



# PREDICTING BANK FAILURE: AN IMPROVEMENT BY IMPLEMENTING A MACHINE-LEARNING APPROACH TO CLASSICAL FINANCIAL RATIOS

Tarun Kumar Chopra  
Final Report

## Introduction

- problem statement
- motivation
- open questions in the domain
- a brief overview of the approach to address the challenges

## Backgrounds

- summary of other related researches (at least 2)
- Pros and Cons
- How the other work is related to the main method

## Methods

- details of the algorithms and methods
- Provide the overall framework figure to implement your work
- Do you have any data preprocessing? If so, describe how you did

## Experiments

- test your methods and show the results with some discussions
- Were you able to reproduce the paper's experiments? Are the results identical? If different, why does it happen?
- Do you have your own experiments with other data? If so, share your observations and analysis to it.
- Share your thoughts about the results and discuss what you think about the model and the solution

## Conclusions

- Add some concluding remarks and possible future work you might want to do

## (Anonymous) Sharing agreement

- Do you agree to share your work as an example for next semester?
- Do you want to hide your name/team if you agree?
- You are fine to say "no" if you don't want.

# 1. INTRODUCTION



---

## I. I PROBLEM



Bank failure is a significant issue for the financial sector, which can lead to severe economic consequences.





## I.2 MOTIVATION

---

machine learning models for predicting bank failure can provide a valuable tool for financial institutions and regulators to monitor and manage risk, protect customers, and maintain economic stability.

Early warning  
system

Risk  
management

Regulatory  
compliance

Customer  
protection

Economic  
stability



---

OPEN QUESTIONS





# Open Questions in Domain

How to Evaluate  
Accuracy of Model  
which can be used to  
predict the Bank  
Failure use case ?

Which All features are  
best suited to solve  
this problem ?

Do we have data to  
train and test ?

---

## 1.3 OVERVIEW



Predicting bank failure is a complex task that requires the consideration of numerous variables and factors. There are several approaches to predicting bank failure, including both quantitative and qualitative methods. It involves several challenges and potential problems.

# Challenges

Data  
availability

Data  
quality

Imbalanced  
data

Feature  
selection

Model  
complexity

Regulatory  
constraints

The background features a series of vertical, slightly curved light streaks in orange, yellow, green, and blue, set against a dark, textured backdrop. A semi-transparent black rectangle is positioned in the lower-middle section, containing the text '2. BACKGROUND' in white, bold, sans-serif font. A thin horizontal line is located directly beneath the text.

## 2. BACKGROUND



A blue ballpoint pen with a silver-colored tip and clip is positioned diagonally on the left side of the image. The background is a document featuring a bar chart with several blue bars of varying heights. A dark blue rectangular box is overlaid on the right side of the image, containing the section title in white text.

## 2.1 SUMMARY OF OTHER RELATED RESEARCHES

---

Predicting bank failure an improvement by implementing a machine-learning approach to classical financial ratios in 2018 where researcher compares the accuracy of two approaches: traditional statistical techniques and machine learning techniques, which attempt to predict the failure of banks.

A sample of 3000 US banks (1438 failures and 1562 active banks) is investigated by two traditional statistical approaches (Discriminant analysis and Logistic regression) and three machine learning approaches (Artificial neural network, Support Vector Machines and k-nearest neighbors).

For each bank, data were collected for a 5-year period before they become inactive. 31 financial ratios extracted from bank financial reports covered 5 main aspects: Loan quality, Capital quality, Operations efficiency, Profitability and Liquidity. The empirical result reveals that the artificial neural network and k-nearest neighbor methods are the most accurate.

The background is a blurred image of a document. It features a line graph with a solid line and a dotted line. There are handwritten numbers: '2.5' on the left, '2.47' on the right, and '20' at the bottom left. A pen is visible in the top right corner.

