and galactic formations. Traditional ML models provide limited interpretability into how they process information to predict such traits, a key disadvantage for scientific validation.

This project aims to bridge this gap by employing SHapley Additive exPlanations (SHAP) values, which provide insight into the contribution of each input feature to model predictions. Grounded in cooperative game theory, SHAP values improve the transparency of ML models and ensure that their predictions are consistent with established astrophysical theories, thus reinforcing scientific integrity.

Through this integration, the project pursues the following goals:

- **Enhance Prediction Accuracy:** Utilize advanced machine learning models to improve the accuracy of stellar luminosity predictions.
- **Interpretability for Astronomers:** Implement interpretability techniques to make model predictions understandable to astronomers.
- **Quantify Uncertainty and Reliability:** Develop methods to quantify the uncertainty associated with each prediction and provide confidence intervals to ensure reliability and robustness.
- **Identifying Uncertain Predictions:** Use uncertainty quantification to identify areas where predictions are less certain and warrant further investigation.

Applying SHAP values in astrophysics is a ground-breaking step in the field of explainable AI, offering a fresh perspective on data interpretation. This report delineates our comprehensive methodology, from data curation to model assessment, with a focus on the interpretability of ML models enhanced by SHAP values.

2 Background

Astrophysics is undergoing a transformative era with the integration of advanced computational methods, expanding the scope of discovery and analysis. These advancements enable in-depth studies on **stellar populations**, **stellar kinematics**, and **intrinsic stellar properties**, which provide detailed insights into the distribution, evolution, and dynamics of stars within galaxies [22].

2.1 Parallax Measurements and Absolute Magnitudes

A fundamental technique in this field is the use of **parallax measurements** for accurately determining stellar distances and enabling the mapping of the cosmos. By measuring the apparent shift in a star's position caused by Earth's orbit around the Sun, parallax provides a direct method to calculate stellar distances. Accurate distance measurements are necessary for calculating **absolute magnitudes**, providing a baseline for comparing stellar luminosities irrespective of their distance from Earth [5].

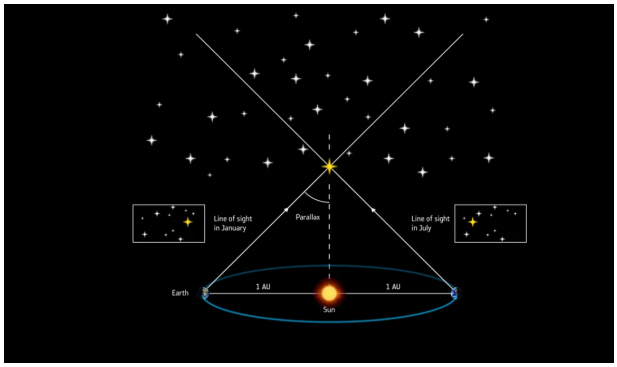

Figure 1: Scientists use parallax to measure far-away distances.

Determinations of **absolute magnitude** are crucial for understanding the true brightness of stars, which is vital for studying stellar and galactic formation and evolution. The **Hertzsprung-Russell (H-R) diagram** graphically represents the relationship between the luminosity and temperature of stars, relying heavily on accurate measurements of absolute magnitudes to establish the evolutionary stages and properties of different stellar populations [15]. By mapping stars onto the H-R diagram, our project can utilize machine learning models to predict stellar parameters and classify stars into their respective stages. This capability allows for precise modelling of stellar evolution.

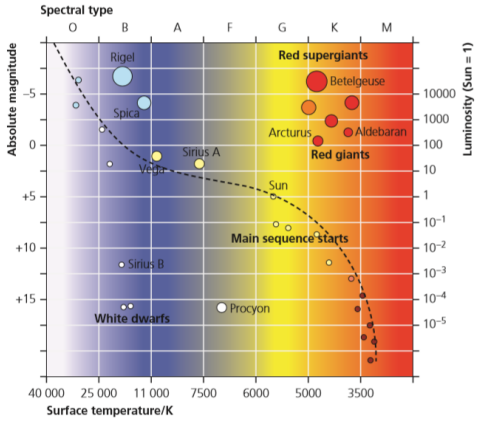

Figure 2: Hertzsprung-Russell Diagram.

2.2 Photometric Analysis

Photometric analysis is essential for determining stellar mass, temperature, and age. By measuring the brightness of stars across different wavelengths, photometric data provides a deep understanding of the Milky Way's composition [2]. For instance, the **G-band mean magnitude** measures a star's brightness in a specific band, providing a baseline for luminosity calculations. The **BP-RP colour index**, indicating the difference between the blue and red photometric bands, provides insights into the star's temperature and composition.

The integrity of astronomical data is assessed by metrics such as the **Renormalized Unit Weight Error (RUWE)** and **astrometric excess noise**. **RUWE** indicates the quality of the astrometric fit, with values close to 1 suggesting good data quality. **Astrometric excess noise** represents the noise in the astrometric solution, impacting the accuracy of luminosity predictions [25].

2.3 Advancements in Machine Learning and SHAP

Advancements in machine learning, especially in model interpretability like **SHAP (SHapley Additive exPlanations)**, have revolutionized the field by enhancing the transparency of machine learning models [4]. SHAP values provide a unified measure of feature importance based on cooperative game theory, ensuring that model outputs are interpretable and trustworthy. This is crucial for scientific discoveries, such as classifying stellar types or predicting stellar parameters based on observational data. In our study, SHAP values were instrumental in interpreting the predictions of our machine learning models, providing insights into the relative importance of different features in determining stellar luminosity.

2.3.1 Mathematical Representation of SHAP Values

The mathematical foundation of Shapley values, as proved by Shapley in 1953, is captured in the following formula [20]:

$$\phi_i(\nu) = \frac{1}{d} \sum_{S \subseteq \Omega \setminus i} \left(\binom{d-1}{|S|} \right)^{-1} (\nu(S \cup i) - \nu(S)) \tag{1}$$

where Ω is the set of all features, d is the total number of features, S is a subset of Ω not containing feature i, and $\nu(S)$ is the model's prediction with features S.

This formula calculates the average contribution of a feature to the model's predictions across all possible subsets of features. It is fundamental for determining feature importance and promoting fairness and accountability in model explanations.

2.4 Machine Learning Models Employed

We compared several advanced machine learning models to predict stellar luminosity, each bringing unique strengths to the task. These models include:

- **XGBoost**: An efficient and scalable implementation of gradient boosting, adept at handling large datasets and complex feature interactions [10].
- **Gradient Boosting**: Similar to XGBoost, builds an ensemble of weak learners and sequentially improves the model by correcting errors [1].
- **Random Forest**: Constructs multiple decision trees and merges them to get a more accurate and stable prediction, robust to overfitting [24].
- **K-Nearest Neighbors (KNN)**: Captures local patterns in the data by considering the closest neighbours to a given point, effective for localized feature relationships [26].
- **Support Vector Regressor (SVR)**: Offers robust performance by finding the hyperplane that best fits the data while minimizing error, suitable for non-linear relationships [21].
- **Neural Networks**: Highly effective for large datasets, capable of capturing complex patterns through multiple layers of interconnected nodes [17].

3 Methods

The methodological rigour of this study underpins the robustness of its conclusions. By systematically sourcing, selecting, and pre-processing data, we ensure a stable foundation for our analysis.

3.1 Data Acquisition

Our study uses the extensive Gaia dataset, acquired from the European Space Agency (ESA), which compiles astrometric and photometric records for over two billion stars, making it a valuable resource for astronomical research. This rich dataset provides detailed positional data and luminosity measures critical for our investigation.

