Abstract

This project investigates the application of genetic algorithms for predicting credit scores, categorizing individuals into "Good," "Poor," and "Standard" credit brackets/categories based on their financial history, salary, and occupation. Unlike traditional methods, our approach leverages genetic algorithms combined with decision tree classifiers to enhance predictive accuracy and robustness. Three models were developed, each employing distinct methodologies to optimize the prediction process and guide the development of a more complex model with higher accuracy. Our best model was able to achieve a remarkable 0.99 F1 score with very minimal miss classifications for all credit brackets which ensures the reliability of generic algorithms in handling the financial institutions' needs when filtering products and assisting their clients.

Table of Contents

1.0 Introduction & Motivation

Our project focuses on predicting the credit score of individuals who are seeking to gain access to specific financial products based on their credit and borrowing history, salary, and occupation.

The prediction process categorizes individuals into 3 credit score categories: “Good”, “Poor” and “Standard” to alleviate the stress and loss of accuracy associated with predicting exact numerical values.

Predicting customers’ credit scores by first gathering all their financial history can show directly what products they are eligible for with setting up appointments with an advisor or the need to go through the convoluted process of submitting a formal credit score checkup to national agencies and without the need to process any highly sensitive personal data such as SIN, home address, phone number(s), etc.

Our model can be integrated with online portals of financial institutions to assess customers rapidly by providing basic questionnaires regarding the income and current debt the client has which significantly reduces the time associated with financial product registration.

2.0 Related Works

Most other approaches similar to our project revolved around using standard classifiers such as KNN, XGB Boost, Naive Bayes, Random Forests, and Decision Trees; furthermore, these approaches revolved around only direct training and testing the model with the same dataset we are utilizing. The accuracy achieved by prior projects ranged from 68% to 81%.

Other approaches revolved around utilizing SMOTE to increase the size of the data set and maximize training efficiency which significantly boosted the range between 71% and 92%.

None of the prior works utilized genetic algorithms or data manipulation methods that mutate the data in any shape or form.

3.0 Methodology

3.1 Data Handling

The dataset [1], used for training and testing, consists of 100 thousand examples of people with different financial standing across different years and months where 80 thousand samples were used for training and 20 thousand samples were used for testing. Both training and testing data frames were picked randomly using the train_test_split() method.

Data cells with null/Nan values in each used numerical attribute are replaced with the median of all the values for that attribute; furthermore, individuals with ages larger than 100 were treated as outliers where their age values were replaced by the median age too.

String attributes were changed to numbers using label encoding while adding a random number between 1 and 100 to each generated number and then shuffling the values to ensure maximum randomization and minimize bias.

Normalization of continuous attributes was also used to minimize the impact of outliers.

3.2 Solutions

We built 3 models utilizing genetic algorithms and harnessing simple, standard, and complex methods to compare the accuracy of predictions effectively. 100 trails were used for all models.

3.2.1 Version-1.0

The first mode consists of:

a. Generating weights randomly for each attribute taken for analysis then each weight is multiplied by the attribute to be summed up, resulting in the fitness function final score.

b. The top 80% of individuals with the lowest scores are considered before performing crossovers and mutations.

c. Crossover is done using a randomized index within the array size of the attributes selected for analysis; the data is switched before and after the index for each randomly selected 2 individuals.

d. Mutation is performed by selecting a random index within the array of chosen attributes. A new, different value is then chosen from the attribute at that index to replace the old value.

e. The mutation rate is set at 0.01, and it is activated by generating a random number between 0 and 1 and verifying if it falls below this rate.

f. After performing the crossover and mutation, the parents' data are updated with the newly generated children, and the weights are randomized again.

g. The weights with the lowest median scores are picked as best.

h. The best weights are multiplied by the same attributes' values in the testing data and summed up, the generated score is then compared with the predicted score for each credit score category.

i. If the score lies within the ranges between the average predicted scores of 2 different categories, the lowest score category is picked.

j. Micro F1 score and confusion matrix was used to assess the accuracy of the prediction to minimize the imbalanced classes in the dataset and gain better insights on the false positives/negatives respectively.

