AUDIT FRAUD RISK

Samuel Chadick, Rebeca Duron, Samantha Friday, Elizabeth Monks, Adam Rodriguez, Jessica Rodriguez, Sachi Wijeratne

PROBLEM STATEMENT

- Current audit processes lack effective tools for firm classification based on historical risk factors, impacting accuracy. Implementing automated logistic regression enhances fraud detection, bolstering financial audit reliability.

ADDRESSING THE PROBLEM

- Mitigating audit risk is pivotal for audit agencies, aligning with strategic objectives. Equipping auditors with machine learning models optimizes resource allocation, precisely forecasting high-risk cases for a more focused and efficient audit approach.

DATA OVERVIEW

- Dataset consists of 10 rows and 776 columns, capturing financial metrics and risk scores for government firms in India.

DATA DICTIONARY

- DISTRICT_SCORE: RISK SCORE OF A DISTRICT IN THE LAST 10YEARS
- HISTORY: FIRMS' AVERAGE HISTORICAL LOSS IN THE LAST 10 YEARS
- MONEY_VALUE: AUDIT MISSTATEMENT AMOUNT
- NUMBER: HISTORICAL DISCREPANCY SCORE
- PARA_A: EXPENDITURE DISCREPANCY
- PARA_B: EXPENDITURE DISCREPANCY
- RISK FLAG: 1 = FRAUD, 0 = NO FRAUD RISK

DATA VISUALIZATION

Financial Discrepancies in Planned Expenditure and Report B (Rs in Crore)

Para_B Average (Audit Fraud Risk 0):

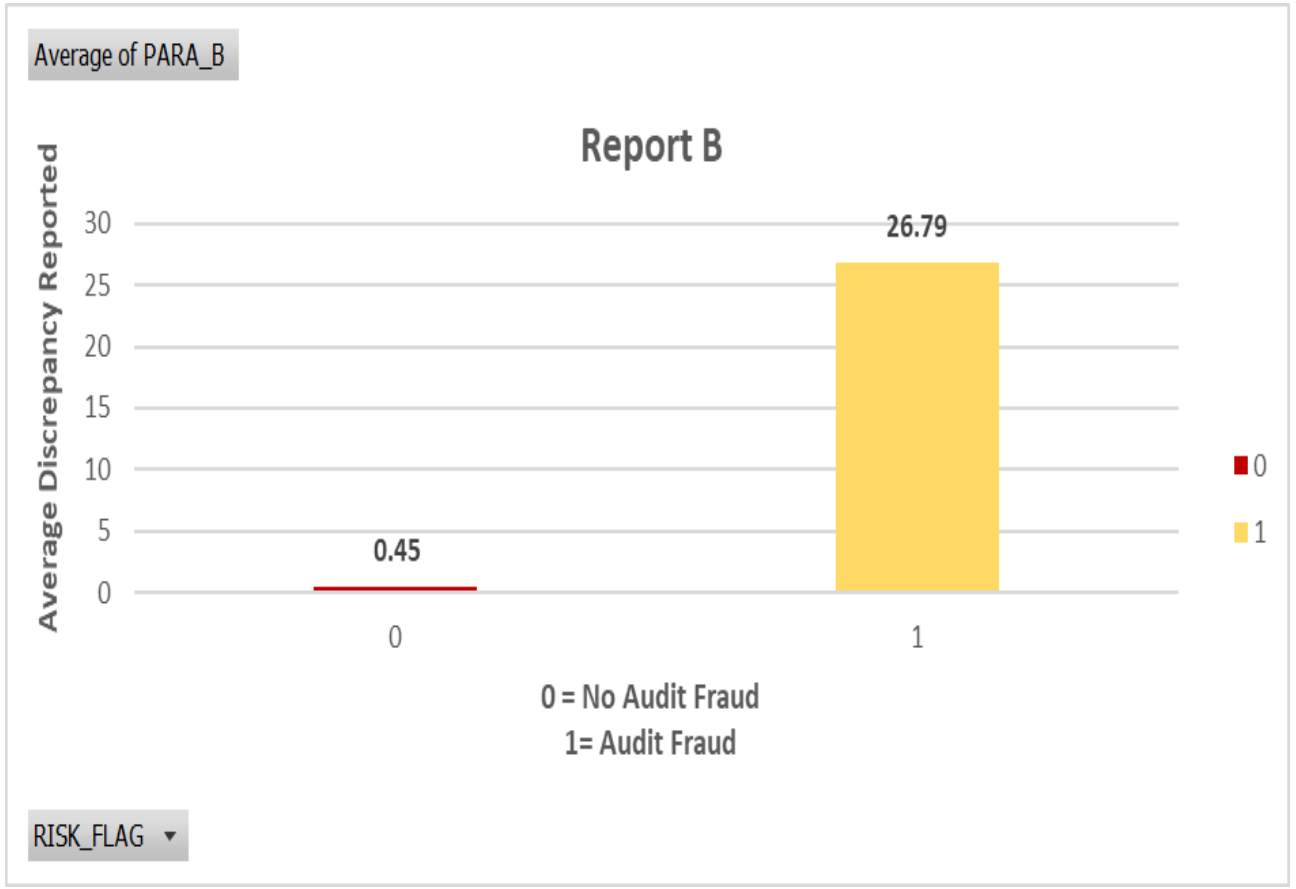
- Discrepancy in planned expenditure: Rs (in crore) with an average score of 0.45