## 2.3 HOW THE OTHER WORK IS RELATED TO THE MAIN METHOD





RESEARCHES SHOWS MACHINE LEARNING TECHNIQUES  
BEST FOR PREDICTING BANK FAILURE



## RELATED RESEARCH LINKS

---

- <https://www.sciencedirect.com/science/article/abs/pii/S0275531917301241#tbl0005>
- <https://www.sciencedirect.com/science/article/pii/S0169207018300359>
- <https://digitalcommons.bryant.edu/cgi/viewcontent.cgi?article=1194&context=eeb>

The background features a dark, almost black, rectangular area in the center. Surrounding this area are several bright, curved, and slightly blurred lines of light in various colors, including orange, yellow, green, and blue. These lines appear to be light trails or reflections, creating a sense of motion and depth. The overall effect is a futuristic or high-tech aesthetic.

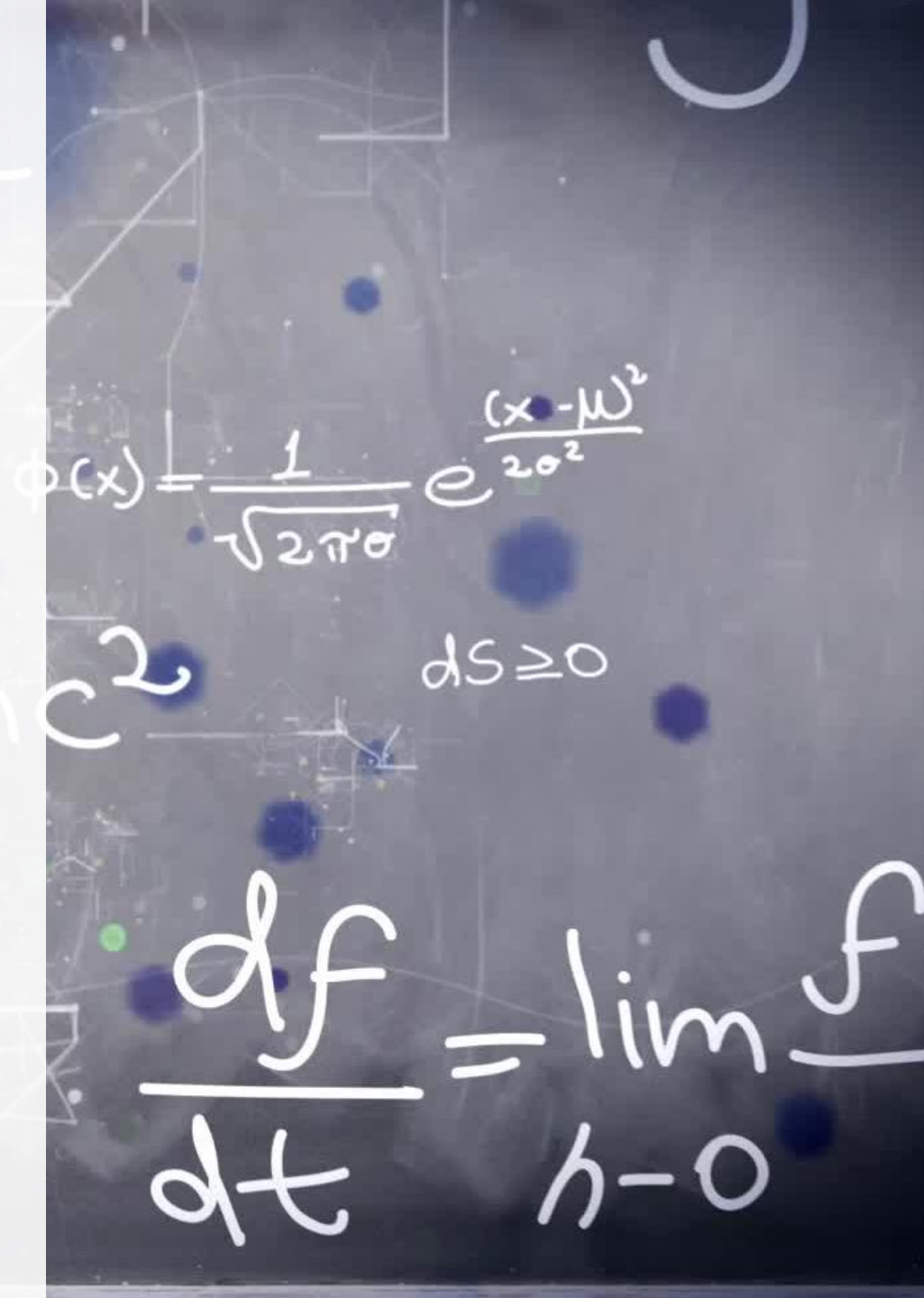
## 3. METHODS

---



## 3.1 THE METHODS USED

1. SGD Classifier
2. Support vector Machine
3. Logistic Regression
4. Kneighbors Classifier
5. Gaussian NB
6. Keras Classifier ( Neural Network)





## 3.2 DETAILS OF THE ALGORITHMS

---



## Model Execution

```
1 def MyModel(X_train, X_test, y_train, y_test):
2     from sklearn.linear_model import SGDClassifier
3     sgd = SGDClassifier(max_iter=1000, tol=1e-3)
4
5     from sklearn.linear_model import LogisticRegression
6     logreg = LogisticRegression(random_state=0)
7
8     from sklearn import neighbors
9     knn = neighbors.KNeighborsClassifier()
