Assignment Problem1:Credit Risk Assessment

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

by

UGGE MAHEESH VARMA **(2303A52163)**

Under the guidance of

Dr. S VAIRACHILAI

Professor, Department of CSE.

Department of Computer Science and Artificial Intelligence

ABSTRACT

The German Credit dataset is widely used in machine learning and explainable AI for evaluating credit risk assessment models. It consists of 1,000 loan applicants, each described by 20 features that capture demographic, financial, and credit history attributes. The task is to classify applicants as having either *good credit* or *bad credit* risk. These features include variables such as account status, loan amount, loan duration, savings, employment history, installment rates, and personal information like age, marital status, and housing situation. This makes the dataset both diverse and realistic for modeling financial decision-making.

In the assignment, the dataset was preprocessed through categorical encoding and scaling to prepare it for supervised learning. A Random Forest Classifier was trained on the data, achieving an accuracy of about 77%. While the model showed strong performance for identifying good credit applicants, it had a lower recall for bad credit, which reflects real-world challenges where false negatives in risk detection can be costly. The confusion matrix revealed imbalances in predictions, highlighting the importance of explainability and interpretability when applying such models in financial systems.

To address interpretability, the Local Interpretable Model-agnostic Explanations (LIME) method was applied to the trained classifier. LIME provided local explanations for individual predictions by identifying which features contributed most to a decision. For example, variables such as account status, age, loan amount, savings, and housing strongly influenced the classification of creditworthiness. These insights demonstrate how explainable AI techniques can supplement black-box models, offering transparency in financial decisions and supporting fairer, more accountable credit risk assessment.

.

INTRODUCTION

The German Credit dataset is a benchmark dataset widely used in the fields of machine learning, data mining, and explainable artificial intelligence (XAI). It was originally created for the purpose of studying credit risk assessment, a crucial task in the financial sector where institutions must decide whether to approve or reject loan applications. The dataset contains information about 1,000 individuals applying for credit, each labeled as either a "good credit risk" or a "bad credit risk." This classification problem provides a realistic and challenging environment for evaluating predictive models, as it involves both numerical and categorical features that capture applicants' financial stability and personal circumstances.

The dataset includes 20 key attributes, such as account balance, duration of credit, payment history, purpose of the loan, savings, employment duration, installment rate, age, personal status, and housing situation. These features reflect a variety of socio-economic factors that financial institutions typically consider when making lending decisions. For example, a stable employment record and sufficient savings may indicate reliability, while high installment rates or poor credit history could signal higher risk. By analyzing these attributes, machine learning models can be trained to predict whether an applicant is likely to repay the loan or default.

One of the main challenges with the German Credit dataset is its imbalance and the subtle overlap of characteristics between good and bad credit applicants. While many features correlate strongly with credit risk, no single attribute alone can determine the outcome, making this a complex, multi-factor decision-making task. Additionally, the cost of misclassification is not equal: approving a high-risk applicant could lead to financial loss, whereas rejecting a low-risk applicant may harm customer relationships. This asymmetry underscores the importance of building accurate and interpretable models that not only perform well statistically but also provide insights into the reasoning behind their decisions.

In recent years, the German Credit dataset has become especially important for research in explainable AI (XAI). Traditional machine learning models, such as random forests or neural networks, can achieve good predictive accuracy but often operate as "black boxes." With tools like LIME and SHAP, researchers can explore which features influence credit decisions most strongly for both individual applicants and the dataset as a whole. This focus on transparency is critical in the financial domain, where accountability, fairness, and regulatory compliance are essential. Thus, the German Credit dataset continues to serve as a valuable resource for studying both predictive modeling and interpretability in risk assessment.

DATASET:

- The German Credit dataset contains 1,000 rows and 21 columns.

```
                status  duration                           credit_history  \
0          ... < 100 DM        6   critical account/other credits existing
1   0 <= ... < 200 DM         48  existing credits paid back duly till now
2  no checking account        12   critical account/other credits existing
3          ... < 100 DM       42  existing credits paid back duly till now
4          ... < 100 DM       24          delay in paying off in the past

              purpose  amount                     savings  \
0  domestic appliances    1169  unknown/no savings account
1  domestic appliances    5951                ... < 100 DM
2           retraining    2096                ... < 100 DM
3     radio/television    7882                ... < 100 DM
4            car (new)    4870                ... < 100 DM

  employment_duration  installment_rate                 personal_status_sex  \
0      ... >= 7 years                 4                       male : single
1  1 <= ... < 4 years                 2  female : divorced/separated/married
2  4 <= ... < 7 years                 2                       male : single
3  4 <= ... < 7 years                 2                       male : single
4  1 <= ... < 4 years                 3                       male : single

  other_debtors  ...                                     property age  \
0          none  ...                                  real estate  67
1          none  ...                                  real estate  22
2          none  ...                                  real estate  49
3     guarantor  ...  building society savings agreement/life insurance  45
4          none  ...                          unknown/no property  53

  other_installment_plans   housing number_credits  \
0                    none       own              2
1                    none       own              1
2                    none       own              1
3                    none  for free              1
4                    none  for free              2

                        job people_liable  telephone foreign_worker  \
0  skilled employee/official            1        yes            yes
1  skilled employee/official            1         no            yes
2       unskilled - resident            2         no            yes
3  skilled employee/official            2         no            yes
4  skilled employee/official            2         no            yes

  credit_risk
0           1
1           0
2           1
3           1
4           0

[5 rows x 21 columns]
```

FIG 1- Dataset

Columns:

1. status (Checking Account Status):

- Represents the status of the applicant's existing checking account.
- Categories indicate account balance or history of overdrafts.
- A critical feature in determining the applicant's financial health.

2. duration (Credit Duration in Months):

- The time period for which the credit is requested.
- Longer durations usually increase the risk of default.
- Helps banks assess repayment capability over time.

3. credit_history:

- Provides past repayment behavior such as "paid back duly" or "critical account."
- A key factor for evaluating credit risk.
- Strong credit history improves chances of loan approval.

4. purpose:

- Indicates the purpose of the credit (e.g., car, radio/TV, furniture, education).
- Certain purposes may carry different risk levels.
- Used by banks to assess the intention behind the loan.

5. amount:

- The total credit amount requested.
- Higher amounts may pose higher risks.
- Needs to be balanced with applicant's income and savings.

6. savings:

- Represents the applicant's savings account balance.
- Categories like "<100 DM," "500–1000 DM," etc.
- Higher savings show financial stability.

7. employment_duration:

- Duration of applicant's employment (e.g., unemployed, <1 year, >7 years).
- Longer employment suggests job stability.
- Strongly related to repayment reliability.

8. installment_rate:

- Installment rate as a percentage of disposable income.
- Higher rates indicate financial burden.
- Helps assess applicant's repayment capacity.

9. personal_status_sex:

- Marital status and gender of the applicant (e.g., male single, female divorced).
- Used historically in credit scoring, though often considered sensitive.
- May indirectly reflect social and financial conditions.

10. other_debtors:

- Indicates if the applicant has guarantors or other debtors.
- Presence of guarantors can reduce credit risk.

- Reflects shared responsibility in repayment.

11. **residence_since:**
    - Number of years the applicant has lived at current residence.
    - Longer residence duration suggests stability.
    - Short stays may indicate financial or social instability.
12. **property:**
    - Type of property owned (e.g., car, real estate, savings, unknown).
    - Owning assets lowers credit risk.
    - Valuable property can serve as collateral.
13. **age:**
    - Age of the applicant in years.
    - Younger applicants may have less credit history.
    - Older applicants may be seen as more stable but close to retirement risk.
14. **other_installment_plans:**
    - Other existing installment plans (e.g., bank, stores, none).
    - Multiple installment plans may signal financial stress.
    - Helps in evaluating overall debt obligations.
15. **housing:**
    - Housing status (e.g., rent, own, free).
    - Owning a house usually reflects stability.
    - Renting may indicate higher financial dependency.
16. **number_credits:**
    - Number of existing credits at this bank.
    - More credits may mean higher financial burden.
    - Helps in risk assessment of cumulative obligations.
17. **job:**
    - Type of job (e.g., skilled, unskilled, management).
    - Higher job categories indicate better earning potential.
    - Strong predictor of repayment ability.
18. **people_liable:**
    - Number of people financially dependent on the applicant.
    - More dependents may increase financial strain.
    - Impacts disposable income available for loan repayment.
19. **telephone:**
    - Indicates whether the applicant has a registered telephone.
    - Historically, telephone ownership was seen as a sign of stability.
    - In modern terms, less predictive but still included.
20. **foreign_worker:**
    - Specifies whether the applicant is a foreign worker.
    - Historically included as a demographic factor.

