PREDICTIVE ANALYSIS ON STANDARD HEALTH DEFICIENCY WITH SUPPORT VECTOR MACHINE USING CONFUSION MATRIX

By

Md. Humayan Kabir

ID: 183-44-173

This report presented in fulfillment of the requirements for the Degree of M.Sc. in Software Engineering

Department of Software Engineering
Daffodil International University

SPRING 2019

LETTER OF APPROVAL

This **Thesis** titled "**Predictive analysis on Standard Health Deficiency with Support Vector Machine using Confusion Matrix**", submitted by **Md. Humayan Kabir**, **ID No.: 183-44-173** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the fulfillment of the requirements for the degree of M.Sc. in Software Engineering and approved as to its style and contents.

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Dr. Md. Asraf Ali — **Internal Examiner 1**
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Mr. Mohammad Khaled Sohel — **Internal Examiner 2**
Asst. Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Prof. Dr. Md. Nasim Akhtar — **External Examiner**
Professor
Department of Computer Science and Engineering
Faculty of Electrical and Electronic Engineering
Dhaka University of Engineering & Technology, Gazipur

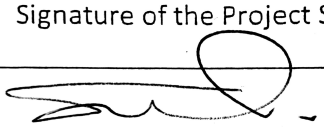
ACKNOWLEDGEMENT

I am grateful to my creator for giving me the opportunity to complete this research work and learn so much. I am thankful to my research supervisor, M. Khaled Sohel, for providing careful guidance starting from selecting the research scope to successfully finalizing the research work. I would also like to thank K. M. Imtiaz-Ud-Din for his valuable comments which was always insightful. Finally, I want to express my gratitude to Professor Dr. Touhid Bhuiyan, head of the Software Engineering faculty, for inspiring us in all means.

THESIS DECLARATION

I hereby declare that, this thesis report is done by me, Md. Humayan Kabir, under the supervision of M. Khaled Sohel, Assistant Professor, Department of Software Engineering, Daffodil International University, in fulfillment of my original work for the degree of M.Sc. in Software Engineering. I am also declaring that neither this thesis nor any part therefore has been submitted else here for the award of M.Sc. or any degree.

Signature of the Project Supervisor



Supervised by

M. Khaled Sohel
Assistant Professor
Department of Software Engineering
Daffodil International University

Submitted by

Md. Humayan Kabir
ID: 183-44-173
Department of Software Engineering
Daffodil International University

ABSTRACT

Standard health deficiency prediction is an essential study not only for people involved in the sector but also for researchers to find new prospect of applications. This research will particularly help preventing deficiency cases before hand with the help of proven methods and measures.

Review of literature suggests a lack of research carried out for studying standard health predictions. As we could not find many researches for forecasting the standard health deficiency, therefore, the aim of this research is to predict standard health deficiency based on scope and severity, and health inspection cycle with different combination of training and testing data set and measure the effectiveness using Confusion Matrix.

We have collected data from an open project, Health Deficiencies, in Healthdata.gov. Out of 354,271 rows were available for download, we had considered first 10,000 rows for this analysis. And for selecting this data row count, different data visualizations helped to understand the data plotting pattern.

For research methodology, we have used Support Vector Machine (SVM) to predict the standard health deficiency. We developed the models with given dataset, trained the models and analyzed against both training and testing datasets along with data plotting visualizations. Finally, by analyzing the in-sample and out-of-sample datasets with Confusion Matrix, we have used Accuracy, Precision, Recall and F-Measures for finding a better dataset combination for training and testing to understand the effectiveness of the prediction model.

Support Vector Machine is a powerful machine learning algorithm that we used in this context for predicting standard health deficiency. Confusion matrix along with different measurements (accuracy, precision, recall & f1-score) helped to make effective comparison between different combination of ratio for training and testing dataset and measured the performance for in-sample and out-of-sample dataset.

