University of Bristol

SCHOOL OF PHYSICS

SCIF30006: Advanced Data Science for Scientific Computing Mini-Project

Comarison of Machine Learning Algorithms for Pulsar Classification

Rory Tidmarsh

April 2, 2025

Abstract

Detection of pulsars can be difficult due to cosmological noise obscuring signals. Machine learning classification algorithms (SVM, BLR and Random Forest) are applied to a dataset for pulsar classification. All models were applied to an imbalanced labelled dataset, achieving high accuracy ($> 98.0\%$) and F1 scores (≥ 0.884), demonstrating their effectiveness at identifying pulsars. When compared to an unsupervised K-means clustering model, the supervised approaches consistently outperformed it across all metrics.

The Random Forest method is recommended, showing a marginally higher cross-validated performance (F1: 0.907) and the AUC score of 0.994 for Receiver Operating Characteristic curves, indicating robust performance across different decision thresholds. Nevertheless, all supervised models performed well and can be used for label prediction on larger databases.

1 Introduction

Pulsars are a type of neutron star that remains after a supernova. Pulsars contain extremely strong magnetic fields and emit strong beams of electromagnetic radiation beams that can be detected by laboratories on earth [1]. However, signals detecting pulsars can be obscured by radio frequency or cosmological noise, making detection difficult. In this report different machine learning algorithms are trained on a labelled in-house database and evaluated for use in predicting labels on larger datasets.

Machine learning is a field of computer science that enables computers to learn from data using algorithms and statistical models, without being explicitly programmed with fixed rules [2]. These algorithms identify patterns and relationships in complex datasets, allowing them to extract meaningful insights and make predictions. Machine learning is widely used in applications such as medical diagnosis, image analysis, fraud detection, classification and language analysis [3].

There are many different algorithms available to solve data problems, with each algorithm suiting different types of problems. Unsupervised learning is a subset of machine learning algorithms that identify patterns in data without any predefined labels or known outcomes. These models allow computers to identify underlying structures such as clusters, trajectories or anomalies without the need for human supervision [4]. Unsupervised learning is commonly used in applications like anomaly detection, dimensionality reduction and clustering [5].

Supervised machine learning algorithms are functions that maps an input to an output. Unlike unsupervised learning, this requires a labelled set of training data where the algorithm maps the input dataset onto the given labels [2]. These supervised models learn to predict outcomes by analysing the relationships between the input data and the target variables, enabling them to make predictions on new, unseen data. Supervised machine learning tasks can be divided into two types of problems: regression and classification problems. Regression tasks involve the algorithm to investigate the relationship in order to predict a continuous numerical value [6]. These modes predict how changes in independent variables affect a dependent variable, making them perfect for predicting price changes, temperatures, sports performances, or the likelihood of responding well to medicinal treatment [4] .

On the other hand, classification problems learn decision boundaries in order to separate data points into different groups, predetermined by their input labels. The simplest case is binary classification where points are classified into 1 or 0 groups ("yes" or "no") with more complex multi-class classification sorting data into many categories. Classification algorithms are used in spam detection, medical diagnosis and document classification [3]. This report focuses on applying classification algorithms to detect pulsars in astronomical data [7], a binary problem where objects are either classified as pulsars or non-pulsars.

1.1 Classification Models

Three supervised machine learning models are considered to asses whether the data is appropriate for classifying pulsars: Binary Logistical Regression (BLR), Support Vector Machines (SVM) and Random Forest models. These models were chosen because of the binary classification required from the dataset, with each having different strengths at capturing relationships in the data. A fourth unsupervised model, K-Means clustering, is also evaluated to compare the performance to the supervised models.

The Binary Logistical Regression (BLR) is a probabilistic model to map binary outcomes [8], like in the scenario we have here. It uses a multi-dimensional sigmoidal cost function to transform to transform input features into a probability between 0 and 1. Given a certain input probability threshold, the values will predicted as 0 or 1, depending on if its probability is below or above the decision threshold, respectively. The model estimates coefficients for each feature that maximise the likelihood of observing the results seen in the training data. These coefficients directly indicate the impact and direction of each feature's influence on the classification outcome. BLR is very interpretable as the coefficients directly represent the importance of each feature, however, it assumes linear decision boundaries and no multicollinearity between features [8]. It is a relativity simple, computationally efficient model that produces accurate estimates when the assumptions are met.

