Loan Approval Prediction using Adversarial Training and Data Science

Dharavath Sai Kiran[1], Avula Dheeraj Reddy[2], Suneetha Vazarla[3], Dileep P[4]

[1,2]B.Tech Student, Department of CSE (Data Science), Malla Reddy College of Engineering and Technology, Hyderabad, India.
[3]Research Scholar, Malla Reddy University, Hyderabad, India.
[4]Professor, Department of CSE (Data Science), Malla Reddy College of Engineering and Technology, Hyderabad, India.

Abstract **-Loan approval is critical decision-making in the financial sector, impacting financial stability and reputation. In recent times, many machine learning models have been introduced. However, these models may be biased towards certain groups of borrowers, resulting in unfair loan approval decisions. So, the financial industry requires a fair and accurate prediction model. This paper proposes a model for loan approval prediction that combines Adversarial Training and Data Science techniques. We develop a model by training with a real-time data set, and testing that shows our model achieves better accuracy and fairness than existing models. Our model demonstrates the potential of Adversarial Training and Data Science for improving the Loan Approval Prediction process. This paper contributes to growing research on Adversarial Training and Data Science techniques in the financial sector.**

Keywords – Loan approval, Adversarial training, Data Science, Fairness, Accuracy, Real-time dataset.

I. INTRODUCTION

The loan approval model is trained on historical data, including features like income, credit score, employment status, etc. If the historical data contains biases against certain groups, the model may learn to replicate those biases and make unfair decisions. However, suppose the model is trained using adversarial training. In that case, it will be forced to consider a broader range of features and learn to be more robust against biased attempts to manipulate the data.

This paper investigates using adversarial training and Data Science for loan approval prediction. We first provide an overview of the field's current state and the challenges associated with developing fair and unbiased loan approval models. We then describe the methodology for implementing Adversarial Training and Data Science in the loan approval context, including selecting appropriate features and constructing the adversarial attacks.

Loan approval is a critical process in the finance sector that determines whether individuals can access financial services. However, traditional loan approval models have been criticized for their lack of accuracy and fairness, which can lead to biased decisions and exclude certain groups of people from financial services. In recent years, Adversarial training and Data Science techniques have emerged as promising approaches for improving the accuracy and fairness of loan approval models.

Adversarial training is a machine learning technique that aims to improve the robustness of models by exposing them to malicious attacks or perturbations. An Adversarial attack is when an attacker deliberately manipulates input data to deceive the model and cause it to make incorrect predictions.

Adversarial training can improve the fairness and accuracy of the model by identifying and mitigating bias in the training data. Mitigating bias is particularly important in loan approval prediction, where bias can result in unfair lending practices.

The Adversarial training process involves training the model on a combination of regular and examples. In this case, the adversarial examples are generated by adding perturbations to the input data designed to trick the model into making incorrect predictions. Training the model on both regular and negative standards, it learns to identify and correct for biases in the data, resulting in a more robust and fairer model.

On the other hand, Data science involves using statistical techniques to extract insights from data and make informed decisions.

By combining these approaches, it may be possible to develop a model with more accurate and fair loan approval models that can mitigate bias and promote financial inclusivity.

This paper proposes a novel framework for Loan Approval Prediction using Adversarial Training and Data Science. We evaluate the effectiveness of our approach on a real-time data set and compare it to existing models. Our results demonstrate that our model may achieve higher accuracy and Data Science for improving loan approval prediction and promoting financial inclusivity. This paper contributes to the growing body of research on Adversarial Training and Data Science techniques in the economic sectors and highlights the need for further investigation.

Some of the currently used approaches for loan approval predictions are:

- Logistic Regression
- Decision Trees
- Ensemble Learning Techniques
- Gradient Boosting

II. LITERATURE REVIEW

According to Authors Prof. Sathish Jayanth Manje, Mr. Dhiraj Majne, Mr. Rahul Pandurang Bhere, and Mr. Rahul Kailas Pawade, the approach proposed for Loan Approval Prediction using Logistic Regression, Decision Tree, and Random Forest techniques [4]. But, in these approaches, only the Logistic Regression approach gives accuracy with default bias.
According to the Author, Shubham Nalwade, Suraj Andhe, Siddhesh Parab, and Prof.Amruths Sankhe, the approach proposed for Loan Approval Prediction using Naïve Bayes, Decision tree, Random Forest, K Nearest Neighbor techniques [5]. But the process gave the result with the same bias and less accuracy.
On the other hand, Adversarial training is a technique that aims to improve the robustness of models by exposing them to adversarial examples or perturbations. Approval models are often trained on historical data, which can contain bias. Adversarial training can help mitigate this bias's effects by exposing the model to those perturbations. The model can learn to be more robust to bias and produce more accurate predictions.