10
11    from sklearn.naive_bayes import GaussianNB
12    nb = GaussianNB()
13
14    from sklearn import svm
15    svm = svm.SVC()
16
17    # list of algorithms to test
18    clfs = [sgd, logreg, knn, nb, svm]
19    # list of algorithm names
20    names = ["SGD", "Logistic Reg", "kNN", "Naive Bayes", "Support Vector Machine"]
21    train_accs = []
22    test_accs = []
23
24    for name, clf in zip(names, clfs):
25        print("{:=^50s}".format(name))
26
27        # TODO 18.3
28        clf.fit(X_train, y_train)
29
30
31        train_score = clf.score(X_train, y_train)
```

```
33
34     test_score = clf.score(X_test, y_test)
35
36     print(f"Train Accuracy: {train_score}\nTest Accuracy: {test_score}")
37     # Track each model/classifier's train and test accuracy
38     train_accs.append(train_score)
39     test_accs.append(test_score)
40
41     t_train = clf.predict(X_train)
42
43
44     t_test = clf.predict(X_test)
45
46     mcc_score = matthews_corrcoef(y_test, t_test)
47     print(f"MCC: {mcc_score}")
48
49
50     #target_names = ['M', 'F', 'I']
51     cm_report = classification_report(y_test, t_test)
52     print(cm_report)
53
54     # TODO 9.8
55     cm = confusion_matrix(y_test, t_test)
56     cm_display = ConfusionMatrixDisplay(cm).plot()
57
58     plt.figure(figsize=(12,4))
59
60     plt.subplot(121)
61     plt.plot(y_train, 'ro')
62     plt.plot(t_train, 'bx')
63     plt.title("Train")
64
```