Support Vector Machines (SVM) uses a multidimensional hyperplane to separate classes. The hyperplane is the decision boundary and is optimised so that the margins between two classes are at a maximum distance apart, hence minimising errors when classifying [2]. The "kernel trick" in SVM transforms data into higher-dimensional space, allowing for the definition of an optimal hyperplane with maximized margin, without the need of explicitly calculating coordinates in this transformed space [9]. Common kernels include Linear, Polynomial and Radial Base Function (RBF), each suited to different patterns in datasets. SVM works well in high dimensional spaces and in areas where there is a clear margin between classes, however, it can be computationally expensive, with long training times for large datasets and requires correct selection of kernels.

Random Forest (RF) is a combination model consisting of multiple decision trees [10]. Decision trees (DT) split the data via different conditions, creating a structure like a flow chart. Each internal node represents a decision on a feature to split data, each branch represents the outcome of that decision, and the leaf nodes represent a label [2]. DT perform repeated partitions in data to maximise the class separation. However, the maximum depth of the DT must be considered as a high depth, resulting in more partitions, can lead to over-fitting and a low depth to under-fitting. RF models are an extension of the DT model by combining many trees. Each tree is trained on bootstrapped samples of the data, considering random features for each partition. The data is finally classified by a majority vote to decide which label a data point has. This ensemble learning approach to decision-making reduces over-fitting, handles non-linear relationships well, is tolerant to noise, and provides feature importance metrics.

In addition to the supervised models, this report examines K-means clustering as an unsupervised learning model. Unlike the previous models, K-means does not require labelled training, instead groups data points based on the similarities in features, minimising the distance of points within clusters, and grouping close data points across all features. Using K=2 clusters, this approach matches the binary nature of the pulsar classification problem. The K-means attempts to naturally separate the data into groups, ideally corresponding to pulsar or non-pulsar clusters. This approach allows a comparison between supervised and unsupervised models and investigates if the data has a natural label divide.

1.2 Performance Metrics

Performance metrics evaluate how well a model predicts the correct classes. Using test data, the model can find out the number of true positives (TP), false positive (FP), true negative (TN) and false negative (FN) predictions from the model.

The accuracy measures the proportion of correctly classified instances (TP and TN) to the total number of predictions (TP+FP+TN+FN). Accuracy incorporates all four outcomes of the confusion matrix (matrix containing all prediction states), however, can be misleading when the dataset is imbalanced. One mistake can (false positive or negative) influence the value much more.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Recall (or sensitivity or true positive rate - TPR) is the ratio of true positive predictions to all actual positives, giving a measure of how well a model identifies positives.

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{2}$$

Precision is the ratio of true positives to all positive predictions, giving a measure of how accurate

Original Feature name	**Abbreviation**
Mean of the integrated profile	Mean_IP
Standard deviation of the integrated profile	StdDev_IP
Excess kurtosis of the integrated profile	Kurtosis_IP
Skewness of the integrated profile	Skew_IP
Mean of the DM-SNR curve	Mean_DM_SNR
Standard deviation of the DM-SNR curve	StdDev_DM_SNR
Excess kurtosis of the DM-SNR curve	Kurtosis_DM_SNR
Skewness of the DM-SNR curve	Skew_DM_SNR
Class label (0 = negative, 1 = positive)	Label

Table 1: *Dataset Feature Names and Their Abbreviations. The "Label" feature is used for classification, only containing discrete values of 0 or 1.*

the positive predictions are. A high precision means that the number of false positives is low and a large proportion of the predictions are accurate.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

The F1 Score is a way of combining the precision and the accuracy into a single value,

$$F_1 = 2 \cdot \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \tag{4}$$

The F1 Score is a very useful metric for analysing datasets when the number of actual true and false values are unbalanced. An F1 score of 1 indicates perfect precision and recall, while a score of 0 suggests that either precision, recall, or both are zero, meaning the model is failing to correctly identify positive cases. The F1 score provides more insight than accuracy in such situations. For example, in an imbalanced dataset where the model correctly predicts the majority case (fails), then the accuracy can still be high. But if the F1 score is low, it suggests that the model struggles to identify the minority cases (positive), which may be more important to detect, especially in the case of astromical data.