3.2.2 Version-2.0

The second model changes consist of:

a. The fitness function consists of multiplying the randomly generated weights by the selected attributes. Then the reciprocal median of the squared differences between the predicted and expected data is taken to generate more consistent scores and reduce the outlier effect.

b. Weights are adjusted after each iteration based on the generated Micro F1 score. If the current score is higher then the previous weights are compared to the last generated weights where the smaller weights on the last executed trial are increased by the median of these weights; furthermore, larger weights are decreased by the median of these weights too while ensuring no value below zero is achieved. If the current score is lower, weights are generated randomly again.

c. The best parent values choice is reversed to take the 80% best highest score values.

d. The choice of best weight values is reversed to take the ones with the highest score.

e. The fitness function now takes both test_x and test_y parameters. Test_x is the attributes and test_y is the credit scores. This ensures that the fitness score reflects the relationship between predictions and actual values.

f. The new fitness function is more robust to outliers, leading to more stable and reliable performance measures.

g. Mutation is now done on two elements instead of one to increase genetic diversity within the population.

h. The new mutation strategy reduces the risk of premature convergence and improves the exploration of the solution space.

3.2.3 Version-3.0

The third model changes consist of:

a. The fitness function consists of a decision tree classifier with a max depth of 10 where it is trained on the 80000 samples and predictions are generated on the same samples labels then the score is calculated using the absolute difference between the prediction and expected values divided by the expected value for all the examples in the dataset.

b. The best 100% scored examples sorted in descending order are taken for crossover and mutation to maximize learning.

c. The continuous attributes were normalized to reduce the effects of outliers on accuracy

d. The accuracy is calculated by averaging the score for the category in the credit scores then a decision tree classifier is set with a depth of 10 and trained on the training data and predictions are generated on the testing data. The score is the absolute difference between the predicted data and the actual testing data is taken and then divided by the actual data.

e. The scores generated are then divided by the average value for each credit score category generated from the training data and checked if they are at least 60% similar. If so, the final prediction for that example in the testing data is then the category in the credit score attribute.

