ANALYZING LOAN APPROVAL PREDICTION: A COMPARATIVE EXAMINATION OF MULTILAYER PERCEPTRON AND SUPPORT VECTOR MACHINES

MOHD ARIFULLAH
MOHD.ARIFULLAH@CITY.AC.UK

ABSTRACT:

This study conducts a comparative analysis of two predictive models for loan approval prediction Multilayer Perceptron (MLP) and Support Vector Machines (SVM). On loan approval dataset, the effectiveness of both methods is evaluated using supervised pattern recognition. Using grid search optimization, several model configurations are investigated and assessed using stratified cross-validation. The efficacy of the model is evaluated using Receiver Operating Characteristic (ROC) curves and confusion metrics. The results show that the MLP algorithm performs better when it comes to predicting loan acceptance.

1. INTRODUCTION

Machine learning approaches are leading the charge in enhancing decision-making processes, especially in the area of loan approvals, as the financial industry undergoes a transition. With a substantial influence on lenders as well as borrowers, there has never been a greater need for quick, accurate, and equitable loan approval procedures. Predictive modeling has the ability to improve decision-making in the loan approval process, reduce default risks, and customize financial products to better meet the needs of clients. This paper seeks to critically evaluate two models in predicting loan approval outcomes: a Multilayer Perceptron (MLP) and a Support Vector Machine (SVM). We delve into various configurations and data distributions to assess their effectiveness in the loan approval prediction task. With the ever-growing volume of loan applications and the complexity of financial data, machine learning models have become instrumental in predicting loan default risk, a key determinant in loan approval decisions.

Following the cleaning and removal of NaN values from the dataset, Section 2 of this paper provides a detailed description of the dataset that was utilized for both testing and training the models. Section 3 delves into the strategies and techniques employed during the implementation stage, offering insights into the methodologies adopted. In Section 4, the outcomes of the implementation, where the performances of the models are critically compared and analyzed. Section 5 concludes the paper, summarizing the key findings and future work.

1.1 MULTILAYER PERCEPTRON (MLP)

An artificial neural network called multi-layer perceptron (MLP) is made up of at least three layers of neurons that are: input, hidden and output. The activation function for every neural unit except the input neurons is non-linear. It trains utilizing a supervised learning algorithm known as backpropagation [1]. This makes it possible to represent complex functions unlike in the case of a single-layer perceptron. MLPs are commonly used for solving tasks which involve supervised learning and some unsupervised learning situations too.

Every node is a neuron that uses a nonlinear activation function, with the exception of the input nodes. MLP trains via a method known as backpropagation. In this process, the network learns from the data by adjusting the weights of connections, based on the error rate obtained in the previous epoch (i.e., each round of forward and backpropagation), aiming to minimize the error rate as much as possible.

MLPs are particularly well-suited for tasks where the relationship between the input data and the output is nonlinear and complex. Choosing a Multilayer Perceptron (MLP) for the loan approval dataset offers numerous advantages due to its ability to model complex, non-linear relationships and interactions between various financial indicators like credit score, income, and employment history.

1.2 Support Vector Machines (SVM)

The SVM is a machine-readable learning technique that can split, retrieve, and identify data [2]. SVMs are a robust set of supervised learning methods used for classification, regression, and outliers' detection. The effectiveness of SVMs in handling classification tasks makes them particularly well-suited for loan approval prediction, which fundamentally revolves around categorizing loan applications as likely to default or not. SVM operates by finding the hyperplane that best divides a dataset into classes. For loan prediction, which is inherently a classification task, SVMs are advantageous because they focus on minimizing classification errors, particularly useful in scenarios where the distinction between classes (e.g., Accepted vs. Rejected) is crucial.

1.3 Dataset

The dataset, sourced from Kaggle, comprises 4,268 entries distributed across 13 columns. It includes three categorical variables, such as Education Level and Self Employed, alongside nine numeric variables. The dataset features a target variable named 'loan_status', which categorizes each entry as either 'Accepted' or 'Rejected'.

1.1. Initial Data Analysis

The initial data analysis starts with the annual income distribution (Fig 1), where the most common incomes cluster between 200,000 to 400,000, peaking in this range before frequencies taper off towards higher incomes, which extend up to nearly 1,000,000. In examining residential assets against loan status (Fig 2), Loan Rejected (status 0) show a median asset value around 1,000,000 with values spread between 500,000 to 1,500,000, whereas defaulters (status 1) not only have a higher median value above 1,250,000 but also a greater spread, ranging from 750,000 to 1,750,000.