Although the F1 score provides a single metric, it cannot capture performance across different classification thresholds. Receiver Operating Characteristic (ROC) curves can be used by plotting the True Positive Rate (recall, equation 2) against the False Positive Rate ($FPR = FP/(FP + TN)$) across various decision thresholds.

A ROC curve approaching the top left of the plot represents a nearly ideal classifier with a high TPR and low FPR across various threshold settings. The Area Under the ROC Curve (AUC) quantifies this numerically. An AUC of 1.0 represents perfect classification, while 0.5 means the performance is comparable to randomly guessing.

2 Analysis

2.1 Exploring the Dataset

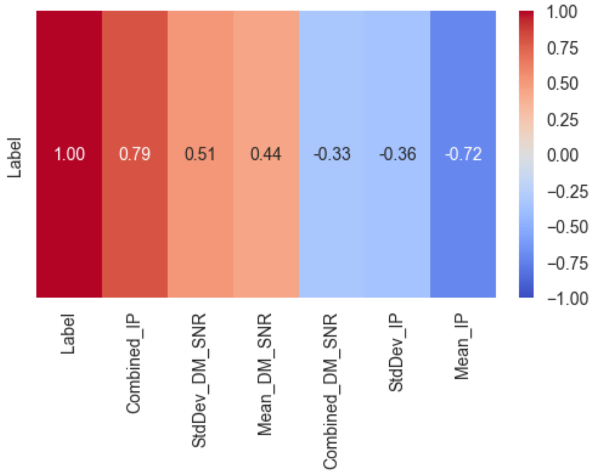


Figure 1: *Correlation with the label Column for Pulsar dataset (see Table 1 for abbreviations). Correlations give an initial insight into the importance of each feature when classifying through supervised machine learning algorithms.*

Upon initial review of the data, eight feature columns were found describing the cosmological object, and one label column identifying if it was a pulsar (1) or not (0), allowing for the entire dataset to be used for training or testing models. The total data set contained 1748 points with only 150 of those being positive labels identifying pulsars, showing that we have an unbalanced dataset. For clarity and readability in figures and later discussion, Table 1 defines the abbreviations of feature names used throughout this report.

The correlation matrix (Appendix A.1) showed multicollinearity amongst features, "Kurtosis_IP" with "Skew_IP" (correlation coefficient of 0.95) and "Kurtosis_DM_SNR" with "Skew_DM_SNR" (correlation coefficient of 0.92). To address this without losing information, two new combined columns were created, "Combined_IP" and "Combined_DM_SNR". Combining them required the application of the standard scalar to the whole data set, ensuring all features are on the same scale with no units.The new combined feature columns were created by calculating the mean value of relevant data points for each entry. The high correlation in features "Kurtosis" and "Skewness" arise from both representing shape properties. Kurtosis represents how heavy-tailed a shape is and Skewness is a measure of symmetry, with a change in one often meaning a change in the other.

This approach to dimensionality reduction aligns with one of the key assumptions for BLR: no multicollinearity. The information from the previous features is retained but the new columns now make BLR a more viable option for classifying this data. For Random Forest models and SVM, although both robust to multicollinearity, this approach reduces complexity in the data and increases model stability, whilst retaining relevant information, allowing for more efficient computations. The feature importance metrics will also remain more meaningful, allowing for direct insight into feature importance within the dataset, something which is less apparent for other techniques like a PCA. Another approach to reduce dimensionality would be to remove features with low correlations to the label column, reducing noise in features that are not important. This was not applied as the dataset is now at six features, reducing further could risk removing valuable information and is not beneficial to the Pulsar dataset.