Para_B Average (Audit Fraud Risk 1):

- Discrepancy in planned expenditure: Rs (in crore) with an average score of 26.79.

Inference:

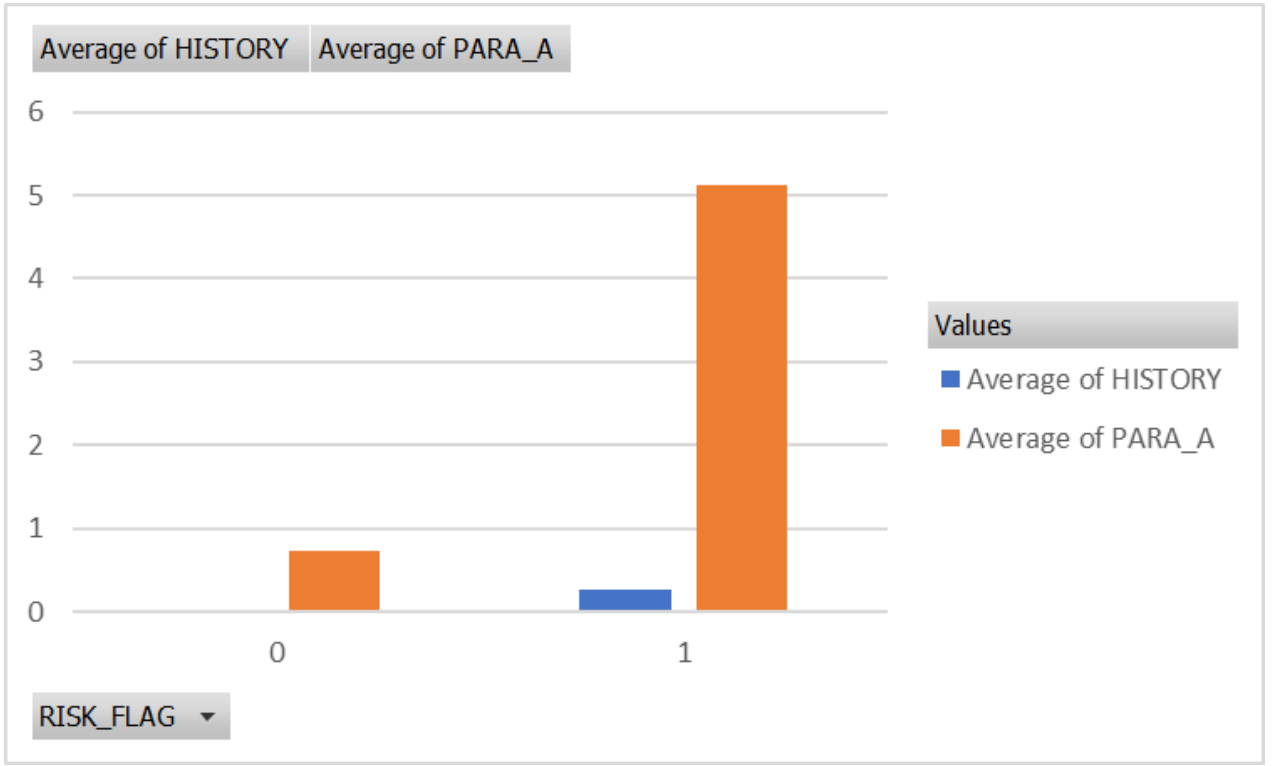
- Report B suggests a considerably higher audit fraud risk (1) compared to the case with a fraud risk of 0, evident in the significant difference in average scores.