4.0 Results & Discussion

4.1 Version-1.0 Results

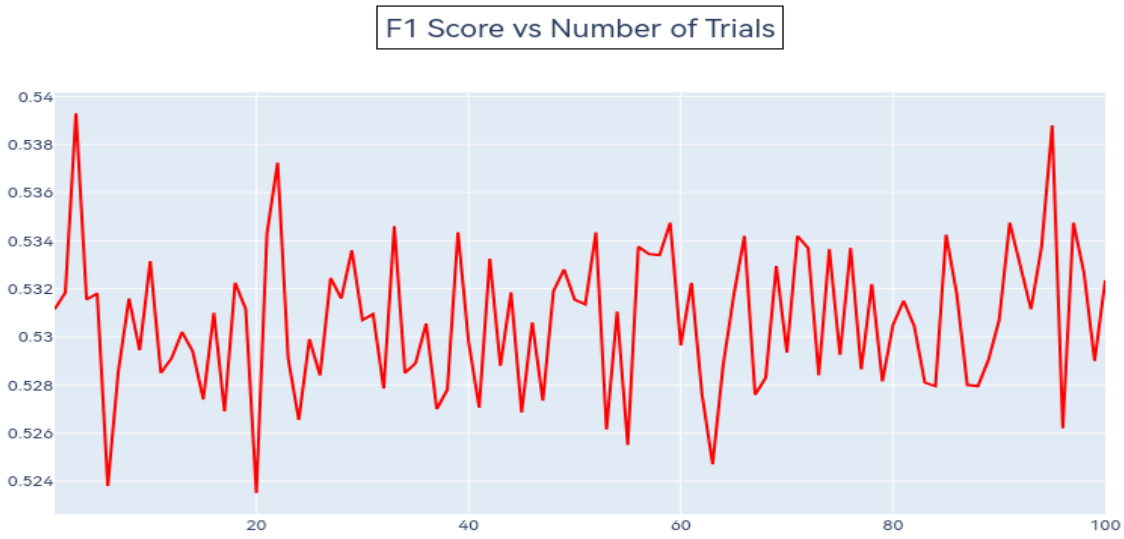

Fig 1: Micro F1 Scores with Respect to the Number of Trials

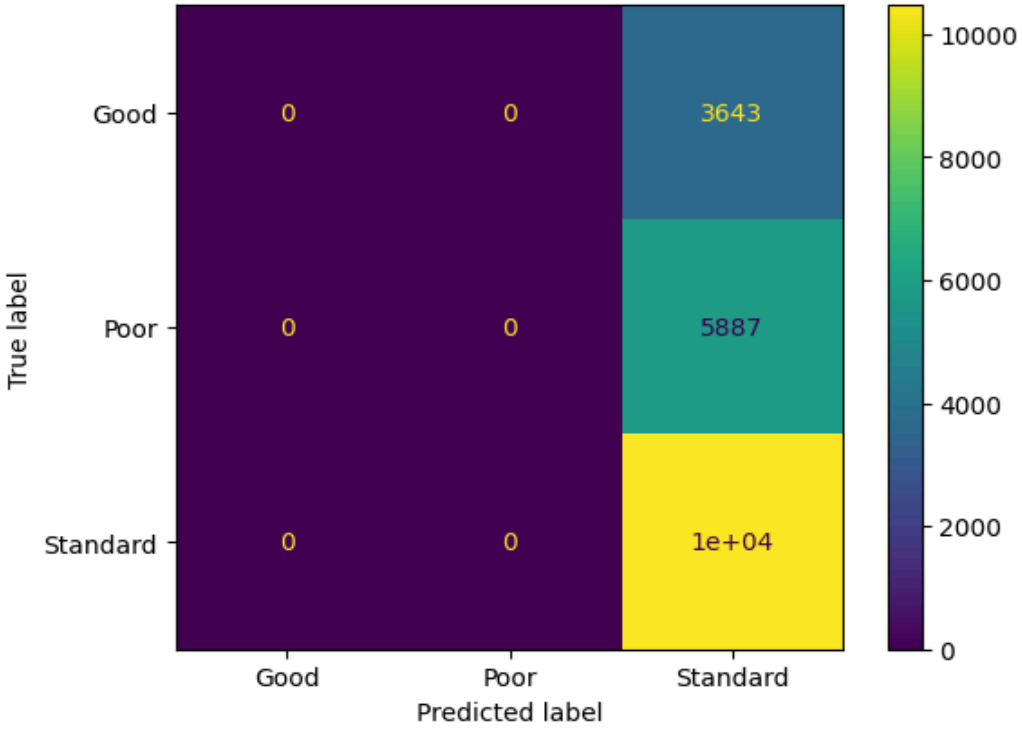

Fig 2: Confusion Matrix Results

4.2 Version-2.0 Results

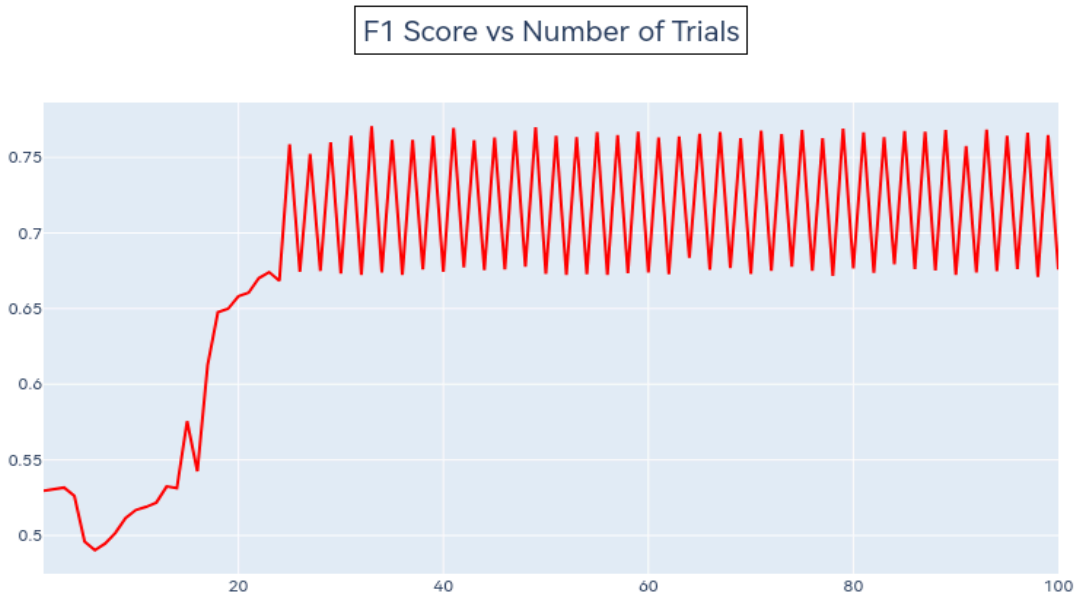

Fig 3: Micro F1 Scores with Respect to the Number of Trials

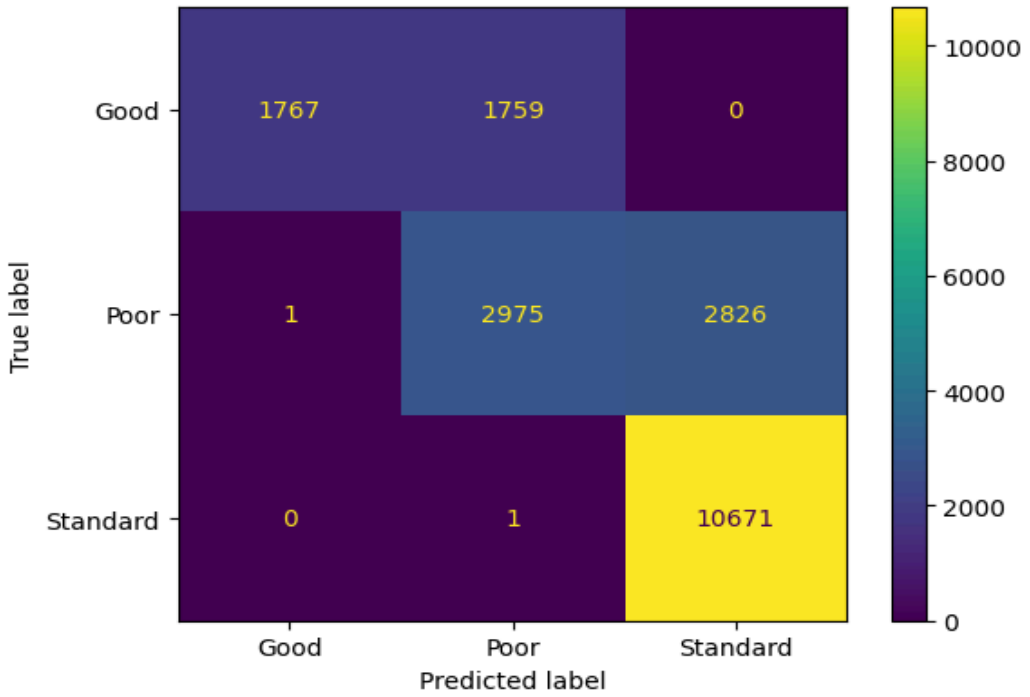

Fig 4: Confusion Matrix Results

4.3 Version-3.0 Results

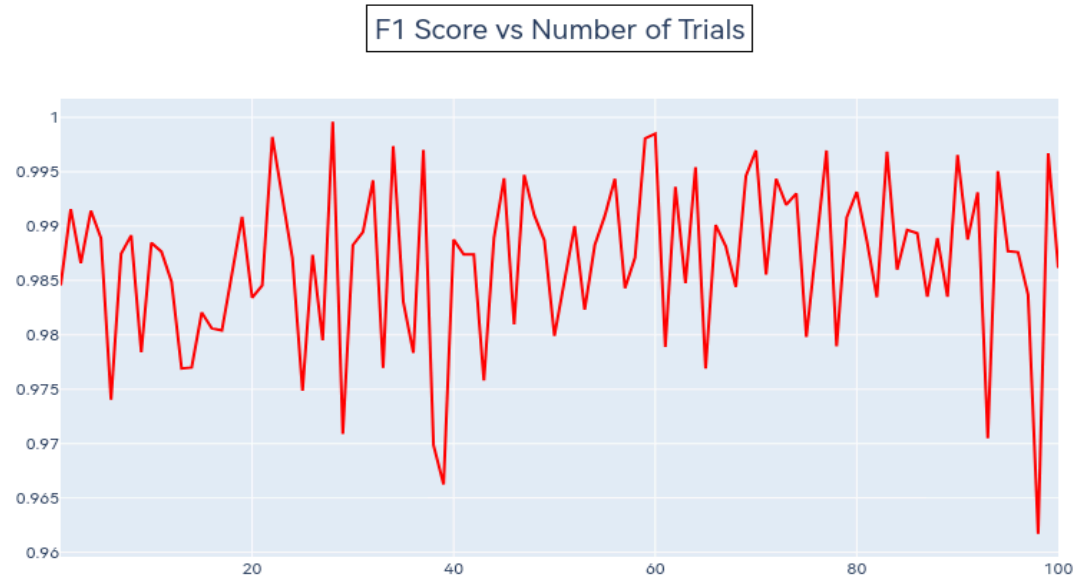

Fig 5: Micro F1 Scores with Respect to the Number of Trials

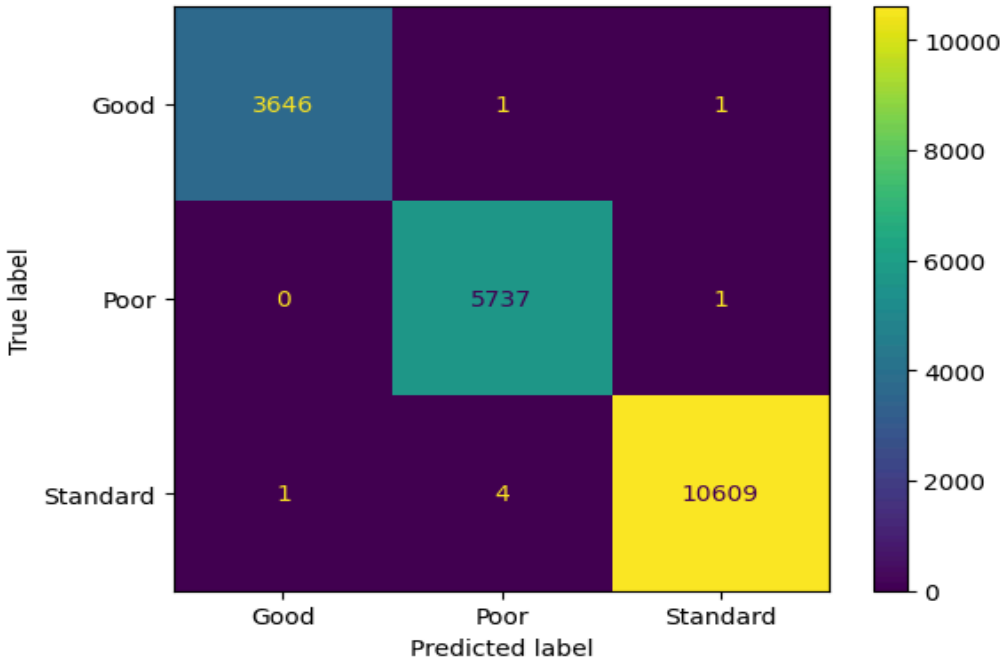

Fig 6: Confusion Matrix Results

4.4 Summary of Findings

The project utilized three distinct models based on genetic algorithms to predict credit scores, each incorporating unique approaches and complexities to enhance accuracy and robustness. Micro F1 scores and confusion matrices were used to calculate the accuracy of predictions.

The first model achieved an F1 score ranging between 0.50 and 0.54 which is considered a balanced prediction; however, the accuracy is not trustable enough outside the most repeated label in the target class which is the “Standard” credit score category where the “Good” and “Poor” labels are only predicted as “Standard”.

