Comparisons of SVM Kernels for Insurance Data Clustering

Irfan Nurhidayat [1], Busayamas Pimpunchat [1*], Samad Noeiaghdam [2, 3], Unai Fernández-Gámiz [4]

[1] Department of Mathematics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand.

[2] Industrial Mathematics Laboratory, Baikal School of BRICS, Irkutsk National Research Technical University, Irkutsk, 664074, Russia.

[3] Department of Applied Mathematics and Programming, South Ural State University, Lenin Prospect 76, Chelyabinsk, 454080, Russia.

[4] Nuclear Engineering and Fluid Mechanics Department, University of the Basque Country UPV/EHU, Nieves Cano 12, Vitoria-Gasteiz, 01006, Spain.

Abstract

This paper will study insurance data clustering using Support Vector Machine (SVM) approaches. It investigates the optimum condition employing the three most popular kernels of SVM, i.e., linear, polynomial, and radial basis kernel. To explore sum insured datasets, kernel comparisons for Root Mean Square Error (RMSE) and density analysis have been provided. It employs these kernels to classify based on sum insured datasets. The objective of this research is to demonstrate to industrial researchers that data grouping may be accomplished in an organized, error-free, and efficient manner utilizing R programming and the SVM approach. In this study, we check the insurance data for the sum insured with statistical methods in the form of Model Performance Evaluation (MPE), Receiver Operating Characteristics (ROC), Area Under Curve (AUC), partial AUC (pAUC), smoothing, confidence intervals, and thresholds. Then, sum insured data are followed up to classify using SVM kernels. This paper finds new ideas for evaluating insurance data using the SVM approach with multiple kernels. This novel research emphasizes the statistical analysis methods for insurance data and uses the SVM method for more accurate data classification. Finally, it informs that this research is a pure finding, and there has never been any research on this subject. This research was conducted using the sum insured data as a sample from the Office of the Insurance Commission (OIC) in Thailand as an independent insurance institution providing actual data.

1- Introduction

Healthcare in Thailand has improved year by year with excellent quality. It is good news for expatriates and students, who also benefit from a health care system ranked among the best in the world. The country launched its universal health care program in 2002, which provides universal health care to all citizens. Several provinces in Thailand already have at least one public and one private hospital. The health system in this country has some criteria, namely government health services, non-governmental health organizations, and other health sectors. The Ministry of Public Health funds all health services for health institutions under the auspices of the government, such as hospitals, medical services, and various health sectors in collaboration with the government [1, 2]. The obstacle faced by health services under the auspices of the government is the existence of long queues, which sometimes makes people prefer private health services. Another consideration for people with upper middle incomes is that they assume that health services in the private sector are better than those in the public. Based on this, the government introduced Universal Health Coverage (UHC) in 2002. With the improvement of the system, now all citizens have access to health care in all health programs, from teeth, eyes, ears, and others. However, other problems arise, namely those related to taxes, mainly for people with lower middle incomes [3, 4].

Insurance is very significant when viewed statistically. High health care can be covered by it. If someone is affected by a disaster or infectious disease, it will not be so burdensome to pay for health care [5-7]. The government's goal is to minimize the gap for every citizen to access the best health care. Therefore, the encouragement for citizens to have insurance continues to be echoed. James et al. (2013), Shah et al. (2017), Sritart et al. (2021) and Rojjananukulpong et al. (2021) [8-11] undertook research to improve the insurance system, and the aim is to develop a direct and context-specific technique to measure inequality in remote areas of Thailand in the hope of assisting in the growth of efficiency in the use of more mobile health clinics. Buchanna et al. (2022) and Marshall et al. (2021) [7, 12] examined the utilization of urban and rural health services to see the differences between the two. The biggest obstacle to emergency transport is Bangkok traffic, where cars do not usually clear the way for an ambulance. Although insurance provides peace of mind for UHC cardholders, there are still difficult to diagnose and requires specialized treatment that can often have devastating financial consequences [13, 14]. Treatments such as significant organ transplants, surgeries, and so forth are treatments that cost much money and require multiple hospital stays of long duration. For such situations, patients should increase their insurance coverage through primary and health insurance to get additional financial protection at an affordable cost. Given the rising cost of health care in Thailand, a medical emergency can quickly deplete your savings. Most insurance companies quote six to ten times your annual salary as an appropriate amount for life insurance. Health insurance aims to provide financial protection if you suffer an illness to protect your savings [1, 2, 15].

In terms of income, inequality is problematic for every individual in every region of Thailand. There is a great demand to boost efficiency and streamline processes in the highly competitive insurance industry. Most carriers suffer from manual process stages, including several repetitive duties such as reviewing a claim entry for completeness and seeking missing information such as a police report from a traffic accident. Handling an underwriting request or claim necessitates matching the request with customer information, typically stored in legacy systems. Finding critical data, such as customer status and the related right to reimbursement, is frequently a barrier to the claims process's end-to-end automation. The convergence of policyholders, insurance personnel, and technology presents tremendous prospects for challenging claims processing. This research offers innovation in insurance data management using the Support Vector Machine (SVM) method. This study examines the sum of insured data classifying by the SVM method, which previously analyzed the data using statistical methods. In the end, we also provide information about the best kernels from some of the chosen kernels. The research methodology used is literature study, namely, studying material from various related sources such as journals, books, and the internet. These research steps are: Sum insured data collection using Microsoft Excel; Perform data analysis with several statistical methods carried out on the R programming; Testing sum insured data using the SVM method with linear, polynomial, and RBF kernels; The classifying results are assumed to be 0 with the sum insured contribution (pay more) and 1 for the regional group with the sum insured contribution (pay less); The RBF kernel is the most accurate for classifying by the SVM method. The flowchart in Figure 1 gives an illustration of our research methodology.
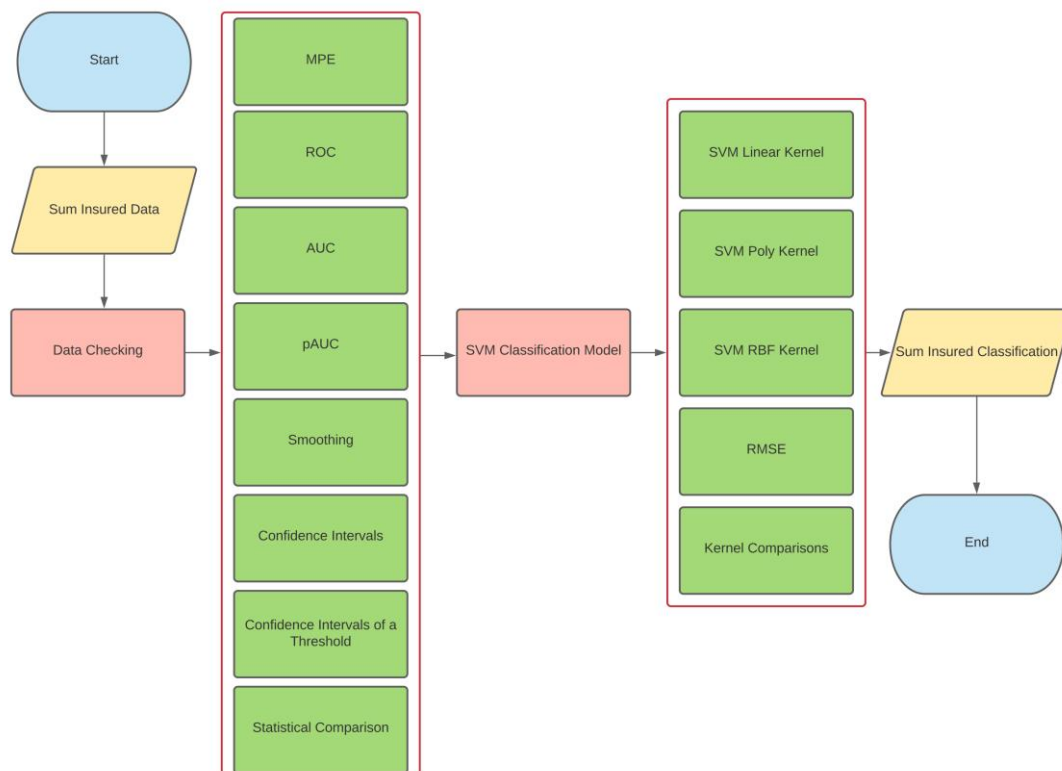

