



Anomaly Detection Challenge

CHALLENGE 1:

Cotton Soil Classification in Satelrite Images

Team: Abracadata

Ishmeet Kaur(03677735)

Mustika Rizki Fitriyanti(03667399)

A decorative network diagram in the top-left corner, featuring a cluster of interconnected nodes. Some nodes are represented by solid grey circles, while others are concentric circles with a grey outline and a white center. These nodes are connected by thin grey lines, forming a complex web structure.

1.

Introduction about the Data

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of interconnected nodes, including solid grey circles and concentric circles with grey outlines, connected by thin grey lines.

I. Introduction

- ◎ The data that is a sub-area of a scene that consist with 82 x 100 pixels.

Each line contains the pixel values in the four spectral bands of each of the 9 pixels in the 3x3 neighbourhood

Training Data

Number Features	36
Number Rows	4435
Classify as Cotton Soil (Y=1)	479
Classify as Normal Soil (Y=0)	3956

Testing Data

Number Features	36
Number Rows	2000

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The diagram is partially cut off by the top and left edges of the slide.

2.

Data Preprocessing

A decorative network diagram in the bottom-right corner, similar to the one in the top-left. It shows a cluster of interconnected nodes and lines, with nodes represented by circles of different sizes and some having concentric rings. The lines are thin and grey. The diagram is partially cut off by the bottom and right edges of the slide.

2.1 Arranging the Data

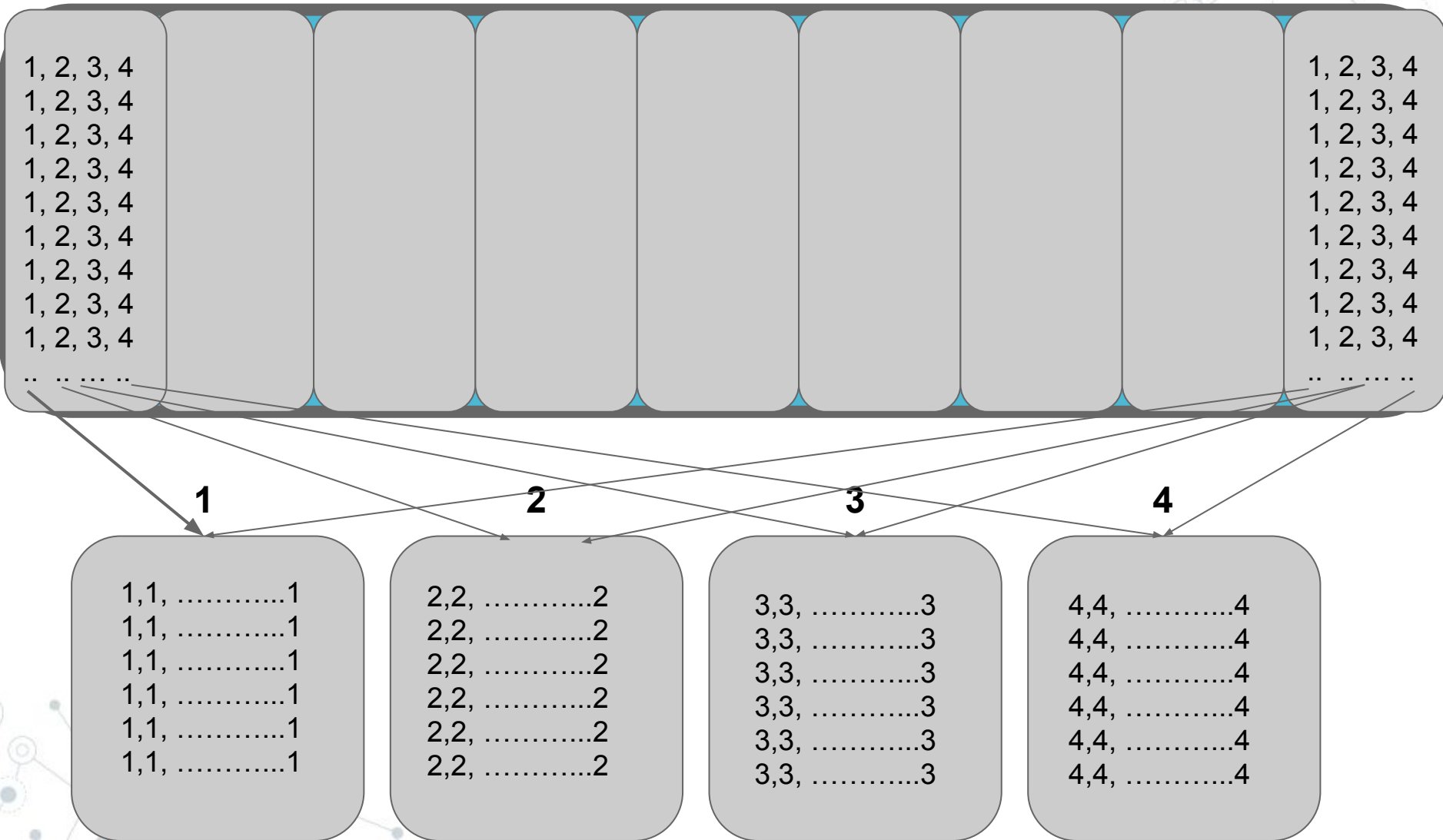
The diagram illustrates the process of arranging data into a 4x4 grid. The top section shows a large container with 16 rows of data, each containing a sequence of 4 numbers (1, 2, 3, 4). The bottom section shows four smaller containers labeled 1, 2, 3, and 4, each containing 8 rows of data. Arrows indicate the mapping of data from the top container to the bottom containers.