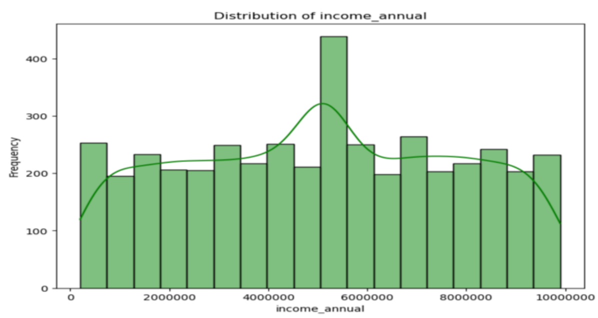

Fig. 1 Distribution of Income.

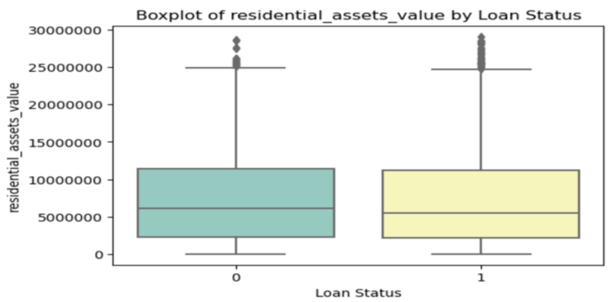

Fig. 2 Boxplot of Residential Assets Value

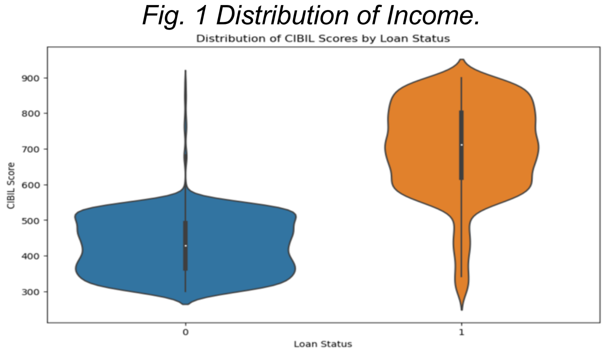

Fig. 3 CIBIL Score Violin Plot.

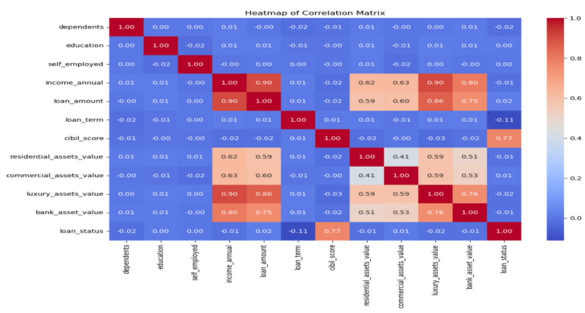

Fig. 4 Correlation Heatmap

The violin plot (Fig. 3) depicted illustrates the distribution of CIBIL scores segmented by loan status, where '0' represents rejected loans and '1' signifies approved loans. The plot clearly shows that approved loans are associated with a wider range of CIBIL scores, with a concentration towards higher scores, indicating good creditworthiness. On the contrary, rejected loans are concentrated within a narrower band of lower scores.

The correlation heatmap (Fig 4) quantifies relationships in the data, with the annual income and loan amount displaying a moderate positive correlation of 0.63, and commercial to residential asset values correlated at 0.51, suggesting that higher values in one asset type are moderately associated with higher values in the other.

2. METHODS

In this part, we describe the architecture and hyperparameters utilized in the construction of the MLP and SVM models, as well as the methods involved in training, validating, and testing them.

2.1 METHODOLOGY

The methodology employed in this project was designed to conduct a thorough assessment and comparison of the predictive abilities of Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models in terms of loan approval. In order to provide a solid analysis, the original dataset was divided into two parts: the training set, which included 80% of the data, and the testing set, which included the remaining 20%. A methodical strategy was used for the model selection phase, utilizing a Grid Search Cross-Validation (GridSearchCV) with a cross validation technique to optimize the hyperparameters of the MLP and SVM models.