- May indirectly relate to stability and long-term residence.

21. credit_risk (Target Variable):

- The target column: good credit (1) or bad credit (0).
- Machine learning models predict this variable.
- Used as the main classification label for credit risk assessment.

PREPROCESSING STEPS (German Credit Dataset):

1. Handling Missing Values:

- The dataset was checked for missing values across all 21 columns.
- No missing values were found, hence no imputation was required.

2. Removing Duplicates:

- Duplicate records were verified in the dataset.
- No duplicate entries were found, ensuring data quality and uniqueness.

3. Outlier Detection and Removal:

- Outliers were analyzed in numerical features such as *age, amount, and duration*.
- Extreme values were reviewed, and those beyond acceptable thresholds were minimized to prevent bias in model training.

4. Encoding Categorical Variables:

- The dataset contained several categorical attributes (e.g., status, credit_history, purpose, savings, employment_duration, personal_status_sex).
- These features were transformed into numeric form using Label Encoding for compatibility with machine learning models.

5. Feature Scaling:

- Numerical features such as *amount, age, duration, and installment_rate* were standardized using StandardScaler.
- Scaling ensured that all features were on a similar range, preventing bias towards attributes with larger numeric values.

6. Splitting the Dataset:

- The data was split into training (80%) and testing (20%) sets.
- This division enabled proper model training and evaluation while avoiding overfitting.

MODEL&PARAMETERS:

1. Model Selection:

- A Random Forest Classifier was chosen as the primary model for credit risk prediction.
- Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs to improve classification accuracy and reduce overfitting.

2. Model Parameters:

- n_estimators = 100 → The model uses 100 decision trees in the forest.
- random_state = 42 → Ensures reproducibility of results.
- Default values were used for other parameters such as max_depth, min_samples_split, and min_samples_leaf.

3. Training and Testing:

- The dataset was split into 80% training and 20% testing sets.
- The model was trained on the training set and evaluated on the test set to measure generalization performance.

4. Performance Metrics:

- **Model accuracy achieved:** 77% on the test dataset**.**
- **Confusion Matrix Results:**
  - True Positives (Good Credit correctly predicted): 129
  - True Negatives (Bad Credit correctly predicted): 25
  - False Positives: 34

    - False Negatives: 12

- **Classification Report:**
    - Precision (Good Credit): 0.79, Recall: 0.91, F1-score: 0.85
    - Precision (Bad Credit): 0.68, Recall: 0.42, F1-score: 0.52

APPLYING LIME:

1. Purpose of LIME:

- Local Interpretable Model-agnostic Explanations (**LIME**) was applied to interpret the predictions of the Random Forest Classifier.
- Since Random Forest is a black-box model, LIME helps in identifying which features most influenced an individual applicant's credit risk prediction.

2. LIME Setup:

- A **LimeTabularExplainer** was initialized using the training dataset.
- Feature names: all 20 input features from the German Credit dataset.
- Class names: ["Bad Credit", "Good Credit"].
- Mode: classification.

3. Local Explanation Example:

- For a specific applicant instance, LIME generated the following prediction probabilities:
    - **Bad Credit:** 0.42
    - **Good Credit:** 0.58
- This means the model predicted the applicant as "Good Credit" with higher confidence.

