Impact of Feature Selection Using Rough Set Theory

Submitted by:

Mushfiqur Rahman

ID: 2015-2-60-021

Md Mohsi Ahammad Himu

ID: 2015-2-60-031

&

Md. Afser Uddin

ID: 2015-2-60-083

In particular fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

Department of Computer Science and Engineering
East West University
Dhaka-1212, Bangladesh

December 5th, 2019

Declaration

We, hereby, declare that the work presented in this thesis is our own work and the outcome of the investigation performed by us under the supervision of our supervisor Dr. Shamim H Ripon, Associate Professor, Department of Computer Science & Engineering, East West University. We also declare with our best knowledge and belief that this work contains neither facts nor material that were previously written or published by another person. We also ensure that no part of this work has been submitted elsewhere for the award of any degree or diploma.

....

(**Dr. Shamim H Ripon**)

Supervisor

....

(**Mushfiqur Rahman)**

.

....

(**Md. Afser Uddin)**

....

(**Md Mohsi Ahammad Himu**)

ABSTRACT

It is not necessary that all feature in a dataset have the same impact on the decision class or on result. Some features have more impact than others and some may not have any impact on the result or some features may have negative impact on the result. It is easier and more efficient to work with cleaner data. But we cannot just remove any feature or data from the dataset randomly. To do so, we have used a feature extraction technique to remove unwanted feature and data from the dataset which helps to make the dataset more accurate and efficient to work with. The technique we have used removes not only the unwanted data and features but also gives us multiple options which contains a list of features with different length. We have used three datasets here which contains Breast Cancer, Hepatitis C and Mushroom information. We have measured the accuracy and the number of successfully measured decision class before applying the feature extraction. To show the significance of the feature extraction technique, we have run the feature extraction technique on them and measured the accuracy and the number of successfully measured decision class again. Then we have compared both result and get to find that after applying feature extraction technique we have a smaller number of features on our dataset and overall accuracy and result improves around 2%-5% depends on the dataset and number of features we have used.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

In the name of Allah, the most beneficent and merciful, we express our sincere gratitude towards the almighty Allah who gave us strength, patience and knowledge to complete this thesis work. After that, we would like to express our gratefulness towards our supervisor, Dr. Shamim H Ripon, Professor, Department of Computer Science & Engineering, East West University for giving us this opportunity to work into the field of machine learning and data mining. Throughout our work, he was always there to guide us and help us to improve more. He gave us moral support and guided in different matters regarding the work. Without his proper guidance, support and encouragement, we wouldn't able to complete this thesis work perfectly. His encouragements, visionaries and thoughtful comments and suggestions, unforgettable support at every stage of our B.Sc. study were simply appreciating and essential. We are also thankful to our parents who supported us mentally. Lastly we would like to thank other faculties of our department and our friends for their support and encouragement regarding our thesis work.

CHAPTER 1

INTRODUCTION

1.1 Introduction & Motivation

In machine learning classification is the main problem. It may be viewed as supervised learning process. The rules learnt from this process are used for classification. For classification data are collected or stored in a database unorganized. Also data are rarely collected for the purpose of mining knowledge in most organizations. A dataset contains a lot of attributes that are not necessary or redundant for rule discovery. If those attributes are not remove, not only the time complexity of the rule discovery process increases, but also quality of the discovered rules may be degraded [1]. Which attribute should be selected and which attribute should be removed is very difficult to choose for anyone. At this point it can be understand there should be a process that can select features that are not redundant and necessary for classification.

Feature selection is the process of selecting subset of relevant features for use in model construction. It helps to make models simple to interpret by researcher, for reducing training time, to avoid the curse of dimensionality and also reduced overfitting.

In this paper, we have proposed an algorithm which is using Rough Set Theory with Reduct for feature selection and evaluated its performance. In this approach, features are selected using rough set theory indiscernibility relation which will lead us to select reducts. Those reducts are the selected features for high performance. Machine learning algorithm like Decision Tree, Neural Network and Support Vector Machine has been used on those reducts. Than performance of every reduct and algorithm has been compared so that which reduct works well when also with which algorithm can be distinguish. Also which algorithm works well on which reducts also compare.

1.2 Objective

The main objectives of our research are as follows:

- Selecting features using a feature selection approach from multidimensional datasets.
- Getting higher accuracy from datasets after feature selection using different machine learning algorithm.
- Performance of algorithms and features has also compared for analyzing the impact of feature selection.

1.3 Contribution

Contribution in our research are as follows:

We have selected three different dataset with different dimensionality and different data size.

Our main focus was to select feature using Rough Set Theory approach with reduct because in most of the cases accuracy of a datasets depends on its features. Also time complexity to train also depends on its features. So selecting the right features Rough Set Theory has been shown much efficiency.

After getting the features from our dataset using rough set theory, we have applied machine learning algorithms- Decision Tree, Neural Network and Support Vector Machine and get the accuracy higher in most of the cases than without selecting features.

We have also compared features of every dataset along with the machine learning algorithms using confusion matrix true positive value.

1.4 Outline

Chapter 1: Chapter 1 introduces importance of feature selection, our motivation to use a feature selection approach, the main objectives of our research, the contributions that we have made regarding the feature selection approach and impact.

Chapter 2: This chapter illustrates the background of our proposed methods and the related works that have been done regarding feature selection using rough set.

Chapter 3: Chapter 3 shows the architectural view of our proposed method.

Chapter 4: This chapter gives overall overview of our dataset.

Chapter 5: This chapter analyzes the results obtained from our proposed methods.

Chapter 6: This chapter does comparative analysis of our dataset.

Chapter 7: The final chapter summarizes the overall work that we have done and also explains the future works that we need to focus on.

CHAPTER 2

LITERATURE REVIEW

2.1 Background

2.1.1 Rough Set

Rough set theory was developed by Zdzislaw Pawlak in the early 1980's. Initially rough set was developed for a finite universe of discourse in which the knowledge base is a partition, which is obtained by any equivalence relation defined on the universe of discourse and discover the hidden data pattern [3]. The main goal of the rough set analysis is induction of approximations of concepts. Rough set constitutes a sound basic for knowledge discovery and data mining. It offers mathematical tools to discover patterns hidden in data. It can be useful for feature selection, feature extraction, data reduction, pattern extraction, automatic classification and learning algorithm [2]. It can identify partial and total dependencies in data, eliminates redundant data, and gives approach to null values, missing data, dynamic data and others.

Basic concepts of rough sets are:

- Information/Decision Systems
- Indiscernibility
- Set Approximation
- Reducts

2.1.1.1 Information/Decision System

In rough set theory, an information system is defined as a pair of (U, A) where U is a non-empty finite set of objects and A is non-empty finite set of attributes such that

$a: U \rightarrow \mathrm{Va}$

For every a ∈ A. Va is called the value set of a. If the attribute set is partitioned into two subsets than one partition is called condition and another one is decision attribute respectively and the system is called decision system/table. Table 2.1.1.1 is an example of information/decision system.

Table 2.1.1.1: Example of Information/Decision System

	Age	LEMS	Walk
X1	16-30	50	yes
X2	16-30	0	no
X3	31-45	1-25	no
X4	31-45	1-25	yes
X5	46-60	26-49	no
X6	16-30	26-49	yes
X7	46-60	26-49	no

2.1.1.2 Indiscernibility

Let $IS = (U, A)$ be an information system, then with any $B \subset A$ there is an associated equivalence relation:

$\mathrm{IND}(B) = \{(x,y) \in U \times U : \forall a \in B, a(x) = a(y)\}$

Where $\mathrm{IND}(B)$ is called B-indiscernibility relation. If $\{(x,y) \in \mathrm{IND}(B)$, then objects x and y are indiscernible from each other by attributes from B. The equivalence classes of the B-indiscernibility relation are donated by $[x]B$. From table 1 indiscernibility relation of condition attributes can be describe. The non-empty subsets of the condition attributes in table 2.1.1.1 are $\{Age\}, \{LEMS\}$ *and* $\{Age, LEMS\}$.