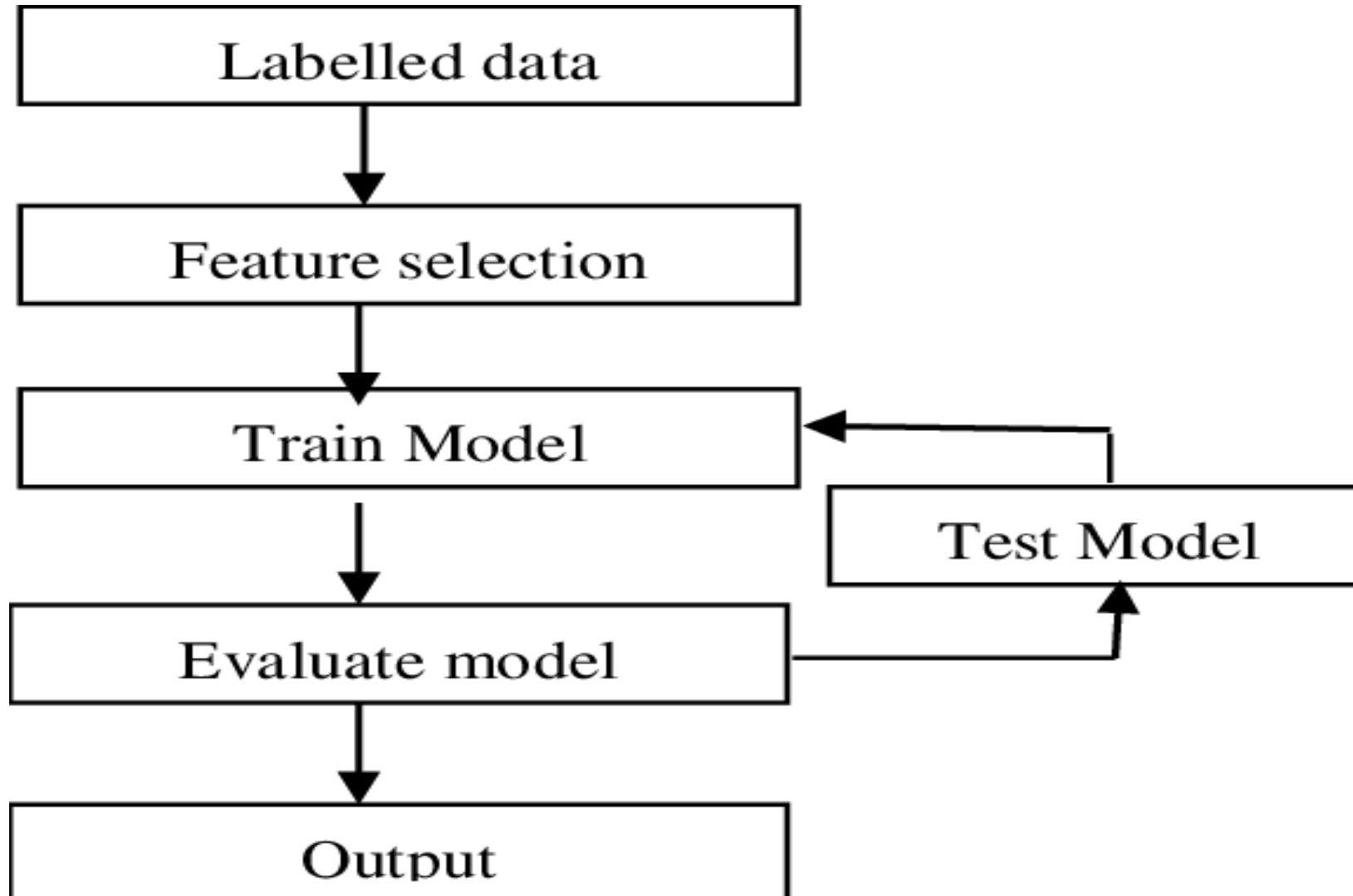
```
65     plt.subplot(122)
66     plt.plot(y_test, 'ro')
67     plt.plot(t_test, 'bx')
68     plt.title("Test")
69     plt.suptitle(name)
70     plt.show()
71     return train_accs, test_accs
```

A close-up, shallow depth-of-field photograph of a person's hands in a workshop. The person is wearing a blue long-sleeved shirt with the sleeves rolled up. They are holding a yellow pencil in their right hand and are in the process of marking a piece of light-colored wood. The background is blurred, showing various workshop tools and equipment. A semi-transparent dark grey rectangular box is overlaid on the lower right portion of the image, containing white text.