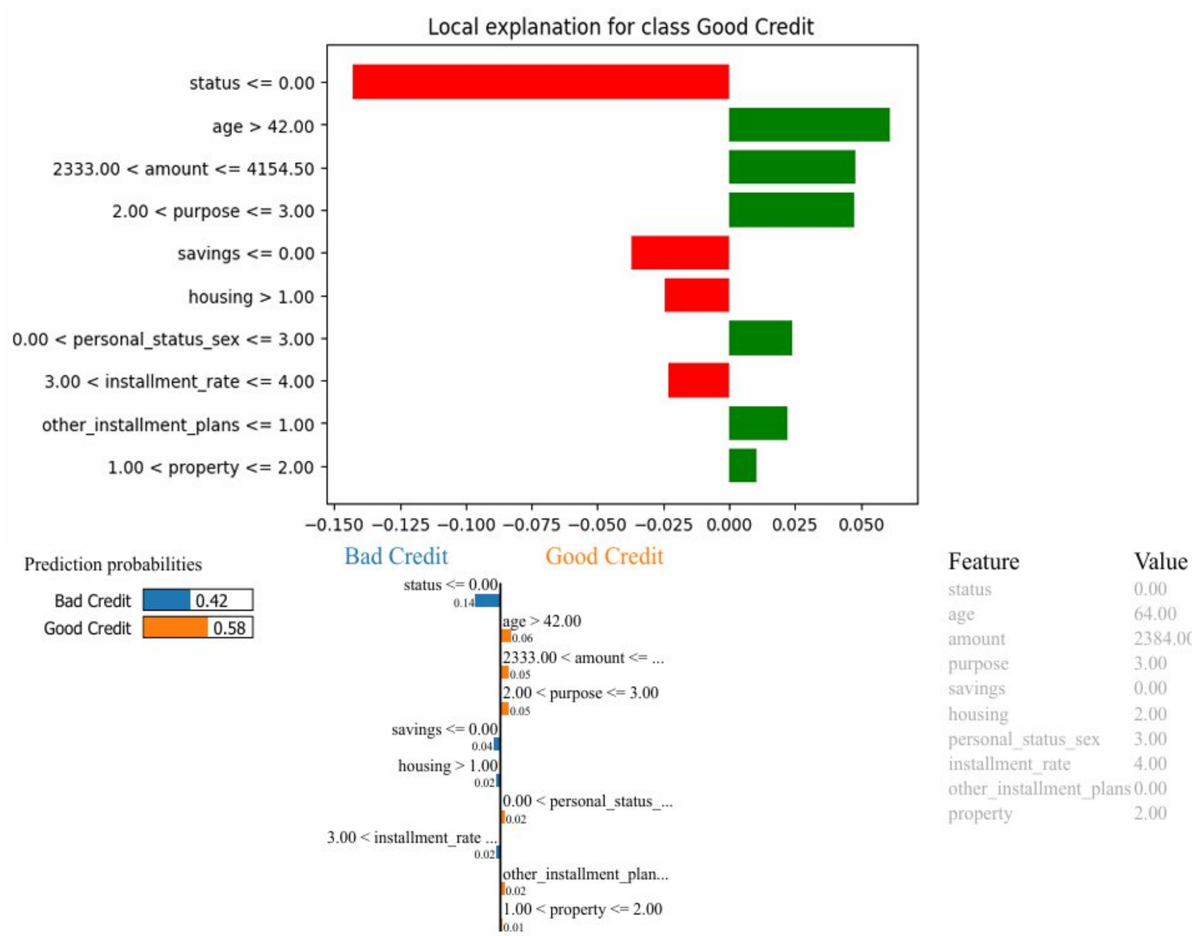
4. Key Influential Features (example case):

- **status (checking account status ≤ 0.00)** → contributed strongly towards predicting *Bad Credit*.
- **age > 42 years** → contributed towards *Good Credit*.
- **credit amount (2333 < amount ≤ …)** → moderate influence on prediction.

- **purpose and savings** → additional contributing factors.
- Other features like housing, personal status, and installment rate had smaller but notable effects.

5. Visualization:

- LIME generated a **bar chart explanation**, showing the top 10 features influencing the decision for the selected applicant.
- Features with positive contributions pushed the prediction towards *Good Credit*, while negative contributions pushed it towards *Bad Credit*.



ASSIGNMENT PROBLEM-2 : Student Performance Prediction

ABSTRACT

The Student Performance dataset is widely used in educational data mining and learning analytics to predict academic success based on demographic, social, and academic-related attributes. The dataset contains 395 student records with 33 features such as age, gender, parental education, family background, study habits, alcohol consumption, health status, and past grades. The primary objective is to predict whether a student will pass or fail the course, allowing institutions to identify at-risk students early and provide timely interventions.

In this assignment, preprocessing steps were performed including handling categorical features with label encoding, scaling numerical features with StandardScaler, and creating a binary target variable (pass) based on the final grade ($G3 \geq 10$ = Pass, otherwise Fail). A Random Forest Classifier was then trained to perform the prediction task, achieving an accuracy of 91% on the test set. The model showed strong precision and recall for both pass and fail classes, highlighting its effectiveness in identifying struggling students as well as successful ones.