Figure 3: The Gaia spacecraft, ESA's mission to chart a three-dimensional map of our Galaxy, the Milky Way.

3.1.1 Data Sourcing

Data was sourced from the comprehensive Gaia Data Release 3 (DR3), which is the most recent release, encompassing precise stellar positional measurements and luminosity across various filters [3].

3.1.2 Data Selection

We concentrated on a densely populated celestial sector at RA 280 degrees and DEC -60 degrees, chosen for its data richness and reliability, thereby bolstering the robustness of our analysis.

3.1.3 Data Query

Utilizing the `astroquery` library, we executed an asynchronous query targeting all stellar data within a specified 0.1-degree square, ensuring efficient data handling and result reproducibility.

3.2 Data Preprocessing

Data preprocessing is a critical step in the analysis pipeline to ensure that the data used in subsequent analyses is clean, relevant, and structured effectively. This section outlines the key preprocessing steps applied to the Gaia DR3 dataset for the computation of stellar properties.

3.2.1 Feature Selection

A comprehensive literature review was conducted to identify key features that are most influential in predicting stellar luminosity, which is essential for understanding galactic structures and evolution [16]. The selected features are listed in Table 1.

Table 1: Selected Features for Stellar Luminosity Prediction

No.	Feature
1	Distance (Derived from Parallax)
2	G-band mean magnitude
3	BP-RP color index
4	Mean magnitude in BP band
5	Mean magnitude in RP band
6	Renormalized Unit Weight Error (RUWE)
7	Astrometric Excess Noise
8	G-band mean flux over error
9	BP/RP excess factor
10	Parallax over error
11	Number of visibility periods used
12	Maximum astrometric excess noise

This feature selection focuses the analysis on the most relevant variables, reducing noise and computational complexity while retaining the most informative aspects of the dataset for stellar luminosity prediction.

3.2.2 Data Cleaning

- **Removing Non-Positive Parallax:** Entries with non-positive parallax values, which can result from measurement errors, data entry issues, or specific coding conventions, were removed to ensure only meaningful parallax measurements are considered [6].
- **Handling Missing Values:** Median imputation was used for selected features with missing values to prepare the dataset for our analysis.

3.2.3 Calculation of Derived Metrics - Feature Engineering

- **Distance Calculation:** Distance to each star was computed using the formula [9]:

$$\text{distance (parsecs)} = \frac{1000}{\text{parallax (milliarcseconds)}} \tag{2}$$

- **Absolute Magnitude Calculation:** The absolute magnitude was calculated using [14]:

$$M = m - 5\log_{10}\left(\frac{d}{10}\right) \tag{3}$$

  where M is the absolute magnitude, m is the apparent magnitude, and d is the distance in parsecs.

3.2.4 Outlier Detection and Removal

This step prevents extreme values from influencing the analysis, leading to more robust results. It is highly useful in astronomical datasets where measurement errors or rare phenomena can produce values not representative of the general stellar population we aim to model.

- **Boxplot Analysis for Outliers:** Initial visualization with boxplots for 'parallax' and 'G-band mean magnitude' to identify outliers.
- **Interquartile Range (IQR) Method:** Outliers were further refined and removed using the IQR method for 'parallax' and 'G-band mean magnitude'.

3.2.5 Feature Transformation

- **Log Transformation:** The 'distance' feature was log-transformed to improve the distribution's normality.

3.3 Model Development

With a clean, well-prepared dataset and well-optimized, we proceeded to develop our models. The goal was to utilize advanced machine-learning techniques to predict stellar luminosity with high accuracy while ensuring the interpretability of the models. The following steps outline the development process.

3.3.1 Data Preparation and Splitting

After the pre-processing of data, the data was partitioned into training (80%) and test (20%) subsets to facilitate model training and evaluation, ensuring that the model's performance could be assessed on unseen data. This splitting method helps in maintaining the integrity of the evaluation process by preventing overfitting and ensuring a fair assessment of the model's generalization capabilities.

3.3.2 Testing of Machine Learning Models

The dataset was trained with various machine learning models, including XGBoost, Gradient Boosting, Random Forest, K-Nearest Neighbors, Support Vector Regressor, and Neural Networks. Each model was assessed for its ability to predict stellar luminosity using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Hyperparameter tuning was also performed via grid search to optimize the models' performance.

3.3.3 Cross-Validation Results

To ensure robust model validation, K-fold cross-validation (K=5) was employed. This technique involves dividing the dataset into five subsets and training the model five times, each time using a different subset as the validation data and the remaining subsets as the training data. This approach helps to ensure consistent performance across different data subsets and further reduces the risk of overfitting.

3.3.4 Selecting the Best Model

The next step involved analyzing feature importance across the different models to understand how each model interprets the data. This analysis was crucial in determining the most suitable model for predicting stellar luminosity. By comparing the performance metrics of each model, we identified the one that provided the highest predictive accuracy and reliability.

3.3.5 Statistical Analysis of Best Model

Once the best-performing model was selected, its performance was further evaluated through a detailed statistical analysis. This included analyzing the distribution of residuals to validate the precision of the predictions. The residual analysis helps in identifying any systematic errors and ensuring that the model's predictions are unbiased and accurate.

3.4 Interpretability with SHAP Values

Interpretability in machine learning models is of utmost importance, especially in fields like astrophysics where understanding the influence of variables on predictions is crucial. SHAP values provide a robust framework for interpreting model predictions, highlighting how each feature contributes to the outcome.

3.4.1 Computation of SHAP Values

The computation of SHAP values involves several key steps:

- Training the model on the dataset to establish a baseline for prediction.
- Utilizing the trained model to forecast outcomes based on the data.
- Employing the SHAP library to calculate the impact of each feature on the model's predictions.

3.4.2 SHAP Methods Used

We employed several SHAP methods tailored to different types of models:

- **KernelSHAP:** Model-agnostic and uses weighted regression, ensuring consistent explanations across various models such as Support Vector Regressor (SVR) and K-Nearest Neighbors (KNN) [12].
- **TreeSHAP:** Optimized for tree-based models, offering efficient computation and detailed insights for models like XGBoost, Gradient Boosting, and Random Forest [30].
- **DeepSHAP:** Combines DeepLIFT and SHAP for deep learning models, providing robust explanations for neural networks [19].

3.4.3 Comparison of Feature Importance Across Models

SHAP analysis was conducted to identify and visualize the most influential features across all models before selecting the best machine learning model for our analysis. This analysis provided insights into how each feature impacts the model's output and offered a clear understanding of the underlying data patterns. We used the SHAP Feature Importance plot [28] to highlight the importance of each feature across different models. This comparative approach ensured that the final model selection was based on a thorough evaluation of feature significance across all models which were considered initially.

3.4.4 Global SHAP Analysis

After selecting the final model, we visualized the SHAP summary plot [27] to identify the most influential features and their effect on model predictions for our best model. Each row represents a feature ordered by importance, with points indicating individual predictions. The horizontal position shows the impact on model output, where rightward shifts indicate higher predictions and leftward shifts indicate lower ones. Color coding with red for high values and blue for low values highlights feature values.

3.4.5 Feature Dependence

To illustrate how individual features impact the model's predictions, we visualized the SHAP dependence plot [11]. It shows the relationship between a feature's value and its SHAP value, indicating how changes in the feature affect the prediction.