### 3.3 FRAMEWORK FIGURE TO IMPLEMENT WORK

---

# Machine Learning Approach





The background features a series of overlapping, stylized geometric shapes, primarily squares and pentagons, in shades of teal, light blue, and dark grey. Each shape contains a white question mark. The shapes are arranged in a way that creates a sense of depth and movement. A small orange horizontal bar is located in the upper right quadrant.

## 3.4 data preprocessing

---

# Data Preprocessing

- Data Ingestion – Reading data from CSV
- Data Cleaning – Checking for Null Values and missing rows
- Data Reduction – Selecting Features which can provide more information to Models

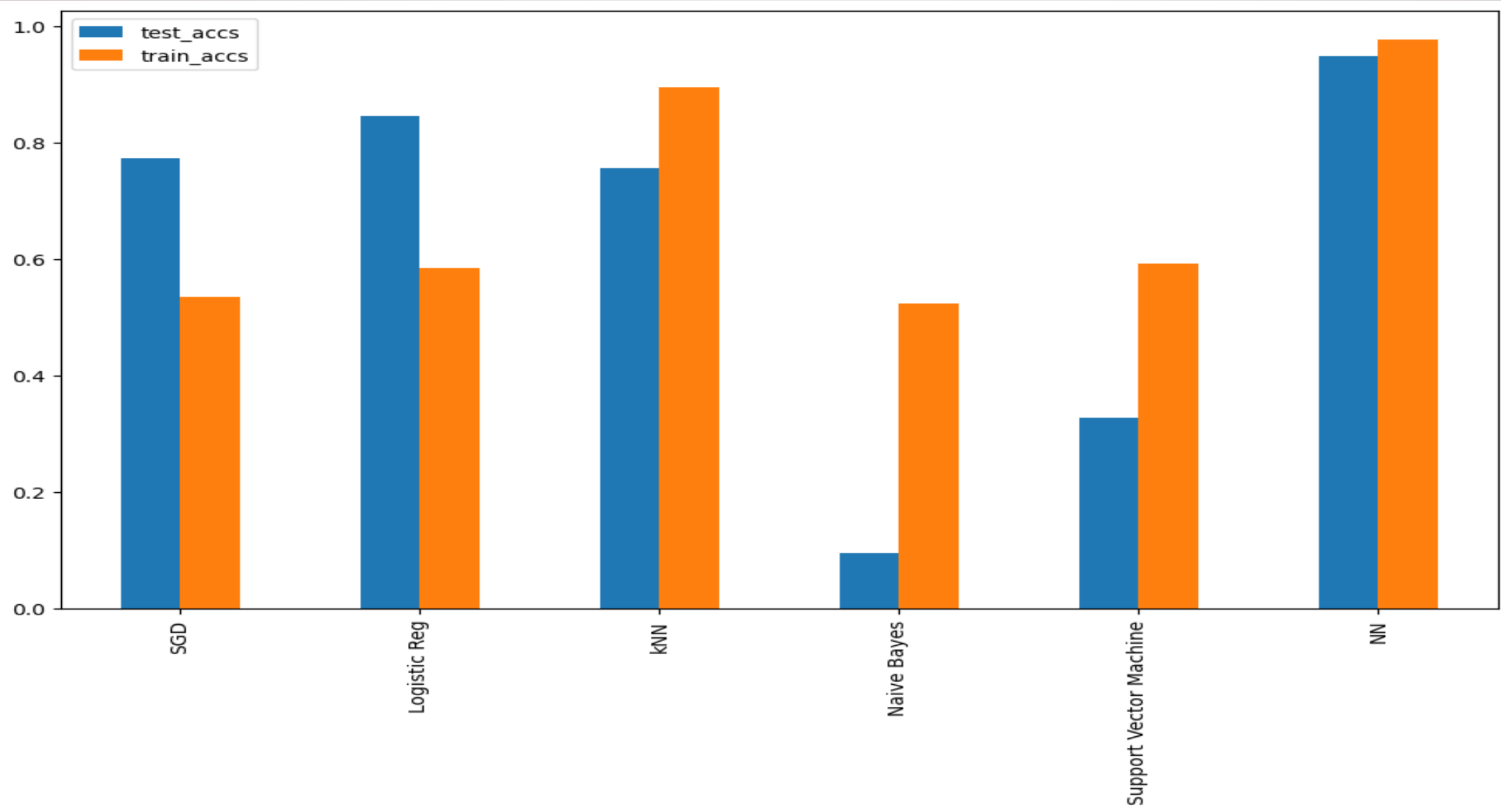
## 4. EXPERIMENTS

---



A background image showing a group of business professionals in a meeting. They are gathered around a table, looking at documents and using mobile devices. The image is partially obscured by a large, semi-transparent circular overlay that frames the central text.

# 4.1 Tested methods and results





# RESULTS

---

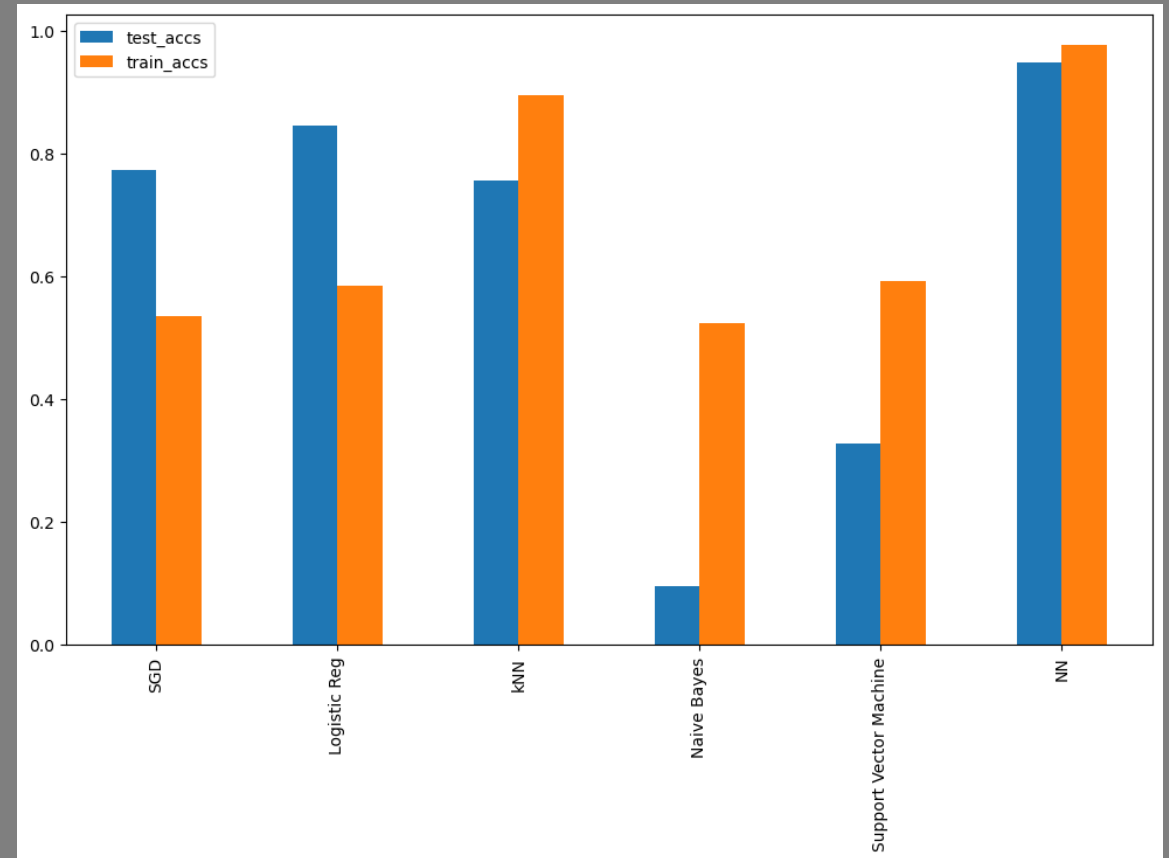
Neural Network ( Keras Classifier) shows best results of Train and Test data for predicting bank failure.



