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

1. 1 PROBLEM

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

1.2 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

2. BACKGROUND

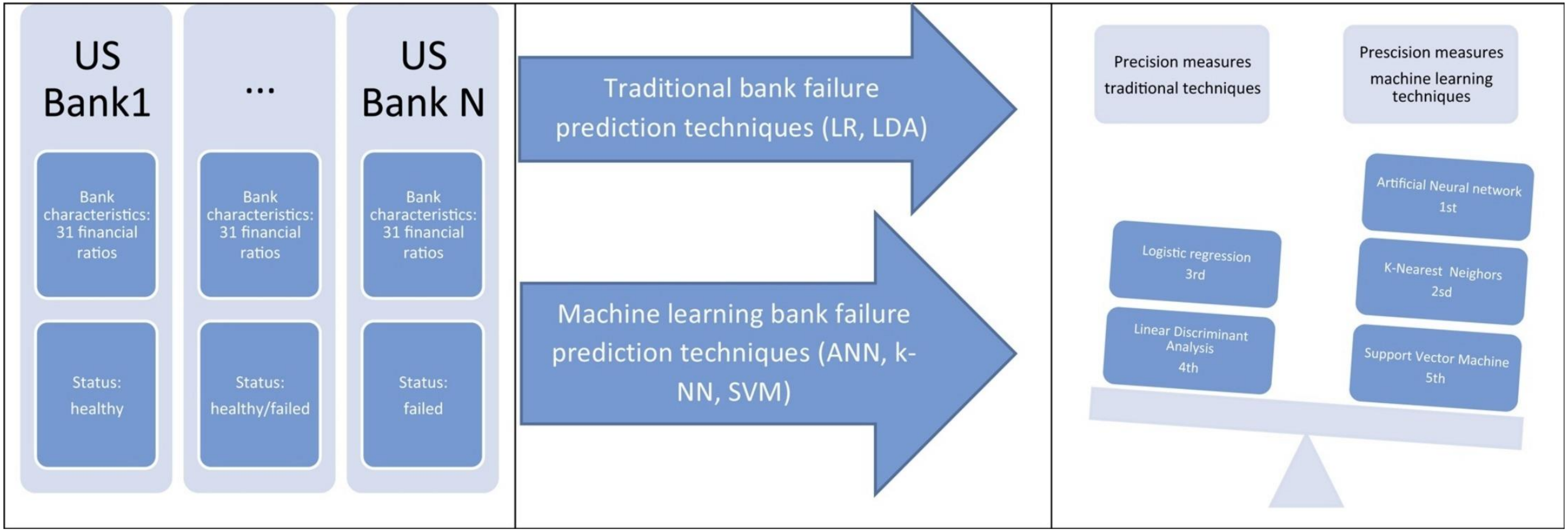
2.1 SUMMARY OF OTHER RELATED RESEARCHES

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.