PARA-A BEHAVIOR AND 10 YEAR HISTORICAL RISK SCORE TRENDS IN COMPARISON WITH RISK STATUS CLASSES FOR A DISTRICT

Similar Pattern in Para A and Para B

- When risk flag is O
  - Both Para A and Para B exhibit lower scores.
- When the risk flag is 1
  - Both Para A and Para B show higher scores.
- This indicates a consistent pattern: Para A and Para B have a similar level of discrepancy corresponding to the risk flag.



COMPARING PARA_A AND PARA_B FOR FRAUD RISK FLAGGED VS. NON-FLAGGED RECORDS

Data Categories

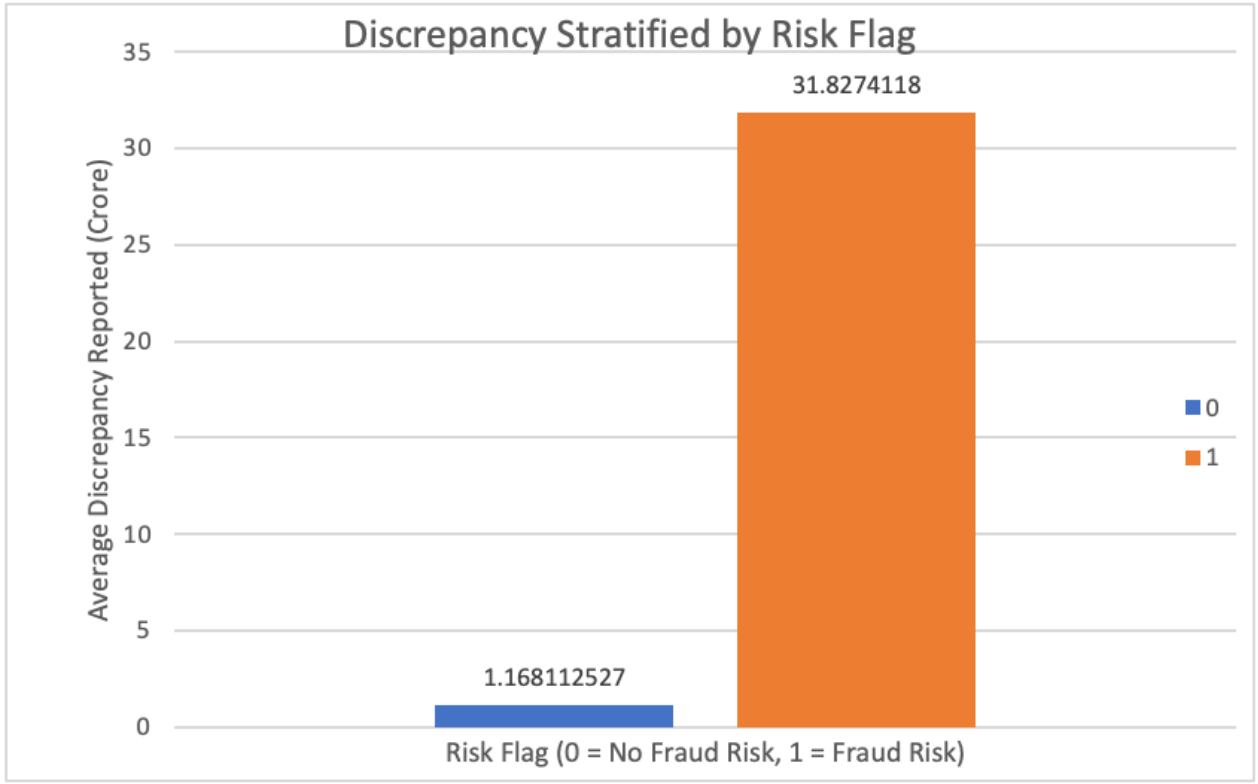
- Records flagged as fraud risk.
- Records not flagged as fraud risk.