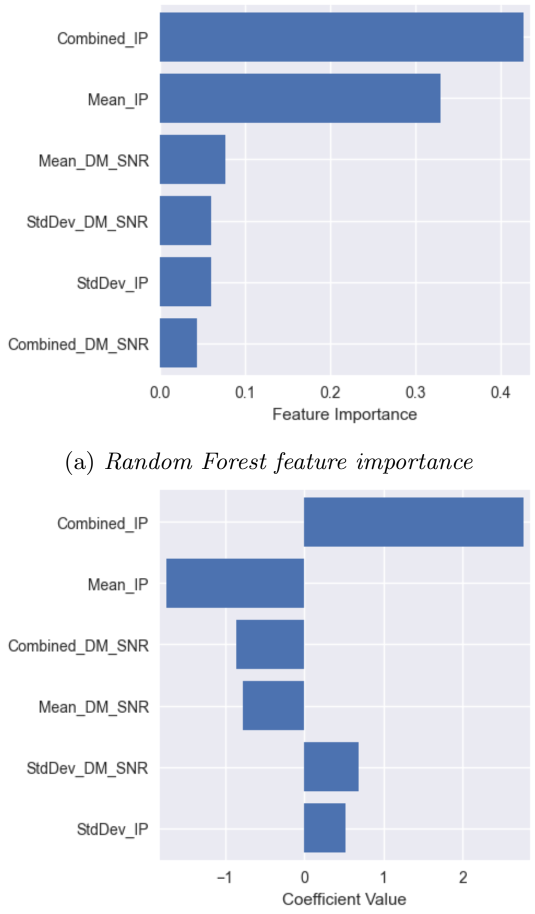
Figure 1 displays the correlation coefficients between each feature and the target label in the dimensionally reduced data set. From this we can expect the "Combined_IP" (correlation coefficient = 0.79) and "Mean_IP" (correlation coefficient = -0.72) to have the largest influence when classifying the data.

The final step when preprocessing the data was to split the data into training (70%) and test subsets (30%). The split used stratified sampling to maintain the distribution of labels across the partition in data, an important step given the imbalanced label counts in the data, and a random state of 42 (this state is used for all random processes).

2.2 Model Performance

Models were trained in a python notebook running python 3.12.8, using Scikit-Learn 1.5.2, and on the same training split of the data. The Random Forest model was run with 100 estimators with a maximum depth of 4, ensuing high accuracy but also reducing over fitting of data. SVM used the RBF kernel as it outperformed the linear kernel in initial testing, capturing the complex relationships between columns more effectively. The BLR model was run with the 'saga' solver and with a maximum of 500 iterations to ensure convergence on predictions. A K-means cluster algorithm was also implemented to compare the performance of unsupervised and supervised approaches to classification. This used 2 clusters to match the binary nature of the classification problem, allowing for comparisons of clusters to actual labels in the Pulsars dataset. All models used the same random state (42) for reproducibility and comparison.

The Random Forest (RF) model and BLR model have feature importance outputs, for BLR this comes from the coefficients in the multidimensional cost function. RF feature importance (Figure 2a) demon-



(a) *Random Forest feature importance*

(b) *BLR cost function coefficients*

Figure 2: *Feature importance for 2a Random Forest and 2b BLR models. Both show that "Combined_IP" and "Mean_IP" are the most important in classifying Pulsars, with other features having small impacts. 2a Feature importance scores represent the fractional contribution of each feature. 2b shows that "Mean_IP" has a large negative effect, with large values in this feature causing negative predictions. See Table 1 for feature abbreviations.*

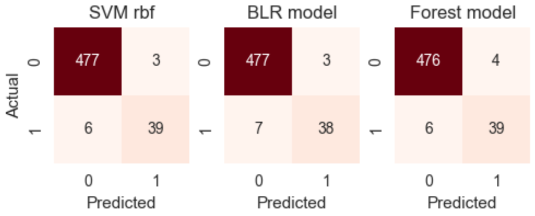


Figure 3: *Confusion matrices on test subset of data for Supervised Machine learning models: SVM, Random Forest and BLR. A classification of* 1 *represents a positive pulsar detection,* 0 *is not a Pulsar. This means* $(0, 0)$ *represents a true negative,* $(1, 1)$ *true positive,* $(0, 1)$ *false negative, and* $(1, 0)$ *false positive.*

Model	Accuracy	Precision	Recall	F1 Score	AUC	F1 Cross Validation
SVM (rbf)	0.983	0.929	0.867	0.897	0.981	0.900
Random Forest	0.981	0.907	0.867	0.886	0.994	0.907
BLR	0.981	0.927	0.844	0.884	0.984	0.894
K Means	0.941	0.622	0.800	0.700		0.661

Table 2: *Model performance metrics for different models applied to the test data. Appendix A.2 provides a clearer visualisation for these performance metrics.*