So,

$IND(\{Age\}) = \{\{X1, X2, X6\}, \{X3, X4\}, \{X5, X7\}\}$

$IND(\{LEMS\}) = \{\{X1\}, \{X2\}, \{X3, X4\}, \{X5, X6, X7\}\}$

$IND(\{Age, LEMS\}) = \{\{X1\}, \{X2\}, \{X3, X4\}, \{X5, X7\}, \{X6\}\}$

2.1.1.3 Set Approximation

Let $T = (U, A)$ and let $B \subset A$ and $X \subset U$. We can approximate X using only the information contained in B by constructing the B-lower and B-upper approximations of X, denoted $B * X$ and $B *X$ respectively, where

$B * X = \{ x|[x]B \subset X\}$

$B * X = \{ x|[x]B \backslash X = \varphi\}$.

B-boundary region of X is defined by,

$BNDB(X) = B * (X) - B * (X)$.

B-outside region of X is defined by,

$U - B * (X)$.

It is consists of those objects that can be with certainty classified as not belonging to X. A set is said to be rough it its boundary region is non-empty, otherwise the set is crisp [4].

The positive region of decision classes $U/IND(D)$ with respect to condition attributes C is denoted by,

$POSc(D) = S\, B * (X)$

2.1.1.4 Reducts

Reducts is the only those attributes that preserve the indiscernibility relation and consequently, set approximation. There are usually several subsets of attributes those which are minimal are called reducts.

The set of attributes R⊂C is called a reduct of C, if $T' = (U, R, D)$ is independent and $POSR(D) = POSC(D)$. in other word, reducts are those minimal subsets who preserve the positive region [5]. From given decision system in table 2.1.1.4 we can extract reducts.

Table 2.1.1.4: Decision System

U	Headache	Muscle Pain	Temp	Flu
U1	Yes	Yes	Normal	No
U2	Yes	Yes	High	Yes
U3	Yes	Yes	Very-high	Yes
U4	No	Yes	Normal	No

U5	No	No	High	No
U6	No	Yes	Very-high	Yes

From table 2.1.1.4 we can get two reducts that preserve the relation of positive region. They are {Muscle-Pain, Temp}, {Headache, Temp}.

2.1.2 Decision Tree

Decision tree is one of the most popular machine learning algorithms. This algorithm is used for both classification and regression problems. In decision tree, each node represents a feature which means attributes, each link represents a decision rule, and each leaf represents an outcome of continuous or categorical values [5]. In Decision Tree the major challenge is to identification of the attribute for the root node in each level. This process is known as attribute selection. We have two popular attribute selection measures; they are information gain, and Gini index [6].

Information Gain: In a decision tree, we use a node for partitioning the training instances into smaller subsets and thus it changes the entropy. Information gain is a measure of this change in entropy. Suppose S is a set of instances, A is an attribute, Sv is the subset of S with A = v, and Values (A) is the set of all possible values of A, then [5]

$$Gain(S,A) = Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|}.Entropy(S_v)$$

Gini Index: Gini Index is a metric which is used to measure how often a randomly chosen element would be incorrectly identified. Here, an attribute with lower Gini index should be preferred. The Formula for the calculation of the Gini Index is given below [6].

$$GiniIndex = 1 - \sum_j p_j^2$$

2.1.3 Artificial Neural Network

Artificial Neural Network (ANN) is a computational model. It is called as neural network. It is based on functions and structures of human biological neural networks [7]. The structure of the ANN affected by a flow of information. Hence ANN changes were based on input and output.

Basically, ANN is a nonlinear statistical data. Which means that in ANN there is a complex relationship between input and output. For that we find different patterns. It also said that neural network is a wide class of nonlinear regression, data reduction and nonlinear dynamic model.

ANN consist of input, output and hidden layers. Transformation of input into valuable output is the main job. Information flows in neural network happens two ways. They are:

Feedforward Networks: In these input signals only travel in one direction without any loop. It flows towards the output layer. It is mostly used in pattern recognition. This network architecture constructed with a single input layer and a single output layer with zero or more hidden layer. The method has two common designs as below -

- At the time of its learning or "being trained"
- At the time of operating normally or "after being trained"

Feedback Networks: In this recurrent or interactive networks can use their internal memory to process sequenced of inputs. Signals can travel in both directions with loops in the network. As of now limited to time series. Typical human brain model.

Architectural components in ANN discussed below-

- **Input Layers, Neurons and Weights:** Neurons or nodes are the basic unit in a neural network. Neurons receive input from the external source. The basic idea here is to compute an output based on associated weight. Weights are given of a neuron based on neurons relative importance compared with other inputs. No finally function is applied to this for computations.
- **Hidden Layers and Output Layers:** Hidden layer is called hidden because it is always isolated from the external world. It takes input from input layer and performs its job. Its job is calculation and transform the result to output nodes. Figure 2.1.3 is an example of a simple Artificial Neural Network.

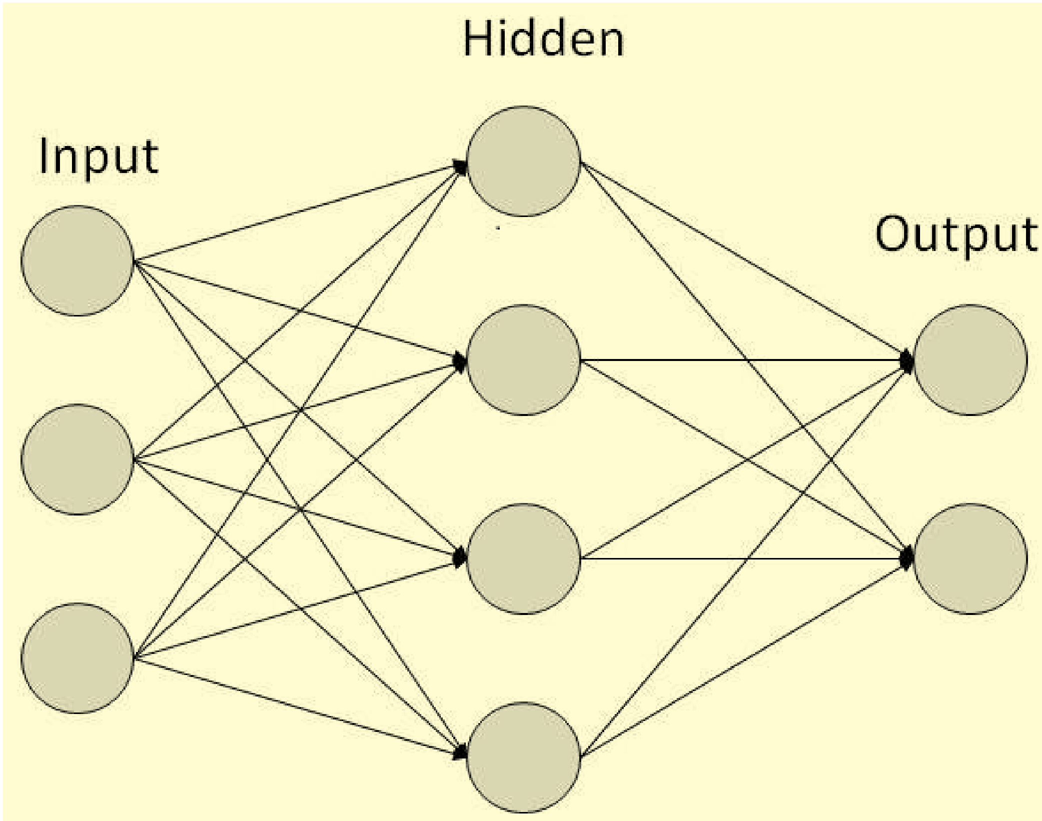

Figure 2.1.3: Simple Artificial Neural Network

Support vector machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression problem. But it mostly used in classification challenges. In SVM algorithm every data is a point in an n-dimensional space where n is number of features. Then, we perform classification by finding the hyper-plane that differentiate each classed from one another. It says that quality and complexity of Support Vector Machine does not depend directly on the dimensionality of the input spaces [9]. From figure 2.1.4 it can be seen that after applying SVM two classes has been differentiate by a hyper-plane.