2.3 HOW THE OTHER WORK IS RELATED TO THE MAIN METHOD



RESEARCHES SHOWS MACHINE LEARNING TECHNIQUES BEST FOR PREDICTING BANK FAILURE

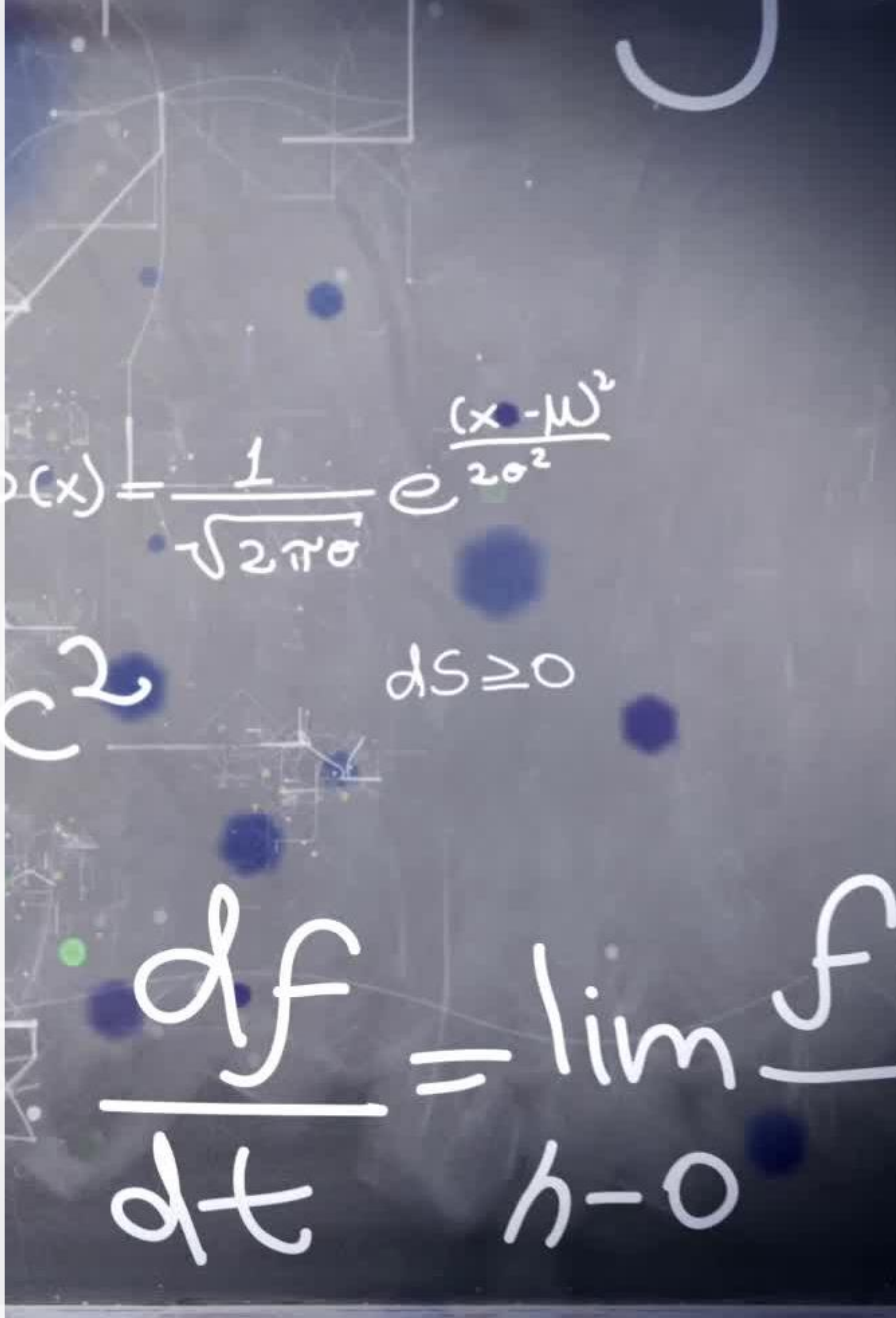
RELATED RESEARCH LINKS

- 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

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

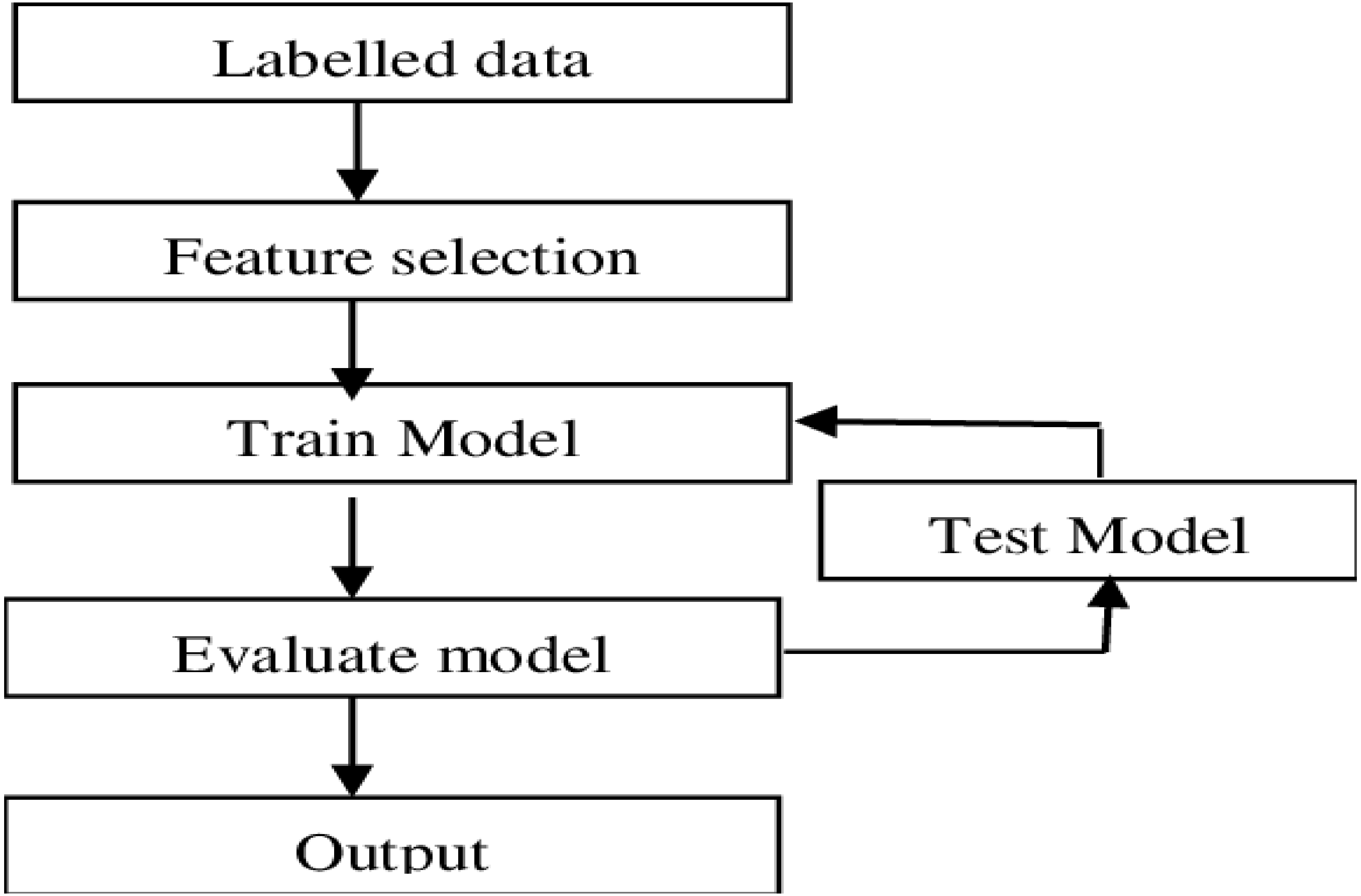
```python
def MyModel(X_train, X_test, y_train, y_test):
    from sklearn.linear_model import SGDClassifier
    sgd = SGDClassifier(max_iter=1000, tol=1e-3)

    from sklearn.linear_model import LogisticRegression
    logreg = LogisticRegression(random_state=0)

    from sklearn import neighbors
    knn = neighbors.KNeighborsClassifier()

    from sklearn.naive_bayes import GaussianNB
    nb = GaussianNB()

    from sklearn import svm
    svm = svm.SVC()

    # list of algorithms to test
    clfs = [ sgd, logreg, knn, nb , svm]
    # list of algorithm names
    names = ["SGD", "Logistic Reg", "kNN", "Naive Bayes","Support Vector Machine"]
    train_accs = []
    test_accs = []

    for name, clf in zip(names, clfs):
        print("{:=^50s}".format(name))

        # TODO 18.3
        clf.fit(X_train, y_train)


        train_score = clf.score(X_train, y_train)
```

```
        test_score = clf.score(X_test, y_test)

        print(f"Train Accuracy: {train_score}\nTest Accuracy: {test_score}")
        # Track each model/classifier's train and test accuracy
        train_accs.append(train_score)
        test_accs.append(test_score)

        t_train = clf.predict(X_train)


        t_test = clf.predict(X_test)

        mcc_score = matthews_corrcoef(y_test, t_test)
        print(f"MCC: {mcc_score}")


        #target_names = ['M', 'F', 'I']
        cm_report  = classification_report(y_test, t_test)
        print(cm_report)

        # TODO 9.8
        cm = confusion_matrix(y_test, t_test)
        cm_display = ConfusionMatrixDisplay(cm).plot()

        plt.figure(figsize=(12,4))

        plt.subplot(121)
        plt.plot(y_train, 'ro')
        plt.plot(t_train, 'bx')
        plt.title("Train")
```

```
        plt.subplot(122)
        plt.plot(y_test, 'ro')
        plt.plot(t_test, 'bx')
        plt.title("Test")
        plt.suptitle(name)
        plt.show()
    return train_accs, test_accs
```

