# Test Results

|  |  |
| --- | --- |
| **K-NN (Log)** | Values |
| Accuracy (%) | 99.602% |
| Confusion Matrix | [182163 308]  [ 488 17041] |

|  |  |
| --- | --- |
| **K-NN (Scaled)** | Values |
| Accuracy (%) | 99.872% |
| Confusion Matrix | [182240 102]  [ 154 17504] |

|  |  |
| --- | --- |
| **Naïve Bayes (Scaled)** | Values |
| Accuracy (%) | 94.8155% |
| Confusion Matrix | [179517 3010]  [ 7359 10114] |

|  |  |
| --- | --- |
| **Naïve Bayes (Log)** | Values |
| Accuracy (%) | 85.069% |
| Confusion Matrix | [154960 27692]  [ 2170 15178] |

|  |  |
| --- | --- |
| **SVC (Scaled) 50% Size - RBF Kernel** | Values |
| Accuracy (%) | 99.772% |
| Confusion Matrix | [91210 66]  [ 162 8562] |

|  |  |
| --- | --- |
| **SVC (Log) 50% Size RBF Kernel** | Values |
| Accuracy (%) | 99.272% |
| Confusion Matrix | [91124 233]  [ 495 8148] |

|  |  |
| --- | --- |
| **SVC (Scaled) 20% Size - Linear Kernel** | Values |
| Accuracy (%) | 96.2025% |
| Confusion Matrix | [36189 272]  [ 1247 2292] |

|  |  |
| --- | --- |
| **SVC (Log) 20% Size - Linear Kernel** | Values |
| Accuracy (%) | 95.385% |
| Confusion Matrix | [36182 316]  [ 1530 1972] |

|  |  |
| --- | --- |
| **SVC (Scaled) 20% Size - Poly Kernel** | Values |
| Accuracy (%) | 99.3425% |
| Confusion Matrix | [36458 45]  [ 218 3279] |

|  |  |
| --- | --- |
| **SVC (Log) 20% Size - Poly Kernel** | Values |
| Accuracy (%) | 98.2075% |
| Confusion Matrix | [36303 160]  [ 557 2980] |

Notes:

* K-NN (Scaled) is the best for accurately predicting fraud, and due to low processing time can be trained against the entire data set
* Log transformed preprocessing negatively effects all models, with the exception of Naïve Bayes, where it improves the ability to predict fraud
* Out of the SVC kernels tested, the RBF kernel was the best at predicting fraud at the lowest processing time cost.
* SVC models took significantly longer to train then Naïve Bayes or K-NN, thus either 50% or 20% of the dataset was used for training.