Our research findings indicate that considering 2 different analyses (A & B) with different ratio of training and testing datasets helped to understand the impact of training on the SVM algorithm where we found that more training data can produce much better predictions. Confusion matrix measurements had proved to be every effective to measure the performance of the SVM algorithm. Finally, for this health deficiency prediction context, Support Vector Matrix is proven to a decent predictor.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF ANALYSIS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Explanation
SVM	Support Vector Machine
RBF	Radial Bias Function
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Background

In USA, hospitals and healthcare providers use different systems and standards for collecting data. Scope and severity, health inspection cycle, standard health deficiency measurements are the few key attributes which are considered essential in the process. Predicting standard health deficiency based on scope and severity, and health inspection cycle is the objective of this research where Support Vector Machine is applied, and Confusion Matrix is used to find a comparative analysis among different combination of training and testing dataset to find an optimal combination.

1.1.1 Scope and severity

A different grading or rating system has been applied by hospitals and nursing homes for Federal surveys in USA. Unlike other healthcare facility types, nursing homes have a different system, a Severity and Scope level from A to L which have standards and conditions of participation for hospitals and other healthcare provider types.

Severity and Scope, which is a system of rating the seriousness of deficiencies. The term Deficiency refers a regulatory requirement for failing to meet any facilities or put officially during the survey as “is found not met”.

This survey system is national for Severity and Scope which is being used by all survey agencies around USA and Centers for Medicare and Medicaid Service (CMS) while Medicare and Medicaid certification survey conducting for nursing homes.

Considering each deficiency, level of harm (severity) to the resident is first measured and then within the nursing home scope of problem is determined by the survey team. Then an alphabetical Severity and Scope value is assigned starting from A to L as a mark of deficiency where "A" rating is least serious and "L" rating is most serious which is like a grade report.

Immediate jeopardy to resident health & safety	Level 4	J	K	L
Actual harm that is not immediate jeopardy	Level 3	G	H	I
No actual harm, with potential for more than minimal harm	Level 2	D	E	F
No actual harm, with potential for minimal harm	Level 1	A	B	C
		Isolated	Pattern	Widespread

Figure 1. Scope Severity Code chart

The survey team first decides the “Level” where deficiency falls in.

Level 1 (the green row) denotes no actual harm though it is potential for minimal harm. To simplify, the situation is with negative impact no more than minor.

Level 2 (the yellow row) again denotes no actual harm though it is potential for more than minimal harm, however, this is not an immediate jeopardy. To simplify, because of the deficient practice, the situation is with no more than minimal physical, mental, and/or psychosocial discomfort. However, to maintain and/or reach ones highest practicable physical, mental and/or psychosocial wellbeing, a potential to compromise the his/her ability.

Level 3 (the orange row) denotes actual harm but with no immediate jeopardy. To simplify, it is a negative or bad result. The resident’s ability to maintain &/or reach greatest practicable well-being was compromised or ended up being clinically compromised, or experienced deterioration or harm.

Level 4 (the red row) denotes serious injury, harm, impairment, or death from the facility’s deficient practice, therefore, immediate action/correction is required. Situation like this, the facility’s practice pattern establishes a reasonable predictability degree of similar actions, situation, practices, or incidents occurring in the future if they do not fix it right now and anything in this row is called Immediate Jeopardy.

After the survey team determining the severity level, then they focus on the scope determining that means how prevalent or widespread the problem is for people, staff and area it affected.

Isolated (the first column) means:

- Limited number of residents are affected
- Limited number of staffs are involved
- An occasional occurrence of the situation or in a very limited number of locations

Pattern (the second column) means

- More than limited number of residents are affected
- More than limited number of staffs are involved
- Situation has occurred in several locations
- Repeated occurrences for the same resident(s) or same deficient practice

Widespread (the third column) means

- The deficiency causing problems are pervasive in the facility or
- A systemic failure which has the potential of affecting a large portion or all the facility's residents

Therefore, considering all these rules and standards together this is the explanation of the scores (grades)

A – **Isolated/Potential for minimal harm** – This deficiency has the potential for causing no more than a minor impact on the resident(s). Least serious rating and is isolated to the fewest number of residents, staff, or occurrences.

B – **Pattern/Potential for minimal harm** – This deficiency has the potential for causing no more than minor negative impact on the resident(s) and was not found to be throughout the facility. Least serious deficiency but affects more than a limited number of residents, staff, or occurrences.