Figure 1. Research steps in classification with SVM kernels

The primary motivation is to assess data quality by decreasing mistakes in insurance data classifications. It offers new applications for categorizing insured data in numerous provinces in Thailand. To analyze Model Performance Evaluation (MPE), Receiver Operating Characteristics (ROC), Area Under Curve (AUC), partial AUC (pAUC), smoothing, confidence intervals, and thresholds have been explained in Yiengprugsawan et al. (2009) [16] and Puenpatom & Rosenman (2008) [17]. At the outset of the talk, it will provide a comparative sum insured statistics from many different years. Then, it attempted to reclassify by utilizing the insurance amount from each province in Thailand. It presumed that Bangkok is an affluent location since residents in the capital have a comparatively high income. This insurance research utilized Thailand's Insurance Commission datasets, used R programming to categorize datasets in simulations, and employed the most commonly used approach, SVM [18-20]. This paper's outline is as follows: Section 1 includes an introduction in basic facts about the subject, study motivation, and a detailed arrangement of the work. Section 2 confirms the amount of sum insured data by comparing statistics for the sum insured from different years—section 3 simulated SVM using three kernels to find the optimal kernel. It selected these three kernels because they have distinctive qualities and almost customarily distributed data, and many academics widely utilize them in SVM analysis. Section 4 displays the tabular results categorization of the sum insured data. Next, it concludes that a province with 0 (poor) results signifies that they must pay high for their premium. In the case of output 1 (good), the payment is much low for their premium. Finally, Section 5 gives conclusions and future work.

2- Validating Data

This section will implement some statistical analysis on the sum insured data used for SVM classification [6, 8]. It initiated the data validation process from the model performance evaluation section to the statistical comparison of the selected data samples. This report includes information from 2011, 2012, and 2014. Proceeding our study of data from each area in Thailand, including using models from 2011, 2012, and 2014, it attempts to reproduce statistical comparisons between sum insured data, shown in Figure 2. It deployed R programming to validate several data checks, separated into four major groups before being validated using additional scientific computations. The R code for testing statistical comparisons is referred to Puenpatom & Rosenman (2008) [17]. While using the three kernels, it identified minimal faults to simulate. It concludes that 0 is the wrong area, representing a high insurance cost, and 1 is the excellent area, i.e., the area with reduced insurance prices.
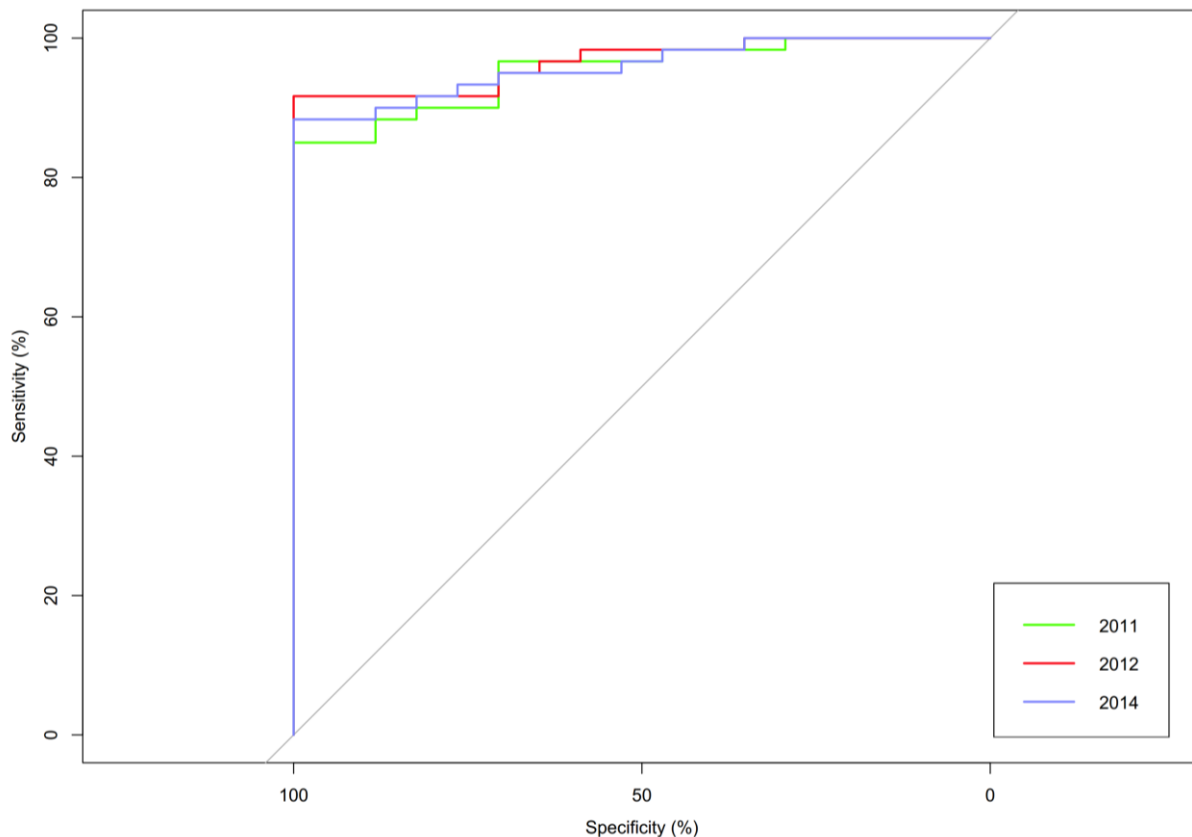

Figure 2. Statistical Comparison

The following sections will examine how SVM kernels work with their programs to execute sum insured data.