Observation

- Average discrepancy for flagged firms for 30 times higher compared to non-flagged firms

Implication

- The visualization highlights a significant disparity in discrepancy levels between firms flagged for fraud risk and those not flagged.



District Risk Analysis

Data Categories:

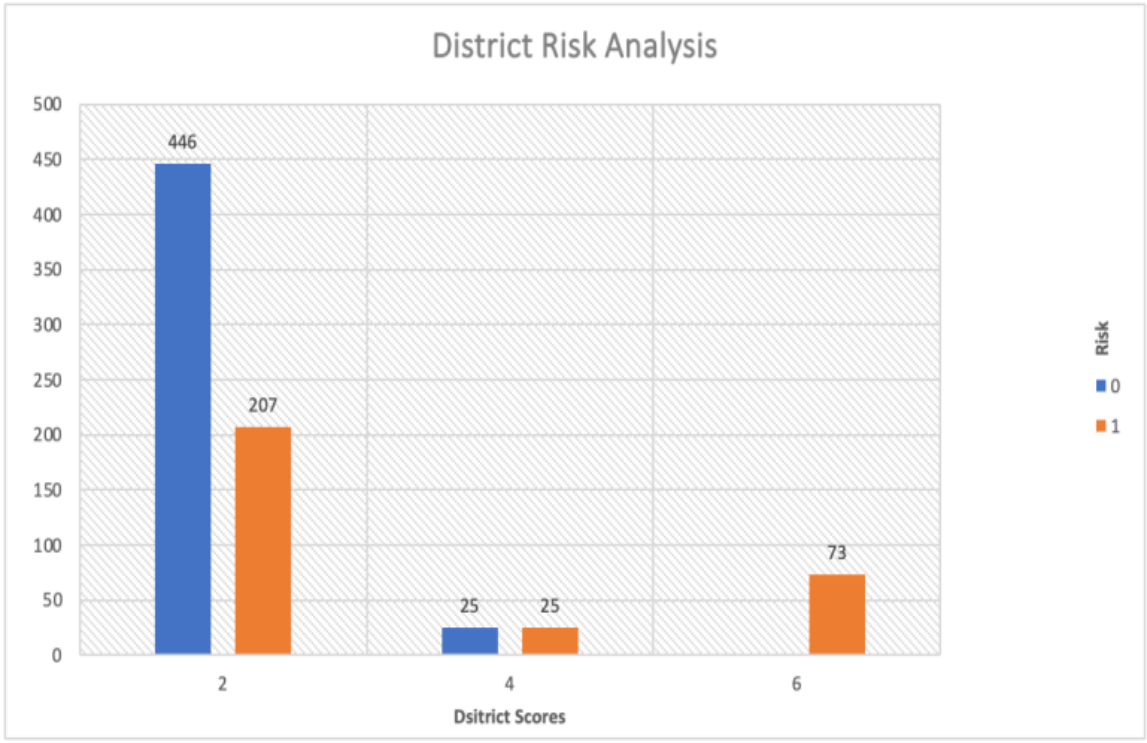
- 3 district scores of 2,4,6
- Each score has numbers of cases

Description:

- District 2: higher incidence of audit fraud risk
- District Score 4 appears to have an equal distribution between audit fraud risk and no audit fraud risk.
- A District Score of 6 and no information provided for Risk Flag.