3.4.6 Local Interpretation Analysis

To explain how the contribution of each feature combines to form the final prediction for a specific instance, we visualized the SHAP waterfall plot [18]. These plots start with a baseline value, typically the average model output, and show how each feature's value pushes the prediction higher or lower, culminating in the final predicted value for a specific instance. By displaying features in order of their impact magnitude, they offer a clear, feature-by-feature breakdown of how the model arrives at its prediction, making them invaluable for model interpretation.

3.4.7 Reliability and Variability of Feature Importance

To assess the reliability and variability of feature importance, we measured how consistent the SHAP values were across different data subsets. This step is crucial to ensure that the identified important features remain influential across various conditions and are not artefacts of specific subsets. To visualize this, we used the SHAP Value Uncertainty plot [7].

3.4.8 Accuracy and Confidence Intervals

Finally, the accuracy and confidence intervals of the model predictions were examined. This provided a comprehensive evaluation of the model's performance and reliability, ensuring that the predictions were robust and trustworthy for astronomical research. The SHAP Prediction Intervals plot [13] has been used to demonstrate this aspect.

4 Results

This section presents detailed analysis results from the Gaia DR3 data, focusing on stellar characteristics in the specified sky region. It includes data summaries, distribution analyses, model performance, and interpretability using SHAP values.

4.1 Stellar Data Summary

The initial data retrieval from Gaia DR3 encompassed 184 stars within the target coordinates (RA: 280 degrees, DEC: -60 degrees). The table below summarizes the main attributes of the retrieved stellar data:

Attribute	Mean	Standard Deviation	Min	Max
Parallax (mas)	0.449	0.890	-3.228	3.644
Phot_G_Mean_Mag	19.065	1.751	11.303	21.579

Table 2: Summary statistics of stellar attributes from Gaia DR3.

4.2 Outlier Analysis and Data Distribution

4.2.1 Boxplot Interpretation

As depicted in Figure 4, the parallax boxplot shows outliers above the upper whisker, suggesting stars with abnormally small parallax readings that imply greater distances. Conversely, the G-band mean magnitude boxplot indicates outliers below the lower whisker, representing unusually bright stars compared to the main stellar population.

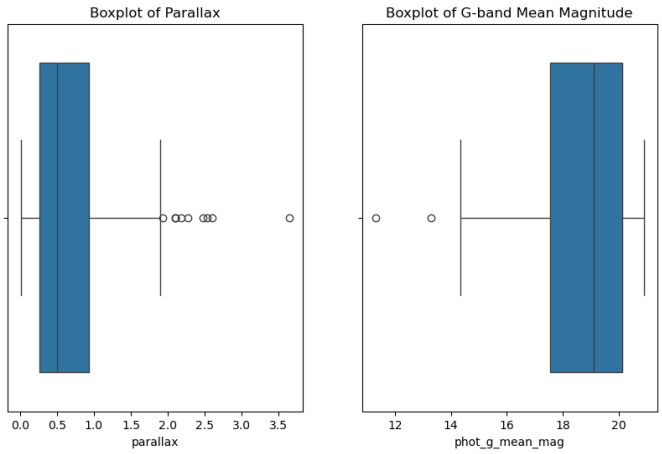

Figure 4: Combined boxplots representing (left) the distribution of parallax values and (right) the G-band mean magnitudes.

4.2.2 Data Characterization Post-Outlier Removal

After removing the outliers, we recalculated key statistics to ensure a robust analysis of the cleaned dataset. Table 3 provides a summary of the main descriptive statistics for parallax and G-band mean magnitude, reflecting the essential properties of the stellar data.

The recalculated mean parallax of 0.713 mas indicates a wide range of distances in our study, while the G-band mean magnitude captures the diverse stellar brightness levels. These metrics are vital for understanding the dynamics of the stellar population in the region. This cleaned dataset serves as the foundation for deeper analytical and modelling efforts.

Table 3: Descriptive statistics of key parameters post-data cleaning.

Parameter	Mean	SD	Min	Max
Parallax (mas)	0.713	0.643	0.012	3.644
Phot G Mean Mag	18.662	1.785	11.303	20.902
Distance (parsecs)	3670.000	7716.200	274.400	83 643.600
Absolute Magnitude	7.074	2.879	−2.331	13.068

4.3 Analysis of Stellar Characteristics

Following the preprocessing of the Gaia DR3 dataset, we analyzed it to understand the distribution of stellar characteristics such as parallax, G-band mean magnitude, distance, and absolute magnitude. The histograms displayed in Figure 5 offer insights into these distributions.

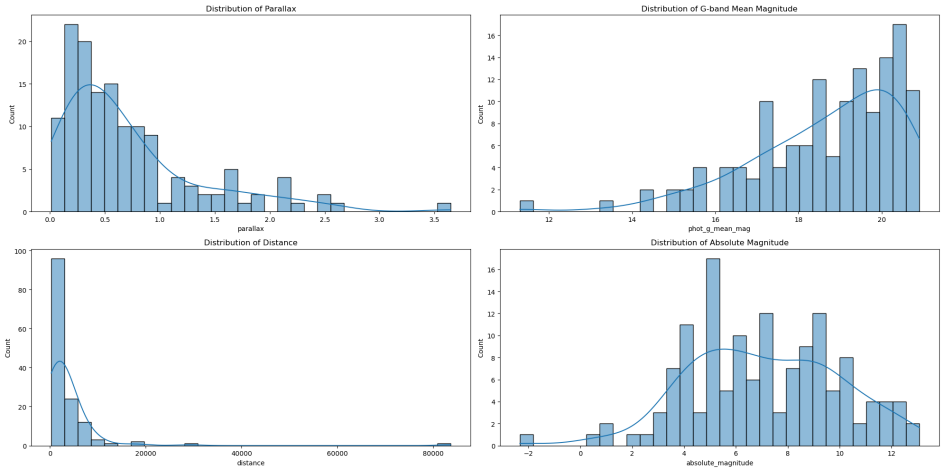

Figure 5: Histograms illustrating the distribution of key astrophysical measurements from the Gaia DR3 dataset. Each panel presents a different attribute, annotated for clarity.

Parallax Distribution: This histogram shows a concentration of lower parallax values, typical for distant stars as seen from Earth. Outliers indicate some nearby stars with unusually high parallax values, warranting further verification.

G-band Mean Magnitude Distribution: The histogram displays a broad range of G-band magnitudes, with peaks suggesting clusters of stars at specific brightness levels, indicative of various stellar types or distances.

Distance Distribution: Illustrating a decrease in star counts with increasing distance, this histogram highlights the challenges in observing distant stars, consistent with the inverse-square law of luminosity.

Absolute Magnitude Distribution: Shows a wide range of luminosities from dwarfs to giants, reflecting the diversity of stellar types and stages within the galaxy, essential for studies of stellar evolution.

These histograms provide key insights into stellar properties and distributions, supporting further analysis and modeling to deepen our understanding of celestial phenomena.

4.4 Distribution Analysis of Log-Transformed Distances

Applying a logarithmic transformation to stellar distances results in a more normalized, normal-like distribution, as illustrated in Figure 6. This transformation addresses the positively skewed nature of astronomical distance data, which often covers a wide range of magnitudes.