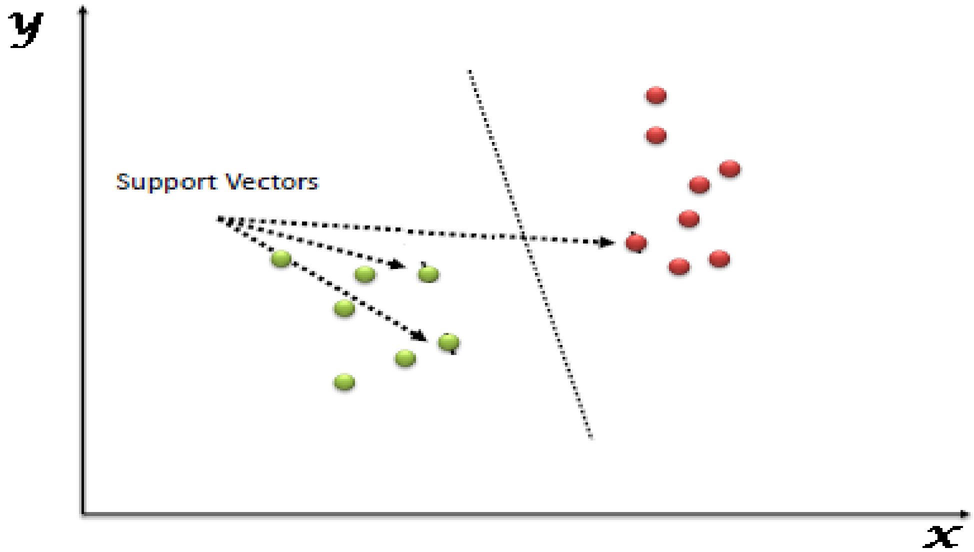

Figure 2.1.4: Support Vector Machine Binary Classification

2.2 Related Work

Thousands of contributions have been made for detecting and selecting features have been conducted by researchers and academics but most of them are in traditional manner.

But with the increase of data and dimensionality of data tradition manner is not a good process. Very few research have been made using rough set. From them in [9] authors proposed an approach where feature selection have been made using rough set along with greedy heuristic. This approach selects only one reduct which has the better accuracy without considering other effect as it is a greedy approach. One of the major contribution of their research is that it can choose only a better reduct set that would not damage the performance of the induction.

Another work in [10] where authors have used three clinical datasets have extracted features using rough set theory with indiscernibility relation and trained the reducts using Backpropagation Neural Network. In their work they have proposed a system merging rough set theory with Neural Network for classification. After applying rough set indiscernibility in every dataset they got minimal subsets of features. From this features a feature set with maximum classification accuracy have been shown as the result. One of the major contribution of this research is that it has done binary classification with superior performance on three different size multidimensional dataset.

A research has been made in [11] where for eliminating irrelevant attributes rough set theory has been applied. Authors built a case-based reasoning models in order to evaluate classification performance of the small attribute set which they got applying rough set theory. In a case-based reasoning system, cases or program modules related to previously developed system are stored in case library, which is represented by a fit data set or a training data set used to train the model [11]. This research show that applying rough set theory for finding small attributes set shows better accuracy in case-based reasoning than with all attributes case-based reasoning system. One of the major development of this research offers that for building high accuracy software quality-matric based classification model rough set theory along with reducts can be applied.

CHAPTER 3

PROPOSED METHOD

In the era of data science lot of work is being going on in every possible sector. For that researchers are collecting more data. And also dimension of this data are getting higher. So for classifying this multi-dimensional data we need a better approach which will reduced and will select the correct features from a dataset and will classify the data correctly. Realizing the situation in our research we have presented a better and efficient approach which will select the correct features and will improve the prediction of classifier. In figure 6 we demonstrate our proposed method.

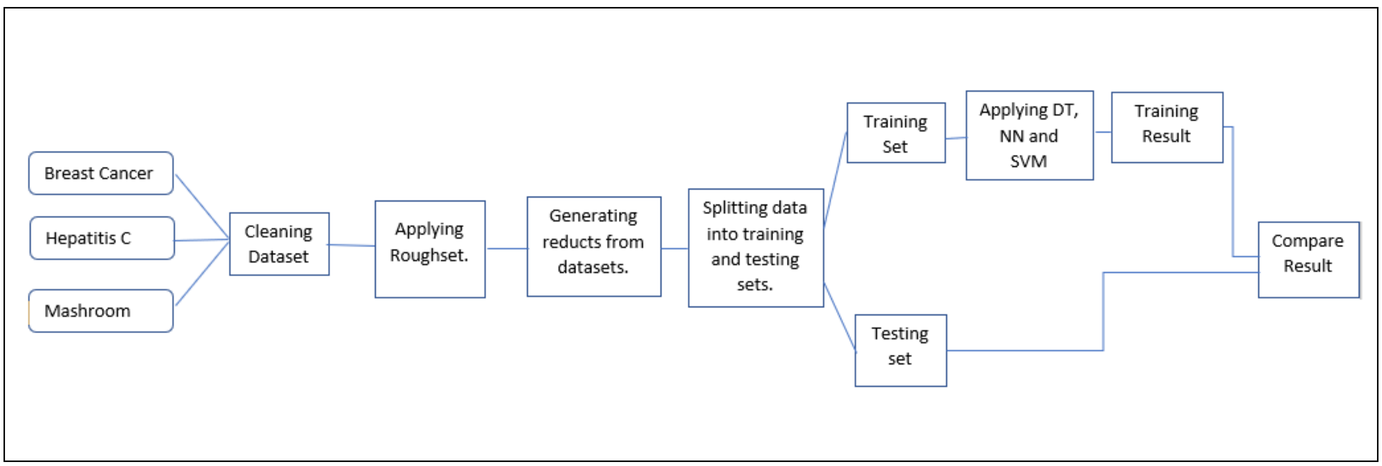

Figure 3.1: Thorough work infrastructure of our proposed method

From figure 3.1 we can described our whole method step by step. Steps of our proposed method-

- First we will extract data from our three datasets.
- Than we will done some preprocessing on our datasets. We will eliminate missing values and then also will normalize our data using min-max normalization.

- Than on every dataset will apply rough set. By applying rough set we will get reduct sets of every dataset. Those every reduct set is our selected feature.
- Than according to features in every reduct set we will construct our data and will split that data every time into training testing in 70:30 ratio. That means 70% will be training data and 30% testing data.
- After that on training data machine learning algorithms – Decision Tree, Artificial Neural Network, Support Vector Machine will be applied to build classifier models.
- Finally, performance of all classification algorithm has been compared along with every reduct set, without reduct set. Also performance of every reduct set and classifier has been measured using confusion matrix.