Insights:

- Specific district scores may indeed relate to the risk flag, with some districts having a higher likelihood of audit fraud risk (Risk Flag 1), while others have a more balanced distribution or incomplete information.



Money Value by District Score

District Score 2:

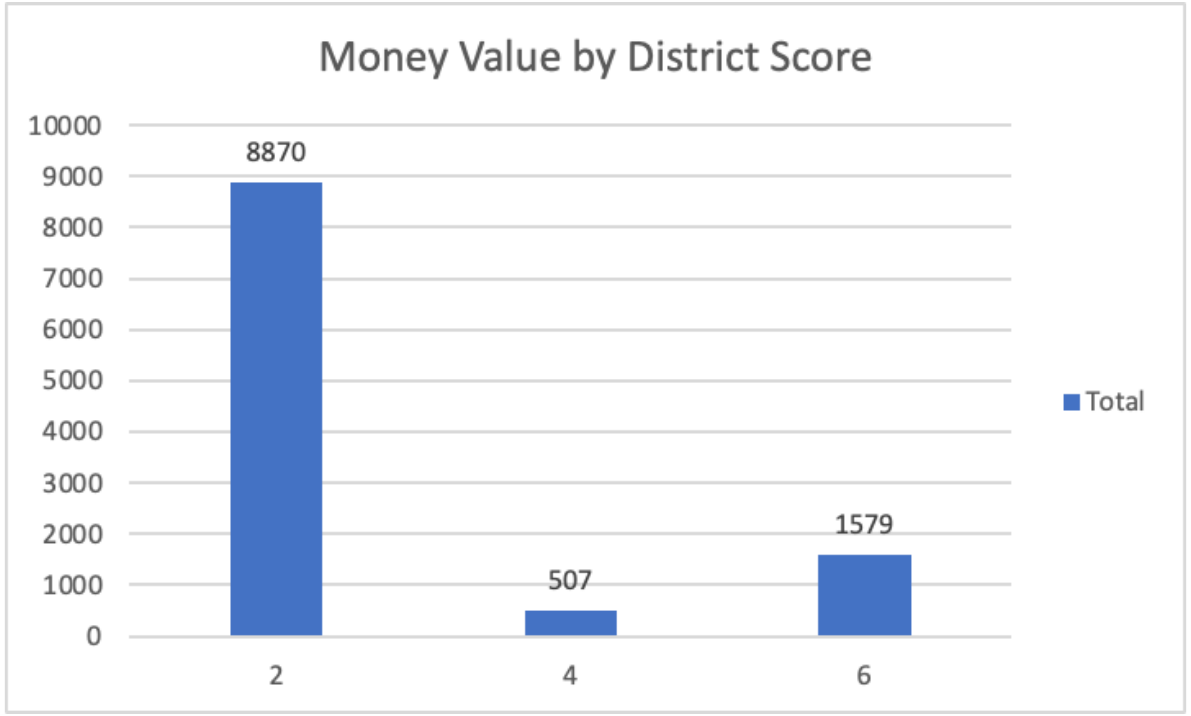
- Historical Risk Score: Moderate
- Money Value: 8,870 Rupees
- Implication: Substantial money in misstatements, suggesting potential financial discrepancies.