In addition, a grid search technique carefully examined several combinations of hyperparameters to identify the most efficient ones based on validation accuracy. Models that were optimized with these parameters were then retrained using the combined training and validation dataset, which improved their predicted accuracy and made full use of the variety of data that was available.

2.2 ARCHITECTURE AND PARAMETERS: MLP

The Multi-Layer Perceptron (MLP) is initialized with a predetermined configuration encompassing distinct architectural characteristics and parameterization: notably, featuring two hidden layers, wherein the initial layer accommodates 100 neurons, followed by a subsequent layer with 50 neurons. The selection of Rectified Linear Unit (ReLU) activation functions for the hidden layers is motivated by their recognized efficacy and computational efficiency in effectuating non-linear transformations within neural network frameworks. Optimization of the network weights is carried out using the **'adam'** solver, renowned for its robustness in handling stochastic optimization.

To rigorously optimize and evaluate the model's performance across various configurations, a comprehensive hyperparameter tuning is conducted via `` `GridSearchCV` ``. This grid search includes an exploration of several architectural variations, encompassing both single-layer (with 50 or 100 neurons) and two-layer (50-50 or 100-100 neurons) configurations. In terms of activation functions, both 'tanh' and 'relu' are assessed to compare their impacts on the model's ability to capture non-linearities. The 'adam' solver is exclusively utilized due to its superior performance in prior applications.

Moreover, the regularization parameter 'alpha', with values set at 0.0001 and 0.05, is tested to determine the optimal balance between model complexity and generalization to new data, thereby mitigating the risk of overfitting. Additionally, the initial learning rates of 0.001 and 0.01 are included in the grid search to fine-tune the learning process, enhancing the speed and stability of convergence.

2.3 ARCHITECTURE AND PARAMETERS USED FOR THE SVM

We employ the Support Vector Machine (SVM) model, to tackle the classification task. Initially, data scaling is achieved using the `StandardScaler` to normalize the training and testing datasets, ensuring that the SVM operates under optimal conditions, given its sensitivity to the scale of input features.

The SVM is configured with a linear kernel to start, with a focus on maintaining simplicity and interpretability in the model's decision boundaries. After scaling, the SVM model is trained on the scaled training data and then used to predict the labels of the scaled test data. The model's performance is evaluated through its accuracy and detailed metrics provided by the classification report, which includes precision, recall, and F1-score for each class.

To further enhance the model's performance and explore its capabilities with different configurations, a `GridSearchCV` is conducted. Additionally, for the polynomial kernel, the degree of the polynomial is varied among 2, 3, and 4. This exhaustive search, performed across a 5-fold cross-validation setup, is aimed at identifying the optimal combination of parameters that maximize the accuracy of the SVM model on the given dataset.

3. RESULTS, FINDINGS & EVALUATION

In the process of model selection, two powerful machine learning algorithms, Support Vector Machine (SVM) and Multilayer Perceptron (MLP), were evaluated with specific parameters and tuned to classify loan statuses effectively. SVM was initially trained using a linear kernel and achieved an accuracy of 92% on the test set. After Hyperparameter tuning via GridSearchCV explored various combinations of parameters such as kernel, regularization, kernel coefficient and kernel degree, ultimately selecting an RBF kernel resulting in an optimized SVM model with a test set accuracy of 94%.

From the results, using the MLP algorithm is the most efficient approach for credit risk assessment [4]. For MLP, the initial model achieved an accuracy of 96% on the test set with max_iterations 300. Through hyperparameter tuning involving variations in network architecture, activation functions, optimizers, and regularization parameters, the best-performing model was identified. This model achieved a test set accuracy of 96.6%.

Both models demonstrated strong predictive performance, with SVM achieving an accuracy of 94% and MLP achieving 96.6% after hyperparameter tuning. The model selection process involved careful consideration of various hyperparameters to optimize the performance.

3.1 ALGORITHM COMPARISON

The comparative examination of the Support Vector Machine (SVM) and Multilayer Perceptron (MLP) models reveals their commendable performance on the test dataset, with the MLP exhibiting a marginally superior performance at the outset. Initially, the SVM attains an accuracy of 0.92, accompanied by precision, recall, and F1 scores of 0.94, 0.93, and 0.94, respectively. Following hyperparameter optimization, these metrics demonstrate a uniform enhancement of 0.02, thereby elevating the SVM's performance to achieve scores of 0.94, 0.95, 0.95, and 0.95, respectively.