CHAPTER 4

DATASET OVERVIEW

For this study three dataset have been selected from UCI machine learning repository. These are Hepatitis C Virus (HCV) for Egyptian patients, Breast Cancer and Mushroom dataset.

Hepatitis C Virus (HCV) has 1385 instances with 29 attributes including a class label. The "Baseline histological staging" is the class label with values {F0, F1, F2, F3, and F4}. These labels represent different prognosis levels of Liver Fibrosis as follows: No Fibrosis (F0), Portal Fibrosis (F1), Few Septa (F2), Many Septa (F3), and Cirrhosis (F4). Table 4.1 describes attributes of the hepatitis dataset.

Table 4.1: Description of hepatitis dataset

Number	Attribute Name	Attribute Values
1	Age	32:61
2	Gender	Male, Female
3	BMI	22:35
4	Fever	Absent, Present
5	Nausea/Vomiting	Absent, Present
6	Headache	Absent, Present
7	Diarrhea	Absent, Present
8	Fatigue & generalized bone ache	Absent, Present
9	Jaundice	Absent, Present
10	Epigastric Pain	Absent, Present
11	WBC	2991:12101
12	RBC	3816422:5018451
13	HGB	2:20

14	Plat	93013:226464
15	AST 1	20:128
16	ALT 1	20:128
17	ALT 4	20:128
18	ALT 12	20:128
19	ALT 24	20:128
20	ALT 36	20:128
21	ALT 48	20:128
22	ALT after 24 w	20:128
23	RNA Base	0:1201086
24	RNA 4	0:1201715
25	RNA 12	0:3731527
26	RNA EOT	0:808450
27	RNA EF	0:808450
28	Baseline histological Grading	1:16
29	Baseline histological staging	F0:F4

The dataset has 336 “F0”, 332 “F1”, 355 “F2” and 362 “F3” instances. The distribution of class label in percentage has been shown in figure 4.1.

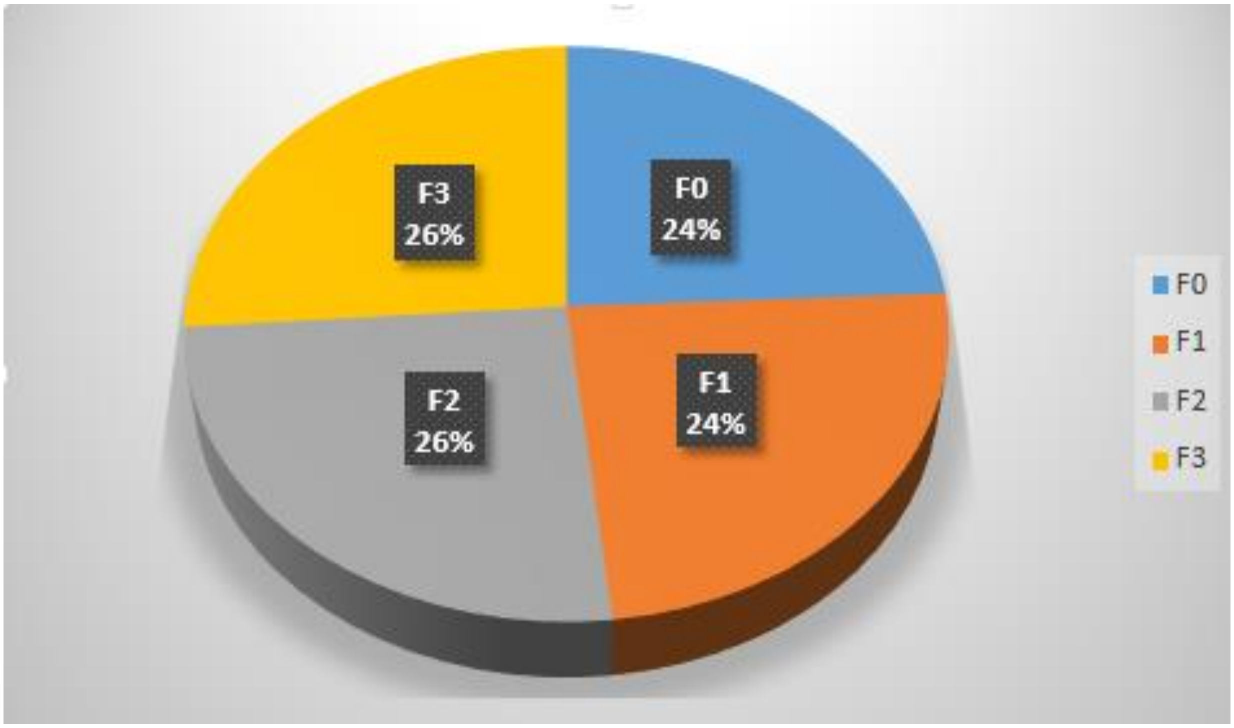

Figure 4.1: Total occurrence of each class in dataset

Breast Cancer has 699 instances with 10 attributes along with 1 class attribute. The class label contains whether a patient's breast tissue is malignant or benign. All attributes have a data type of integer value ranging from 1 to 10. It holds 16 samples with missing value. All the missing values belong to the attribute of bare nucleoli. Here in class label attribute numeric value "2" is for malignant and "4" is for benign. Table 4.2 describes all the attributes of breast cancer dataset.

Table 4.2: Description of breast cancer dataset

Number	Attribute Name	Attribute Values	Missing Values
1	Clump thickness	1:10	0
2	Uniformity of cell size	1:10	0
3	Uniformity of cell shape	1:10	0
4	Marginal adhesion	1:10	0
5	Single epithelial cell size	1:10	16

6	Bare nucleoli	1:10	0
7	Bland chromatin	1:10	0
8	Normal nucleoli	1:10	0
9	Mitosis	1:10	0
10	Class	2,4	0

Among 699 instances benign class has 458 instances which is 66% of whole data and malignant class has 241 instances which is 34%. Total occurrence of this classes in percentage has been shown in figure 4.2.