3- SVM Kernels

The SVM linear kernel, the SVM polynomial kernel, and the SVM RBF kernel are employed in simulations. While conducting the simulation, we apply the collected data on the sum insured for 2011, 2012 and 2014. For instance, in 2011, the data belong the total sum insured data from every province in Thailand. Thailand is split into four regions: the central, the north, the northeast, and the south. It grouped the 77 areas into categories based on their geographical location. There were 26 parts in the central regions, 17 in the north, 20 in the northeast, and 14 in the south. It uses the same technique to allocate the 2012 and 2014 data. It intend to discover the optimal kernel for categorizing insured data and information for future study by evaluating the data using SVM with three alternative kernels. Assuming that level 0 is the recommended area with the obligation to pay more sum insured, while level 1 is considered better than level 0, the sum insured can be paid less. The simulation section will begin by designing a sum insured database in Microsoft Excel, then employ it in R programming through the SVM classification by utilizing caret packages, including the linear kernel, polynomial, and RBF. All numerical experiments are on Windows 10 Pro, Intel Core i5-7500 CPU, 8GB RAM, 64 bits. The following is the R code for SVM linear testing

```
# SVM Linear

library(readxl)
library(caret)
X3y_data1 <- read_excel("D:/SVM Code in R/3y_data.xlsx",
                sheet = "Sheet2", range = "A1:E27")
View(X3y_data1)
str(X3y_data1)
intrain1 <- createDataPartition(y = X3y_data1$Lv, p= 0.7, list = FALSE)
training1 <- X3y_data1[intrain1,]
testing1 <- X3y_data1[-intrain1,]
dim(training1);
dim(testing1);
anyNA(X3y_data1)
summary(X3y_data1)
training1[["Lv"]] = factor(training1[["Lv"]])
trctrl1 <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
svm_Linear1 <- train(Lv ~., data = training1, method = "svmLinear",
            trControl=trctrl1,
            preProcess = c("center", "scale"),
            tuneLength = 50)
svm_Linear1
test_pred1 <- predict(svm_Linear1, newdata = testing1)
test_pred1
confusionMatrix(table(test_pred1, testing1$Lv))
grid1 <- expand.grid(C = c(0,0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2,5))
set.seed(123)
svm_Linear_Grid1 <- train(Lv ~., data = training1, method = "svmLinear",
            trControl=trctrl1,
            preProcess = c("center", "scale"),
            tuneGrid = grid1,
            tuneLength = 50)
svm_Linear_Grid1
plot(svm_Linear_Grid1, lwd=2, cex = 2.5, col = "red", bg = "black",  pch=21)
svm_Linear_Grid1$bestTune
test_pred_grid1<- predict(svm_Linear_Grid1, newdata = testing1)
test_pred_grid1
confusionMatrix(table(test_pred_grid1, testing1$Lv))
```

Figure 3 is identified utilizing the SVM linear kernel from each region in Thailand for the sum insured data with samples in 2011, 2012 and 2014.
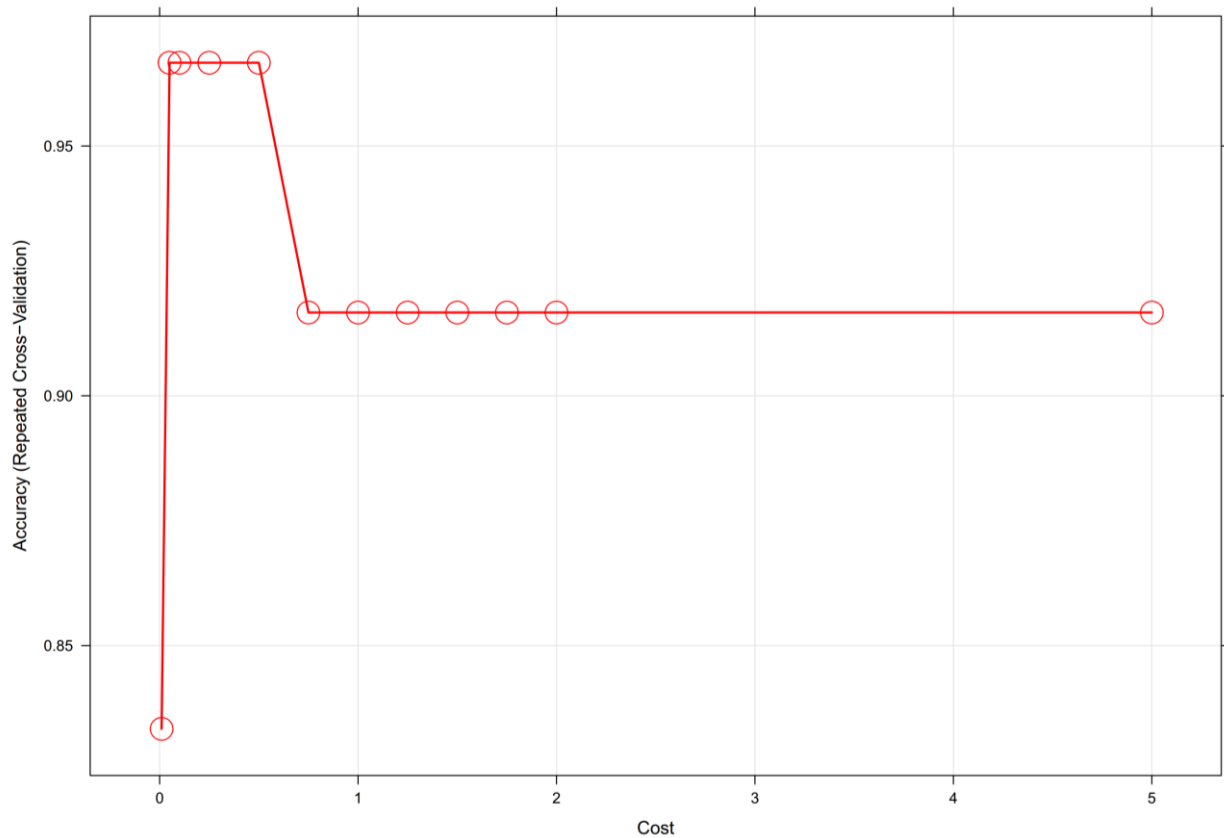