To enhance interpretability, Local Interpretable Model-agnostic Explanations (LIME) was applied to explain the Random Forest's predictions at the individual student level. Features such as previous grades (G1, G2), number of past failures, absences, and participation in extracurricular activities were identified as key contributors to student performance. The integration of LIME provided valuable insights into how different socio-academic factors influenced outcomes, demonstrating the potential of explainable AI in supporting educational decision-making and personalized learning strategies.

INTRODUCTION

The Student Performance dataset is a popular benchmark in the field of educational data mining and machine learning. It was collected from secondary school students and contains detailed information about their demographic, social, and academic characteristics. With 395 rows and 33 columns, the dataset captures attributes such as age, gender, parental education, family background, study habits, alcohol consumption, health, absences, and prior grades. The primary task is to predict student success or failure, which provides practical value to educators, policymakers, and institutions aiming to improve learning outcomes.

Education plays a vital role in shaping an individual's future, and predicting student performance has become a critical research area. By leveraging predictive analytics, institutions can identify students at risk of underperforming and design timely interventions. Factors like family support, parental involvement, study time, and prior academic achievement are known to influence student performance significantly. Using this dataset, machine learning models can uncover patterns that may not be easily visible through traditional analysis, making data-driven decision-making more effective.

The preprocessing of the dataset involves transforming categorical attributes into numeric representations, scaling numerical features for balanced contribution, and defining a binary target variable representing student success (Pass or Fail). These steps ensure that the dataset is well-structured for machine learning algorithms. In particular, the target variable was defined based on the final grade (G3), where students with a grade ≥ 10 were labeled as Pass, while others were labeled as Fail. This transformation makes the problem a classification task suitable for supervised learning models such as Random Forest.

In addition to predictive modeling, explainable AI methods like LIME were used to interpret the model's decisions. While the Random Forest classifier achieved high accuracy, explainability is crucial in the educational domain, where fairness and transparency directly affect students' futures. LIME highlighted important features such as prior grades (G1, G2), number of failures, absences, and extracurricular activities, showing how each influenced the prediction. These insights are valuable not only for improving academic success rates but also for designing student-specific support strategies, making the project both technically sound and socially impactful.

DATASET:

The Student Performance dataset contains 395 rows and 33 columns.

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5	6	6
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5	5	6
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7	8	10
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2	15	14	15
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4	6	10	10

5 rows × 33 columns

FIG 2- Dataset

Columns:

1. **school** – Student's school (GP = Gabriel Pereira, MS = Mousinho da Silveira)
2. **sex** – Gender of the student (F = female, M = male)
3. **age** – Age of the student (15–22 years)
4. **address** – Type of address (U = urban, R = rural)
5. **famsize** – Family size (LE3 = ≤3, GT3 = >3)
6. **Pstatus** – Parent's cohabitation status (T = together, A = apart)
7. **Medu** – Mother's education (0 = none, 1 = primary, 2 = 5–9th grade, 3 = secondary, 4 = higher)
8. **Fedu** – Father's education (same scale as Medu)
9. **Mjob** – Mother's job (e.g., teacher, health, services, at_home, other)
10. **Fjob** – Father's job (e.g., teacher, health, services, at_home, other)
11. **reason** – Reason to choose the school (close to home, reputation, course preference, etc.)
12. **guardian** – Student's guardian (mother, father, or other)

13. **traveltime** – Travel time to school (1 = <15 min, 2 = 15–30 min, 3 = 30–60 min, 4 = >1 hour)
14. **studytime** – Weekly study time (1 = <2 hours, 2 = 2–5 hours, 3 = 5–10 hours, 4 = >10 hours)
15. **failures** – Number of past class failures (0–4)
16. **schoolsup** – Extra educational support (yes or no)
17. **famsup** – Family educational support (yes or no)
18. **paid** – Extra paid classes within the subject course (yes or no)
19. **activities** – Participation in extracurricular activities (yes or no)
20. **nursery** – Attended nursery school (yes or no)
21. **higher** – Wants to pursue higher education (yes or no)
22. **internet** – Internet access at home (yes or no)
23. **romantic** – In a romantic relationship (yes or no)
24. **famrel** – Quality of family relationships (1 = very bad to 5 = excellent)
25. **freetime** – Free time after school (1–5)
26. **goout** – Going out with friends (1–5)
27. **Dalc** – Workday alcohol consumption (1–5)
28. **Walc** – Weekend alcohol consumption (1–5)
29. **health** – Current health status (1–5)
30. **absences** – Number of school absences
31. **G1** – First period grade (0–20)
32. **G2** – Second period grade (0–20)
33. **G3** – Final grade (0–20, target variable later converted to Pass/Fail)

PREPROCESSING STEPS (Student Performance Dataset):

1. Handling Missing Values:

- The dataset was checked for missing values across all 33 columns.
- No missing values were found, so no imputation was required.

