**CHAPTER 5**

**RESULT AND EVALUATION**

**5.1. SIMULATION SCENARIO**

The proposed model has been designed for the classification of the kidney based health examination data from nearly 400 patients in the given dataset. This data contains the various parameters, which includes blood pressure, age, red blood cell count, white blood cell count, etc. In this thesis, the SVM and KNN classifiers are applied to the dataset in order to obtain the results.

1. A document or an individual entity is broken on the basis of quantitative and qualitative variables.
2. The quantitative and qualitative variables are handled differently in order to create the balanced version of these variables.
3. Finally, both of the feature matrixes, both obtained from quantitative and qualitative, are combined together to create the feature matrix.

Afterwards, the data is divided into training and testing dataset, which is done using the random selection by creating the random number series. The cross validation split works on the 75:25 ratio, which divided the 400 samples into 300 for training and 100 for testing.

6.2. **PERFORMANCE PARAMETERS**

Evaluation of the system has been done using standard metrics Recall Precision and F1-Measure.

**5.2.1. Recall**

Recall is the test of the probability of the accuracy, which indicates the performance of the proposed model in the presence of the false negative cases. The false negative cases depict the falsely detected case from the data entries. In recall, the accuracy of the proposed model has been analyzed in the presence of false negative cases:

(1)

**5.2.2. Precision**

The precision depicts the accuracy of the model in the presence of the false positive cases. The accuracy of the model depicts the overall impact of the false positive cases, which rejects positive cases. A positive case in our case is when the data entry contains the certain set of parameters from one of the registered category, but returns the false result for such entries.

(2)

**5.2.3. F1-Measure**

**The F1-Measure is the cumulative parameter to assess the overall impact of the precision and recall in the case to study the overall impact of the false positive and false negative cases over the overall accuracy assessed from the preliminary statistical parameters. The F1-score value is represented in the range of 0 to 1 or 0 to 100, decided as per the maximum ranges of the precision and recall. The following equation is utilized to measure the F1-measure:**

Where R is recall, and p or P is precision.

**5.2.4. Accuracy**

The overall accuracy is the analysis of the proposed model in the terms of overall accuracy, which is computed by dividing the total number of true cases (including true negative and true positive), by all of the cases.

(4)

**5.3. RESULTS OF KNN CLASSIFICATION**

The KNN classifier has been analyzed for the various performance parameters in the given scenario under the 25 rotations. The results are obtained for the statistical type 1 and type 2 errors, which have been shown in the following table. The type 1 and type 2 features analyzed under this analysis involves the true positive, true negative, false positive and false negative cases.

**Table 5.1: Statistical error based analysis of KNN**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | True Positive | True Negative | False Positive | False Negative |
| 1 | 59 | 37 | 3 | 1 |
| 2 | 67 | 30 | 2 | 1 |
| 3 | 63 | 35 | 2 | 0 |
| 4 | 67 | 29 | 3 | 1 |
| 5 | 62 | 37 | 1 | 0 |
| 6 | 58 | 39 | 3 | 0 |
| 7 | 62 | 38 | 0 | 0 |
| 8 | 54 | 46 | 0 | 0 |
| 9 | 56 | 40 | 3 | 1 |
| 10 | 60 | 34 | 4 | 2 |
| 11 | 55 | 40 | 4 | 1 |
| 12 | 62 | 37 | 1 | 0 |
| 13 | 52 | 42 | 5 | 1 |
| 14 | 60 | 36 | 2 | 2 |
| 15 | 56 | 40 | 1 | 3 |
| 16 | 57 | 39 | 2 | 2 |
| 17 | 55 | 44 | 1 | 0 |
| 18 | 63 | 35 | 1 | 1 |
| 19 | 58 | 38 | 1 | 3 |
| 20 | 59 | 38 | 0 | 3 |
| 21 | 55 | 42 | 3 | 0 |
| 22 | 64 | 34 | 1 | 1 |
| 23 | 66 | 32 | 2 | 0 |
| 24 | 62 | 35 | 1 | 2 |
| 25 | 64 | 32 | 1 | 3 |

**Figure 5.1: Statistical error based analysis of KNN**

The results of KNN are obtained in the form of various performance parameters as per shown in the following tables. The statistical analysis includes the parameters of accuracy, precision, recall and f1 error.