III.METHODOLOGY

The technique this paper approaches is Adversarial Training and Data Science.

Data Collection:

First, we obtained our Loan Data Set from Kaggle.
Source link of data set:
https://www.kaggle.com/datasets/burak3ergun/loan-data-set

Which included features loan id, gender, married, dependents, education, self-employed, applicant income, loan amount, loan amount term, credit history, property area, loan status, and total income.

Data Preprocessing:

We perform data preprocessing steps to eliminate null values and missing data. And we take inputs as
x: The input features, in the form of pandas Data Frame
y: The target variable, in the form of a pandas Series
In Data preprocessing, we print the data frame, list of columns in the data set, summary statistics, shape, unique values, mean, Median, mode, and correlation matrix of the data collection, and find any missing deals and fill null values with an average of the concern column.

Adversarial Training:

Apply Adversarial training to the model by generating adversarial examples that specifically target and challenge the model's learned biases.

Model Training:

We split the data set for training and testing. We train the model using a training data set and test the trained model using test data set to evaluate the model's performance. We consider the version using metrics such as accuracy, precision, recall and f1 score, and we draw a confusion matrix table.

Prediction:

Making predictions using the test data set.

Model Evaluation:

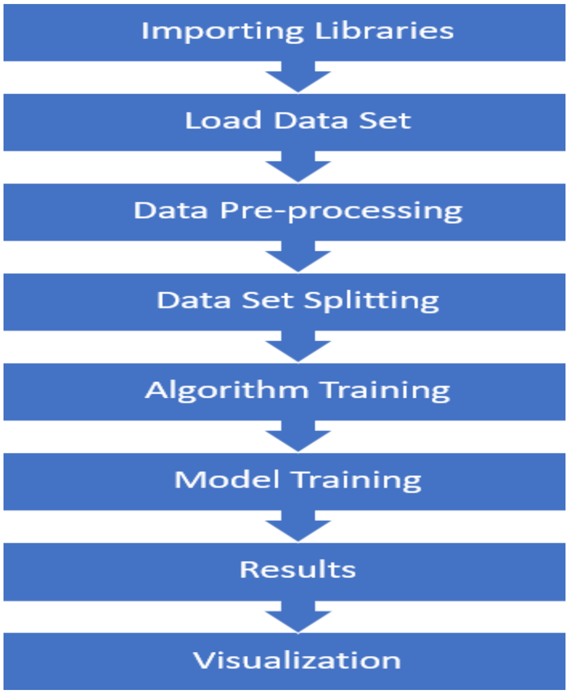
Evaluate the model's performance using accuracy, precision, recall, f1 score and confusion matrix. The confusion matrix table shows model performance by comparing actual values of the data with predicted values. It gives a more detailed picture of how well the model is performing by showing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).
True positives (TP) are the cases where the model correctly predicted a positive outcome (the loan was approved) when the work was positive.
True negatives (TN) are the cases where the model correctly predicted a negative outcome (the loan was rejected) when the actual work was negative.
False positives (FP) are the cases where the model predicted a positive outcome (the loan was approved) when the actual work was negative (the loan was rejected).
False negatives (FN) are the cases where the model predicted a negative outcome (the loan was rejected) when the actual work was positive (the loan was approved).

Deployment:

Deploy the final model on a real-world system.