C – **Widespread/Potential for minimal harm** – This deficiency has the potential to affect a large portion or all the residents but has the potential for causing no more than a minor negative impact on the resident(s). Least serious deficiency but was found to be widespread throughout the facility and/or has the potential to affect a large portion or all the residents.

D – **Isolated/Minimal harm or potential for actual harm** – This deficiency is one that results in minimal discomfort to the resident or has the potential to negatively affect the person's ability to achieve his/her highest functional status. This is a less serious (but not lowest level) deficiency and is isolated to the fewest number of residents, staff, or occurrences.

E – **Pattern/Minimal harm or potential for actual harm** – This deficiency is one that results in minimal discomfort to the resident or has the potential (not yet realized) to negatively affect the person's ability to achieve his/her highest functional status. This deficiency was not found to be throughout the facility. This is a less serious (but not lowest level) deficiency and affects more than a limited number of residents, staff, or occurrences.

F – **Widespread/Minimal harm or potential for actual harm** – This deficiency is one that results in minimal discomfort to the resident or has the potential (not yet realized) to negatively affect the residents' ability to achieve his/her highest functional status. This is a less serious (but not lowest level) deficiency but was found to be widespread throughout the facility and/or has the potential to affect a large portion or all the residents.

G – **Isolated/Actual harm** – This deficiency results in a negative outcome that has negatively affected the person's ability to achieve his/her highest functional status. This is a more serious deficiency but is isolated to the fewest number of residents, staff, or occurrences.

H – **Pattern/Actual harm** – This deficiency results in a negative outcome that has negatively affected the person's ability to achieve his/her highest functional status. This deficiency was not found to be throughout this facility. This is a more serious deficiency and affects more than a limited number of residents, staff, or occurrences.

I – **Widespread/Actual harm** – This deficiency results in a negative outcome that has negatively affected the residents' ability to achieve his/her highest functional status. This is a

more serious deficiency that was found to be widespread throughout the facility and/or has the potential to affect a large portion or all the residents.

J – **Isolated/Immediate Jeopardy** – This deficiency is one which places the residents in immediate jeopardy as it has caused (or is likely to cause) serious injury, harm, impairment, or death to a resident receiving care in the facility. Immediate corrective action is necessary when this deficiency is identified. This is the most serious deficiency although it is isolated to the fewest number of residents, staff, or occurrences.

K – **Pattern/Immediate Jeopardy** – This deficiency is one which places the residents in immediate jeopardy as it has caused (or is likely to cause) serious injury, harm, impairment, or death to a resident receiving care in the facility. Immediate corrective action is necessary when this deficiency is identified. This deficiency was not found to be throughout the facility. This is the most serious deficiency and affects more than a limited number of residents, staff, or occurrences.

L – **Widespread/Immediate Jeopardy** – This deficiency is one which places the residents in immediate jeopardy as it has caused (or is likely to cause) serious injury, harm, impairment, or death to a resident receiving care in the facility. Immediate corrective action is necessary when this deficiency is identified. This is the most serious deficiency and was found to be widespread throughout the facility and/or has the potential to affect a large portion or all the residents.

1.1.2 Health inspection cycle

Based on outcomes from state health inspections, the health Inspections cycle is measure and based on the number, scope, and severity of deficiencies identified during the three most recent annual inspection surveys, the ratings for the health inspection domain are calculated. Scope and severity are the weight factors for all deficiency findings. The number of revisits required is also considered to ensure the correction of that deficiency's identification during the health inspection survey.

According to their scope and severity, Health Inspection Results, points are assigned to individual health deficiencies, therefore, widespread and more serious deficiencies receive more points. For instance, a "past non-compliance" status deficiency with "immediate jeopardy" (i.e., J-, K- or L-level) is associated with a G-level deficiency are assigned. Here, Life Safety survey deficiencies are not included in the rating calculations.

1.1.3 Health deficiencies

For body development and preventing diseases, body requires several different vitamins and minerals that are crucial which are often referred as micronutrients and are not produced naturally in the body, therefore, need to get them from the diet. When the body doesn't absorb or get from food the necessary amount of a nutrient, a nutritional deficiency occurs. A variety of health problems can result from these deficiencies including digestion problems, skin disorders, stunted or defective bone growth, and even dementia.