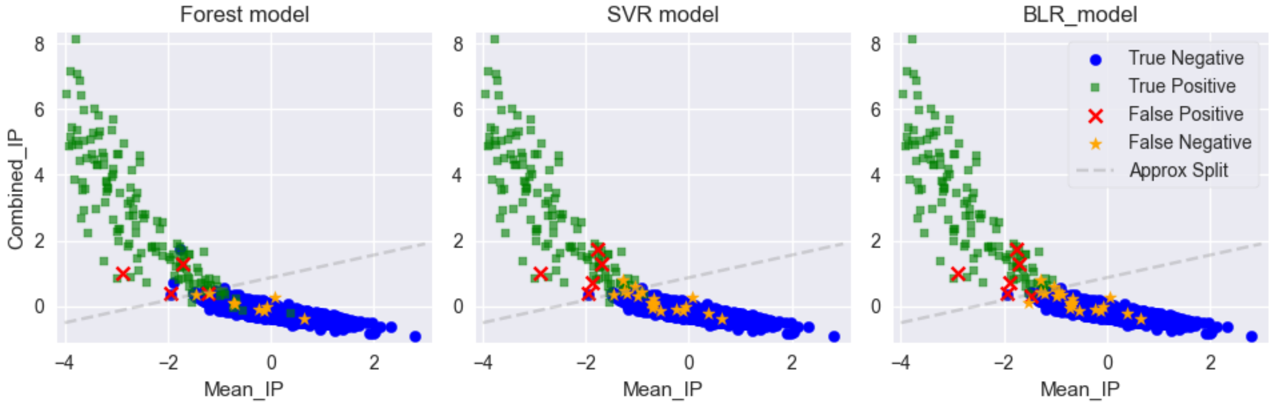


Figure 4: *Classification of whole pulsar dataset for features "Combined_IP" and "Mean_IP" for different supervised machine learning algorithms. See Table 1 for feature abbreviations.*

strates the prediction from the correlation matrix (Figure 1), "Combined_IP" is the most important column, followed by "Mean_IP". The feature importance scores represent the fractional contribution to the models predictive ability based on the mean decrease in impurity across all trees.

For the BLR model (Figure 2b), the coefficient values represent log-odds ratios, describing both the size and direction of the features effect on the classification labels. The large positive coefficient of 2.77 for "Combined_IP" demonstrates that a unit increase in this standardised feature increases the odds of pulsar classification by approximately a factor of 16 ($\times e^{2.77}$). On the other hand, the large negative coefficient of -1.74 for "Mean_IP" decreases the odds of a Pulsar classification, multiplying the odds by 0.18. The signs and sizes of these columns aligns with correlations observed in the correlation matrix (Figure 1, with the highly correlated or inversely correlated columns having more importance on classification than other feature columns.

The confusion matrices (Figure 3) provides an insight into each model's predictive performance for different classification outcomes. The confusion matrix compares the predictions from the model with the actual labels, displaying the TN (non-pulsars correctly predicted), FP (non-pulsars classified as pulsars), FN (pulsars classified as non-pulsars) and TP (pulsars correctly predicted). This shows similar predictive performance on the testing data for all models, small misclassification rates (FP and FN) and correctly identifying a majority split of non-pulsars (TN). The accuracy on this test data can then be calculated (Equation 1), resulting in: 0.983 for SVM, 0.981 for both BLR and Random Forest models. Accuracies can vary when changing the random state of the split, but this trend of high accuracy (> 0.980) is consistent.

This consistency across the confusion matrices shows that the data is very suitable for classification purposes. The similarity also shows that data characteristics, rather than model-specific features, aid the classification the most. The near linear separation observed in the two most important features, "Combined_IP" and "Mean_IP" (shown in Figure 4) provides an explanation of the consistent classification results.

The feature visualisation reveals that misclassification typically to occur when data points fall on the opposite side of the near-linear decision bounties to the majority of similarly labelled points. These outliers represent cases where the most important features in classification may predict one label but other characteristics mean the label is actually different. By inspection, the Random Forest model shows slightly better handling of these boundary cases due to its ability to handle more complex relationships. The BLR model depends more on this linear classification with a clear linear separation of positives and negative detections on opposing sides of the barrier.

The near linear boundary between Pulsar and non-Pulsar classes means that more simple model, like BLR, can be used to classify the data and still achieve high accuracy. The more complex models (Random Forest and SVM), while capable of capturing more complex trends, don't significantly improve the performance due to to this fundamental data feature.