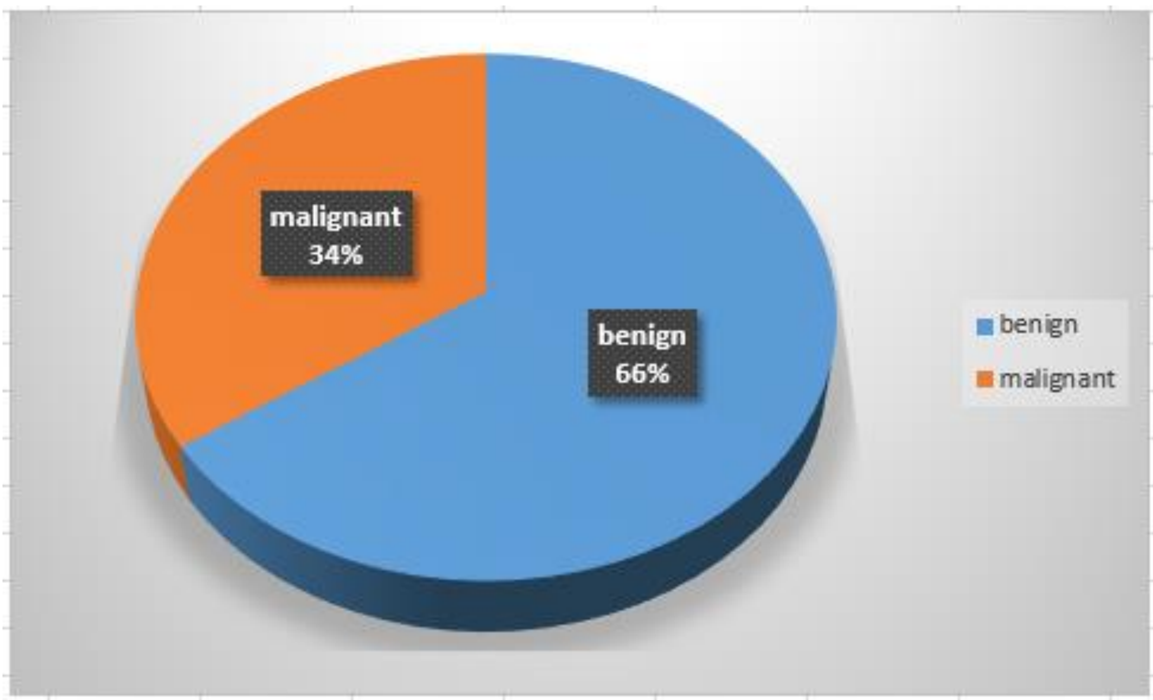

Figure 4.2: Total occurrence of each class in dataset

Mushroom dataset has 8124 instances with 22 attributes along with 1 class attribute with missing value of 2480. The class label with values {u, g, m, d, p, l,w}. There are different values of classifier as follows: universal (u), gray (g), musty (m), distant (d), paths (p), large (l), and white (w). Table 4.3 describes attributes of mushroom dataset.

Table 4.3: Description of hepatitis dataset

Number	Attribute Name	Attribute Values
1	Cap Shape	bell, conical, convex, flat, knobbed, sunken
2	cap-surface	fibrous, grooves, scaly, smooth
3	cap-color	brown, buff, cinnamon, gray, green, pink, purple, red, white, yellow
4	bruises	bruises, no
5	odor	almond, anise, creosote, fishy, foul, musty, none, pungent, spicy
6	gill-attachment	attached, descending, free, notched
7	gill-spacing	close, crowded, distant
8	gill-size	broad, narrow
9	gill-color	black, brown, buff, chocolate, gray, green, orange, pink, purple, red, white, yellow
10	stalk-shape	enlarging, tapering
11	stalk-root	bulbous, club, cup, equal, rhizomorphs, rooted, missing
12	stalk-surface-above-ring	fibrous, scaly, silky, smooth
13	stalk-surface-below-ring	fibrous, scaly, silky, smooth
14	stalk-color-above-ring	brown, buff, cinnamon, gray, orange, pink, red, white, yellow
15	stalk-color-below-ring	brown, buff, cinnamon, gray, orange, pink, red, white, yellow
16	veil-type	partial, universal
17	veil-color	brown, orange, white, yellow
18	ring-number	none, one, two
19	ring-type	cobwebby, evanescent, flaring, large,

		none, pendant, sheathing, zone
20	spore-print-color	black, brown, buff, chocolate, green, orange, purple, white, yellow
21	population	abundant, clustered, numerous, scattered, several, solitary
22	habitat	grasses, leavel, meadows, paths, urban, waste, woods

Among 8124 instances universal(u) has 368 instance which is 5%, gray has 2148 which is 26%, musty has 292 which is 4%, distant has 3148 which is 39%, paths has 1144 which is 14%, large has 832 which is 10% and white(w) has 192 which is 2%. Total occurrence of this classes in percentage has been shown in figure 4.3.

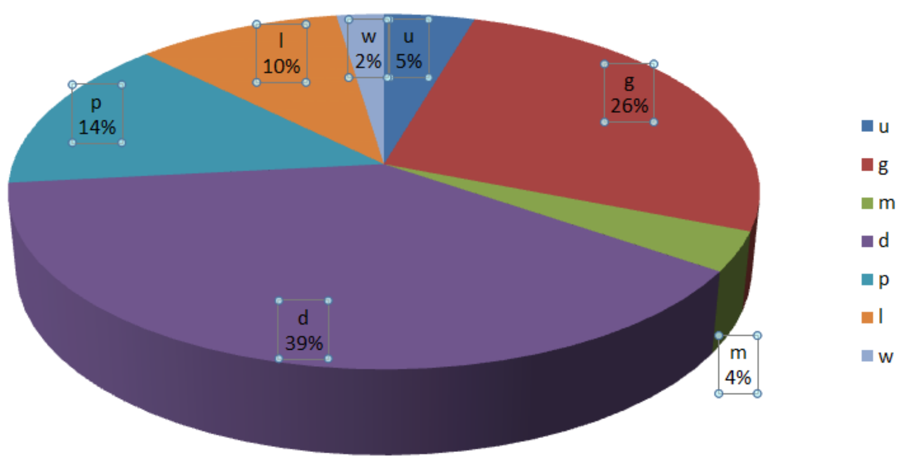

Figure 4.3: Total occurrence of each class in dataset

CHAPTER 5

RESULT ANALYSIS

After applying rough set on each dataset, we have multiple reduct sets for every dataset. They are as follows:

Table 5.1: Reduct for Breast Cancer dataset

Number of attributes	1	3	4	6	7	8
Number of reducts	1	1	3	2	3	12

From table 5.1 we can see that Breast Cancer dataset has total 29 reducts with different size of attributes. As from table 1 we can demonstrate that there are reduct of length 1, 3, 4, 6, 7 and 8. Table 5.2 shows reducts for Hepatitis C dataset.

Table 5.2: Reduct for Hepatitis C dataset

Number of attributes	8	9	12	13	15	16	17	18	19	20	21	22	24	25	26	27
Number of reducts	2	7	5	3	9	8	6	5	12	9	11	16	15	13	14	2

From table 5.2 we can see that Hepatitis C dataset has total 70 reducts. Length of this reducts are 8, 9, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 26 and 27.

Table 5.3: Reduct for Mushroom dataset

Number of attributes	1	3	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Number of reducts	1	5	4	7	11	10	9	13	14	12	7	11	6	12	14	13	5	8	12

Table 5.3 illustrates the total reduct set for Mushroom dataset. There are total 98 reduct set. Length of this reduct sets are between 1 to 21 out of 22 attributes.

Then applying machine learning algorithm on dataset with reduct and without reduct we got the accuracy of each dataset. First, we have applied classification algorithm on dataset without reduct. Table 5.4 accuracy of every dataset without reduct has been shown.

From table 5.4 we can illustrate that after applying Decision tree, Neural Network and Support Vector Machine (SVM) on Breast cancer dataset we got the accuracy respectively 91.707%, 92.19% and 93.12%. After applying on Hepatitis C we got the accuracy of Decision tree is 24.75%, Neural Network 25.7211% and SVM 22.83%. From Mushroom dataset Decision tree accuracy is 53.62%, Neural Network 66.17% and SVM 66.37% without reduct.

Table 5.4: Accuracy of each dataset without reduct

Datasets	Decision Tree	Neural Network	SVM
Breast Cancer	91.707%	92.19%	93.12%
Hepatitis C	24.75%	25.7211%	22.83%
Mushroom	53.62%	66.19%	66.37%

Then we have applied Decision tree, Neural Network and SVM on our dataset with reduct. Every reduct shows the better accuracy than without reduct. Table 5.5 shows the average accuracy of every dataset with reduct.