3.3 FRAMEWORK FIGURE TO IMPLEMENT WORK

Machine Learning Approach

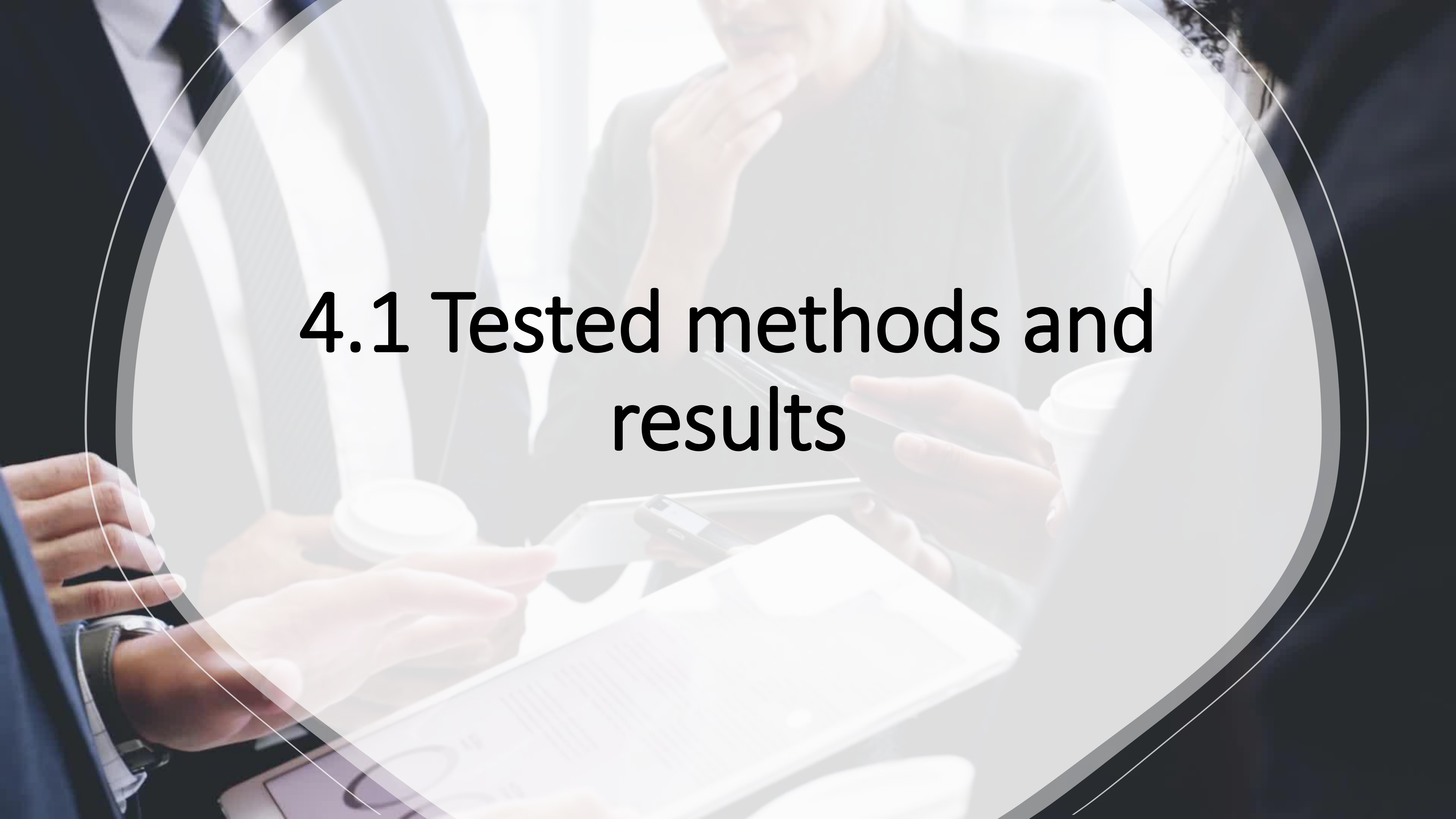


3.4 data preprocessing

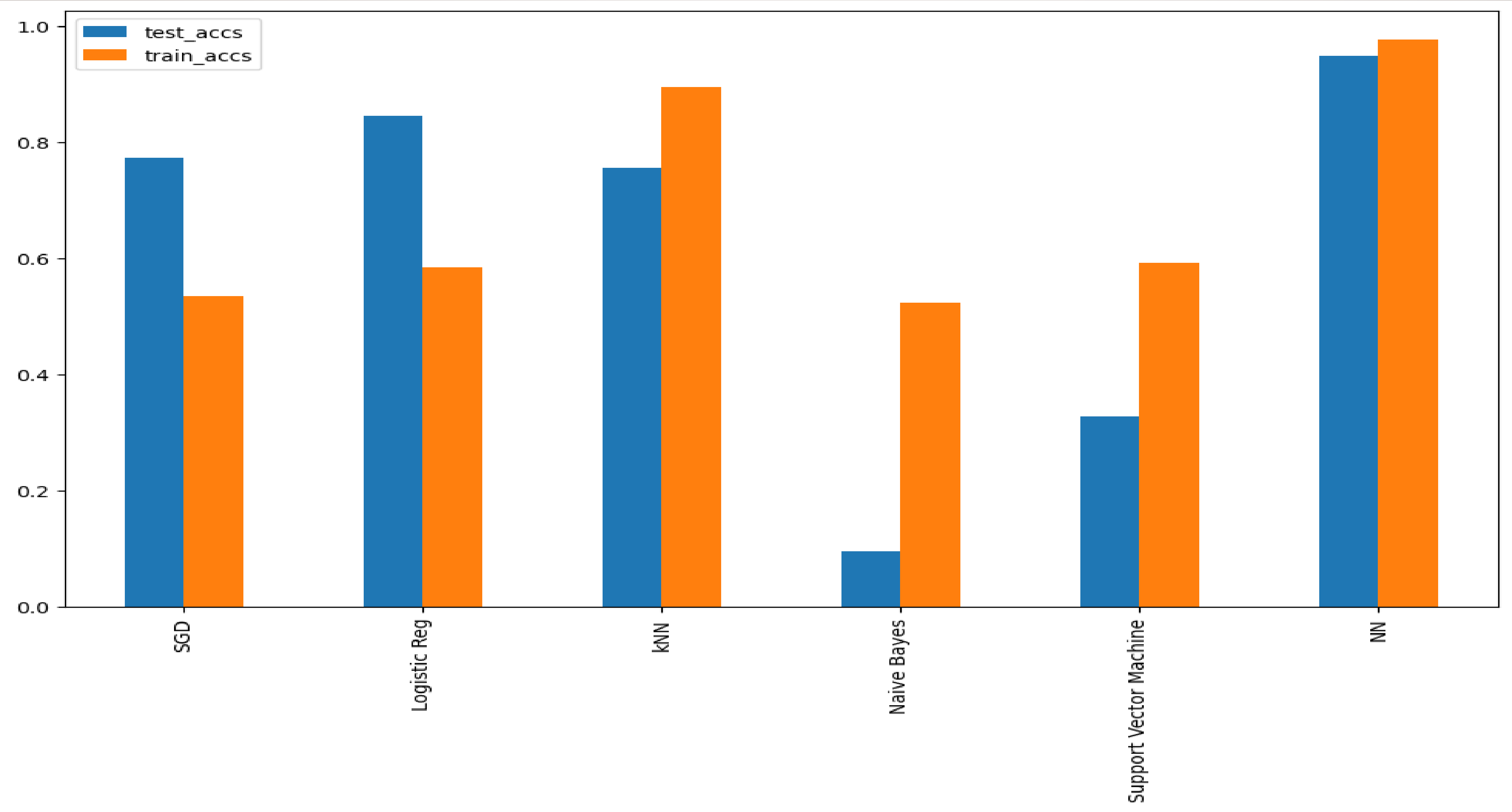
Data Preprocessing

- 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



4.1 Tested methods and results



1.0
0.8
0.6
0.4
0.2
0.0
test_accs
train_accs
SGD
Logistic Reg
kNN
Naive Bayes
Support Vector Machine
NN