Figure 3. SVM Linear Kernel

Figure 3 shows the data of sum insured in Thailand with a 90% of accuracy and a tuning parameter of C, also known as the cost that defines the probable misclassifications. It imposes a penalty on the model to make a mistake, i.e., the greater value of C, the less likely the SVM linear kernel algorithm would misclassify a point [21]. In this scenario, caret generates an SVM linear classifier with C equal to 1. It can input several C values to find an optimal C value that optimizes model cross-validation accuracy. The following is the R code for SVM polynomial testing

```
# SVM Polynomial

library(readxl)
library(caret)
X3y_data1 <- read_excel("D:/SVM Code in R/SVM Polynomial Kernel/3y_data.xlsx",
            sheet = "Sheet2", range = "A1:E27")
View(X3y_data1)
str(X3y_data1)
intrain1 <- createDataPartition(y = X3y_data1$Lv, p= 0.7, list = FALSE)
training1 <- X3y_data1[intrain1,]
testing1 <- X3y_data1[-intrain1,]
dim(training1);
dim(testing1);
anyNA(X3y_data1)
summary(X3y_data1)
training1[["Lv"]] = factor(training1[["Lv"]])
trctrl1 <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
svm_Poly1 <- train(Lv ~., data = training1, method = "svmPoly",
```

```
            trControl=trctrl1,
            preProcess = c("scale"))
svm_Poly1
test_pred1 <- predict(svm_Poly1, newdata = testing1)
test_pred1
confusionMatrix(table(test_pred1, testing1$Lv))


grid1 <- expand.grid(C = c(0.5,0.6,0.7,0.8,0.9),degree=c(1,2,3,4), scale=c(2,4,8,10))


set.seed(123)
svm_Poly_Grid1 <- train(Lv ~., data = training1, method = "svmPoly",
            trControl=trctrl1,tuneGrid=grid1,preProcess = c("scale"))
sel.poly
svm_Poly_Grid1
plot(svm_Poly_Grid1, lwd=2, cex = 2.5, col = "red", bg = "black",  pch=21)
svm_Poly_Grid1$bestTune
test_pred_grid1<- predict(svm_Poly_Grid1, newdata = testing1)
test_pred_grid1
confusionMatrix(table(test_pred_grid1, testing1$Lv))
```

Thus the SVM classifiers using non-linear kernels, i.e., polynomial or radial basis functions can be produced using the mentioned code. The caret package could quickly compute the polynomial and the radial basis function SVM non-linear models (Figure 4).
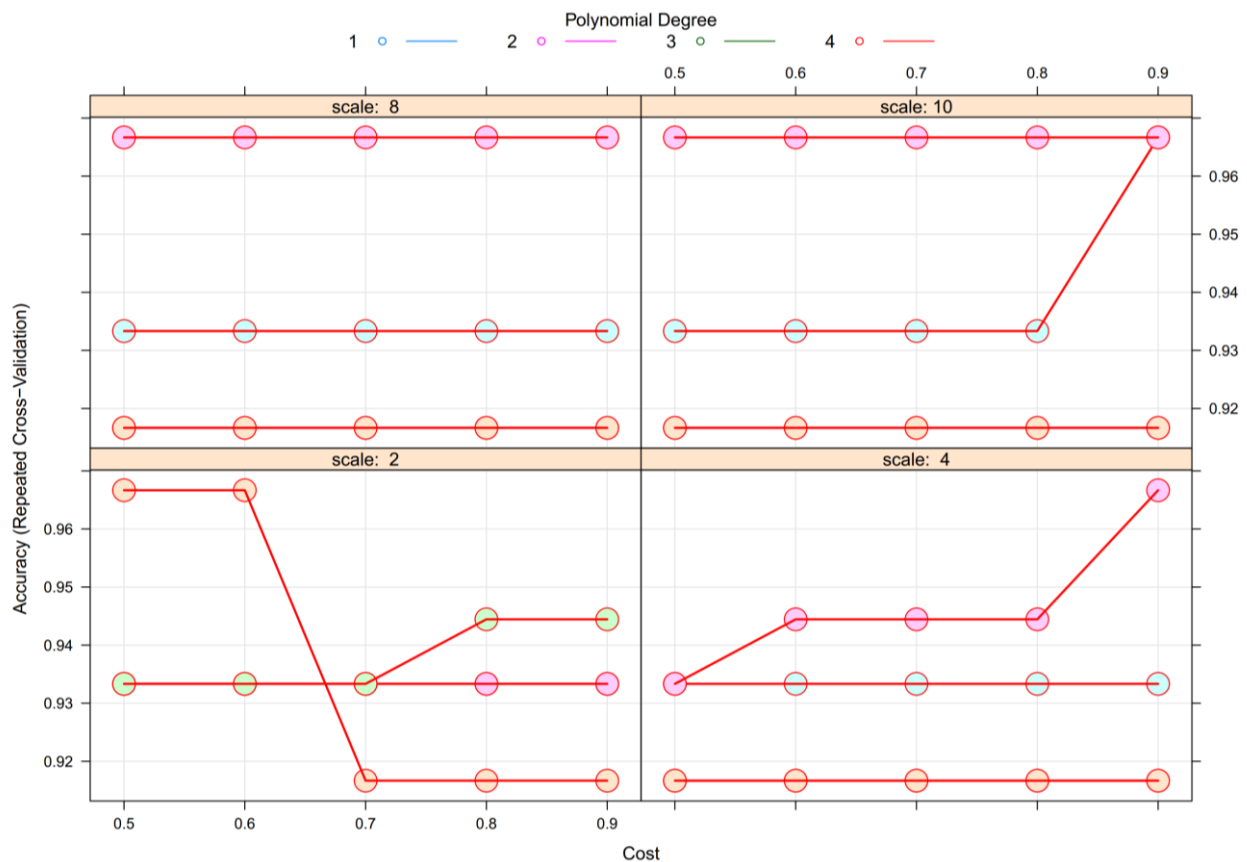