The Receiver Operating Characteristic (ROC) curves (Figure 5) give further insight into the models' performance. All three supervised models perform well,

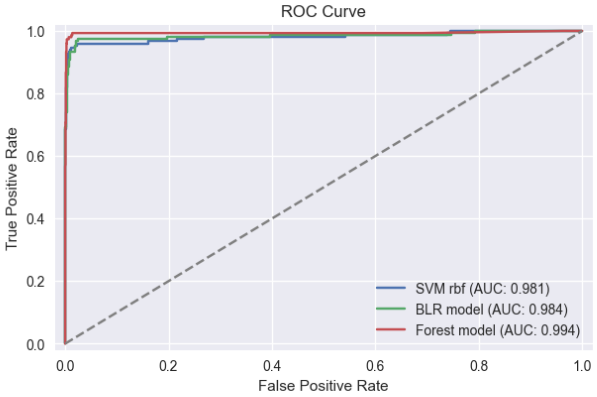


Figure 5: *Receiver Operating Characteristic (ROC) curves for classifying pulsars using 3 different supervised machine learning models: SVM (RBF kernel), BLR and Random Forest. Random Forest had the highest Area Under Curve (AUC) value, demonstrating that it is the strongest at predicting Pulsars, with a high True Positive Rate across all thresholds. However, all models show strong predictions with all AUC values* > 0.98.

with each curve reaching into the top left corner of the plot, indicating that all models maintain a high TPR (recall) with low FPR across different decision thresholds. The Random Forest model performs marginally better (AUC= 0.994) than the SVM (AUC= 0.981) and the BLR (AUC= 0.984).

Comparing performance metrics (shown in Table 2, or visualised in Appendix A.2) performance of each model, when applied to the test data, can be compared. All supervised models show high accuracies (≥ 9.81), but considering we have an unbalanced dataset (150 Pulsars out of 1748 total object) investigating the F1 score is more useful to consider the performance concerning identifying pulsars.

The SVM achieves the highest F1 score (0.897) due to the highest balance between precision (0.929) and recall (0.867). The Random Forest and BLR models show similar F1 scores (0.886 and 0.884 respectively), with Random Forest having a slightly lower precision but matching SVM's recall. This demonstrates that all models can identify Pulsars well, despite the imbalanced dataset. This is important in astrological data as pulsars are infrequent amongst cosmological objects and having a model accurately identifies pulsars amongst noise is essential.

In order to gain a more comprehensive insight, the cross-validation scores evaluate how each model performs across different splits of the data, creating a more generalised score. The cross-validation score was evaluated using 5 test sets on the entire dataset, focussing on the F1 score, getting a more generalised assessment on the performance with imbalanced classes. Random Forest achieves the highest mean across the 5 tests (0.907), followed by SVM (0.900) and BLR (0.894). All supervised models perform excellently in this metric, showing that they all classify the minority case of Pulsars well.

The two clusters output by the K-means were compared to the known labels, allowing for the calculation of descriptive metrics as if it were a binary classification algorithm. The unsupervised K-means approach performs substantially lower across all metrics, with an F1 score of 0.700 and a precision of 0.622, while maintaining a reasonably high recall (0.800). The much lower cross-validated F1 score (0.662) further demonstrates that this model does not generalise well compared to the supervised methods. The performance gap shown here shows the huge advantage of using labelled training data in order to classify Pulsars.

3 Conclusion

The Pulsar database can be used effectively to train a supervised machine learning model that can predict labels for a larger database. Consistently high performance in classification is seen across the three supervised models used: SVM, Random Forest and BLR. All models maintained high accuracy (> 0.98) and a high F1 score (≥ 0.884), demonstrating models can identify Pulsars effectively despite the imbalanced dataset (150 Pulsars out of 1748 samples). The Random Forest model has a superior AUC score (0.994) when considering the ROC curve, indicating that the dataset has great discriminatory features. The "Combined_IP" and "Mean_IP" features show a near-linear separation of labels, explaining why even simple models like BLR work well.

The preprocessing approach of applying the standard scalar and combining correlated columns reduces the dimensionality of the data, whilst maintaining interpretability. This addresses the issues of multicollinearity, allowing for the BLR model to be applied.

The unsupervised learning approach using K-means with two clusters (to mimic binary Pulsar non-Pulsar classification) consistently underperformed. The F1 score (0.700) is comparatively low compared to all of the supervised learning approaches, demonstrating that assigning these clusters as labels is an ineffective classification approach.