Table 5.5: Accuracy of each dataset with reduct

Datasets	Decision Tree	Neural Network	SVM
Breast Cancer	95.15%	94.12%	95.12%
Hepatitis C	28.36%	27.64%	25.96%
Mushroom	63.00%	70.26%	68.31%

From table 5.5 we can see that average accuracy of breast cancer dataset after applying Decision tree, Neural Network and SVM is 95.15%, 94.12% and 95.12% respectively. Also Hepatitis C dataset also shown better performance than without reduct. Accuracy of Hepatitis C with reduct as follows, for decision tree 28.36%, for Neural Network 27.64% and for SVM 25.96%. For Mushroom dataset accuracy of Decision tree is 63.00%, Neural Network 70.26% and SVM 68.31%.

CHAPTER 6
Performance Comparison

6.1 Comparative Analysis

For evaluating the performance of reduct sets and algorithms we have generated confusion matrix for every datasets reducts. Table 6.1 illustrates that which reduct set has successfully determine classifiers after applying classification algorithm on Breast cancer algorithm.

Table 6.1.1: Result of successfully determining each class from Breast Cancer dataset:

Class	Actual Result	**All Attributes**			1			3			4			6			7			8		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
Benign	130	**125**	**125**	**125**	123	123	123	126	126	126	126	127	125	124	126	125	126	127	126	125	128	126
Malignant	75	**67**	**66**	**70**	59	59	59	66	69	66	67	65	65	68	68	70	67	72	69	69	70	71

On the table 6.1.1 we have shown the number of successfully determined class after applying rough set and using reduct sets to determine the decision class. Here, we can see that except the first reduct almost every other reduct set has better performance than the original dataset with all attributes. Among them reduct set with 8 attribute has the best result for neural network which is able to determine 128, 70 Benign and Malignant class.

Table 6.1.2: Result of successfully determining each class from Hepatitis C dataset

Class	Actual Result	**Without Reduct**			8 (R29)			9 (R25)			12(R26)			13(R11)			15(R23)			16(R33)		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
F0	104	**27**	**31**	**4**	24	13	53	30	24	25	15	27	2	27	16	5	23	27	2	24	23	8
F1	110	**14**	**20**	**3**	31	22	1	21	21	2	29	17	11	30	21	7	31	19	0	26	25	2
F2	93	**28**	**27**	**43**	23	27	23	26	18	34	17	23	41	34	28	41	22	26	58	27	24	39
F3	109	**37**	**27**	**45**	31	32	31	22	26	31	35	25	42	29	36	43	33	33	39	29	35	36

Table 6.1.3: Number of successfully determining each class from Hepatitis C dataset

Class	Actual Result	**Without Reduct**			17(R17)			18 (R28)			19(R15)			20(R24)			21(R30)			22(R35)		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
F0	104	**27**	**31**	**4**	21	18	0	31	25	2	26	21	2	19	27	2	29	37	0	29	20	2
F1	110	**14**	**20**	**3**	29	27	0	21	25	5	19	29	1	15	23		20	24	1	26	23	5
F2	93	**28**	**27**	**43**	23	19	53	20	28	35	29	24	48	34	25	39	18	29	48	26	30	46
F3	109	**37**	**27**	**45**	27	29	40	28	34	43	32	35	41	39	39	44	24	39	44	30	32	37

Table 6.1.4: Number of successfully determining each class from Hepatitis C dataset

Class	Actual Result	**Without Reduct**			24 (R16)			25(R62)			26(66)			27(R39)		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
F0	104	**27**	**31**	**4**	22	28	27	22	23	27	28	20	6	34	22	7
F1	110	**14**	**20**	**3**	20	20	4	20	26	5	14	22	3	22	25	4
F2	93	**28**	**27**	**43**	22	19	48	25	22	47	27	18	43	23	31	46
F3	109	**37**	**27**	**45**	32	27	34	42	35	38	37	37	37	26	44	43

We have divided the result of successfully determining each class from Hepatitis C dataset into three table which are 6.1.2, 6.1.3 and 6.1.4. On these table we have shown the number of successfully determined class after applying rough set and using reduct sets to determine the decision class. It has four decision class which are F0, F1, F2 and F3. From the table it is visible that not all the reduct sets has better performance than the original dataset. Though some a good number of reduct set are able to perform better than the original dataset.

Table 6.1.5: Number of successfully determining each class from Mushroom dataset

Class	Actual Result	All Attributes			1			3			5			6			7			8		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
u	116	29	63	50	29	63	50	30	63	50	31	63	50	31	63	50	30	63	50	32	63	50
g	572	263	321	339	262	321	339	261	321	339	260	321	339	260	321	339	261	321	339	258	321	339
m	94	588	640	665	588	640	665	588	640	665	588	640	665	588	640	665	588	640	665	588	640	665
d	716	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
p	179	36	54	63	37	54	63	36	54	63	36	54	63	36	54	63	36	54	63	37	54	63
l	17	3	45	38	3	45	38	4	45	38	4	45	38	4	45	38	4	45	38	4	45	38

Table 6.1.6: Number of successfully determining each class from Mushroom dataset

Class	Actual Result	All Attributes			9			10			11			12			13			14		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
u	116	29	63	50	31	63	50	71	77	50	87	89	60	64	91	122	72	82	122	31	63	50
g	572	263	321	339	261	321	339	672	624	714	333	321	339	325	322	346	337	361	362	261	321	339
m	94	588	640	665	588	640	665	635	640	683	623	640	665	631	640	665	614	640	667	588	640	665
d	716	17	17	17	17	17	17	50	54	17	17	17	17	50	64	28	50	55	28	17	17	17
p	179	36	54	63	36	54	63	63	63	63	63	63	63	36	54	63	49	58	63	36	54	63
l	17	3	45	38	3	45	38	41	45	81	40	81	81	41	45	43	41	45	43	3	45	38

Table 6.1.7: Number of successfully determining each class from Mushroom dataset

Class	Actual Result	All Attributes			15			16			17			18			19			20		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
u	116	29	63	50	79	63	122	73	70	73	76	78	114	82	69	109	68	63	122	64	74	109
g	572	263	321	339	333	350	339	332	354	339	315	377	383	324	362	339	315	357	339	315	352	339
m	94	588	640	665	614	640	665	603	644	667	623	646	665	623	651	665	603	644	665	603	652	665
d	716	17	17	17	17	17	17	42	56	63	48	79	63	17	17	17	37	63	63	17	17	17
p	179	36	54	63	58	54	63	50	64	63	51	61	63	51	62	63	37	67	92	36	64	92
l	17	3	45	38	28	45	81	22	49	81	32	46	48	21	49	43	24	47	38	5	49	43

Table 6.1.8: Number of successfully determining each class from Mushroom dataset

Class	Actual Result	All Attributes			21			22			4		
		DT	NN	SVM	DT	NN	SVM	DT	NN	SVM	DT	NN	SVM
u	116	29	63	50	60	63	84	60	63	84	30	63	50
g	572	263	321	339	292	340	339	292	340	339	263	321	339
m	94	588	640	665	589	640	665	589	640	665	588	640	665
d	716	17	17	17	17	17	17	17	17	17	17	17	17
p	179	36	54	63	36	54	63	36	54	63	35	54	63
l	17	3	45	38	5	45	43	5	45	43	3	45	38

On the above three table we have shown the actual classification result on the dataset and the number of classifications with all attributes using Decision tree, Neural Network and SVM. We have also included the result after applying roughest on the dataset and using the reduct sets.

6.2 Accuracy Comparison:

In this section we have made a comparison between the overall performance which includes comparison between accuracy of the dataset and class identification.