Figure 4. SVM Polynomial Kernel

The caret package automatically chooses the ideal model tuning parameters and optimizing model accuracy. In these tries, the scales are 2, 4, 8, 10 with degrees 1, 2, 3, 4 and the final values after executing the program are degree is equal to 1, scale is 0.1 and C is 0.25 with 92% accuracy [21]. The R code for SVM RBF testing is presented as follows:

```
# SVM RBF
library(readxl)
library(caret)
X3y_data1 <- read_excel("D:/SVM Code in R/SVM RBF Kernel/3y_data.xlsx",
             sheet = "Sheet2", range = "A1:E27")
View(X3y_data1)
str(X3y_data1)
intrain1 <- createDataPartition(y = X3y_data1$Lv, p= 0.7, list = FALSE)
training1 <- X3y_data1[intrain1,]
testing1 <- X3y_data1[-intrain1,]
dim(training1);
dim(testing1);
anyNA(X3y_data1)
summary(X3y_data1)
training1[["Lv"]] = factor(training1[["Lv"]])
trctrl1 <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
svm_Radial1 <- train(Lv ~., data = training1, method = "svmRadial",
             trControl=trctrl1,
             preProcess = c("center", "scale"),
             tuneLength = 100)
svm_Radial1
test_pred1 <- predict(svm_Radial1, newdata = testing1)
test_pred1
confusionMatrix(table(test_pred1, testing1$Lv))
grid1 <- expand.grid(sigma = c(0, 0.01, 0.04, 0.06, 0.08, 0.1, 0.5, 0.9),
             C = c(0, 0.1, 0.5, 0.75,
             1, 1.5, 2,5))
set.seed(123)
svm_Radial_Grid1 <- train(Lv ~., data = training1, method = "svmRadial",
             trControl=trctrl1,
             preProcess = c("center", "scale"),
             tuneGrid = grid1,
             tuneLength = 100)
svm_Radial_Grid1
plot(svm_Radial_Grid1, lwd=2, cex = 2.5, col = "red", bg = "black",  pch=21)
svm_Radial_Grid1$bestTune
test_pred_grid1<- predict(svm_Radial_Grid1, newdata = testing1)
test_pred_grid1
confusionMatrix(table(test_pred_grid1, testing1$Lv))
```

It routinely utilizes the caret package while doing SVM analysis with different kernels. The program seems efficient since it categorizes data to reduce the error and improve the accuracy.

In Figure 5, the tuning parameter sigma is an ideal model at 0.2714456 with C=16 and 96% of accuracy. Histograms and stem and leaf plots both assist in providing an overview of metrics of the central tendency and symmetry of observational data. The code for RMSE testing is presented in *R* as follows:
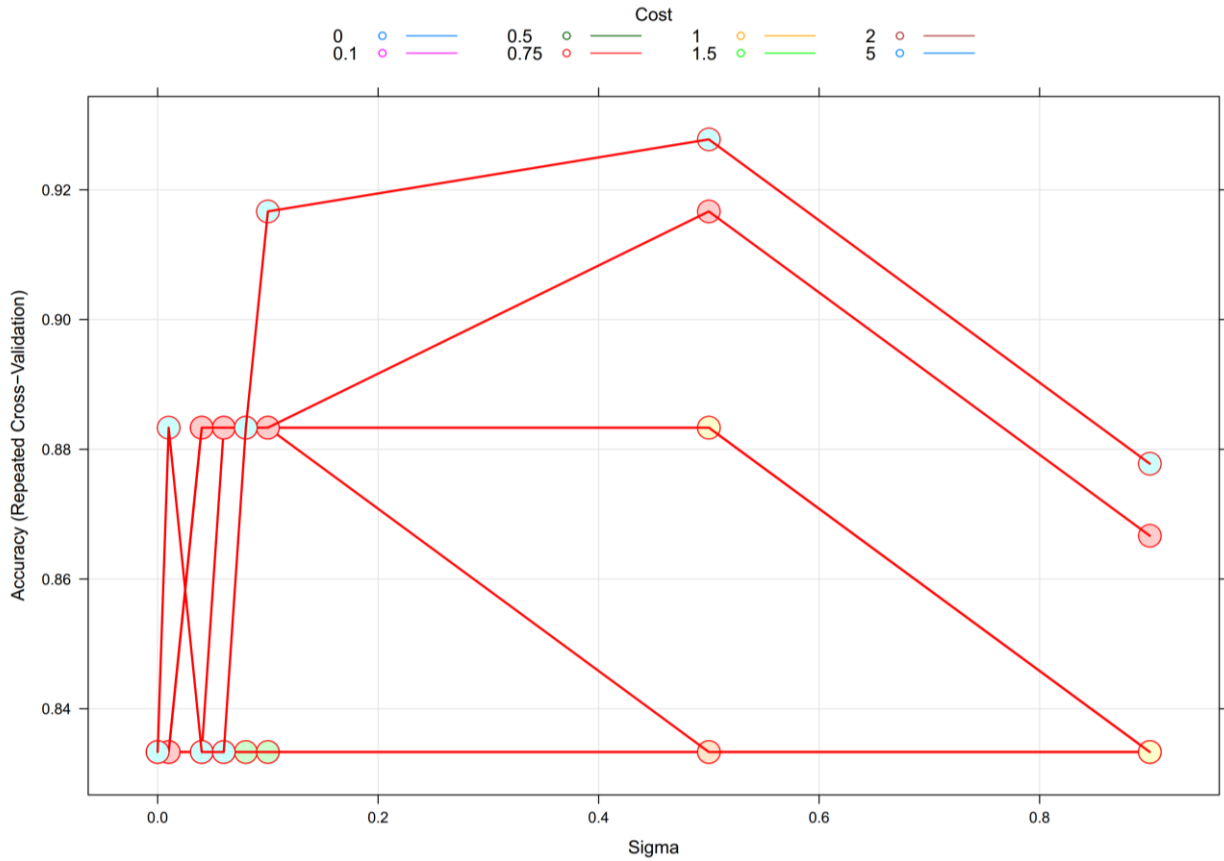
Figure 5. SVM RBF Kernel

```
# RMSE
library(caret)
library(caTools)
library(ggplot2)
library(lattice)
library(pROC)

kernelcomp <- read_excel("D:/SVM Code in R/3y_data.xlsx", sheet = "Sheet2", range = "A28:E45")
View(kernelcomp)
str(kernelcomp)

intrain <- createDataPartition(y = kernelcomp$year, p= 0.7, list = FALSE)
x <- kernelcomp[intrain,]
y <- kernelcomp[-intrain,]

dim(x);
dim(y);
anyNA(kernelcomp)
summary(kernelcomp)
x[["year"]] = factor(x[["year"]])
trctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3)

#Pre-Compute CV folds so it can use the same ones for all models
CV_Folds <- createMultiFolds(y, k = 10, times = 5)
df <- kernelcomp
#Using x, y arguments
```

```
L_model <- train(x = df , y = df$year, method='svmLinear')
P_model <- train(x = df , y = df$year, method='svmPoly')
R_model <- train(x = df , y = df$year, method='svmRadial')


#Fit a Linear SVM
L_model <- train(year ~.,  data = df, method="svmLinear")


#Fit a Poly SVM
P_model <- train(year ~., data = df, method="svmPoly")


#Fit a Radial SVM
R_model <- train(year ~., data = df, method="svmRadial")


#Compare 3 models:
resamps <- resamples(list(Linear = L_model, Poly = P_model, Radial = R_model))
summary(resamps)


cols = list(col=c("blue","green","red"), pch=c(2,4,8), lwd = 2)
bwplot(resamps, metric = "RMSE", par.settings = list(plot.symbol=cols,box.rectangle = cols, box.dot = cols, box.umbrella=cols))
```

Box and Whisker Plots, also known as BoxPlot or Box-Plot, are another graphical presentation that may summarize more specific information about the distribution of observed data values (box-plot). As the name implies, Box and Whisker is a form that consists of a Box (box) and a Whisker. A BoxPlot SVM based on each kernel is shown in Figure 6 of the Root Mean Square Error (RMSE).