# BEST MODEL COMPARISON

After comparing 6 different models for predicting bank failure Keras Classifier (Neural Networks) shows best results.

Logistic regression and SGD produce better test accuracy as well





4.2 Were you able to reproduce the paper's experiments? Are the results identical? If different, why does it happen?



# Results and Metrics Evaluations

With Multi Model evaluation I were able to find out which model provide better accuracy and results but due to shortage of authentic data ( Bankscope data vs open source data ) not able to confirm our results

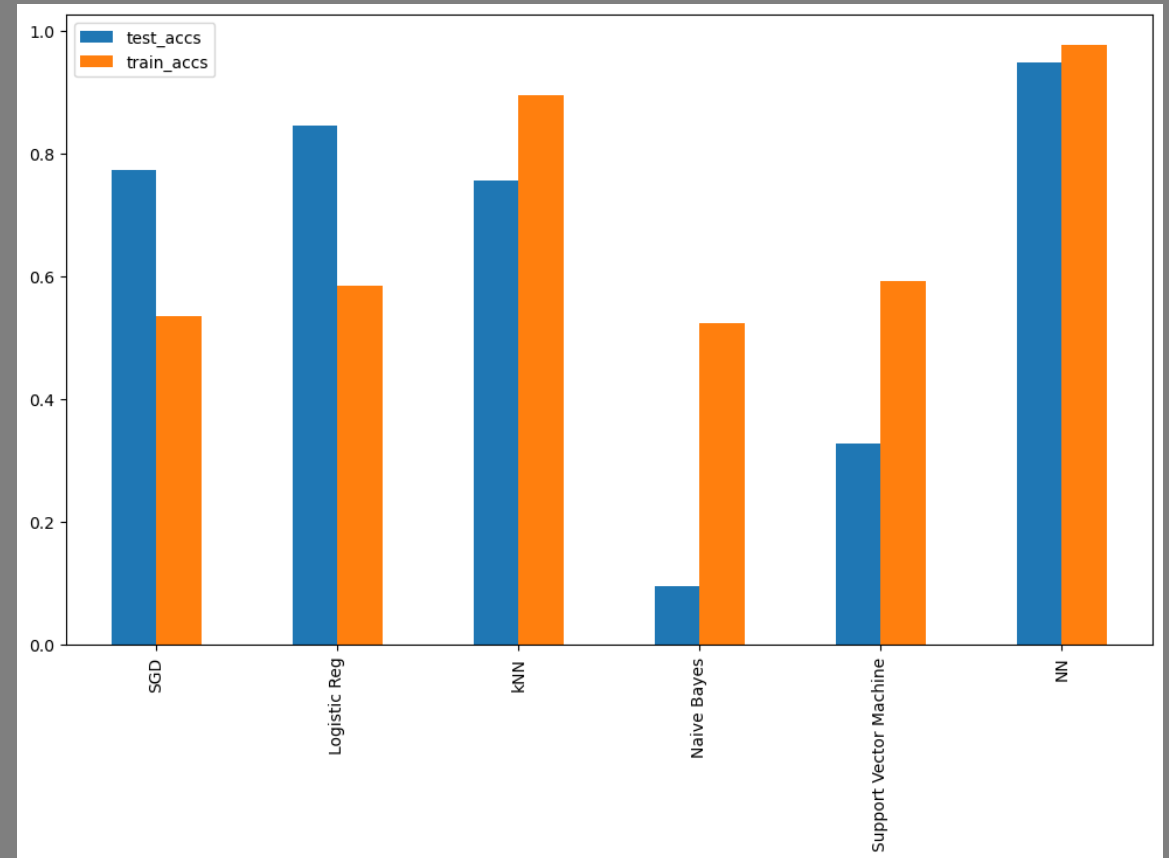


4.3 Own experiment with other data.  
Observations and analysis.

# BEST MODEL COMPARISON

After comparing 6 different models for predicting bank failure Keras Classifier (Neural Networks) shows best results.