Top Container Data (16 rows):

- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4
- 1, 2, 3, 4

Bottom Containers Data (8 rows each):

- Container 1:** 1, 1, 1
- Container 2:** 2, 2, 2
- Container 3:** 3, 3, 3
- Container 4:** 4, 4, 4



2.2 Handling Missing Values

Replacement Method	Description
Mean	Replacement of the missing values was done both row and column wise.
Median	The missing values were replaced by the median value both along the row and column.
Mode	The missing values were replaced by the most frequent occurring element in both row and column wise.
Min/Max	Minimum and maximum values for both row and column were used to replace missing values.
Interpolation	The missing values were estimated by interpolation.
Per Spectral Value	The mean , median and mode values for every spectral of the image were used to replace the missing values per spectral.



3.

Creation and Evaluation of the Models

Naive Bayes

Decision Trees

K-Nearest Neighbours

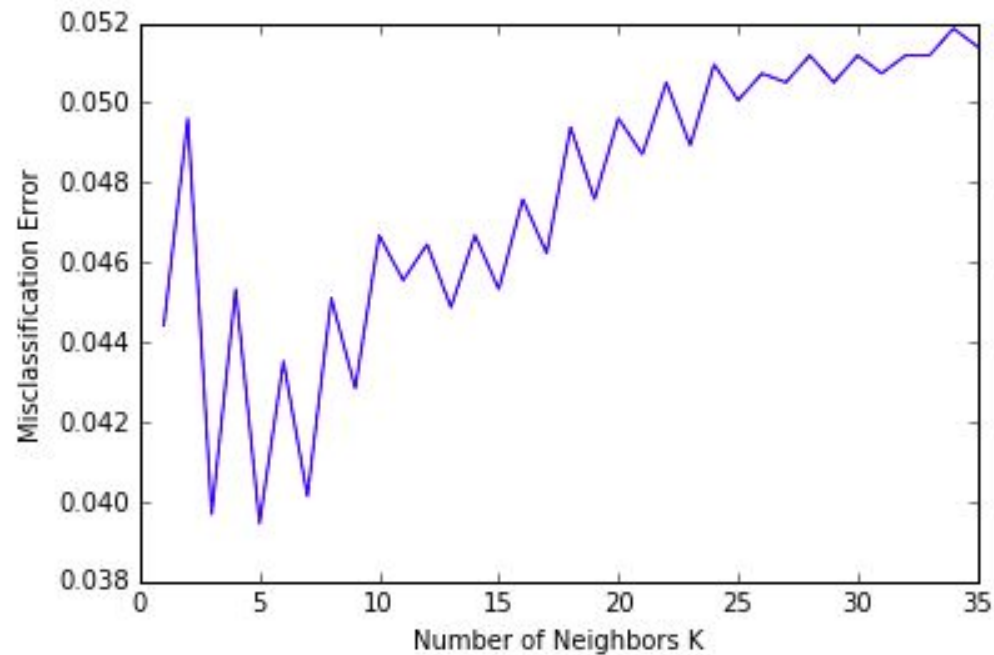
LogisticRegression

Random Forest

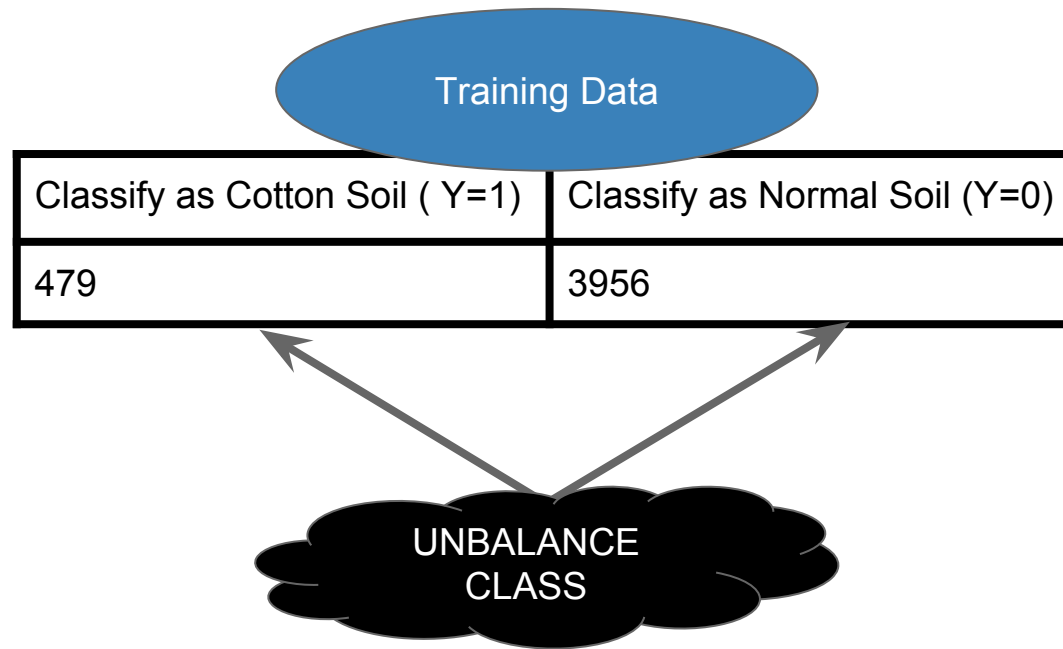


3.2 Result 1: Implementation of Classification Algorithms

Calculating the misclassification error, we found that the best accuracy for K-Nearest Neighbors could be achieved when the $k=5$.