BoxPlot graphically presents a summary of a sample distribution that can characterize the shape of the data distribution (skewness), central tendency measures and observational data spread (diversity). The results of Mean Absolute Error (MAE), RMSE and R-Squared are obtained from Figure 6.
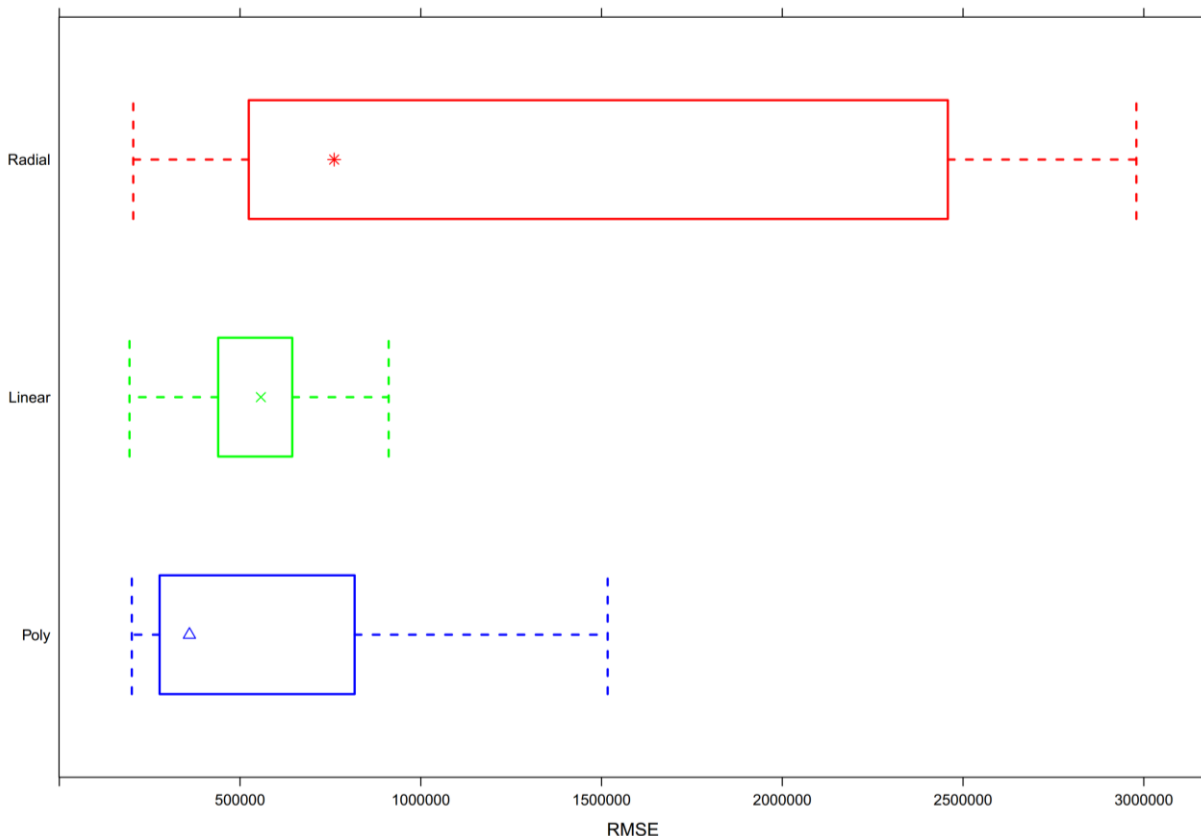

Figure 6. BloxPlot of RMSE for each of the Kernels

The code for checking the kernel comparisons in R is presented as follows

```
# RMSE vs. Density
 densityplot(resamps, metric = "RMSE", par.settings = list(superpose.line = list(col = c("blue","green","red"), lwd = 2)), plot.points=FALSE,
        auto.key=TRUE)
```

An RMSE simulation of density will then be studied using Table 1 by exhibiting the comparison between kernels as shown below (Figure 7).

Table 1. Statistical Test

Mean Absolute Error			
Test	**Linear**	**Polynomial**	**RBF**
Minimum	132382.3	185818.8	147140.2
Q1	297849.3	245008.6	366284.5
Median	363194.9	294245.8	571921.3
Mean	375530.5	393851.5	839716.9
Q3	451604.8	542913.3	1375325.6
Maximum	660683.5	929560.5	1939014.6
Root Mean Square Error			
Test	**Linear**	**Polynomial**	**RBF**
Minimum	194206.7	200609.8	204577.2
Q1	439472.3	277504.1	524122.5
Median	557629.4	360014.1	760343.5
Mean	535727.1	557193.3	1388152.9
Q3	644474.7	816996.6	2458012.4
Maximum	911203.4	1516963.0	2979495.5
R-Squared			
Test	**Linear**	**Polynomial**	**RBF**
Minimum	0.085548456	0.246117584	0.006402295
Q1	0.8738624	0.9188833	0.4533769
Median	0.9691322	0.9669956	0.6216420
Mean	0.8748125	0.8986097	0.5792569
Q3	0.9900580	0.9779445	0.7720732
Maximum	0.9999064	0.9975390	1.0000000

The comparison of the three kernels in Figure 7 based on RMSE and density concluded that the radial kernel is superior to the other kernels, and the SVM RBF kernel is the simulation approach with the highest accuracy.