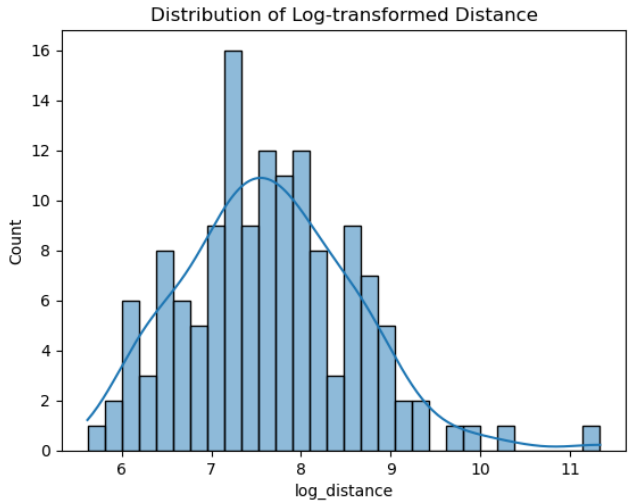

Figure 6: Histogram of log-transformed distances, indicating a normal-like distribution with a single peak.

The log transformation simplifies the data's scale, making the data more suitable to linear modelling techniques, which typically assume normally distributed inputs. The single peak observed in the histogram suggests a homogeneous sample distribution, facilitating further statistical and machine-learning analyses. This methodological step is crucial for ensuring that distance-related variables integrate effectively into predictive models and analyses.

4.5 Model Performance Analysis

The performance of our selected machine learning models was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Table 4 summarizes the results.

Model	**RMSE**	**MAE**	**R^2**
XGBoost	0.4761	0.3433	0.9676
Random Forest	0.5648	0.3165	0.9545
Gradient Boosting	0.4691	0.2806	0.9686
Neural Network	0.1642	0.1179	0.9962
Support Vector Regressor	0.0526	0.0446	0.9996
K-Neighbors	0.8024	0.5992	0.9081

Table 4: Model Performance Comparison.

The Support Vector Regressor (SVR) demonstrated the highest performance with the lowest RMSE and MAE, and the highest R^2, indicating excellent predictive accuracy and minimal error. The Neural Network also performed exceptionally well, showing high accuracy and low error rates. Gradient Boosting and XGBoost models demonstrated good predictive capability, though not as strong as SVR and Neural Networks. The Random Forest model, while effective, was less accurate than the previously mentioned models. The K-Nearest Neighbors (KNN) model showed the lowest performance, suggesting it is less suitable for this prediction task than the other models. These findings emphasize the importance of selecting the appropriate model for astrophysical predictions to ensure accurate and reliable results.

4.6 SHAP Feature Importance Analysis

The SHAP Feature Importance plot showcases the influence of various features across all the different machine learning models. As we can clearly see from Figure 7, the log-transformed distance consistently emerges as the most influential feature across all models. Both Blue Photometer Mean

Magnitude and Green Photometer Mean Magnitude are also significant, highlighting their importance in determining stellar brightness.

Model-specific insights reveal that the Support Vector Regressor (SVR) places high importance on the log-transformed distance and photometric magnitudes, with moderate emphasis on other features such as BP-RP color index and BP-RP excess factor. Similarly, Neural Networks also prioritize the log-transformed distance and photometric magnitudes but show a higher importance for noise-related features such as Maximum Five-Dimensional Astrometric Sigma. XGBoost and Random Forest models emphasize photometric magnitudes significantly, with the Blue and Green Photometer Mean Magnitudes being the most critical, while also considering the log-transformed distance as an important feature. Gradient Boosting and K-Nearest Neighbors exhibit a balanced importance across distance and photometric features, with a slightly higher emphasis on photometric measurements.

Overall, the bar plot of SHAP feature importance provides a comparative view of feature significance across different models, helping to understand how each model utilizes the features for predicting stellar luminosity. We then selected the best model for further analysis via SHAP.

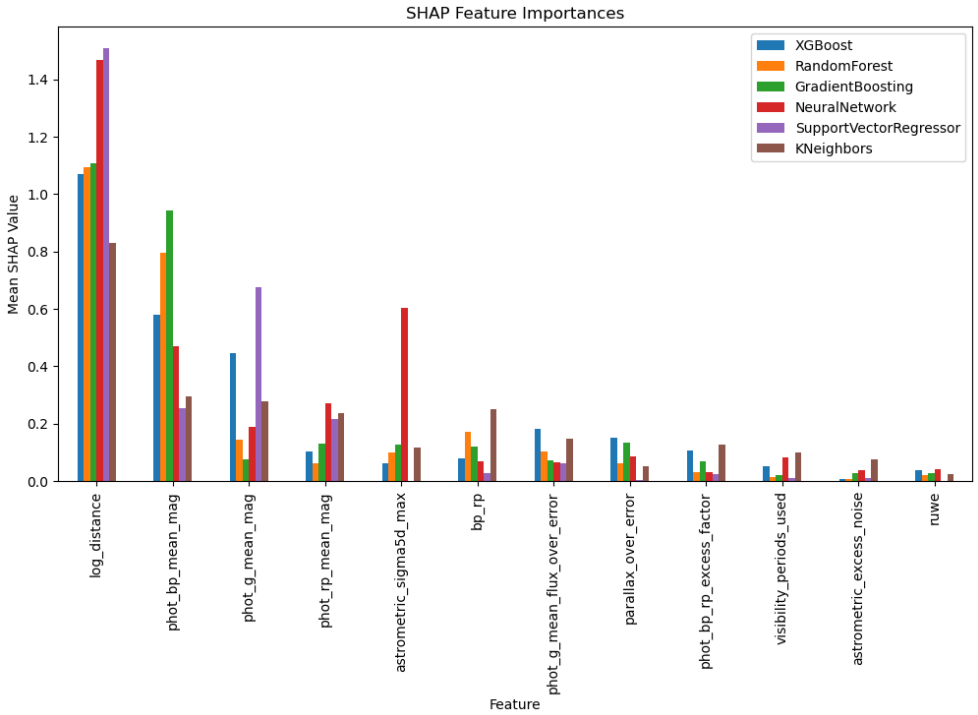

Figure 7: SHAP Feature Importances plot showing the influence of various features on the model's predictions.

4.7 Statistical Validation of Support Vector Regression: Evaluating Our Best Model

To ensure the robustness and reliability of our best model i.e. the SVR model [8], we conducted a series of statistical validations. The results of these validations are presented in the following analysis.

4.7.1 Learning Curve Analysis

The training and cross-validation scores converge as the training set size increases, indicating high accuracy with minimal variance [29]. This shows that the SVR model generalizes well to unseen data.

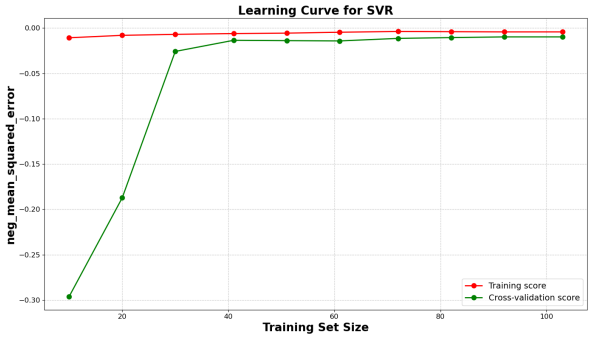

Figure 8: Learning Curve for SVR.

4.7.2 Residuals Distribution

A roughly symmetric distribution centred around zero is formed, indicating unbiased predictions. Most errors are close to zero, though some larger residuals suggest occasional outliers [23]. This highlights the model's good performance while also indicating areas for improvement, such as investigating outliers and refining feature engineering.