The recommendation to researchers is that a Random Forest classifier model with a maximum depth of 4 can be trained on this data for label prediction on larger datasets. This models strong cross-validation and AUC performance shows that it can perform well across various decision thresholds on new astronomical data. However, if a less computationally intense algorithm is required both BLR and SVM provide excellent alternatives. For optimal results, the preprocessing should include feature standardization and combine highly correlated columns for kurtosis and skewness for both the integrated profile and the DM_SNR.

References

[1] Teruaki Enoto, Shota Kisaka, and Shinpei Shibata. “Observational diversity of magnetized neutron stars”. en. In: *Reports on Progress in Physics* 82.10 (Sept. 2019). Publisher: IOP Publishing, p. 106901. ISSN: 0034-4885. DOI: `10.1088/1361-6633/ab3def`.

[2] Batta Mahesh et al. “Machine learning algorithms-a review”. In: *International Journal of Science and Research (IJSR).* 9.1 (2020), pp. 381–386.

[3] Pramila P. Shinde and Seema Shah. “A Review of Machine Learning and Deep Learning Applications”. In: *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).* Aug. 2018, pp. 1–6. DOI: `10.1109/ICCUBEA.2018.8697857`.

[4] Tammy Jiang, Jaimie L. Gradus, and Anthony J. Rosellini. “Supervised Machine Learning: A Brief Primer”. In: *Behavior Therapy* 51.5 (Sept. 2020), pp. 675–687. ISSN: 0005-7894. DOI: `10.1016/j.beth.2020.05.002`.

[5] Samreen Naeem et al. “An unsupervised machine learning algorithms: Comprehensive review”. In: *International Journal of Computing and Digital Systems* (2023).

[6] Pradeep Singh. *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools, and Applications.* en. John Wiley & Sons, Feb. 2022.

[7] M. J. Keith et al. “The High Time Resolution Universe Pulsar Survey – I. System configuration and initial discoveries”. In: *Monthly Notices of the Royal Astronomical Society* 409.2 (Dec. 2010), pp. 619–627. ISSN: 0035-8711. DOI: `10.1111/j.1365-2966.2010.17325.x`.

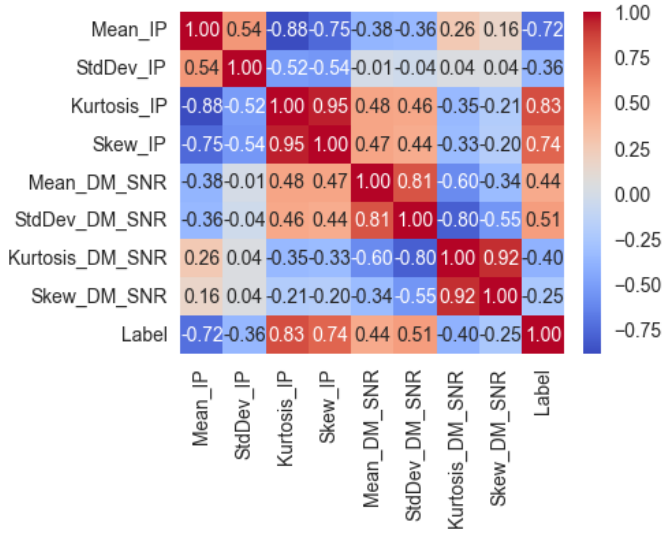
[8] Jenine K Harris. “Primer on binary logistic regression”. In: *Family Medicine and Community Health* 9.Suppl 1 (Dec. 2021). ISSN: 2305-6983. DOI: `10.1136/fmch-2021-001290`.

[9] Dirk Valkenborg et al. “Support vector machines”. English. In: *American Journal of Orthodontics and Dentofacial Orthopedics* 164.5 (Nov. 2023). Publisher: Elsevier, pp. 754–757. ISSN: 0889-5406, 1097-6752. DOI: `10.1016/j.ajodo.2023.08.003`.

[10] Yanli Liu, Yourong Wang, and Jian Zhang. “New Machine Learning Algorithm: Random Forest”. en. In: *Information Computing and Applications.* Ed. by Baoxiang Liu, Maode Ma, and Jincai Chang. Berlin, Heidelberg: Springer, 2012, pp. 246–252. DOI: `10.1007/978-3-642-34062-8_32`.