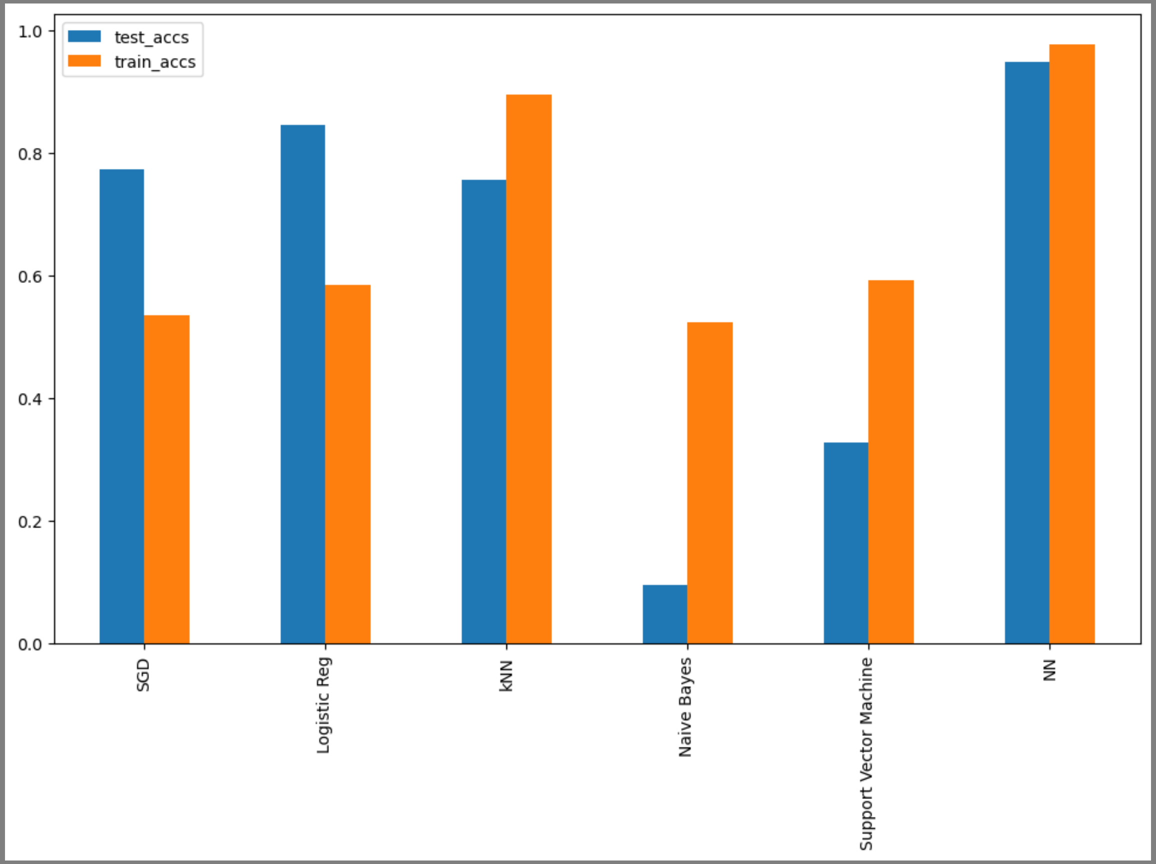
RESULTS

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

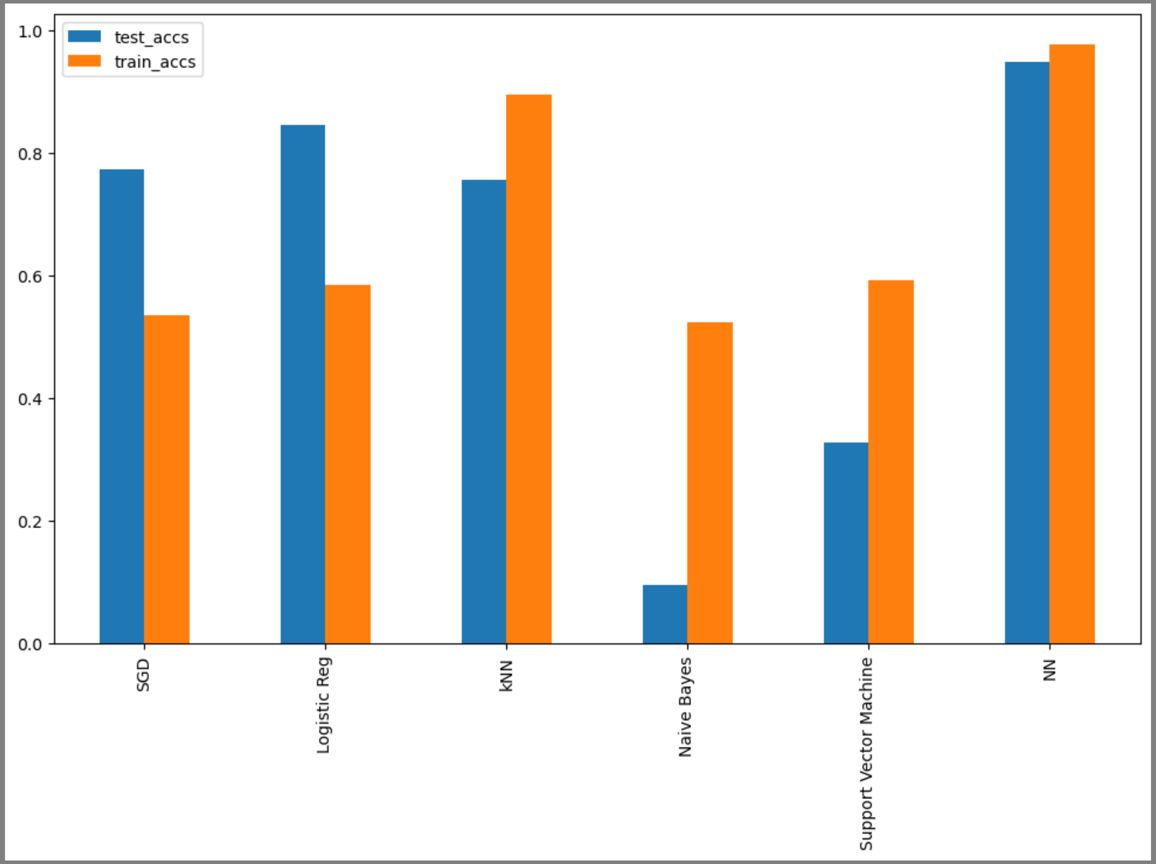
With Multi Model evaluation I were able to find out which model provide better accuracy and results but due to shortage of authentic data (Bankscope data vs open source data) not able to confirm our results