Algorithm

Step 1: Start by Importing Libraries
Step 2: Load the data set
Step 3: Print the Data frame
Step 4: Print the list of columns
Step 5: Check for missing, null values in the Data frame and fill in missing values with a mean of each column.
Step 6: Print the summary statistics of the Data frame
Step 7: Print the shape of the Data frame
Step 8: Print unique values
Step 9: Print the mean, median, and mode of columns in the Data frame
Step 10: Calculate the correlation of the Data frame
Step 11: Create a heatmap of the correlation matrix
Step 12: Create a box plot
Step 13: Split the Data frame into the feature matrix x and the target vector y
Step 14: map categorical values into binary labels
Step 15: Split the data into training and testing sets
Step 16: Define a function to create Adversarial examples and train the model on them
Step 17: Use the test data set for prediction
Step 18: Plot the Confusion matrix
Step 19: Calculate accuracy, precision, recall, and f1-score
Step 20: Print the calculated accuracy, precision, recall, and f1-score




```
list(df.columns)
```

```
['0',
 'Loan_ID',
 'Gender',
 'Married',
 'Dependents',
 'Education',
 'Self_Employed',
 'ApplicantIncome',
 'CoapplicantIncome',
 'LoanAmount',
 'Loan_Amount_Term',
 'Credit_History',
 'Property_Area',
 'Loan_Status',
 ' Total_Income ']
```

Fig 1: Flow chart of Methodology

Fig 2: List of columns in Data Frame

IV. IMPLEMENTATION

The loan approval prediction model using Adversarial Training and Data Science can be implemented with the cooperation of financial service providers like banks. But it is challenging to implement in real-life scenarios as banks or financial service providers are unwilling to share the information of the users or customers due to the competition in the market and legal reasons. Banks or financial service providers have to protect the customer's information.

Furthermore, implementing a machine learning model in a real-life scenario requires careful consideration of various ethical, legal, and regulatory issues, including privacy and transparency.

Overall, the loan approval prediction model using adversarial training and Data Science has the potential to significantly improve the loan approval process prediction.

And make it more efficient and accurate in real-life scenarios, but it requires careful and responsible deployment. This approach is implemented in Python.

```
df
```

	0	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term
0	1	LP001002	Male	No	0	Graduate	No	5849	0.0	128	360
1	2	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	66	360
2	3	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	120	360
3	4	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	141	360
4	5	LP001008	Male	No	0	Graduate	No	6000	0.0	267	360
...	...	...	...	...	...	...	...	...	...	...	...
454	455	LP002453	Male	Yes	3+	Graduate	No	7085	0.0	113	360
455	456	LP002455	Male	No	0	Graduate	No	3859	0.0	100	360
456	457	LP002459	Female	Yes	2	Graduate	No	4301	0.0	93	360
457	458	LP002467	Male	Yes	0	Graduate	No	3708	2569.0	162	180
458	459	LP002472	Male	Yes	1	Graduate	No	4354	0.0	150	360