In contrast, the MLP starts with stronger metrics—an accuracy of 0.96, and precision, recall, and F1 score of 0.97. After tuning, its accuracy increases to 0.966, and recall to 0.98, while precision and F1 score hold steady at 0.97. This enhancement in recall indicates an improved capability of the MLP to identify relevant cases, further solidifying its superior performance compared to the SVM for this dataset.

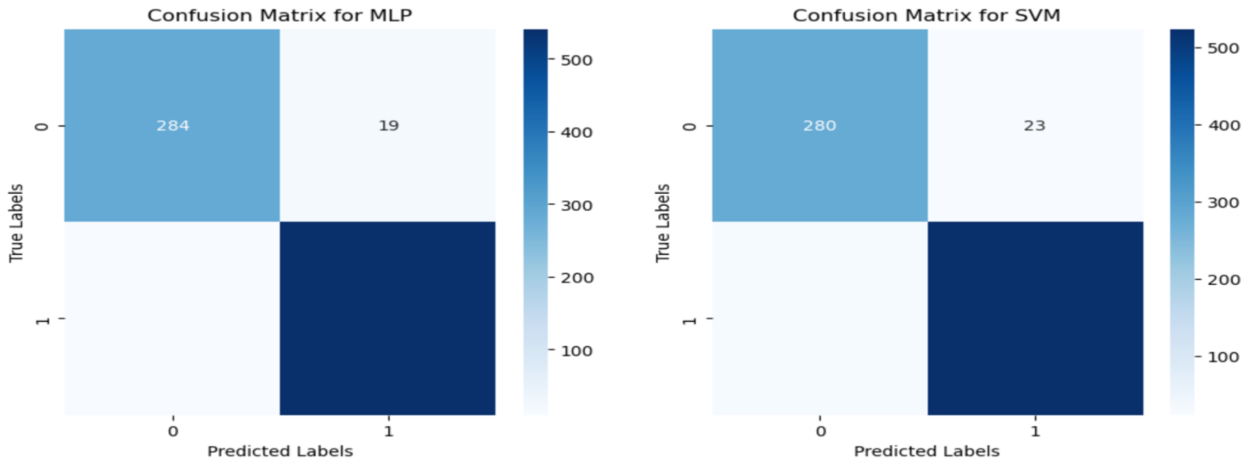

Fig 5 Confusion Metrices

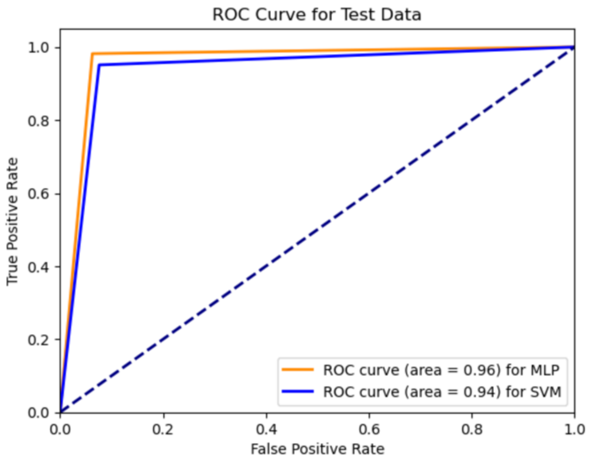

Fig 6 ROC Curves Test Data

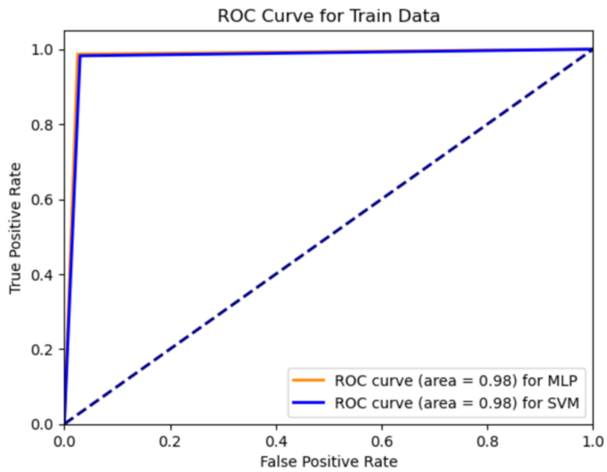

Fig 7 ROC Curve Train Data

The confusion matrices (Fig 5) and Receiver Operating Characteristic (ROC) (Fig 6 & Fig 7) curves for the MLP and SVM provide a comparative insight into their performance. For the MLP, the confusion matrix indicates an impressive specificity with 284 true negatives and a sensitivity with no false negatives for the positive class, whereas the SVM shows slightly less specificity with 280 true negatives and a marginally higher rate of false positives at 23. Neither model registers any false negatives, demonstrating high sensitivity. In terms of ROC curves, the test data shows the MLP with an AUC of 0.96, slightly outperforming the SVM's AUC of 0.94, suggesting the MLP is more effective at classifying the positive class across various thresholds. The ROC curves for the training data reflect identical AUC values of 0.98 for both models, indicating extremely high performance on the data they were trained on. However, the slightly higher test AUC for the MLP implies it generalizes better to unseen data compared to the SVM.

4. Conclusion

In this paper, our comparative analysis of the Multilayer Perceptron (MLP) and Support Vector Machine (SVM) models revealed that the MLP holds a significant advantage in performance metrics. Supervised machine learning techniques for classification automate customer evaluation process in a faster and efficient manner [4]. Financial institutions and candidates alike are interested in knowing the rationale behind a denial. Moving forward, it would be prudent to focus on accruing a more extensive dataset to mitigate the costs associated with misclassification a step that is particularly important given the critical nature of accurate predictions in this domain. In addition to accuracy, it is imperative to incorporate various other performance metrics when evaluating model efficacy, including the confusion matrix, training error, testing time, and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve.

