Contents

[Dataset Statistics 2](#_Toc58356136)

[Accuracy Score: Overview 3](#_Toc58356137)

[<LR: Accuracy Score> 4](#_Toc58356138)

[<SVM: Accuracy Score> 6](#_Toc58356139)

[Confusion Matrix-LR: Overview 8](#_Toc58356140)

[<Confusion Matrix-LR: Original VS New> 9](#_Toc58356141)

[Confusion Matrix-SVM: Overview 11](#_Toc58356142)

[<Confusion Matrix-SVM: Original VS New> 12](#_Toc58356143)

# Dataset Statistics

Original Dataset (0: 700, 1:300)

Bootstrap 100 Dataset (0:967, 1:33)

Bootstrap 1,000 Dataset (0:958, 1:42)

Bootstrap 5,000 Dataset (0:965, 1:35)

Bootstrap 10,000 Dataset (0:965, 1:35)

Bootstrap 10,000 & 0.5 from the Original Dataset (0:472, 1:28)

Random Oversampling Dataset (0:700, 1:700)

Random Under-Sampling Dataset (0:300, 1:300)

Adding Gaussian Noise on the original Dataset (0:700, 1:300)

# Accuracy Score: Overview

Accuracy score is based on the original test set

|  |  |  |
| --- | --- | --- |
| Accuracy Score | LR | SVM |
| Original Dataset | 76% | 76.5% |

|  |  |  |
| --- | --- | --- |
| Bootstrap 100 iterations | 44% | 34.5% |
| Bootstrap 1000 iterations | 39.5% | 34.5% |
| Bootstrap 5000 iterations | 42.5% | 34.5% |
| Bootstrap 10,000 iterations | 54.5% | 34.5% |
| Bootstrap 10,000 iterations & 0.5 of Original Dataset | 51% | 34.5% |
| Random Oversampling from the Original Dataset | 76.5% | 79% |
| Random Under-sampling from the Original Dataset | 74.5% | 77.5% |
| Gaussian Noise on numerical variables | 77% | 79.5% |
| Mean Accuracy Score | 57.44% | 51.06% |
| SD Accuracy Score | 16.09% | 22.87% |

# <LR: Accuracy Score>

|  |  |
| --- | --- |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Bootstrap 100 iterations | 44% |
| Mean | 60% |
| SD | 22.63% |

|  |  |
| --- | --- |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Bootstrap 1000 iterations | 39.5% |
| Mean | 57.75% |
| SD | 25.81% |

|  |  |
| --- | --- |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Bootstrap 5000 iterations | 42.5% |
| Mean | 59.25% |
| SD | 23.69% |

|  |  |
| --- | --- |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Bootstrap 10,00 iterations | 54.5% |
| Mean | 65.25% |
| SD | 15.20% |

|  |  |
| --- | --- |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Bootstrap 10,000 iterations & 0.5 of Original Dataset | 51% |
| Mean | 63.50% |
| SD | 17.68% |

|  |  |
| --- | --- |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Random Oversampling from the Original Dataset | 76.5% |
| Mean | 76.25% |
| SD | 0.35% |

|  |  |
| --- | --- |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Random Under-sampling from the Original Dataset | 74.5% |
| Mean | 75.25% |
| SD | 1.06% |
| Accuracy Score | LR |
| Original Dataset | 76% |
| Gaussian Noise on numerical variables | 77% |
| Mean | 76.50% |
| SD | 0.71% |

# <SVM: Accuracy Score>

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Bootstrap 100 iterations | 34.5% |
| Mean | 55.5% |
| SD | 29.7% |

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Bootstrap 1000 iterations | 34.5% |
| Mean | 55.5% |
| SD | 29.7% |

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Bootstrap 5000 iterations | 34.5% |
| Mean | 55.5% |
| SD | 29.7% |

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Bootstrap 10,00 iterations | 34.5% |
| Mean | 55.5% |
| SD | 29.7% |

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Bootstrap 10,000 iterations & 0.5 of Original Dataset | 34.5% |
| Mean | 55.5% |
| SD | 29.7% |

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Random Oversampling from the Original Dataset | 79% |
| Mean | 77.75% |
| SD | 1.77% |

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Random Under-sampling from the Original Dataset | 77.5% |
| Mean | 77% |
| SD | 0.71% |

|  |  |
| --- | --- |
| Accuracy Score | SVM |
| Original Dataset | 76.5% |
| Gaussian Noise on numerical variables | 79.5% |
| Mean | 78% |
| SD | 2.12% |

# Confusion Matrix-LR: Overview

Confusion Matrix is based on the original test set

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bootstrap 100 iterations | 44 | 44 | 94 | 18 |
| Bootstrap 1000 iterations | 54 | 25 | 113 | 8 |
| Bootstrap 5000 iterations | 40 | 45 | 93 | 22 |
| Bootstrap 10,000 iterations | 22 | 87 | 51 | 40 |
| Bootstrap 10,000 iterations & 0.5 of Original Dataset | 30 | 72 | 66 | 32 |
| Random Oversampling from the Original Dataset | 47 | 106 | 32 | 15 |
| Random Under-sampling from the Original Dataset | 47 | 102 | 36 | 15 |
| Gaussian Noise on numerical variables | 20 | 134 | 4 | 42 |
| Mean | 38.00 | 76.88 | 61.13 | 24.00 |
| SD | 12.55 | 37.14 | 37.14 | 12.55 |