Fig 3: Data Frame

```
sns.boxplot(x=df['ApplicantIncome'])
```

<AxesSubplot:xlabel='ApplicantIncome'>

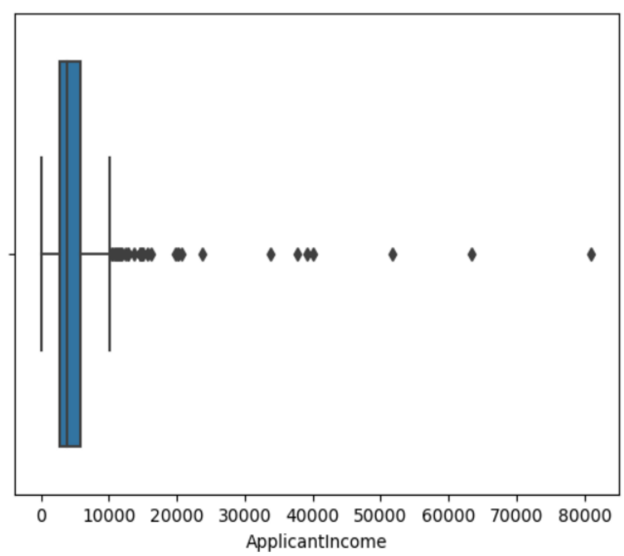


Fig 4: Box plot of Applicant Income

V. RESULTS

When we run the algorithm, the results are accuracy, precision, recall, and f1 score. And it plots the confusion matrix table.

```
cm = confusion_matrix(y_test, y_pred_binary)
```

```
sns.heatmap(cm, annot=True, cmap='Blues', fmt='g')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

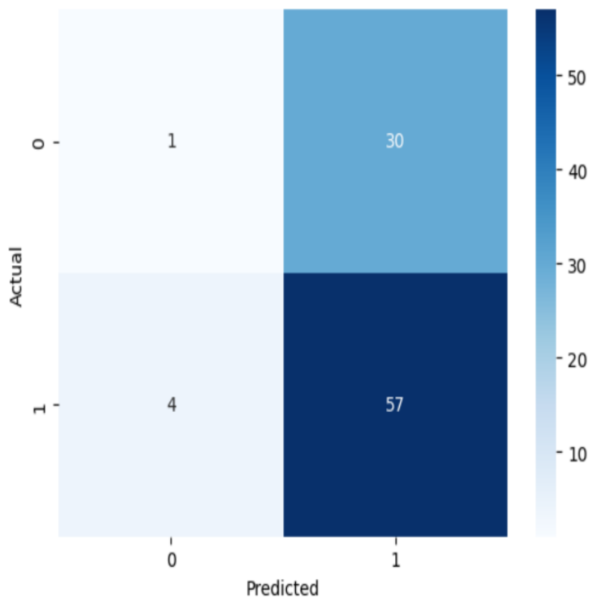


Fig 5: Confusion Matrix table

Accuracy measures the proportion of correct predictions made by the model out of all the predictions made. It is calculated as the number of accurate predictions divided by the total number of predictions

Accuracy = (TP + TN) / (TP + TN + FP +FN)

Precision measures the proportion of true positives (correctly predicted positive samples) out of all the samples predicted as positive by the model. It is calculated as the number of true positives divided by the sum of true and false positives.

Precision = TP / (TP + FP)

Recall measures the proportion of true positives (correctly predicted positive samples) out of all the positive samples. It is calculated as the number of true positives divided by the sum of true and false negatives.

Recall = TP / (TP + FN)

F1-score is the harmonic mean of precision and recall and provides a single metric that balances the tradeoff between precision and recall.

F1-score = 2 * (precision * recall) / (precision + recall)

The confusion matrix table shows the model's performance by comparing the actual with predicted class labels. It gives a detailed picture of how well the model is performed by showing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FP).

The results are obtained here when the 0.2 sizes of the test data set is used. If we increase the data set size, the model will be well-trained and may give more accuracy compared to this.

The confusion matrix table shows model performance by comparing actual values of the data with predicted values. It gives a more detailed picture of how well the model is performing by showing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

True positives (TP) are the cases where the model correctly predicted a positive outcome (the loan was approved) when the work was positive.

True negatives (TN) are the cases where the model correctly predicted a negative outcome (the loan was rejected) when the actual work was negative.

False positives (FP) are the cases where the model predicted a positive outcome (the loan was approved) when the actual work was negative (the loan was rejected).

False negatives (FN) are the cases where the model predicted a negative outcome (the loan was rejected) when the actual work was positive (the loan was approved).

By analyzing the confusion matrix, we can calculate evaluation metrics such as accuracy, precision, recall, and F1 score, which provides a more comprehensive understanding of the model performance.

```
accuracy = accuracy_score(y_test, y_pred_binary)
precision = precision_score(y_test, y_pred_binary)
recall = recall_score(y_test, y_pred_binary)
f1 = f1_score(y_test, y_pred_binary)
```

```
print('Accuracy:', accuracy)
print('Precision:', precision)
print('Recall:', recall)
print('F1-score:', f1)
```