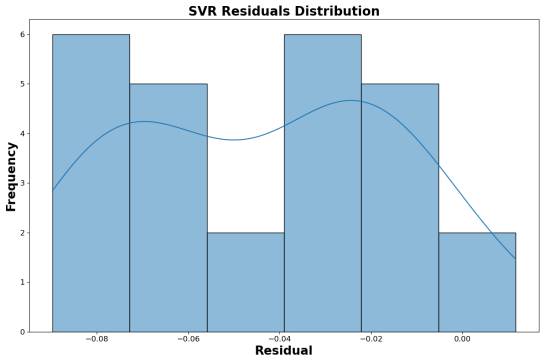

Figure 9: Residuals Distribution of SVR.

4.8 SHAP Summary Plot Analysis

The Global SHAP Feature Importance plot highlights the impact of each feature on the predictions made by the Support Vector Regressor (SVR) model.

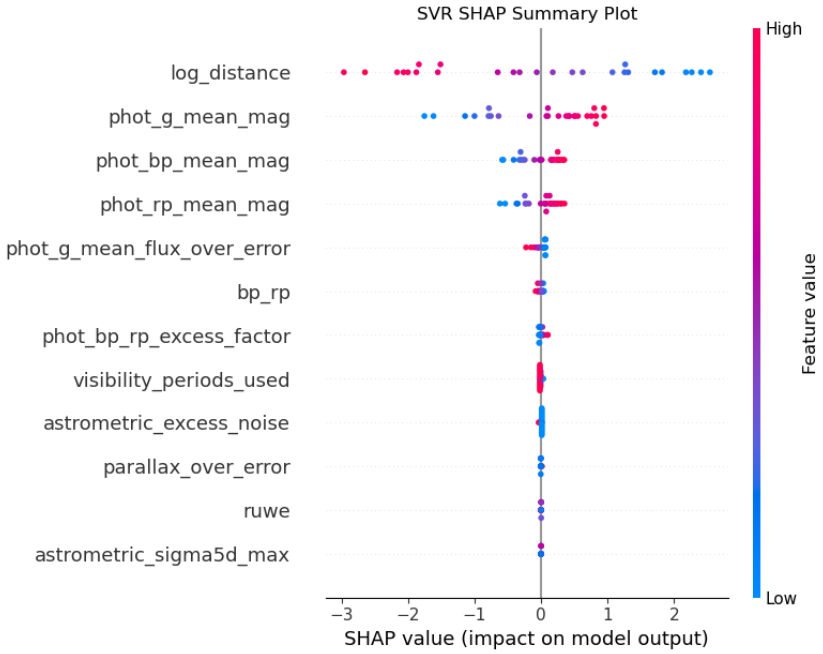

Figure 10: Global SHAP Feature Importance for SVR.

The log-transformed distance is the most influential feature, indicating its critical role in predicting stellar luminosity. Green Photometer Mean Magnitude and Blue Photometer Mean Magnitude are also significant contributors, representing the brightness of stars in different photometric bands. The Red Photometer Mean Magnitude, though slightly less influential, still significantly impacts the model's predictions.

Moderately important features include the ratio of the Green Photometer Mean Flux to its Error, BP-RP colour index, and BP/RP Excess Factor, which refine the predictions by providing information about the precision of flux measurements along with the star's temperature and composition.

Other features, such as the Green Photometer Mean Flux Over Error and the Blue Photometer-Red Photometer Color Index, display concentrated SHAP values around zero, implying a lesser impact on the model's output. Features like the Renormalized Unit Weight Error and the Maximum Five-Dimensional Astrometric Sigma show the least impact, with SHAP values closely clustered around zero, indicating minimal influence.

4.9 SHAP Dependence Plot Analysis

The SHAP Dependence plots illustrate the relationship between the individual features and their SHAP values, showing how each feature impacts the Support Vector Regressor (SVR) model's predictions.

The plot for the **log-transformed distance** shows a strong inverse linear relationship between the feature value and its SHAP value. As the log-transformed distance increases, the SHAP value decreases, indicating that closer stars with higher log-transformed distance values positively impact predicted luminosity.

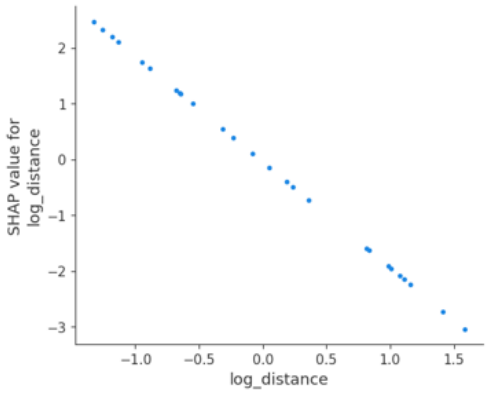

Figure 11: SHAP Dependence Plot for Log-Transformed Distance.

The **Green Photometer Mean Magnitude** plot exhibits a clear positive linear relationship with SHAP values. Higher values lead to higher SHAP values, signifying that brighter stars in the green photometric band positively influence the model's predictions.

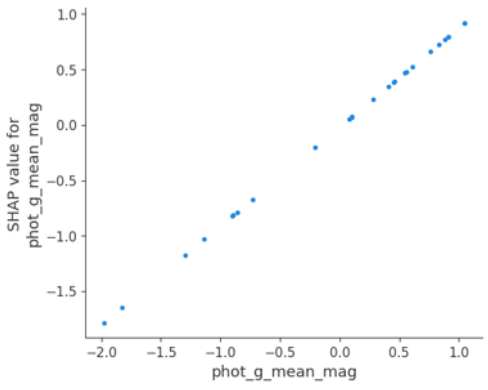

Figure 12: SHAP Dependence Plot for Green Photometer Mean Magnitude.

The **Blue Photometer Mean Magnitude** and **Red Photometer Mean Magnitude** also display positive linear relationships with their SHAP values. Brighter stars in these photometric bands contribute positively to predicted luminosity, emphasizing the importance of photometric measurements in determining stellar luminosity.

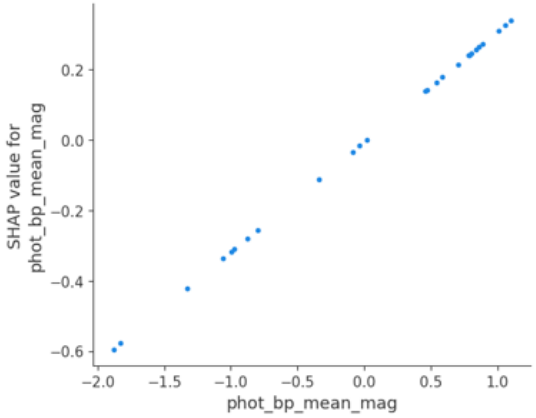

Figure 13: SHAP Dependence Plot for Blue Photometer Mean Magnitude.

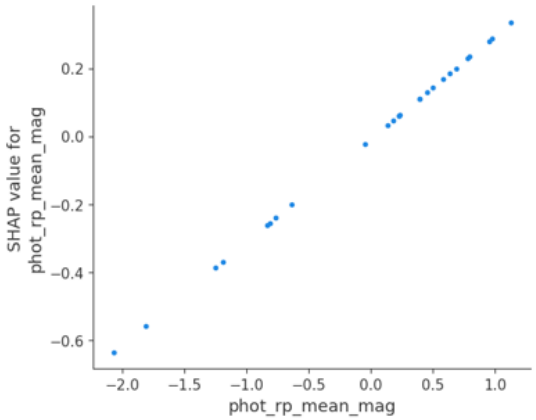

Figure 14: SHAP Dependence Plot for Red Photometer Mean Magnitude.