For the future, we intend to evaluate a deep learning neural network performance for credit scoring modeling [5]. Additionally, experimenting with a broad spectrum of hyperparameter tuning strategies could substantially refine model performance, an approach that seems especially promising for algorithms like MLP and SVM and potentially other ensemble methods. Such future efforts are poised to not only bolster the MLP's already impressive results but also to provide a broader comparative framework that includes enhanced versions of other robust classifiers.

5. References

[1]. V. -E. Neagoe, A. -D. Ciotec and G. -S. Cucu, "Deep Convolutional Neural Networks Versus Multilayer Perceptron for Financial Prediction," 2018 International Conference on Communications (COMM), Bucharest, Romania, 2018, pp. 201-206, doi: 10.1109/ICComm.2018.8484751. keywords: {Neurons;Training;Machine learning;Convolutional neural networks;Multilayer perceptrons;Training data;Data models;financial prediction;credit scoring;Multilayer Perceptron (MLP);Deep Learning (DL);Convolutional Neural Networks (CNN)},

[2]. P. Tumuluru, L. R. Burra, M. Loukya, S. Bhavana, H. M. H. CSaiBaba and N. Sunanda, "Comparative Analysis of Customer Loan Approval Prediction using Machine Learning Algorithms," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2022, pp. 349-353, doi: 10.1109/ICAIS53314.2022.9742800. keywords: {Support vector machines;Measurement;Machine learning algorithms;Profitability;Forestry;Prediction algorithms;Risk management;Random Forest;Machine Learning;K-Nearest Neighbor;Support Vector Machine;Logistic Regression},

[3]. Y. Abakarim, M. Lahby and A. Attioui, "Towards An Efficient Real-time Approach To Loan Credit Approval Using Deep Learning," 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC), Rabat, Morocco, 2018, pp. 306-313, doi: 10.1109/ISIVC.2018.8709173. keywords: {Biological neural networks;Real-time systems;Deep learning;Predictive models;Logistics;Support vector machines;Machine learning;Deep Learning;Loan approval;Real-time data},

[4]. Tetteh, M., Puthussery, A., Yamunakrishnan, Subbulakshmi, S. (2024). Credit Risk Assessment with Madaline and Multilayer Perceptrons. In: Reddy, V.S., Wang, J., Reddy, K. (eds) Soft Computing and Signal Processing. ICSCSP 2023. Lecture Notes in Networks and Systems, vol 864. Springer, Singapore. https://doi.org/10.1007/978-981-99-8628-6_47