Depends on one's age, the amount of each nutrient should be consumed. In USA, many foods that can be bought in the grocery store like cereals, bread, and milk are fortified with

nutrients which are very helpful to prevent nutritional deficiency. However, our body is unable to absorb certain nutrients sometimes even if we are consuming them, therefore, it is possible to be deficient in any of the nutrients our body needs.

1.1.4 Support vector machine (SVM)

A Support Vector Machine (SVM) algorithm is a discriminative classifier used for separating hyperplane. To simplify, provided with training data, this algorithm results an optimal hyperplane that used for categorizing new examples.

This algorithm is based on decision plant which define decision boundaries. A set of objects having different class membership is being separated by decision plane. A schematic example is shown in the illustration Figure-2. In this instance, the objects belong either to class GREEN or RED. On the right side, a separating line defines a boundary of which all objects are GREEN, and all objects are RED to the left where any new object (white circle) falling to the right is labeled, either classified as GREEN or RED.

Therefore, the objective of the support vector machine algorithm is to identify a hyperplane in an N-dimensional space where “N” presents the number of features that distinctly classifies the data points. There are many possible hyperplanes that could be chosen to separate the two classes of data points. The objective is to find a plane which has the maximum margin like the maximum distance between data points of both classes.

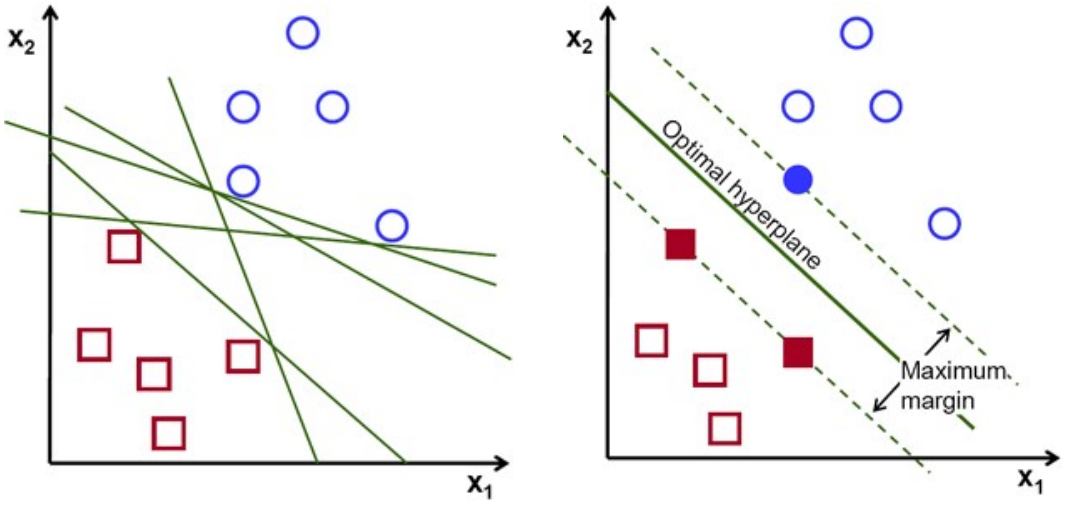


Figure 2. Possible SVM hyperplanes

Hyperplanes and Support Vectors

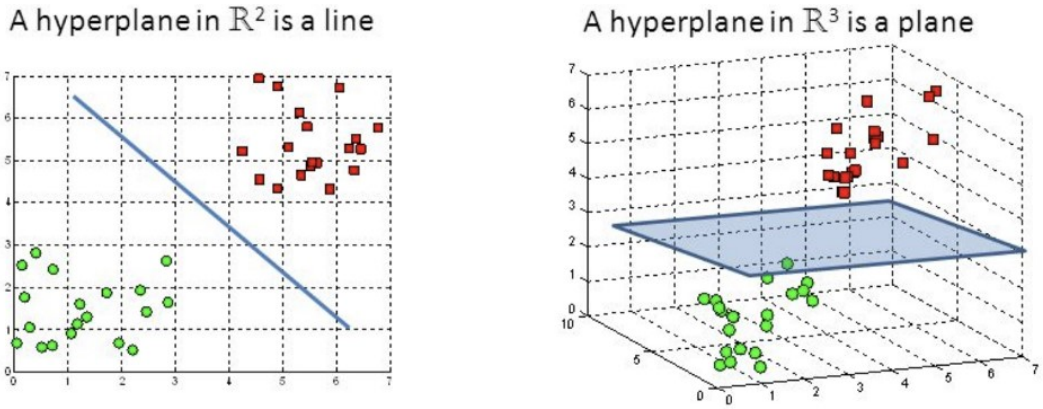


Figure 3. SVM Hyperplanes in 2D and 3D feature space