Overall, the SHAP Dependence plots provide valuable insights into feature interactions within the SVR model. The observed linear relationships for key features like **log-transformed distance** and **photometric magnitudes** align with astronomical principles, validating the model's predictive behaviour and reinforcing the importance of these features in determining stellar luminosity. These plots ensure the reliability and interpretability of the SVR model's predictions.

4.10 SHAP Waterfall Plot Analysis

The SHAP waterfall plot, as we can see in Figure 15, offers insights into the breakdown of feature contributions to the model's prediction for a specific instance, where $f(x) = 6.587$. This value represents the adjusted prediction after accounting for the contributions of individual features from the baseline model prediction of 6.923.

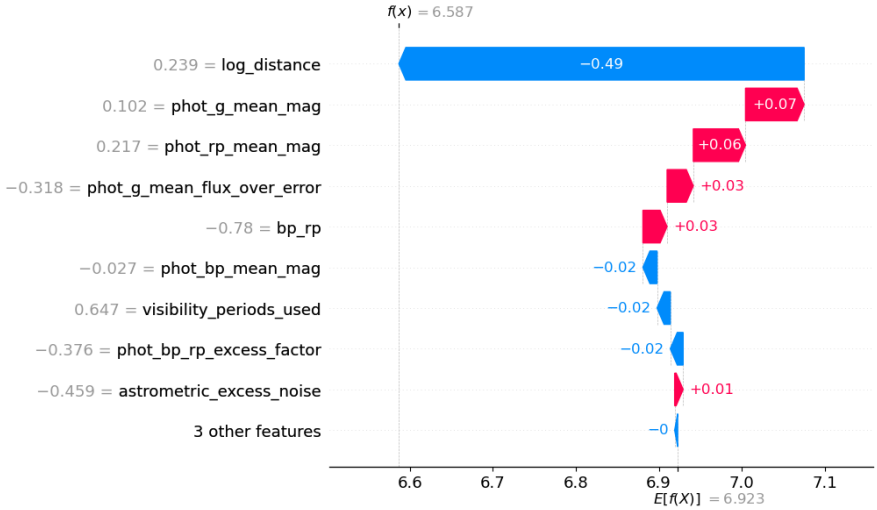

Figure 15: Waterfall plot showing SHAP values for a specific prediction.

The most significant feature is the log-transformed distance, which reduces the model's prediction by approximately 0.49 units. This indicates that stars closer to the observer, having higher log-transformed distance values, lead to lower predicted values. In contrast, G-band mean and RP-band mean magnitude have positive contributions, increasing the prediction by around 0.07 and 0.06 units, respectively.

Other features such as G-band mean flux over error and BP-RP colour index also contribute positively, each adding 0.03 to the prediction. The plot reveals a complex interaction among features, with some having opposing effects. For instance, while log-transformed distance pushes the prediction down, several magnitude-related features push it up. Smaller negative impacts come from features such as BP-band mean magnitude, visibility periods used, and BP-RP excess factor.

4.11 SHAP Value Uncertainty Analysis

The SHAP Value Uncertainty Analysis plot, as we can see in Figure 16, for our Support Vector Regressor (SVR) model, provides crucial insights into the reliability and consistency of features in predicting stellar luminosity.

This analysis is valuable as it quantifies not just the importance of each feature, but also the uncertainty associated with its impact across different data subsets. The visualization reveals the impact and consistency of features in our SVR model for stellar luminosity prediction. The x-axis shows mean SHAP values (feature importance), while the y-axis represents SHAP value uncertainty (variability of importance).

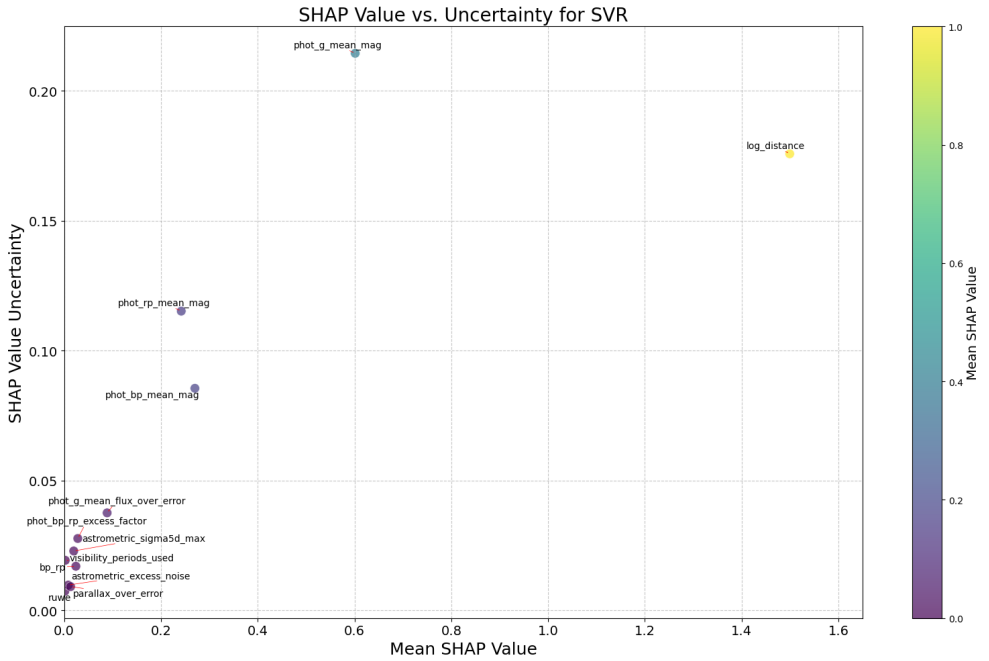

Figure 16: SHAP Value Uncertainty Analysis for SVR model features.

Log-transformed distance emerges as the most influential predictor, with the highest mean SHAP value and low uncertainty, indicating its consistent impact across data subsets. Interestingly, while the distance feature shows slightly higher uncertainty than other features, this uncertainty remains relatively low given its high importance. Photometric measurements (Green, Blue, and Red Photometer Mean Magnitudes) show significant importance with low uncertainties, demonstrating their stable contributions.

Most features exhibit less than 0.05 uncertainty values, signifying the model's overall stability. Features near the origin like the BP-RP colour index have lower importance but maintain low uncertainty, suggesting consistent and smaller, influences on predictions. The low uncertainties across key features enhance confidence in the model's robustness and generalizability, strengthening the scientific validity of our machine-learning approach in stellar luminosity prediction.

4.12 SHAP Prediction Intervals with SVR

The SHAP Prediction Intervals plot analyzes the accuracy and confidence of predictions made by the Support Vector Regressor (SVR) model. This plot displays the true values of absolute magnitude (blue line), mean SHAP predicted values (orange line), and 95% SHAP prediction intervals (grey shaded area) for a subset of samples.