```
Accuracy: 0.6521739130434783
Precision: 0.6593406593406593
Recall: 0.9836065573770492
F1-score: 0.7894736842105262
```

Fig 6: Evaluation Metrics

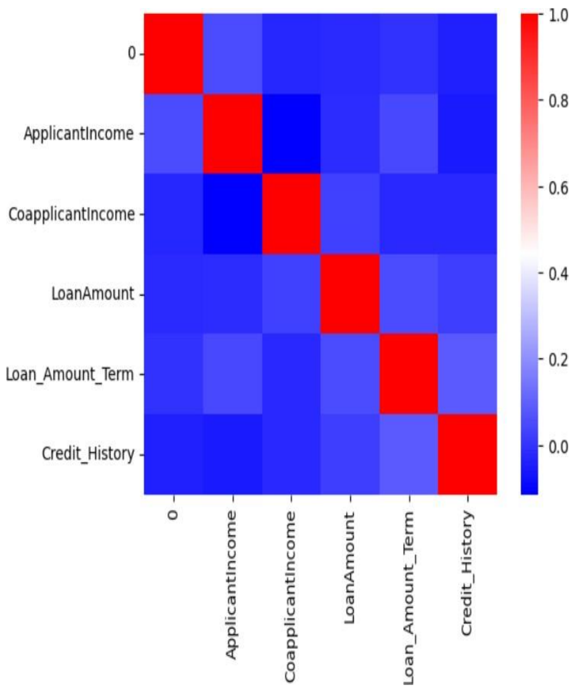


Fig 7: Plotting of Correlation

VI. CONCLUSION

This paper proposes an approach to predict loan approval using Adversarial Training and Data Science techniques. We described the methodology, which collects and preprocesses the data, creates the model and uses adversarial training to improve the model's performance. We implemented the model in Python and evaluated its performance.

The algorithm reaches 65% accuracy. The accuracy may increase when the model is trained with vast amounts of data as input to it. We also drew a confusion matrix to understand the model's performance better.

VII. FUTURE ENHANCEMENTS

As we couldn't reach our goal of 100% accuracy in predicting loan approvals, there is room for some improvement here. We may incorporate additional data sources in the future, potentially improving the model's performance.

We may use other adversarial training techniques, such as adversarial autoencoders (AAEs) could lead to further improvements.

We may explore Deep Learning architectures for better results.

VIII. REFERENCES

[1] Ms. Kathe Ruthika Pramod, Ms. Panhale Sakshi Dattatray, Ms. Avhad Pooja Prakash, Ms. Dapse Punnam Laxman,
Mr. Ghorpade Dinesh B. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm". International Journal of Creative Research Thoughts (IJCRT), volume 9, Issue 6 June 2021.

[2] Dr.C K Gomathy, Ms. Charulatha, Mr. Aakash,
Ms. Sowjanya. "The Loan Prediction using Machine Learning". International Research Journal of Engineering and Technology (IRJET), volume 8, Issue 10 Oct 2021.

[3] Ms. Pidikiti Supriya, "Loan Prediction using Machine Learning Models," International Journal of Engineering and Techniques, Mar-Apr 2019.

[4] Prof. Sathish Jaywant Manje, Mr. Dhiraj Raghunath Manje, Mr. Rahul Pandurang Bhere, and Mr. Rahul Kailash Pawade. "Loan Prediction Model using Logistic Regression, Decision Tree, Random Forest". International Journal for Research in Engineering Application & Management (IJREAM), volume 07, Issue May 2021.

[5] Mr. Shubam Nalawade, Mr. Suraj Andhe, Mr. Siddhesh Parab, and Prof. Amruta Sanke. "Loan Approval Prediction". International Research Journal of Engineering and Technology (IRJET). Volume 09, Issue 04 April 2022.

[6] Mr. Rajiv Kumar, Vinod Jain, Prem Sagar Sharma. "Prediction of Loan Approval using Machine Learning". International Journal of Advanced Science and Technology, 2019.

[7] Ms. Pidikiti Supriya, "Loan Prediction using Machine Learning Models". International Journal of Engineering and Technology, 2019.

[8] Mr. Mohammad Ahmad Sheik, "An Approach of Loan about Machine Learning Algorithm", Proceedings of the International Conference on Electronics and Sustainable Communication System, 2020.

[9] Vijayakamal, M., & Vasumathi, D. (2022). A Novel Approach to Detect Anomalies in Business Process Event Logs Using Deep Learning Algorithm. In Soft Computing and Signal Processing: Proceedings of 3rd ICSCSP 2020, Volume 2 (pp. 363-374). Springer Singapore.

[10] "An Approach for Prediction of Loan about Machine Learning Algorithm," Mohammad Ahmad Sheik, Proceedings of the International Conference on Electronics and Sustainable Communication Systems (2020).