Logistic regression and SGD produce better test accuracy as well







## 4.4 Results and the solution

# Results and Metrics Evaluations

With Multi Model evaluation, I am able to find out which model provide better accuracy and results but due to shortage of authentic data ( Bank scope data vs opensource data ) not able to confirm my results.

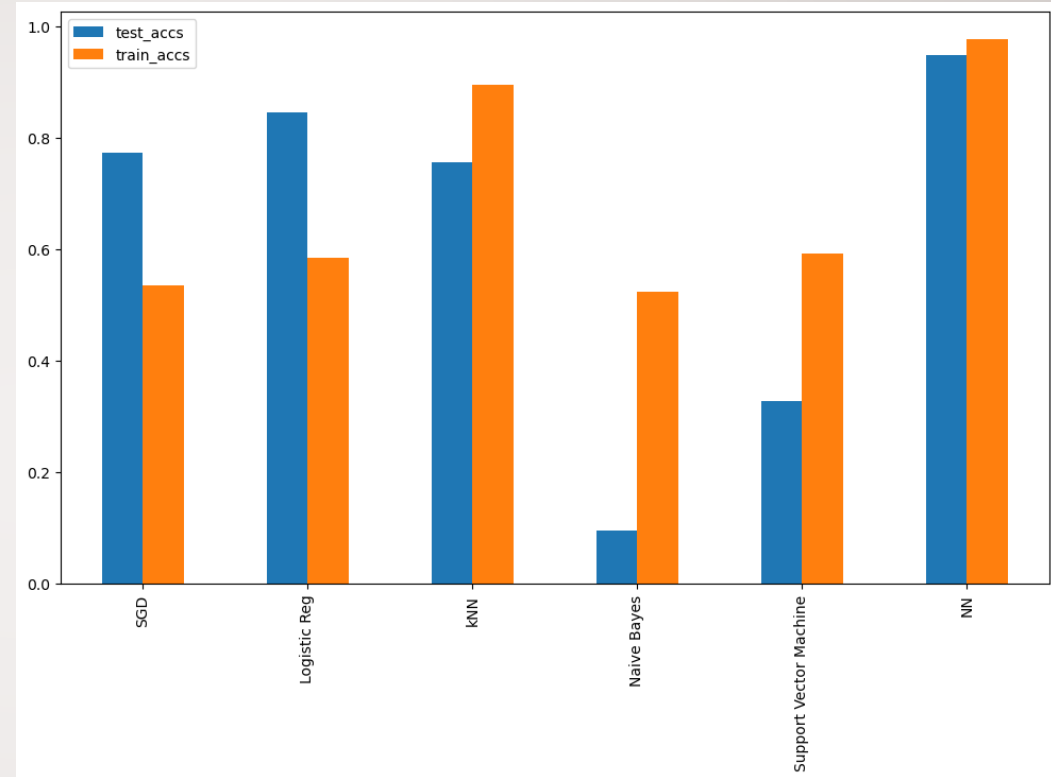


---

CONCLUSIONS



- After comparing 6 different models for predicting bank failure Keras Classifier (Neural Networks) shows best results.
- Logistic regression and SGD produce better test accuracy as well



## References:

- Hong Hanh Le, Jean-Laurent Viviani : Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios. Version of Record 17 April 2018 in Science Direct.
- Periklis Gogas, Theophilos Papadimitriou, Anna Agrapetidou from Democritus University of Thrace, Department of Economics, Greece: Forecasting bank failures and stress testing: A machine learning approach. Version of Record 24 April 2018 in Science Direct.
- Armen Eghian : Comparing Machine Learning Techniques for Predicting Bank Failure.