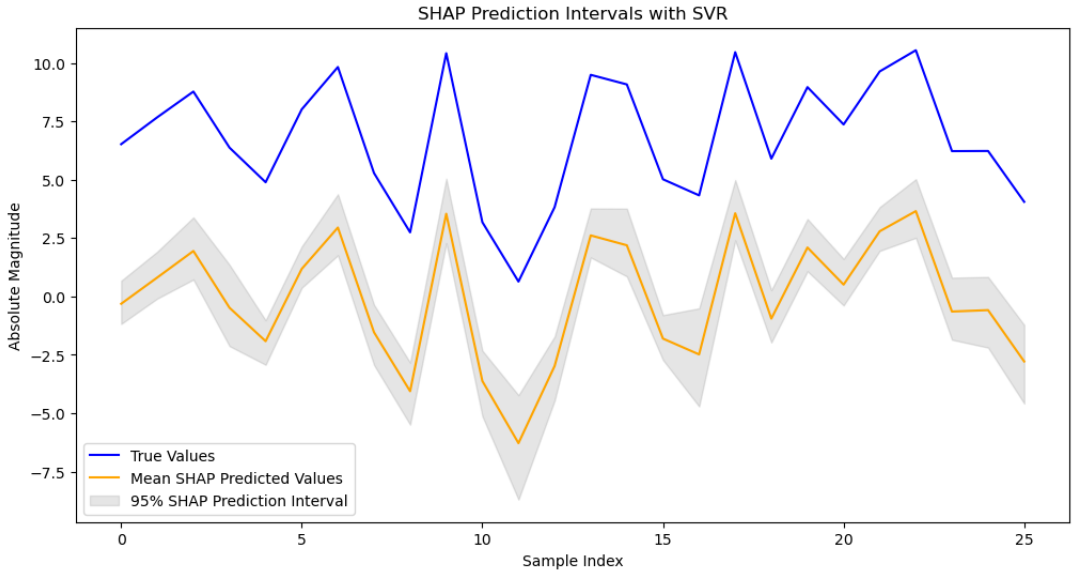

Figure 17: SHAP Prediction Intervals with SVR.

The blue line represents actual observed values, while the orange line shows the SVR model's predicted values. The close alignment between these lines indicates that the model accurately captures underlying patterns in the data. The grey shaded area represents the 95% prediction intervals, showing the range within which true values are expected to fall with 95% confidence. The generally narrow intervals show the model's high confidence in its predictions, although some variability exists where the model is less certain.

Overall, the SHAP Prediction Intervals plot highlights the SVR model's accuracy and reliability in predicting stellar luminosity. The close match between true and predicted values, along with narrow prediction intervals, underscores the robustness of the model's predictions, reinforcing its suitability for this application. To estimate SHAP value uncertainty, we used bootstrapping techniques, resampling the training data multiple times. This provided a robust measure of uncertainty, ensuring the reliability of our findings.

5 Conclusion

This research project has highlighted the powerful integration of advanced machine-learning techniques with astrophysical data analysis to predict stellar luminosity. By utilising the extensive dataset from the Gaia DR3 mission and testing it on a variety of machine-learning models, we not only achieved the best-performing model but also the influence of each feature across the models.

The Support Vector Regressor (SVR) emerged as the most effective model, showcasing exceptional performance with the lowest Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), along with the highest Coefficient of Determination (R^2) among all tested models. This excellent performance shows the SVR's ability to capture complex patterns within astronomical data.

The application of SHAP (SHapley Additive exPlanations) values was instrumental in explaining the 'black box' nature of our machine learning models. SHAP analysis consistently identified the log-transformed distance as the most influential feature in predicting stellar luminosity across all models, aligning with established astrophysical principles and validating the model's scientific accuracy. Photometric measurements (Green, Blue, and Red Photometer Mean Magnitudes) also demonstrated significant and stable importance, underscoring their critical role in determining stellar brightness.

The SHAP Dependence plots revealed clear linear relationships between key features and their impact on predictions, providing interpretable insights that align with astronomical theories. Additionally, the low uncertainties observed in the SHAP Value Uncertainty Analysis for key features bolstered our confidence in the model's robustness and generalizability.

This project has advanced our capability to predict stellar luminosity and paved the way for transparent and interpretable machine-learning applications in astrophysics. Integrating SHAP values with traditional models offers a promising approach to bridge the gap between complex computational methods and scientific interpretability.

To summerize, this research project demonstrates the potential of explainable AI in enhancing our understanding of stellar properties. By combining the predictive power of machine learning with the interpretability offered by SHAP values, we have developed a robust and scientifically consistent approach to stellar luminosity prediction. This methodology not only improves our ability to analyze extensive astronomical datasets but also ensures that the insights gained are interpretable and aligned with established astrophysical principles, thereby contributing to the ongoing evolution of computational astrophysics.

5.1 Future Work

In future research, several key areas should be explored to further enhance the predictive power and interpretability of our models, particularly with the application of SHAP values:

- **Enhanced Feature Engineering with SHAP:** Future research could utilize SHAP values to identify and incorporate additional relevant features from the Gaia DR3 dataset or other astronomical databases. This could involve features such as stellar age, metallicity, and radial velocity, which may provide deeper insights and improve model accuracy.
- **Advanced Uncertainty Quantification Methods:** Developing and integrating advanced methods for uncertainty quantification, such as Bayesian inference, ensemble methods, and probabilistic modelling, will help provide more reliable predictions. Combining these techniques with SHAP values can offer robust uncertainty estimates, crucial for scientific research and decision-making processes.
- **Integration of SHAP with Temporal Analysis:** Future research could integrate SHAP analysis with temporal data to understand how feature importance and model predictions change over time. This approach can provide deeper insights into dynamic systems in astrophysics, allowing for the prediction of stellar behavior and changes with greater accuracy and interpretability.

By pursuing these future directions, we can continue to improve the accuracy, interpretability, and applicability of machine learning models in astrophysics, ultimately advancing our understanding of the universe.

Acknowledgements

I extend my deepest gratitude to Prof. Dino Sejdinovic, whose invaluable guidance was pivotal in successfully completing this project focused on Shapley values for explaining machine learning models. Prof. Sejdinovic's profound expertise greatly enriched my understanding and approach, enabling me to effectively explore and apply sophisticated analytical techniques.

A Appendices

The codebase for our report can be found on the following GitHub link: https://github.com/nikaps99/StellarLuminosityPredictor/blob/main/Project_Code.ipynb

References

[1] Jung Min Ahn, Jungwook Kim, and Kyunghyun Kim. Ensemble machine learning of gradient boosting (xgboost, lightgbm, catboost) and attention-based cnn-lstm for harmful algal blooms forecasting. *Toxins*, 15(10), 2023.

[2] C. R. J. Alfredo. Galactic archaeology: Tracing the milky way's formation and evolution through stellar populations. *J Electrical Electron Eng*, 2(4):559–564, 2023.

[3] C. Babusiaux, C. Fabricius, S. Khanna, T. Muraveva, C. Reylé, F. Spoto, A. Vallenari, X. Luri, F. Arenou, M. A. Alvarez, F. Anders, T. Antoja, E. Balbinot, C. Barache, N. Bauchet, D. Bossini, D. Busonero, T. Cantat-Gaudin, J. M. Carrasco, C. Dafonte, S. Diakite, F. Figueras, A. Garcia-Gutierrez, A. Garofalo, A. Helmi, O. Jimenez-Arranz, C. Jordi, P. Kervella, Z. Kostrzewa-Rutkowska, N. Leclerc, E. Licata, M. Manteiga, A. Masip, M. Monguio, P. Ramos, N. Robichon, A. C. Robin, M. Romero-Gomez, A. Saez, R. Santovena, L. Spina, G. Torralba Elipe, and M. Weiler. Gaia data release 3: Catalogue validation. *Astronomy & Astrophysics*, 674:A32, 2023.

[4] Dillon Bowen and Lyle Ungar. Generalized shap: Generating multiple types of explanations in machine learning. *arXiv*, June 2020.

[5] Li Chao, Zhang Wen-hui, and Lin Ji-ming. Study of star/galaxy classification based on the xgboost algorithm. *Chinese Astronomy and Astrophysics*, 43(4):539–548, 2019.