4- Insurance Data Clustering

Classification using SVM with multiple related kernels is performed because SVM classification is the best method for doing in-depth research with current data to reduce mistakes. As a result, all of the data verification it has done jointly through simulations is to inform prudence and correctness. Further analysis may be performed by introducing premium data or modifying the categorization algorithm. The significant findings of this investigation are presented in Table 2.
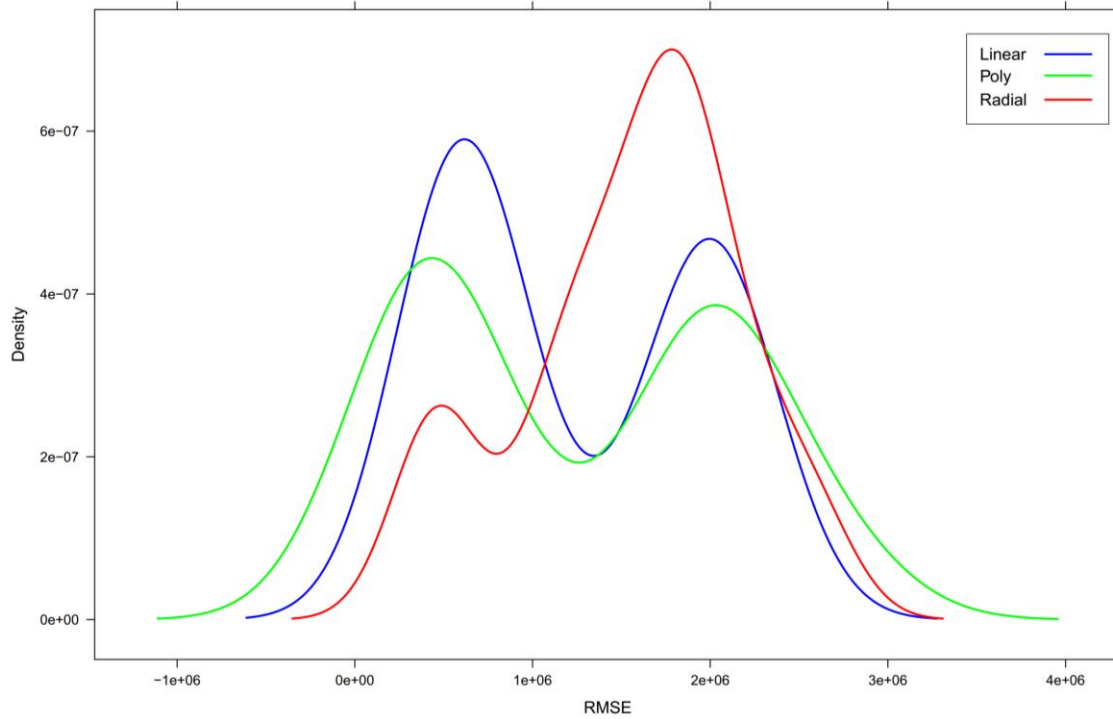
Figure 7. RMSE vs. Density based on Kernel Comparisons

Table 2. Sum Insured Classification

	Central Regions		
No.	Central Province	Sum Insured Level	Recommendation
1	Bangkok	0	Pay More
2	Kanchanaburi	1	Pay Less
3	Chanthaburi	1	Pay Less
4	Chachoengsao	1	Pay Less
5	Chonburi	0	Pay More
6	Chai Nat	1	Pay Less
7	Trat	1	Pay Less
8	Nakhon Nayok	1	Pay Less
9	Nakhon Pathom	1	Pay Less
10	Nonthaburi	0	Pay More
11	Pathum Thani	0	Pay More
12	Prachuap Khiri Khan	1	Pay Less
13	Prachinburi	1	Pay Less
14	Phetchaburi	1	Pay Less
15	Phra Nakhon Si Ayutthaya	1	Pay Less
16	Rayong	1	Pay Less
17	Ratchaburi	1	Pay Less
18	Lopburi	1	Pay Less
19	Samut Prakan	0	Pay More
20	Samut Songkhram	1	Pay Less
21	Samut Sakhon	1	Pay Less
22	Sa Kaeo	1	Pay Less
23	Saraburi	1	Pay Less
24	Sing Buri	1	Pay Less
25	Suphan Buri	1	Pay Less
26	Ang Thong	1	Pay Less

North Regions			
No.	**North Province**	**Sum Insured Level**	**Recommendation**
1	Kamphaeng Phet	1	Pay Less
2	Chiang Rai	0	Pay More
3	Chiang Mai	0	Pay More
4	Tak	1	Pay Less
5	Nakhon Sawan	0	Pay More
6	Nan	1	Pay Less
7	Phetchabun	1	Pay Less
8	Phrae	1	Pay Less
9	Phayao	1	Pay Less
10	Phichit	1	Pay Less
11	Phitsanulok	0	Pay More
12	Mae Hong Son	1	Pay Less
13	Lampang	1	Pay Less
14	Lamphun	1	Pay Less
15	Sukhothai	1	Pay Less
16	Uttaradit	1	Pay Less
17	Uthai Thani	1	Pay Less

Northeast Regions			
No.	**Northeast Province**	**Sum Insured Level**	**Recommendation**
1	Kalasin	1	Pay Less
2	Khon Kaen	0	Pay More
3	Chaiyaphum	1	Pay Less
4	Nakhon Phanom	1	Pay Less
5	Nakhon Ratchasima	0	Pay More
6	Buriram	1	Pay Less
7	Maha Sarakham	1	Pay Less
8	Mukdahan	1	Pay Less
9	Yasothon	1	Pay Less
10	Roi Et	1	Pay Less
11	Loei	1	Pay Less
12	Sisaket	1	Pay Less
13	Sakon Nakhon	1	Pay Less
14	Surin	1	Pay Less
15	Nong Khai	1	Pay Less
16	Nong Bua Lamphu	1	Pay Less
17	Amnat Charoen	1	Pay Less
18	Udon Thani	0	Pay More
19	Ubon Ratchathani	0	Pay More
20	Bueng Kan	1	Pay Less

South Regions			
No.	South Province	Sum Insured Level	Recommendation
1	Krabi	1	Pay Less
2	Chumphon	1	Pay Less
3	Trang	1	Pay Less
4	Nakhon Si Thammarat	0	Pay More
5	Narathiwat	1	Pay Less
6	Pattani	1	Pay Less
7	Phang Nga	1	Pay Less
8	Phatthalung	1	Pay Less
9	Phuket	0	Pay More
10	Yala	1	Pay Less
11	Ranong	1	Pay Less
12	Songkhla	0	Pay More
13	Satun	1	Pay Less
14	Surat Thani	0	Pay More

Insurance firms can utilize the findings of this categorization to determine whether provinces receive more or less of the sum guaranteed. Based on Table 2, for the Chiang Mai (0) area, residents with original domiciles in Chiang Mai must make high insurance payments. Meanwhile, residents with Lampang (1) identity cards pay low insurance premiums. The determination of the average to determine the area to pay high or low contributions is flexible and is determined by the insurance company that issues the average policy for the sum insured.

5- Conclusion

An error developed when the quantity of the sum insured in a specific location was not calculated accurately. In this case, the SVM method provided a more accurate and effective alternative step in classifying the sum insured, following a thorough examination of all three prominent SVM kernels. It concluded that the SVM approach should only use the radial kernel for future research. The essential aspect was that the radial kernel had a low RMSE and a high density, so the output of the resultant data analysis was close to the likelihood of correctness with the realization of the area of each zone. In Table 2, the results of data analysis for each region in Thailand were given. In the payment of the sum insured premium, a person's decision to pay more or less was based on the person's address on the identity card. Even if a person traveled to work in another area in Thailand, the contribution was still decided based on the original address of residence. For example, a student from Pattani (Thailand), whose area was at level 1, implied that the insurance coverage was small. If moving to Bangkok for studying, where the area was at level 0, then the sum insured charged to the student was still at level 1 because the original address on the resident's identity card was from the Pattani area. However, if a student changed a residential address permanently, the sum insured charged would be at level 0 according to the Bangkok area. This new study offered a detailed and accurate classification of the insurance data where the primary basis was that there were still some imperfections in the classification of data in Thailand. It tried to find renewable methods that were more effective and efficient for data clustering by presenting the findings in this paper. Further studies of this manuscript could be expanded by comparing the SVM method with other classification methods or by replacing the sum insured data with other insurance data, such as analysis of the insurance stock market.