A Appendix

A.1 Correlation matrix



	Mean_IP	StdDev_IP	Kurtosis_IP	Skew_IP	Mean_DM_SNR	StdDev_DM_SNR	Kurtosis_DM_SNR	Skew_DM_SNR	Label
Mean_IP	1.00	0.54	-0.88	-0.75	-0.38	-0.36	0.26	0.16	-0.72
StdDev_IP	0.54	1.00	-0.52	-0.54	-0.01	-0.04	0.04	0.04	-0.36
Kurtosis_IP	-0.88	-0.52	1.00	0.95	0.48	0.46	-0.35	-0.21	0.83
Skew_IP	-0.75	-0.54	0.95	1.00	0.47	0.44	-0.33	-0.20	0.74
Mean_DM_SNR	-0.38	-0.01	0.48	0.47	1.00	0.81	-0.60	-0.34	0.44
StdDev_DM_SNR	-0.36	-0.04	0.46	0.44	0.81	1.00	-0.80	-0.55	0.51
Kurtosis_DM_SNR	0.26	0.04	-0.35	-0.33	-0.60	-0.80	1.00	0.92	-0.40
Skew_DM_SNR	0.16	0.04	-0.21	-0.20	-0.34	-0.55	0.92	1.00	-0.25
Label	-0.72	-0.36	0.83	0.74	0.44	0.51	-0.40	-0.25	1.00

Figure 6: *Correlations of all features directly from the Pulsars dataset.*

A.2 Performance Results

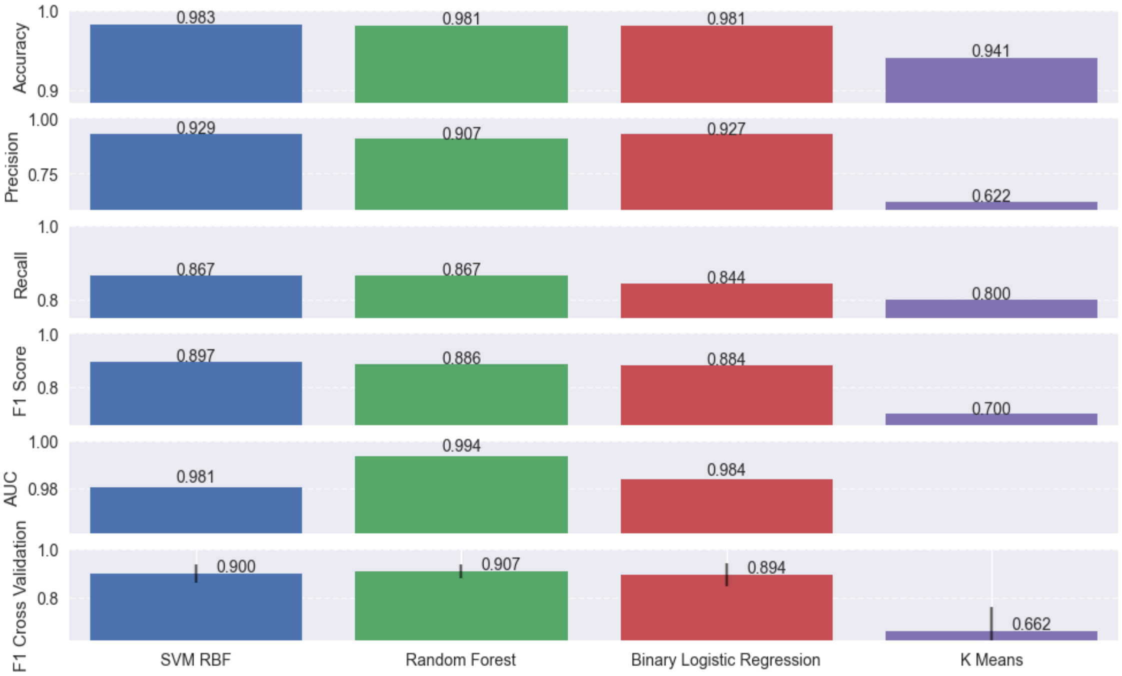


Figure 7: *Classification performance metrics on the entire dataset for three supervised machine learning models: SVM (rbf kernel), Random Forest and BLR; and an unsupervised model: K-Means. K-Means has no score for AUC this metric requires probability estimates which are not produced by unsupervised models.*