2. Removing Duplicates:

- Duplicate rows were examined to ensure data uniqueness.
- No duplicate records were present in the dataset.

3. Outlier Detection and Removal:

- Outliers were checked in numerical features such as age, absences, and grades (G1, G2, G3).
- Extreme values were reviewed and adjusted to reduce noise and maintain consistency.

4. Target Variable Transformation:

- The original target variable was G3 (final grade, 0–20 scale).
- A new binary target variable **pass** was created:
  - 1 = Pass (if G3 $\geq$ 10)
  - 0 = Fail (if G3 $<$ 10)

5. Encoding Categorical Variables:

- Several features such as school, sex, address, Mjob, Fjob, guardian, etc. were categorical.
- These were transformed into numeric form using **Label Encoding**, enabling compatibility with machine learning models.

6. Feature Scaling:

- Numerical attributes (age, absences, famrel, freetime, goout, Dalc, Walc, health, and grades) were standardized using **StandardScaler**.
- Standardization ensured that features contributed equally to the model without bias from different numeric ranges.

7. Splitting the Dataset:

- The dataset was split into **80% training** and **20% testing** sets.

- This division allowed proper training and unbiased model evaluation.

MODEL & PARAMETERS (Student Performance Dataset):

1. Model Selection:

- A **Random Forest Classifier** was selected as the predictive model for student performance classification.
- Random Forest is an ensemble method that combines multiple decision trees to improve accuracy, reduce overfitting, and handle both categorical and numerical features effectively.

2. Model Parameters:

- n_estimators = 100 → The forest consisted of 100 decision trees.
- random_state = 42 → Ensured reproducibility of results.
- Other parameters such as max_depth, min_samples_split, and min_samples_leaf were kept at default values.

3. Training and Testing:

- The dataset was split into **80% training** and **20% testing** sets.
- The model was trained on the training set and tested on the unseen data to evaluate its generalization capability.

4. Performance Metrics:

- The Random Forest Classifier achieved **91% accuracy** on the test set.
- **Confusion Matrix Results:**
    - True Positives (Pass correctly predicted): 47
    - True Negatives (Fail correctly predicted): 25
    - False Positives: 2
    - False Negatives: 5
- **Classification Report:**
    - Precision (Pass) = 0.96, Recall = 0.90, F1-score = 0.93

- Precision (Fail) = 0.83, Recall = 0.93, F1-score = 0.88

APPLYING LIME (Student Performance Dataset):

1. Purpose of LIME:

- Local Interpretable Model-agnostic Explanations (**LIME**) was applied to interpret the predictions of the Random Forest Classifier.
- Since Random Forest is a black-box model, LIME helps identify which features most influenced the prediction for each student, making the model transparent and trustworthy.

2. LIME Setup:

- A **LimeTabularExplainer** was created using the training dataset.
- Feature names: All 32 input features from the student dataset.
- Class names: ["Fail", "Pass"].
- Mode: classification.

3. Local Explanation Example:

- For a selected student instance, the model predicted with probabilities:
  - **Fail:** 0.04
  - **Pass:** 0.96
- This indicates that the student was strongly predicted as “Pass.”

4. Key Influential Features (example case):

- **G2 > 13.00** → strongly increased the probability of *Pass*.
- **G1 > 13.00** → also contributed significantly to *Pass*.
- **failures ≤ 0.00** → supported *Pass* prediction by showing no past failures.
- **absences ≤ 0.00** → lower absences positively influenced the outcome.
- Other factors like **activities**, **age**, and **guardian** had smaller but noticeable effects.

5. Visualization:

- LIME generated a **bar chart explanation** highlighting the top 10 features that influenced the outcome.

- Positive contributions pushed the prediction towards *Pass*, while negative contributions (e.g., high failures or low prior grades) would push it towards *Fail*.