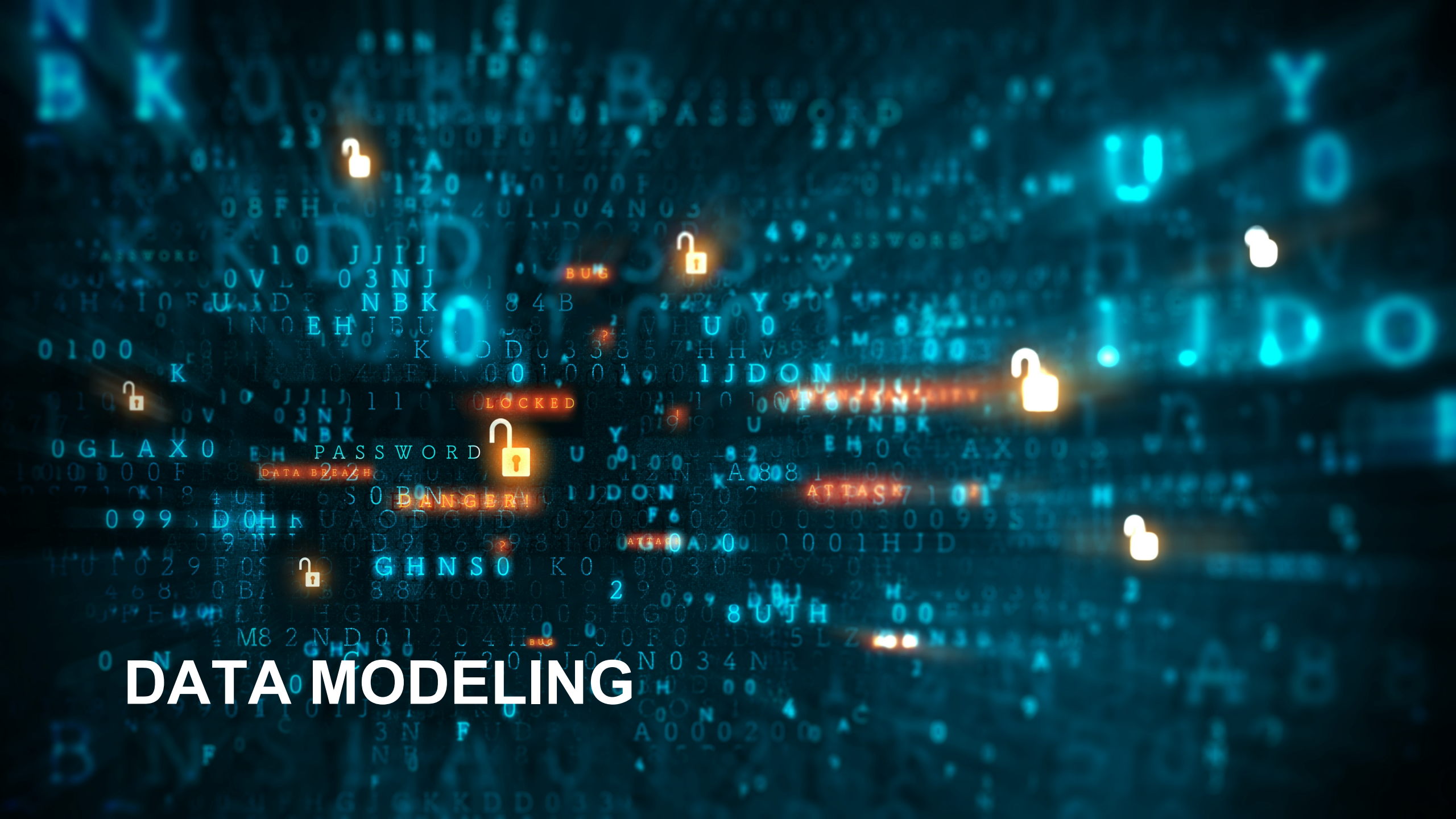
District Score 4:

- Historical Risk Score: Higher
- Money Value: 507 Rupees
- Implication: Despite a higher risk score, the money involved in misstatements is relatively low.

District Score 6:

- Historical Risk Score: Intermediate
- Money Value: 1,579 Rupees
- Implication: Falls between others, indicating moderate historical risk and a moderate amount of money in misstatements.




DATA MODELING

K-NEAREST NEIGHBORS CLASSIFICATION MODEL EVALUATION

accuracy: 92.67%

	true false	true true	class precision
pred. false	138	14	90.79%
pred. true	3	77	96.25%
class recall	97.87%	84.62%	

Model Selection: Chose k-nearest neighbors for predicting fraud risk.

Optimal k-Value:

- Determined optimal k-value through trial-and-error.
- Selected k = 5 for the model.