**Table 5.2: Statistical parameter based analysis of KNN**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | Accuracy | Precision | Recall | F1 Measure |
| 1 | 96 | 95.16129 | 98.33333 | 96.72131 |
| 2 | 97 | 97.10145 | 98.52941 | 97.81022 |
| 3 | 98 | 96.92308 | 100 | 98.4375 |
| 4 | 96 | 95.71429 | 98.52941 | 97.10145 |
| 5 | 99 | 98.4127 | 100 | 99.2 |
| 6 | 97 | 95.08197 | 100 | 97.47899 |
| 7 | 100 | 100 | 100 | 100 |
| 8 | 100 | 100 | 100 | 100 |
| 9 | 96 | 94.91525 | 98.24561 | 96.55172 |
| 10 | 94 | 93.75 | 96.77419 | 95.2381 |
| 11 | 95 | 93.22034 | 98.21429 | 95.65217 |
| 12 | 99 | 98.4127 | 100 | 99.2 |
| 13 | 94 | 91.22807 | 98.11321 | 94.54545 |
| 14 | 96 | 96.77419 | 96.77419 | 96.77419 |
| 15 | 96 | 98.24561 | 94.91525 | 96.55172 |
| 16 | 96 | 96.61017 | 96.61017 | 96.61017 |
| 17 | 99 | 98.21429 | 100 | 99.0991 |
| 18 | 98 | 98.4375 | 98.4375 | 98.4375 |
| 19 | 96 | 98.30508 | 95.08197 | 96.66667 |
| 20 | 97 | 100 | 95.16129 | 97.52066 |
| 21 | 97 | 94.82759 | 100 | 97.34513 |
| 22 | 98 | 98.46154 | 98.46154 | 98.46154 |
| 23 | 98 | 97.05882 | 100 | 98.50746 |
| 24 | 97 | 98.4127 | 96.875 | 97.6378 |
| 25 | 96 | 98.46154 | 95.52239 | 96.9697 |

**Figure 5.2: Statistical parameter based analysis of KNN**

The overall accuracy and analysis based comparison has been performed in the following table, which makes it easier to understand the total deficiency on the scale of 0 to 100. Hence, the accuracy based analysis has been conducted between accuracy and error only in this table.

**Table 5.3: Overall Accuracy & Error based analysis for KNN**

|  |  |  |
| --- | --- | --- |
| Index | Accuracy | Error |
| 1 | 96 | 4 |
| 2 | 97 | 3 |
| 3 | 98 | 2 |
| 4 | 96 | 4 |
| 5 | 99 | 1 |
| 6 | 97 | 3 |
| 7 | 100 | 0 |
| 8 | 100 | 0 |
| 9 | 96 | 4 |
| 10 | 94 | 6 |
| 11 | 95 | 5 |
| 12 | 99 | 1 |
| 13 | 94 | 6 |
| 14 | 96 | 4 |
| 15 | 96 | 4 |
| 16 | 96 | 4 |
| 17 | 99 | 1 |
| 18 | 98 | 2 |
| 19 | 96 | 4 |
| 20 | 97 | 3 |
| 21 | 97 | 3 |
| 22 | 98 | 2 |
| 23 | 98 | 2 |
| 24 | 97 | 3 |
| 25 | 96 | 4 |

**5.4. RESULTS OF SVM CLASSIFICATION**

The SVM classifier has been analyzed for the various performance parameters in the given scenario under the 25 rotations. The results are obtained for the statistical type 1 and type 2 errors, which have been shown in the following table. The type 1 and type 2 features analyzed under this analysis involves the true positive, true negative, false positive and false negative cases.

**Table 5.4: Statistical error based analysis of SVM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | True Positive | True Negative | False Positive | False Negative |
| 1 | 57 | 42 | 1 | 0 |
| 2 | 64 | 36 | 0 | 0 |
| 3 | 65 | 34 | 1 | 0 |
| 4 | 64 | 34 | 2 | 0 |
| 5 | 58 | 41 | 1 | 0 |
| 6 | 67 | 29 | 4 | 0 |
| 7 | 61 | 38 | 0 | 1 |
| 8 | 60 | 40 | 0 | 0 |
| 9 | 57 | 43 | 0 | 0 |
| 10 | 63 | 32 | 5 | 0 |
| 11 | 59 | 41 | 0 | 0 |
| 12 | 66 | 34 | 0 | 0 |
| 13 | 56 | 44 | 0 | 0 |
| 14 | 66 | 33 | 1 | 0 |
| 15 | 66 | 33 | 1 | 0 |
| 16 | 59 | 39 | 2 | 0 |
| 17 | 66 | 34 | 0 | 0 |
| 18 | 54 | 45 | 1 | 0 |
| 19 | 53 | 45 | 2 | 0 |
| 20 | 57 | 42 | 1 | 0 |
| 21 | 59 | 40 | 1 | 0 |
| 22 | 56 | 44 | 0 | 0 |
| 23 | 66 | 33 | 1 | 0 |
| 24 | 63 | 35 | 2 | 0 |
| 25 | 56 | 44 | 0 | 0 |