The second model improved upon the first by enhancing the fitness function to consider iterative weight adjustment plus randomization, achieving an F1 score ranging between 0.48 and 0.77. The confusion matrix analysis for this model showed a significant improvement in handling miss classifications of labels without maximal frequency where the “Good” and “Poor” credit score labels are predicted correctly 50% of the time and the “Standard” highest frequency label is getting predicted almost 100% of the time. This shows significant improvement to the first model and with above-average accuracy.

The third model utilized a decision tree classifier-based fitness function with a maximum depth of 10, focusing on maximizing learning from the best-scored examples. This model demonstrated superior performance with an F1 score ranging between 0.96-0.99, reducing prediction errors and providing reliable credit score categorizations. The confusion matrix shows that “Good”, “Poor” and “Standard” are predicted at almost 100% accuracy without bias toward one label or the other.

Overall, the third model exhibited the best performance among the three, significantly reducing prediction errors and providing reliable credit score categorizations for real-world use. The project's innovative use of genetic algorithms combined with decision tree classifiers proved to be more effective than traditional methods, enhancing predictive accuracy and robustness. These findings underscore the importance of advanced algorithmic approaches in financial predictive analytics and suggest practical applications for integrating these models into financial institutions' online portals to streamline credit evaluations.

5.0 Conclusion

The project successfully demonstrated the potential of genetic algorithms in predicting credit scores by efficiently categorizing individuals into credit score brackets (“Good”, “Poor”, and “Standard”). The three models developed showcased varying degrees of accuracy and robustness, with the third model proving the most effective. Key outcomes include enhanced prediction accuracy through the use of genetic algorithms, particularly with decision tree classifiers, which significantly improved credit score predictions. The models managed imbalanced datasets effectively, reducing the impact of outliers and improving the reliability of predictions. Furthermore, the developed models can be integrated into financial institutions' online portals to streamline credit evaluations, reducing the need for extensive and sensitive data processing. The findings underscore the importance of continuous refinement in algorithmic approaches to enhance predictive capabilities.

6.0 Future Works

- Employing PCA to reduce data size to only the most attributes with the crucial needed information from the original dataset to boost processing time.
- Employing further techniques to remove or replace corrupt data and outliers to further increase the accuracy of predictions and performance.
- Utilizing different classifiers as fitness functions such as Random Forests, Logistic Regression, and SVC.
- Harnessing the power of SMOTE to increase the size of the dataset to increase the efficiency of the learning process by the algorithm.

7.0 References

[1] R. Paris, “Credit score classification,” *www.kaggle.com*, 2022. https://www.kaggle.com/datasets/parisrohan/credit-score-classification

8.0 Appendix

Below is a link for our source code:

Future Forecasters, “ECE 470_Project”, *https://colab.research.google.com,* 2024. https://colab.research.google.com/drive/1C_90YqNFxd2sMLGPBO9M6zPS-FtZMRQl