Performance Metrics:

- Accuracy Score: 92.87%
- Recall Score: 84.62%
- AUC Score: 0.957

Model Comparison: Among three models created, k-NN performed the poorest.

LOGISTIC REGRESSION MODEL: SUPERIOR PERFORMANCE IN FRUAD RISK PREDICTION

Model Selection: Chose logistic regression for binary target variable.

Performance Metrics:

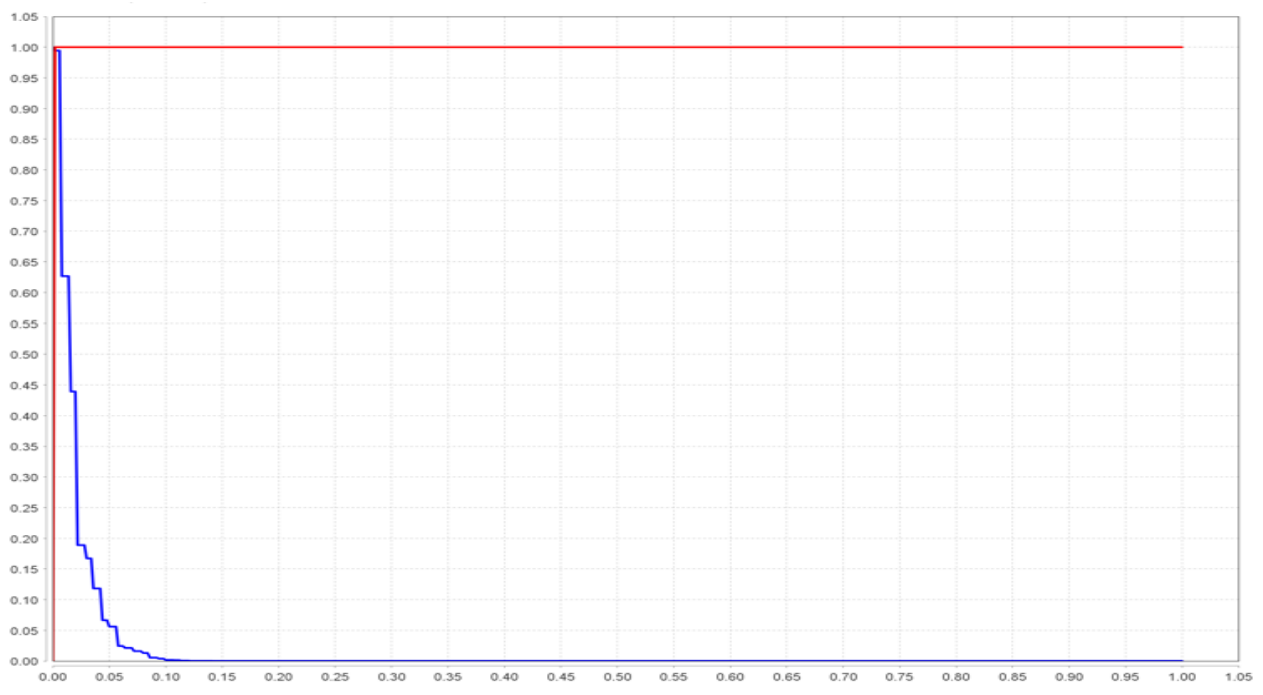
- Accuracy: 99.57%
- Recall: 100.00%
- AUC: 1.000

Confusion Matrix: Only one error: false positive for fraud risk.

Summary: Logistic regression excelled in predicting fraud risk classes.

accuracy: 99.57%

	true false	true true	class precision
pred. false	140	0	100.00%
pred. true	1	91	98.91%
class recall	99.29%	100.00%	



recall: 100.00% (positive class: true)

	true false	true true	class precision
pred. false	140	0	100.00%
pred. true	1	91	98.91%
class recall	99.29%	100.00%	

RANDOM FOREST MODEL: ROPBUST CLASSIFICATION PERFORMANCE

Model Selection: Chose random forest for superior classification over single decision trees.

Tuning: Optimal model achieved through trial-and-error.