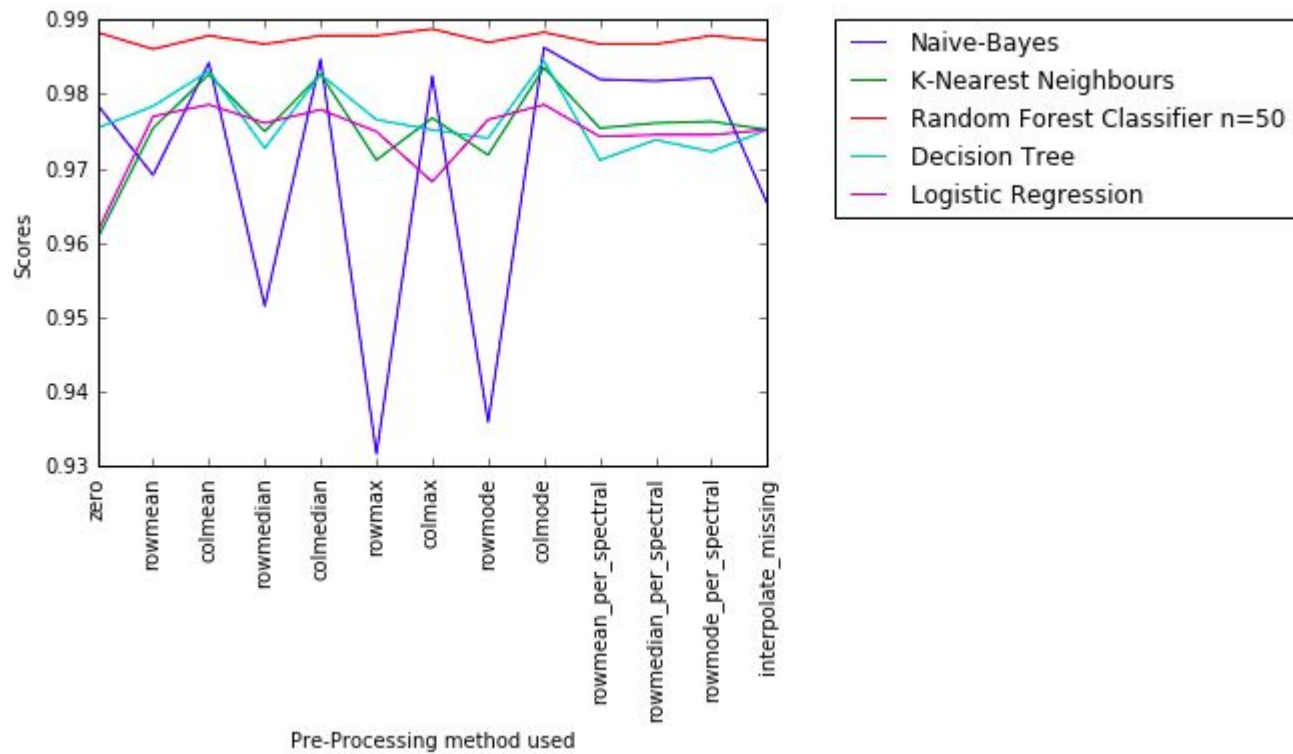


3.1 Using Stratified K-Cross Validation



Stratification will ensure that the percentages of each class in the entire data will be the same or close within each individual fold.

3.2 Result 1: Implementation of Classification Algorithms



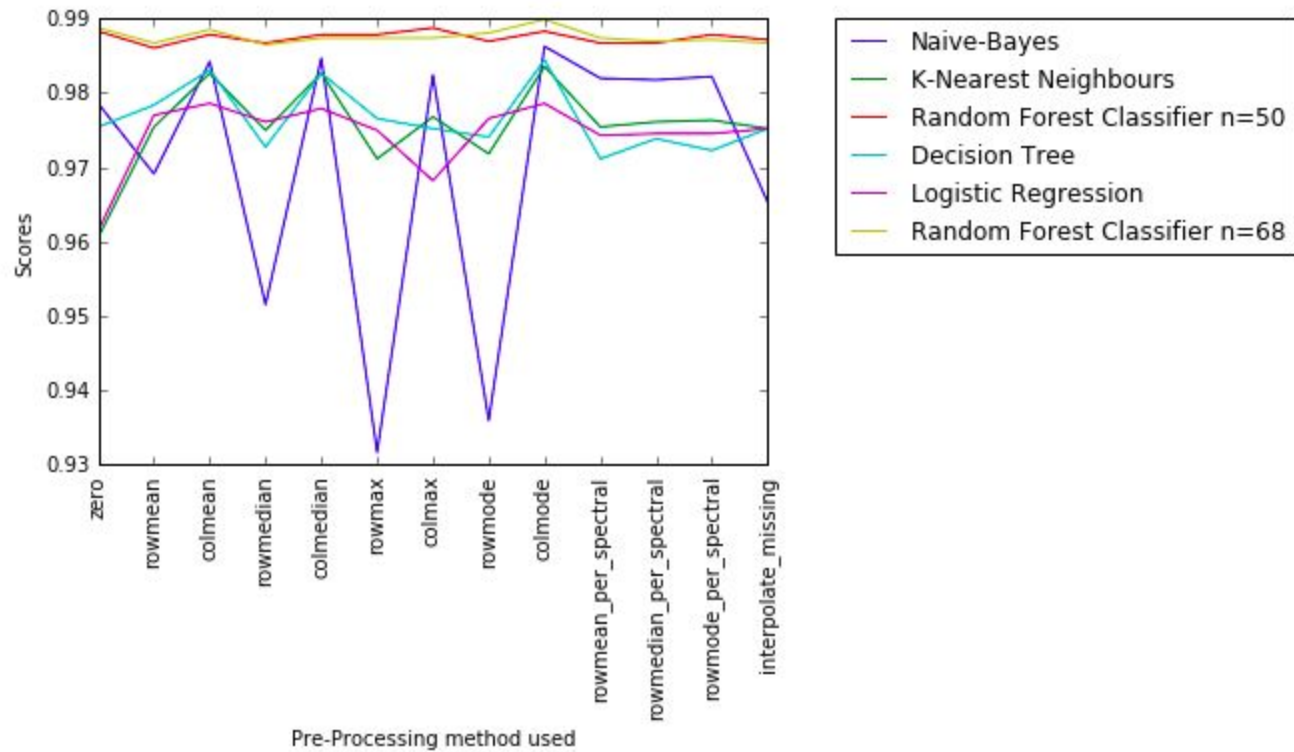
A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are larger and have concentric circles, suggesting different levels of connectivity or importance. The lines are thin and gray, creating a mesh-like structure.

4.

Result Summary

A decorative network diagram in the bottom-right corner, similar to the one in the top-left, showing a cluster of interconnected nodes and lines. The nodes are small circles, some with concentric circles, and the lines are thin and gray.

4.1 Final Accuracy Comparison



4.2 Kaggle Score

Method	Public Score	Private Score
Random Forest colmode n=68	0.98121	0.98953
Random Forest colmedian n=68	0.97693	0.99421
Random Forest colmean n=50	0.97693	0.99009