6- Declarations

6-1- Author Contributions

Conceptualization, I.N., B.P., and S.N.; methodology, I.N., B.P., S.N., and U.F.-G.; software, I.N., B.P., S.N., and U.F.-G.; validation, I.N., B.P., S.N., and U.F.-G.; formal analysis, I.N., B.P., S.N., and U.F.-G.; investigation, I.N., B.P., S.N., and U.F.-G.; resources, I.N., B.P., and S.N.; data curation, I.N., B.P., S.N., and U.F.-G.; writing—original draft preparation, I.N., B.P., S.N., and U.F.-G.; writing—review and editing, I.N., B.P., S.N., and U.F.-G.; visualization, I.N., B.P., S.N., and U.F.-G.; supervision, B.P., S.N., and U.F.-G.; project administration, B.P., S.N., and U.F.-G.; funding acquisition, B.P., S.N., and U.F.-G. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

Data was obtained from Office of the Insurance Commission (OIC) in Thailand and available on request from the corresponding author.

6-3- Funding

6-4- Acknowledgements

6-5- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Tangcharoensathien, V., Tisayaticom, K., Suphanchaimat, R., Vongmongkol, V., Viriyathorn, S., & Limwattananon, S. (2020). Financial risk protection of Thailand's universal health coverage: Results from series of national household surveys between 1996 and 2015. International Journal for Equity in Health, 19(1), 1–12. doi:10.1186/s12939-020-01273-6.

[2] Paek, S. C., Meemon, N., & Wan, T. T. H. (2016). Thailand's universal coverage scheme and its impact on health-seeking behavior. SpringerPlus, 5(1), 1–16. doi:10.1186/s40064-016-3665-4.

[3] Suraratdecha, C., Saithanu, S., & Tangcharoensathien, V. (2005). Is universal coverage a solution for disparities in health care? Findings from three low-income provinces of Thailand. Health Policy, 73(3), 272–284. doi:10.1016/j.healthpol.2004.11.019.

[4] Carrin, G., Waelkens, M. P., & Criel, B. (2005). Community-based health insurance in developing countries: A study of its contribution to the performance of health financing systems. Tropical Medicine and International Health, 10(8), 799–811. doi:10.1111/j.1365-3156.2005.01455.x.

[5] Pauly, M. V., Zweifel, P., Scheffler, R. M., Preker, A. S., & Bassett, M. (2006). Private health insurance in developing countries. Health Affairs, 25(2), 369–379. doi:10.1377/hlthaff.25.2.369.

[6] Zanaty, E. A., & Afifi, A. (2011). Support vector machines (SVMs) with universal kernels. Applied Artificial Intelligence, 25(7), 575–589. doi:10.1080/08839514.2011.595280.

[7] Buchanna, G., Premchand, P., & Govardhan, A. (2022). Classification of Epileptic and Non-Epileptic Electroencephalogram (EEG) Signals Using Fractal Analysis and Support Vector Regression. Emerging Science Journal, 6(1), 138-150. doi:10.28991/ESJ-2022-06-01-011.

[8] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning, 112, 1-18. Springer, New York, United States. doi:10.1007/978-1-0716-1418-1.

[9] Shah, T. I., Milosavljevic, S., & Bath, B. (2017). Measuring geographical accessibility to rural and remote health care services: Challenges and considerations. Spatial and Spatio-Temporal Epidemiology, 21, 87–96. doi:10.1016/j.sste.2017.04.002.

[10] Sritart, H., Tuntiwong, K., Miyazaki, H., & Taertulakarn, S. (2021). Disparities in healthcare services and spatial assessments of mobile health clinics in the border regions of Thailand. International Journal of Environmental Research and Public Health, 18(20), 1–24. doi:10.3390/ijerph182010782.

[11] Rojjananukulpong, R., Ahmad, M. M., & Saqib, S. E. (2021). Disparities in physical accessibility among rural Thais under universal health coverage. American Journal of Tropical Medicine and Hygiene, 105(3), 837–845. doi:10.4269/ajtmh.20-1627.

[12] Marshall, A. I., Kantamaturapoj, K., Kiewnin, K., Chotchoungchatchai, S., Patcharanarumol, W., & Tangcharoensathien, V. (2021). Participatory and responsive governance in universal health coverage: An analysis of legislative provisions in Thailand. BMJ Global Health, 6(2), 1–10. doi:10.1136/bmjgh-2020-004117.

[13] Sumriddetchkajorn, K., Shimazaki, K., Ono, T., Kusaba, T., Sato, K., & Kobayashi, N. (2019). Universal health coverage and primary care, Thailand. Bulletin of the World Health Organization, 97(6), 415–422. doi:10.2471/BLT.18.223693.

[14] Yiengprugsawan, V., Carmichael, G. A., Lim, L. L. Y., Seubsman, S., & Sleigh, A. C. (2010). Has universal health insurance reduced socioeconomic inequalities in urban and rural health service use in Thailand? Health and Place, 16(5), 1030–1037. doi:10.1016/j.healthplace.2010.06.010.

[15] Yiengprugsawan, V., Seubsman, S. ang, & Sleigh, A. C. (2012). Health, Well-being, and Social Indicators among Monks, Prisoners, and Other Adult Members of an Open University Cohort in Thailand. Journal of Religion and Health, 51(3), 925–933. doi:10.1007/s10943-010-9410-3.

[16] Yiengprugsawan, V., Lim, L. L. Y., Carmichael, G. A., Seubsman, S. A., & Sleigh, A. C. (2009). Tracking and Decomposing Health and Disease Inequality in Thailand. Annals of Epidemiology, 19(11), 800–807. doi:10.1016/j.annepidem.2009.04.009.

[17] Puenpatom, R. A., & Rosenman, R. (2008). Efficiency of Thai provincial public hospitals during the introduction of universal health coverage using capitation. Health Care Management Science, 11(4), 319–338. doi:10.1007/s10729-008-9057-8.

[18] Chandanachulaka, S. (2020). Thailand: Country report on children's environmental health. Reviews on Environmental Health, 35(1), 71–77. doi:10.1515/reveh-2019-0082.

[19] Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12(1), 1–8. doi:10.1186/1471-2105-12-77.

[20] SIB Swiss Institute of Bioinformatics (2022). PROC: Display and analyze ROC curves in R and S+. Available online: https://web.expasy.org/pROC/ (accessed on March 2022).

[21] Campbell, J.J. (2022). SVM with CARET: Loading required R packages. Available online: https://rpubs.com/uky994/593668 (accessed on February 2022).