# <Confusion Matrix-LR: Original VS New>

Original Dataset & Bootstrap 100 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Bootstrap 100 iterations | 44 | 44 | 94 | 18 |
| Mean | 38 | 82 | 56 | 24 |
| SD | 8.49 | 53.74 | 53.74 | 8.49 |

Original Dataset & Bootstrap 1000 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Bootstrap 1000 iterations | 54 | 25 | 113 | 8 |
| Mean | 43 | 72.50 | 65.50 | 19.00 |
| SD | 15.66 | 67.18 | 67.18 | 15.66 |

Original Dataset & Bootstrap 5000 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Bootstrap 5000 iterations | 40 | 45 | 93 | 22 |
| Mean | 36.00 | 82.50 | 55.50 | 26.00 |
| SD | 5.66 | 53.03 | 53.03 | 5.66 |

Original Dataset & Bootstrap 10,000 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Bootstrap 10,000 iterations | 22 | 87 | 51 | 40 |
| Mean | 27.00 | 103.50 | 34.50 | 35.00 |
| SD | 7.07 | 23.33 | 23.33 | 7.07 |

Original Dataset & Bootstrap 10,000 iterations & 0.5 of Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Bootstrap 10,000 iterations & 0.5 of Original Dataset | 30 | 72 | 66 | 32 |
| Mean | 31.00 | 96.00 | 42.00 | 31.00 |
| SD | 1.41 | 33.94 | 33.94 | 1.41 |

Original Dataset & Random Oversampling from the Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Random Oversampling from the Original Dataset | 47 | 102 | 36 | 15 |
| Mean | 39.50 | 111.00 | 27.00 | 22.50 |
| SD | 10.61 | 12.73 | 12.73 | 10.61 |

Original Dataset & Random Under-sampling from the Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Random Under-sampling from the Original Dataset | 47 | 102 | 36 | 15 |
| Mean | 39.50 | 111.00 | 27.00 | 22.50 |
| SD | 10.61 | 12.73 | 12.73 | 10.61 |

Original Dataset & Gaussian Noise from the Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - LR | TP | TN | FP | FN |
| Original Dataset | 32 | 120 | 18 | 30 |
| Gaussian Noise on numerical variables | 20 | 132 | 6 | 42 |
| Mean | 26.00 | 126.00 | 12.00 | 36.00 |
| SD | 8.49 | 8.49 | 8.49 | 8.49 |

# Confusion Matrix-SVM: Overview

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Bootstrap 100 iterations | 62 | 7 | 131 | 0 |
| Bootstrap 1000 iterations | 62 | 7 | 131 | 0 |
| Bootstrap 5000 iterations | 62 | 7 | 131 | 0 |
| Bootstrap 10,000 iterations | 62 | 7 | 131 | 0 |
| Bootstrap 10,000 iterations & 0.5 of Original Dataset | 62 | 7 | 131 | 0 |
| Random Oversampling from the Original Dataset | 48 | 110 | 28 | 14 |
| Random Under-sampling from the Original Dataset | 47 | 102 | 36 | 15 |
| Gaussian Noise on numerical variables | 26 | 133 | 5 | 36 |
| Mean | 54.25 | 45.13 | 92.88 | 7.75 |
| SD | 13.38 | 53.47 | 53.47 | 13.38 |

# <Confusion Matrix-SVM: Original VS New>

Original Dataset & Bootstrap 100 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Bootstrap 100 iterations | 62 | 7 | 131 | 0 |
| Mean | 47 | 64 | 74 | 15 |
| SD | 21.21 | 80.61 | 80.61 | 21.21 |

Original Dataset & Bootstrap 1000 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Bootstrap 1000 iterations | 62 | 7 | 131 | 0 |
| Mean | 47 | 64 | 74 | 15 |
| SD | 21.21 | 80.61 | 80.61 | 21.21 |

Original Dataset & Bootstrap 5000 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Bootstrap 1000 iterations | 62 | 7 | 131 | 0 |
| Mean | 47 | 64 | 74 | 15 |
| SD | 21.21 | 80.61 | 80.61 | 21.21 |

Original Dataset & Bootstrap 10,000 iterations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Bootstrap 5000 iterations | 62 | 7 | 131 | 0 |
| Mean | 47 | 64 | 74 | 15 |
| SD | 21.21 | 80.61 | 80.61 | 21.21 |

Original Dataset & Bootstrap 10,000 iterations & 0.5 of Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Bootstrap 10,000 iterations & 0.5 of Original Dataset | 62 | 7 | 131 | 0 |
| Mean | 47 | 64 | 74 | 15 |
| SD | 21.21 | 80.61 | 80.61 | 21.21 |

Original Dataset & Random Oversampling from the Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Random Oversampling from the Original Dataset | 48 | 110 | 28 | 14 |
| Mean | 40.00 | 115.50 | 22.50 | 22.00 |
| SD | 11.31 | 7.78 | 7.78 | 11.31 |

Original Dataset & Random Under-sampling from the Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Random Under-sampling from the Original Dataset | 47 | 102 | 36 | 15 |
| Mean | 39.50 | 111.50 | 26.50 | 22.50 |
| SD | 10.61 | 13.44 | 13.44 | 10.61 |

Original Dataset & Gaussian Noise from the Original Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Confusion Matrix - SVM | TP | TN | FP | FN |
| Original Dataset | 32 | 121 | 17 | 30 |
| Gaussian Noise on numerical variables | 26 | 133 | 5 | 36 |
| Mean | 29.00 | 127.00 | 11.00 | 33.00 |
| SD | 4.24 | 8.49 | 8.49 | 4.24 |