To help classifying the data points, Hyperplanes are decision boundaries. Data points can be attributed to different classes by falling on either side of the hyperplane. Moreover, the number of features depends upon the dimension of the hyperplane. The hyperplane is just a line

if the number of input features is 2. The hyperplane becomes a two-dimensional plane if the number of input features is 3. Nonetheless, when the number of features exceeds 3, it becomes difficult to imagine.

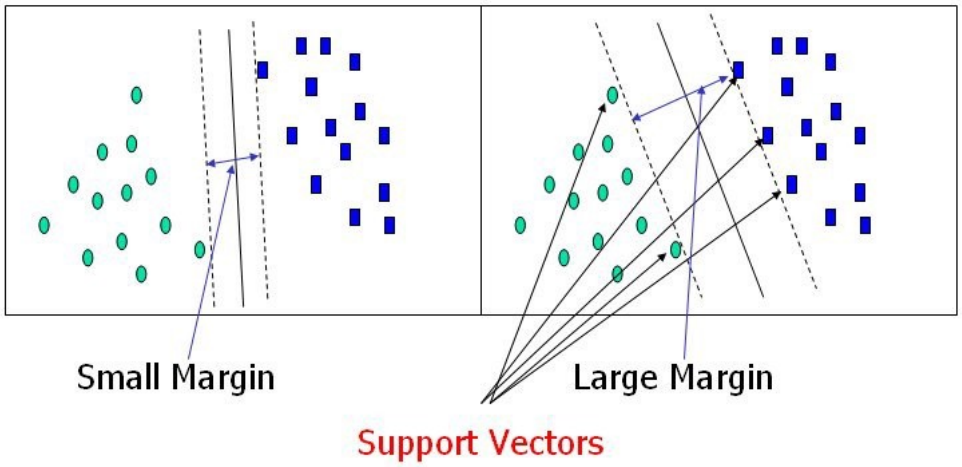


Figure 4. Support Vectors

The closer data points to the hyperplane and influence the position and orientation of the hyperplane are the Support vectors. We maximize the margin of the classifier by using these vectors.

1.1.5 Confusion matrix

A confusion matrix is a tabular representation that often used to describe the classification model performance on a set of testing data. Where the related terminology can be confusing, the confusion matrix itself is relatively simple to understand.

Where output can be two or more classes, it is a performance measurement for machine learning classification problem. A tabular representation of 4 different combinations of predicted and actual values, is extremely useful for measuring Recall, Precision, Specificity and Accuracy.

n=165	**Predicted: NO**	**Predicted: YES**
Actual: NO	50	10
Actual: YES	5	100

Figure 5. Confusion matrix example A

We can have information from Figure-5

- "YES" and "NO" are two possible predicted classes where if we predict the presence of a disease, for example, "YES" would mean they have the disease, and "NO" would mean they don't have the disease.
- The classifier made total 165 predictions.
- The classifier predicted "YES" 110 times, and "no" 55 times out of those 165 cases,
- 60 patients in the sample do not the disease where 105 patients have.

Here are definitions of the most basic terms

- True Positives (TP): we predicted YES, they have the disease, and they do have the disease.
- True Negatives (TN): We predicted NO, and they don't have the disease.