- Criteria: Accuracy
- Number of Trees: 50

Performance Metrics:

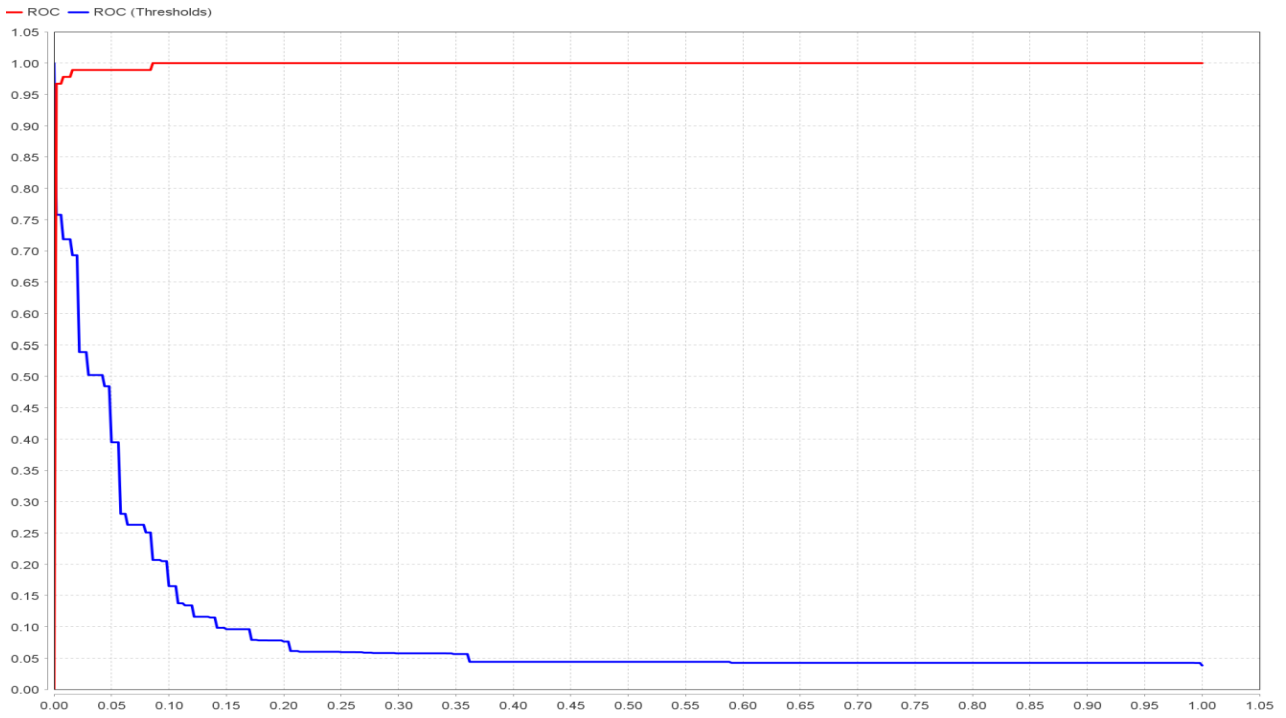
- Accuracy: 98.71%
- Recall: 98.90%
- AUC: 0.999

Comparison: Outperformed k-NN, slightly below logistic regression.

accuracy: 98.71%

	true false	true true	class precision
pred. false	139	1	99.29%
pred. true	2	90	97.83%
class recall	98.58%	98.90%	

AUC: 0.999 (positive class: true)



recall: 98.90% (positive class: true)

	true false	true true	class precision
pred. false	139	1	99.29%
pred. true	2	90	97.83%
class recall	98.58%	98.90%	

CONCLUSION

Objective: Optimize auditor time for fraud risk assessment.

Key indicators: district/sector scores, spending discrepancies, and large missed statement values.

Best Model: Logistic regression excelled in accuracy, recall, and AUC.

Recommendation: Implement logistic regression for efficient fraud risk predictions.

Outcome: Streamline auditing with a data-driven approach, focusing on identified risk factors.


THANK YOU...