**Figure 5.3: Statistical error based analysis of SVM**

The results of SVM are obtained in the form of various performance parameters as per shown in the following tables. The statistical analysis includes the parameters of accuracy, precision, recall and f1 error.

**Table 5.5: Statistical parameter based analysis of SVM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | Accuracy | Precision | Recall | F1 Measure |
| 1 | 99 | 98.27586 | 100 | 99.13043 |
| 2 | 100 | 100 | 100 | 100 |
| 3 | 99 | 98.48485 | 100 | 99.23664 |
| 4 | 98 | 96.9697 | 100 | 98.46154 |
| 5 | 99 | 98.30508 | 100 | 99.1453 |
| 6 | 96 | 94.3662 | 100 | 97.10145 |
| 7 | 99 | 100 | 98.3871 | 99.18699 |
| 8 | 100 | 100 | 100 | 100 |
| 9 | 100 | 100 | 100 | 100 |
| 10 | 95 | 92.64706 | 100 | 96.18321 |
| 11 | 100 | 100 | 100 | 100 |
| 12 | 100 | 100 | 100 | 100 |
| 13 | 100 | 100 | 100 | 100 |
| 14 | 99 | 98.50746 | 100 | 99.24812 |
| 15 | 99 | 98.50746 | 100 | 99.24812 |
| 16 | 98 | 96.72131 | 100 | 98.33333 |
| 17 | 100 | 100 | 100 | 100 |
| 18 | 99 | 98.18182 | 100 | 99.08257 |
| 19 | 98 | 96.36364 | 100 | 98.14815 |
| 20 | 99 | 98.27586 | 100 | 99.13043 |
| 21 | 99 | 98.33333 | 100 | 99.15966 |
| 22 | 100 | 100 | 100 | 100 |
| 23 | 99 | 98.50746 | 100 | 99.24812 |
| 24 | 98 | 96.92308 | 100 | 98.4375 |
| 25 | 100 | 100 | 100 | 100 |

**Figure 5.4: Statistical parameter based analysis of SVM**

For the SVM classification, the overall accuracy and analysis based comparison has been performed in the following table, which makes it easier to understand the total deficiency on the scale of 0 to 100. Hence, the accuracy based analysis has been conducted between accuracy and error only in this table.

**Table 5.6: Overall Accuracy & Error based analysis for SVM**

|  |  |  |
| --- | --- | --- |
| Index | Accuracy | Error |
| 1 | 99 | 1 |
| 2 | 100 | 0 |
| 3 | 99 | 1 |
| 4 | 98 | 2 |
| 5 | 99 | 1 |
| 6 | 96 | 4 |
| 7 | 99 | 1 |
| 8 | 100 | 0 |
| 9 | 100 | 0 |
| 10 | 95 | 5 |
| 11 | 100 | 0 |
| 12 | 100 | 0 |
| 13 | 100 | 0 |
| 14 | 99 | 1 |
| 15 | 99 | 1 |
| 16 | 98 | 2 |
| 17 | 100 | 0 |
| 18 | 99 | 1 |
| 19 | 98 | 2 |
| 20 | 99 | 1 |
| 21 | 99 | 1 |
| 22 | 100 | 0 |
| 23 | 99 | 1 |
| 24 | 98 | 2 |
| 25 | 100 | 0 |

**5.5. COMPARATIVE ANALYSIS**

The comparison between the classification accuracy, precision, recall, f1 error and statistical errors has been conducted in this section. The comparison includes the average, standard deviation, median, minimum and maximum values. The comparative analysis is supposed to show the clear analysis, and helps to declare the best algorithm. Hence, the averaging factors play the key roles to distinguish the performance. The following table compares both classification, KNN & SVM on the basis of the true type errors from the statistical errors. The SVM is known to produce the more number of true positive on the average than KNN, although the maximum number of is true negative is higher in case of KNN than SVM.