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

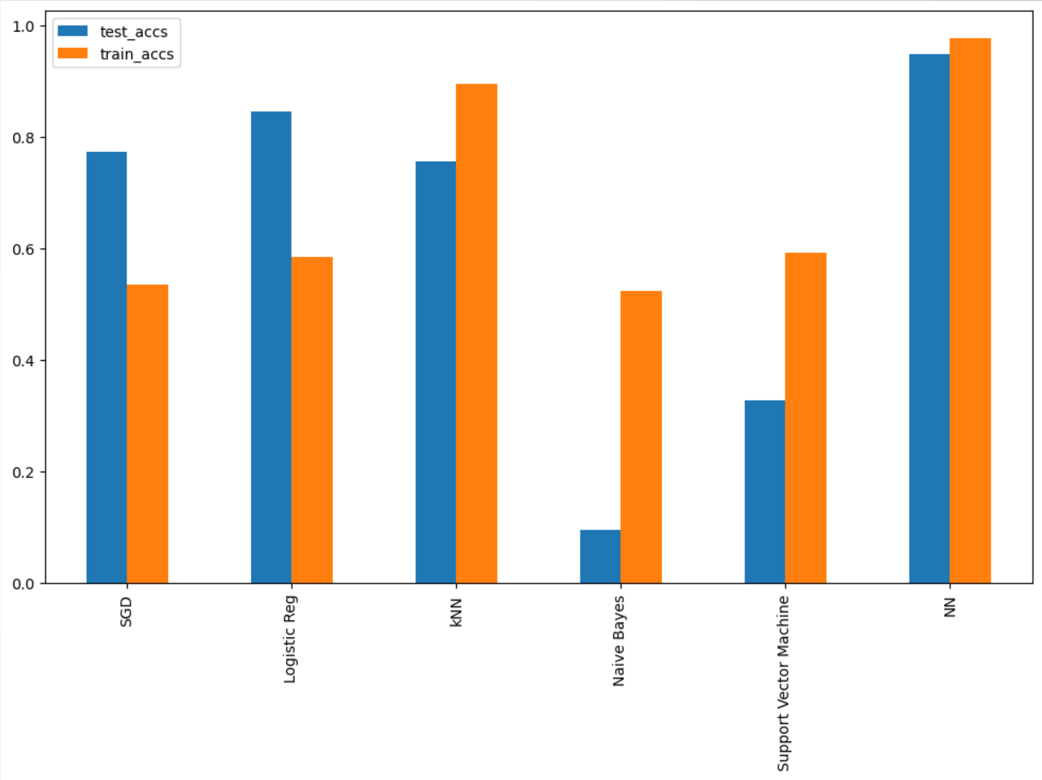
Results and Metrics Evaluations

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 fromDemocritus 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.