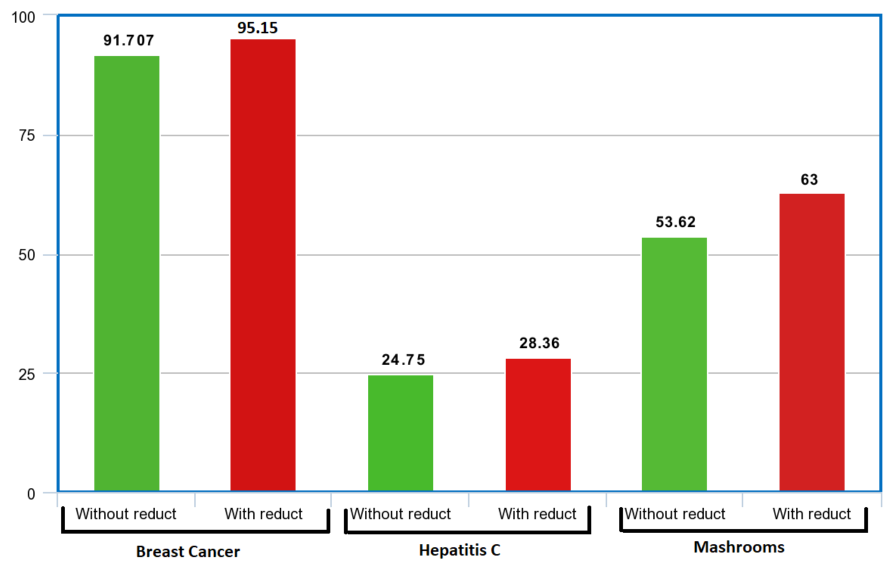

Figure 6.2.1: Difference between accuracy before and after using reduct (Decision Tree)

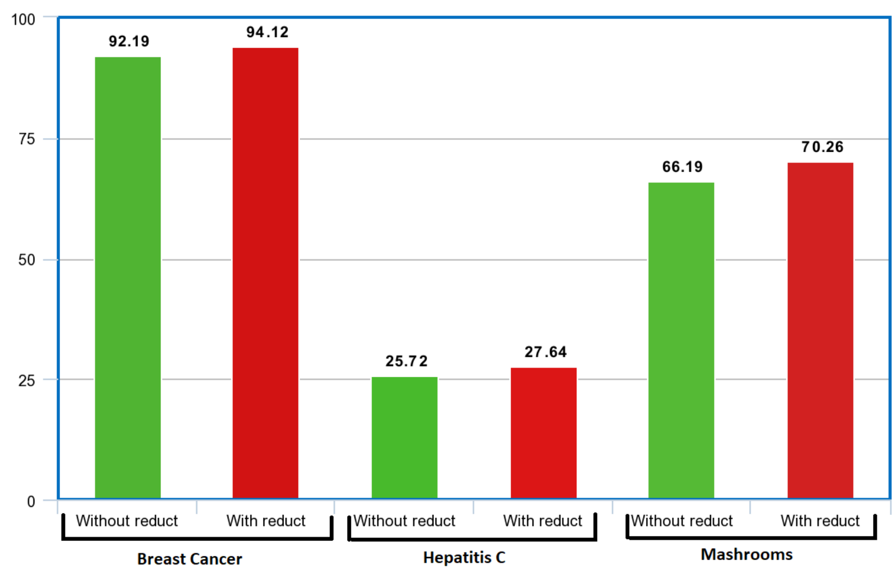

Figure 6.2.2: Difference between accuracy before and after using reduct (Neural Network)

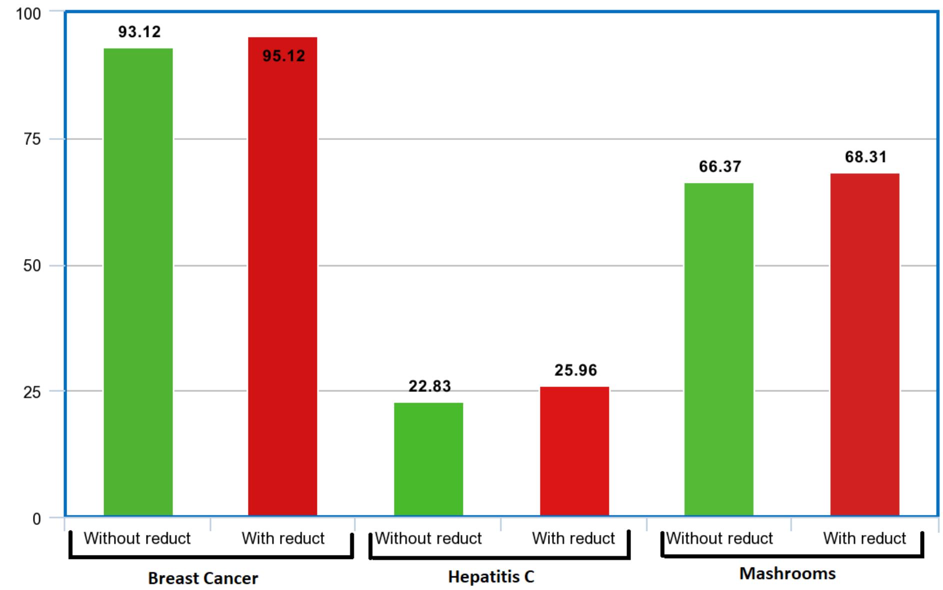

Figure 6.2.3: Difference between accuracy before and after using reduct (Neural Network)

On the figure 6.2.1, 6.2.2 and 6.2.3 we have shown difference on accuracy with and without reduct on three different algorithms. Figure 6.2.1, Figure 6.2.2, and Figure 6.2.3 are for Decision tree. Neural Network and SVM respectively. On those figures we can see that there is a slightly increase in accuracy after using reduct sets. Difference between accuracy varies on datasets and algorithm. For the Breast Cancer dataset accuracy increases about 2%-3% on every algorithm. For Hepatitis C dataset accuracy increases around 2%-4% and for Mushroom dataset accuracy increases around 4%-10% based on algorithm.

6.3 Best result in each algorithm:

In this section we have compared and shown the result of each class detection with and without using reduct sets.

On the table 6.3.1 we can see the difference of class detection number between reduct sets and original dataset. "Actual Result" column contains the actual number of each class in all datasets. "All attributes" column contains the number of each class after applying three algorithms on the dataset containing all the attributes. And the remaining column contains the number of classes after using reduct sets and applying Decision tree, Neural Network and SVM.

Table 6.3.1: Breast Cancer dataset.

Class	Actual Results	All attributes			Algorithms		
		DT	**NN**	**SVM**	DT	NN	SVM
Benign	130	**125**	**125**	**125**	126 (3,4,6,7)	128 (8)	126 (3,7,8)
Malignant	75	**67**	**66**	**70**	69 (8)	72 (7)	71 (8)

Table 6.3.1 shows the comparison between reduct set and original Breast Cancer data. From the table it is visible that were able to determine 126, 69 Benign and Malignant class respectively. Neural Network was able to determine 128, 72 and SVM was able to determine 126, 71 Benign and malignant class respectively.

Table 6.3.2: Hepatitis C dataset.

Class	Actual Results	All attributes			Algorithms		
		DT	**NN**	**SVM**	DT	NN	SVM
F0	104	**27**	**31**	**4**	34 (27)	34 (21)	53 (1)
F1	110	**14**	**20**	**3**	31 (8,15)	29 (19)	11 (12)
F2	93	**28**	**27**	**43**	34 (13,20)	31 (27)	58 (15)
F3	109	**37**	**27**	**45**	42 (25)	44 (27)	44 (20,21)

Table 6.3.2 shows the result comparison based on Hepatitis C dataset. Hepatitis C dataset had four feature which are F0, F1, F2 and F3. Applying decision tree, we were able to determine 34,

31, 34 and 42 F0, F1, F2 and F3 class respectively. Neural network was able to determine 34, 29, 31, 44 class and SVM determined 53, 11, 58, 44 F0, F1, F2 and F3 class respectively.