**Table 5.7: Comparative Analysis of KNN and SVM based on type 1 parameters**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameters | SVM | | KNN | |
| TP | TN | TP | TN |
| Mean | 60.72 | 38.2 | 59.84 | 37.16 |
| Median | 60 | 39 | 60 | 37 |
| Max | 67 | 45 | 67 | 46 |
| Min | 53 | 29 | 52 | 29 |

**Figure 5.5: Comparative Analysis of KNN and SVM based on type 1 parameters**

The SVM and KNN has been compared on the basis of false type parameters, which includes the false positive and false negative parameters. The following table shows the higher performance of SVM on the basis of average false positive cases, which is 1.04 in comparison to 1.88 for KNN. This pattern is identical with lower difference on the basis of false negative.

**Table 5.8: Comparative Analysis of KNN and SVM based on type 2 parameters**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameters | SVM | | KNN | |
| FP | FN | FP | FN |
| Mean | 1.04 | 0.04 | 1.88 | 1.12 |
| Median | 1 | 0 | 2 | 1 |
| Max | 5 | 1 | 5 | 3 |
| Min | 0 | 0 | 0 | 0 |

**Figure 5.6: Comparative Analysis of KNN and SVM based on type 2 parameters**

The SVM shows the higher performance than KNN on the basis of accuracy, precision and recall based parameters. The following table shows the higher mean for all accuracy, precision and recall than KNN, which clearly signifies the higher performance.

**Table 5.9: Comparative Analysis of KNN and SVM based on Accuracy, Precision and Recall**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parameters | SVM | | | KNN | | |
| Accuracy | Precision | Recall | Accuracy | Precision | Recall |
| Mean | 98.92 | 98.37481 | 99.93548 | 97 | 96.94921 | 98.18315 |
| Median | 99 | 98.50746 | 100 | 97 | 97.10145 | 98.4375 |
| Max | 100 | 100 | 100 | 100 | 100 | 100 |
| Min | 95 | 92.64706 | 98.3871 | 94 | 91.22807 | 94.91525 |

**Figure 5.7: Comparative Analysis of KNN and SVM based on Accuracy, Precision and Recall**

Similarly, the SVM shows the higher performance than KNN on the basis of F1 error and overall error based parameters. The following table shows the lower mean for both F1 error and overall error than KNN, which clearly signifies the higher performance.

**Table 5.10: Comparative Analysis of KNN and SVM based on F1 error and overall error**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameters | SVM | | KNN | |
| F1 Measure | Overall Error | F1 Measure | Overall Error |
| Mean | 99.13926 | 1.08 | 97.54074 | 3 |
| Median | 99.23664 | 1 | 97.47899 | 3 |
| Max | 100 | 5 | 100 | 6 |
| Min | 96.18321 | 0 | 94.54545 | 0 |

**Figure 5.8: Comparative Analysis of KNN and SVM based on F1 error and overall error**

**CHAPTER 6**

**CONCLUSION**

**6.1. CONCLUSION**

The healthcare systems are in the development all across the globe to centralize the health records of all citizens in order to facilitate the flexible provision to all health facilities to the historical health records whenever required. These systems are known to integrate the data of all of the possible diseases together on one platform. These systems can be also used to predict the chances of appearance of disease on the basis of the healthcare analysis done periodically specifically in the developed nations. In this thesis, the work has been performed on the chronic kidney diseases. The chronic kidney diseases are known to claim a number of lives every year across the globe. Hence it becomes important to implement such system, which can predict the chances of developing the disease in person specific. The support vector machine (SVM) and k-nearest neighbor (KNN) based models are used to predict the chronic kidney diseases. The dataset includes 35 features, out of which some of them are quantitative (both continuous and discrete) and other all categorical (qualitative). Different modules are developed to handle the categorical and quantitative variables in the dataset to avoid the problems related to the column dominance, execution errors, etc. The quantitative variables undergo the maximum minimum scaling, which is known to convert the data values to 0-1 scale. In this thesis, the SVM has been found better than KNN on the basis of nearly all of the parameters. The SVM has been recorded with 98.92% (mean) and 99% (median) of accuracy, which is significantly higher than KNN’s 97% (both mean and median). Also SVM outperformed KNN on the basis of precision by (98.37% mean) and recall (99.94% mean) against 96.95% (precision mean) and 98.18% (recall mean)**.**

**6.2. FUTURE WORK**

In the future, the deep learning classification can be utilized to improve the overall classification performance. Also, the optimization algorithms can be used to create the more balanced and advanced feature descriptors to obtain the higher accuracy.