[5]. Damrongsakmethee, T., Neagoe, VE. (2019). C4.5 Decision Tree Enhanced with AdaBoost Versus Multilayer Perceptron for Credit Scoring Modeling. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds) Computational Statistics and Mathematical Modeling Methods in Intelligent Systems. CoMeSySo 2019 2019. Advances in Intelligent Systems and Computing, vol 1047. Springer, Cham. https://doi.org/10.1007/978-3-030-31362-3_22

INTERMEDIATE RESULTS

In assessing the efficacy of machine learning models such as Support Vector Machines (SVM) and Multilayer Perceptrons (MLP), it is observed that the SVM achieved an accuracy of 92%, while the MLP attained 96%. These metrics reflect the proportion of predictions each model made correctly in relation to the actual labels within the test dataset.

When tuning a Multilayer Perceptron (MLP) using 5-fold cross-validation, a noted decrease in accuracy suggests several potential factors. Cross-validation, by dividing the dataset into multiple subsets, can highlight overfitting that single-split evaluations may overlook, revealing a model's performance across a more varied data set and possibly explaining the reduced accuracy. Additionally, the decrease might reflect inadequacies in the hyperparameter settings which, while seemingly optimal in isolated tests, fail to generalize across the diverse subsets presented during cross-validation. This issue could be exacerbated by a dataset that is either too small or noisy, where the model struggles to learn generalizable features effectively.

Furthermore, discrepancies in class distributions across folds, particularly in imbalanced datasets, can also lead to variations in performance. These insights necessitate a reevaluation of the model's configuration and the consideration of more sophisticated data preprocessing or techniques to manage class imbalance.

GLOSSARY:

1. Supervised Learning: This approach involves training models on labeled data, where the desired output labels are known, allowing the model to learn a mapping from inputs to outputs. It is applicable to tasks such as regression and classification.

2. Unsupervised Learning: Unlike supervised learning, unsupervised learning involves training models on data without labeled outcomes. Here, the goal is to identify underlying patterns or distributions in the data, useful in clustering and association tasks.

3. Backpropagation: A fundamental method in neural networks, backpropagation is used for training by iteratively adjusting the weights of the network, minimizing the error between the predicted and actual outputs by propagating the error backward through the network.

4. Recall: Recall, another name for sensitivity, is a metric used to assess a model's accuracy in identifying all pertinent instances in a dataset. The ratio of true positives to the total of false negatives and true positives is how it is defined.

5. Precision: As the ratio of true positives to the total number of instances predicted as positive (sum of true positives and false positives), precision quantifies the accuracy of the model's positive predictions.

6. Adam Solver: The 'adam' solver is an optimisation technique that is mostly used to train deep learning models. It is renowned for its ability to handle sparse gradients and adapt learning rates.

7. GridSearchCV: A technique used to optimize model parameters through an exhaustive search over a specified parameter grid. This method is implemented using cross-validation to determine the configuration that achieves the best performance metrics.

8. Hyperparameters: These configuration parameters govern how machine learning models are built and how they learn. Hyperparameters, in contrast to model parameters, need to be established before training because they cannot be learned from the data.

9. ReLU (Rectified Linear Unit): A widely used activation function in neural networks that is very good at adding non-linearity to the model without changing the convolution layer's receptive fields.

10. C (Regularization Parameter): A hyperparameter that controls the amount of regularization in models such as Logistic Regression and Support Vector Machines, helping to prevent overfitting by penalizing larger values of model parameters.

11. Kernel: For SVM to work with complicated datasets, a kernel function converts the input data into a higher-dimensional space so that classes can be separated linearly.

12. Confusion Matrix: A table that shows how well a classification model performs on a set of test data that has known true values. It makes it simple to visualise how well an algorithm performs.

13. ROC (Receiver Operating Characteristic) Curve: A visual representation of the true positive rate versus the false positive rate that shows how a binary classifier system can be diagnostically assessed as its discrimination threshold is changed.

14. AUC (Area Under the Curve): A measurement of the model's degree of separability obtained from the ROC curve. Better model performance in differentiating between the classes is indicated by a higher AUC.

15. Robust Classifiers: Classifiers that maintain high levels of accuracy and consistency under varying conditions in the input data, including noise, missing values, and data anomalies.