Table 6.3.3: Mushroom dataset.

Class	Actual Results	All attributes			Algorithms		
		DT	**NN**	**SVM**	DT	NN	SVM
U	116	36	54	63	63(10,11)	67(19)	92(19,20)
G	572	263	321	339	601(10)	624(10)	661(10)
M	94	3	45	38	50(10,12,13)	79(17)	63(16,19)
D	716	588	640	665	635 (8)	640 (20)	683 (10)
P	179	67	48	55	87 (11)	91 (12)	122 (12,13,15,19)
L	17	9	9	10	11(10,12,13)	15(11)	15(16,17)

Mushroom dataset has six attributes and they are u, g, m, d, p and I. All algorithms that ran on reduct set were able to determine more class the original dataset. Among three algorithm SVM was able to perform better on this dataset.

6.4 Best algorithm in each class:

Class	Actual Result	All attributes			Best Algorithm on each reduct & class					
		DT	**NN**	**SVM**	1	3	4	6	7	8
Benign	130	**125**	**125**	**125**	All	NN	NN	NN	NN	NN
Malignant	75	**67**	**66**	**70**	All	NN	DT	SVM	NN	SVM

Table 6.4.1: Breast Cancer dataset.

Table 6.4.2: Hepatitis C dataset.

Class	Actual Results	All Attributes			Best Algorithm on each reduct & class															
		DT	**NN**	**SVM**	8	9	12	13	15	16	17	18	19	20	21	22	24	25	26	27
F0	104	**27**	**31**	**4**	SVM	DT	NN	DT	NN	DT	DT	DT	DT	NN	NN	DT	NN	SVM	DT	DT
F1	110	**14**	**20**	3	DT	DT, NN	DT	DT	DT	DT	DT	NN	NN	NN	NN	DT	DT, NN	NN	NN	NN
F2	93	**28**	**27**	**43**	NN	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	NN
F3	109	**37**	**27**	**45**	NN	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	DT	DT, NN, SVM	NN

Table 6.4.3: Mushroom dataset.

Class	Actual Result	All attributes					Best Algorithm on each reduct & class															
		DT	**NN**	**SVM**	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
u	116	36	54	63	NULL	NULL	SVM	SVM	SVM	SVM	SVM	SVM	SVM	ALL	ALL	SVM	SVM	SVM	SVM	NN	SVM
g	572	263	321	339	NULL	NULL	SVM	SVM	SVM	SVM	SVM	SVM	SVM	DT	SVM	SVM	SVM	DT	NN	NN	SVM
m	94	3	45	38	NULL	NULL	NN	NN	NN	NN	NN	NN	NN	SVM	NN SVM	NN	NN	NN	SVM	SVM	SVM
d	716	588	640	665	NULL	NULL	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM
p	179	67	48	55	NULL	NULL	NN	NN	NN	NN	NN	NN	NN	NN	NN	SVM	SVM	NN	SVM	DT SVM	SVM
l	17	9	9	10	NULL	NULL	ALL	ALL	ALL	ALL	ALL	ALL	ALL	NN	ALL	NN	55	ALL	ALL	SVM	NN

On the table 6.4.1, 6.4.2 and 6.4.3 we have shown the best algorithm for each class according to our dataset and reduct sets. For the Breast cancer dataset to detect "Benign" class Neural Network will work best for any all reduct sets. And for "Malignant" class SVM works best for maximum number of reduct sets. Now on the Hepatitis C dataset, for class F0 and F1 Decision tree and for class F3, F4 SMV works best for maximum number of reducts. And on mushroom dataset to detect U, G, D class SVM will work best most of the time and for class M, P Neural network is the best option. And other class I will work on all algorithm.

From all the discussion and tables, we can clearly see that after using reduct sets we have improved the accuracy and the number of class detection in each dataset. We can conclude using reduct sets not only improved the accuracy and class detection but also eliminate the unnecessary attributes from datasets and make it easier to work with.

CHAPTER 7

CONCLUSION & FUTURE WORK

Feature selection is one of the most important process before constructing a model. It reduces the time complexity of training set and also removed the unnecessary attributes. So for detecting features correctly and accurately we have proposed an approach using rough set theory. It is impossible to identify and remove all inappropriate and unwanted feature from a dataset. The technique we have used gives us a list of reduct sets which contains only the appropriate list of features. We get the list by applying rough set with reduct on it. That list contains multiple number of reduct sets with different length. After running multiple algorithm on all dataset with and without applying rough set theory we have compared the overall result. Though all the reduct sets didn't give us promising result but most of reduct sets were able to perform better than the original dataset and were able to give us increase in performance and result which are promising enough to prove the significance of using rough set.

We have used roughest theory to for feature selection. There are other methods too for feature selection. In future we have plan to use other methods to get reduct set and compare those with roughest theory. Apart from this, we have used only three dataset which has only either numeric or alphabetic data. We have plan to apply roughest on a dataset which has both numeric and alphabetic data and compare the performance.

BIBLIOGRAPHY

[1] Zhong, N., Dong, J. & Ohsuga, S. Journal of Intelligent Information Systems (2001) 16: 199.

[2] Pawlak, Z. International Journal of Computer and Information Sciences (1982) 11: 341.

[3] P. Ramasubramanian, K. Iyakutti, P. Thangavelu, G. Jegadeeswari Jeya and S. Shameera Begam, "Data mining techniques for teaching result analysis using rough set theory," *2008 International Conference on Computing, Communication and Networking*, St. Thomas, VI, 2008, pp. 1-8.

[4] M. Zhang and J. T. Yao, "A rough sets based approach to feature selection," *IEEE Annual Meeting of the Fuzzy Information, 2004. Processing NAFIPS '04.*, Banff, Alberta, Canada, 2004, pp. 434-439 Vol.1.

[5] "Chapter 4: Decision Trees Algorithms – Deep Math Machine learning.ai – Medium." [Online]. Available: https://medium.com/deep-math-machine-learning-ai/chapter-4-decision-trees-algorithms-b93975f7a1f1. [Accessed: 05-Apr-2019].

[6] "Decision Tree Introduction with example - GeeksforGeeks." [Online]. Available: https://www.geeksforgeeks.org/decision-tree-introduction-example/. [Accessed: 05-Apr-2019].

[7] Adel El-Shahat (February 28th 2018). Introductory Chapter: Artificial Neural Networks, Advanced Applications for Artificial Neural Networks, Adel El-Shahat, IntechOpen, DOI: 10.5772/intechopen.73530.Availablefrom: https://www.intechopen.com/books/advanced-applications-for-artificial-neural-networks/introductory-chapter-artificial-neural-networks

[8] Zhong, N., Dong, J. & Ohsuga, S. Journal of Intelligent Information Systems (2001) 16: 199.

[9] Suykens, J. & Vandewalle, J. Neural Processing Letters (1999) 9: 293.

[10] Kindie Biredagn Nahato, Khanna Nehemiah Harichandran, and Kannan Arputharaj, "Knowledge Mining from Clinical Datasets Using Rough Sets and Backpropagation Neural Network," Computational and Mathematical Methods in Medicine, vol. 2015, Article ID 460189, 13 pages, 2015.

[11] Khoshgoftaar, Taghi & Bullard, Lofton & Gao, Kehan. (2009). ATTRIBUTE SELECTION USING ROUGH SETS IN SOFTWARE QUALITY CLASSIFICATION. International Journal of Reliability, Quality and Safety Engineering - IJRQSE. 16.