[6] Chang Chen and A. Pullen. Removing interlopers from intensity mapping probes of primordial non-gaussianity. *Monthly Notices of the Royal Astronomical Society*, 2021.

[7] Santiago Cifuentes, Leopoldo Bertossi, Nina Pardal, Sergio Abriola, Maria Vanina Martinez, and Miguel Romero. The distributional uncertainty of the shap score in explainable machine learning, 2024.

[8] L. Delchambre, C. A. L. Bailer-Jones, I. Bellas-Velidis, R. Drimmel, D. Garabato, R. Carballo, D. Hatzidimitriou, D. J. Marshall, R. Andrae, C. Dafonte, E. Livanou, M. Fouesneau, E. L. Licata, H. E. P. Lindstrøm, M. Manteiga, C. Robin, A. Silvelo, A. Abreu Aramburu, M. A. Álvarez, J. Bakker, A. Bijaoui, N. Brouillet, E. Brugaletta, A. Burlacu, L. Casamiquela, L. Chaoul, A. Chiavassa, G. Contursi, W. J. Cooper, O. L. Creevey, A. Dapergolas, P. de Laverny, C. Demouchy, T. E. Dharmawardena, B. Edvardsson, Y. Frémat,

P. García-Lario, M. García-Torres, A. Gavel, A. Gomez, I. González-Santamaría, U. Heiter, A. Jean-Antoine Piccolo, M. Kontizas, G. Kordopatis, A. J. Korn, A. C. Lanzafame, Y. Lebreton, A. Lobel, A. Lorca, A. Magdaleno Romeo, F. Marocco, N. Mary, C. Nicolas, C. Ordenovic, F. Pailler, P. A. Palicio, L. Pallas-Quintela, C. Panem, B. Pichon, E. Poggio, A. Recio-Blanco, F. Riclet, J. Rybizki, R. Santoveña, L. M. Sarro, M. S. Schultheis, M. Segol, I. Slezak, R. L. Smart, R. Sordo, C. Soubiran, M. Süveges, F. Thévenin, G. Torralba Elipe, A. Ulla, E. Utrilla, A. Vallenari, E. van Dillen, H. Zhao, and J. Zorec. Gaia data release 3. apsis. iii. non-stellar content and source classification. *Astronomy & Astrophysics*, 674:A31, 2023.

[9] Stephen J. Edberg. Studies of a population of stars: Distances and motions, 2010.

[10] Fangzhou Guo, Jie Lin, Xiaofeng Wang, Xiaodian Chen, Tanda Li, Liyang Chen, Qiqi Xia, Jun Mo, Gaobo Xi, Jicheng Zhang, Qichun Liu, Xiaojun Jiang, Shengyu Yan, Haowei Peng, Jialian Liu, Wenxiong Li, Weili Lin, Danfeng Xiang, Xiaoran Ma, and Yongzhi Cai. Minute-cadence observations of the LAMOST Fields with the TMTS – V. Machine learning classification of TMTS catalogues of periodic variable stars. *Monthly Notices of the Royal Astronomical Society*, 528(4):6997–7015, 02 2024.

[11] Munir Hiabu, Joseph T. Meyer, and Marvin N. Wright. Unifying local and global model explanations by functional decomposition of low dimensional structures, 2023.

[12] Maxwell Hogan, Nabil Aouf, Phillippa Spencer, and Jay Almond. Explainable object detection for uncrewed aerial vehicles using kernelshap. In *2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 136–141, 2022.

[13] Tianzhi Huang, Dejin Le, Lili Yuan, Shoujia Xu, and Xiulan Peng. Machine learning for prediction of in-hospital mortality in lung cancer patients admitted to intensive care unit. *PLoS ONE*, 18(1):e0280606, 2023.

[14] D. W. Hughes. The introduction of absolute magnitude (1902 - 1922). *Journal of Astronomical History and Heritage*, 9(2):173–179, 2006.

[15] N. Langer and R. P. Kudritzki. The spectroscopic hertzsprung-russell diagram. *Astronomy & Astrophysics*, 564:A52, April 2014.

[16] JCS Leite. A machine learning approach to determine stellar atmospheric parameters using spectral data. 2023.

[17] Hui Li, Rong-Wang Li, Peng Shu, and Yu-Qiang Li. Machine learning-based identification of contaminated images in light curve data preprocessing. *Research in Astronomy and Astrophysics*, 24(4):045025, apr 2024.

[18] Junyao Li, Yuxiang Luo, Meina Dong, Yating Liang, Xuejing Zhao, Zhaoming Ge, and Yafeng Zhang. Tree-based risk factor identification and stroke level prediction in stroke cohort study. *BioMed Research International*, 2023:7352191, 2023.

[19] Yi-Ching Lin and Fang Yu. Deepshap summary for adversarial example detection. In *2023 IEEE/ACM International Workshop on Deep Learning for Testing and Testing for Deep Learning (DeepTest)*, pages 17–24, 2023.

[20] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*, 2017.

[21] P. Venkata Mahesh, S. Meyyappan, and Rama Koteswara Rao Alla. Support vector regression machine learning based maximum power point tracking for solar photovoltaic systems. *International Journal of Electrical and Computer Engineering Systems*, 14(1):100–108, Jan. 2023.

[22] Lorenza Nanni, Daniel Thomas, James Trayford, Claudia Maraston, Justus Neumann, David R. Law, Lewis Hill, Annalisa Pillepich, Renbin Yan, Yanping Chen, and Dan Lazarz. iMaNGA: mock MaNGA galaxies based on IllustrisTNG and MaStar SSPs – II. The catalogue. *Monthly Notices of the Royal Astronomical Society*, 522:5479–5499, May 2023.

[23] Kefeng Ning, Min Liu, Mingyu Dong, and Zhansong Wu. Robust ls-svr based on variational bayesian and its applications. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 2920–2926, 2014.

[24] P M Plewa. Random forest classification of stars in the Galactic Centre. *Monthly Notices of the Royal Astronomical Society*, 476(3):3974–3980, 02 2018.

[25] Atila Poro, Mahya Hedayatjoo, Maryam Nastaran, Mahshid Nourmohammad, Hossein Azarara, Sepideh AlipourSoudmand, Fatemeh AzarinBarzandig, Razieh Aliakbari, Sadegh Nasirian, and Nazanin Kahali Poor. Estimating the absolute parameters of w uma-type binary stars using gaia dr3 parallax. *New Astronomy*, 110:102227, 2024.

[26] A Sheik Abdullah, Devika Rajasekar, and A Rahul. A machine learning approach for the identification and estimation of pulsars upon statistical analysis. In *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, pages 1–12, 2023.

[27] Zehua Shuai, Tae J. Kwon, and Qian Xie. Using explainable ai for enhanced understanding of winter road safety: insights with support vector machines and shap. *Canadian Journal of Civil Engineering*, 0(0):null, 0.

[28] Huanjing Wang, Qianxin Liang, John T. Hancock, and Taghi M. Khoshgoftaar. Feature selection strategies: a comparative analysis of shap-value and importance-based methods. *Journal of Big Data*, 11(1):44, 2024.

[29] Ping Wang, Sanqing Tan, Gui Zhang, Shuang Wang, and Xin Wu. Remote sensing estimation of forest aboveground biomass based on lasso-svr. *Forests*, 13(10), 2022.

[30] Peng Yu, Chao Xu, Albert Bifet, and Jesse Read. Linear treeshap, 2023.