- False Positives (FP): We predicted YES, but they don't actually have the disease, also known as Type I error.
- False Negatives (FN): We predicted NO, but they actually do have the disease, also known as Type II error

Here, by adding the terms to the confusion matrix, and adding the row and column totals:

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Figure 6. Confusion matrix example B

For a binary classifier, below measurements are commonly computed from a confusion matrix

- **Accuracy**: Overall, how often is the classifier correct?
  - (TP+TN)/total = (100+50)/165 = 0.91

- **Misclassification Rate**: Overall, how often is it wrong?
    - (FP+FN)/total = (10+5)/165 = 0.09
    - equivalent to 1 minus Accuracy
    - also known as "Error Rate"

- **True Positive Rate**: When it's actually yes, how often does it predict yes?
    - TP/actual yes = 100/105 = 0.95
    - also known as "Sensitivity" or "Recall"

- **False Positive Rate**: When it's actually no, how often does it predict yes?
    - FP/actual no = 10/60 = 0.17

- **True Negative Rate**: When it's actually no, how often does it predict no?
    - TN/actual no = 50/60 = 0.83
    - equivalent to 1 minus False Positive Rate
    - also known as "Specificity"

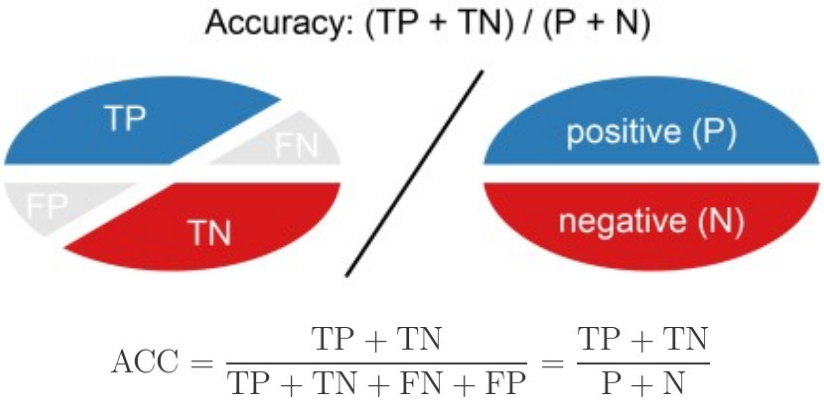
- **Precision**: When it predicts yes, how often is it correct?
    - TP/predicted yes = 100/110 = 0.91

- **Prevalence**: How often does the yes condition actually occur in our sample?
    - actual yes/total = 105/165 = 0.64

1.1.6 Accuracy

Accuracy (ACC) is calculated as the total number of the dataset is divided by the number of all correct predictions. The worst accuracy is 0.0 where the best accuracy is 1.0.

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$



$$ACC = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}$$

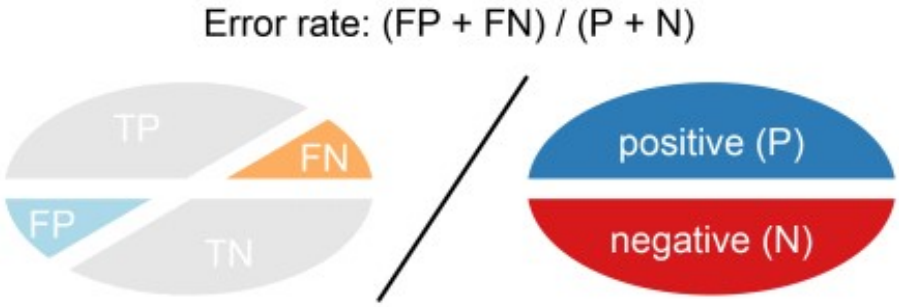
Figure 7. Accuracy equation

Accuracy is calculated as the total number of two correct predictions (TP + TN) divided by the total number of a dataset (P + N).

Nonetheless, accuracy assumes equal costs for both kinds of errors. A 99% accuracy can be excellent, good, mediocre, poor or terrible depending upon the problem at the same time.

1.1.7 Misclassification or error rate

The number of all incorrect predictions divided by the total number of the dataset is the Error rate (ERR) calculation. The worst is 1.0 whereas the best error rate is 0.0.



$$\mathrm{ERR} = \frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}} = \frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{P} + \mathrm{N}}$$

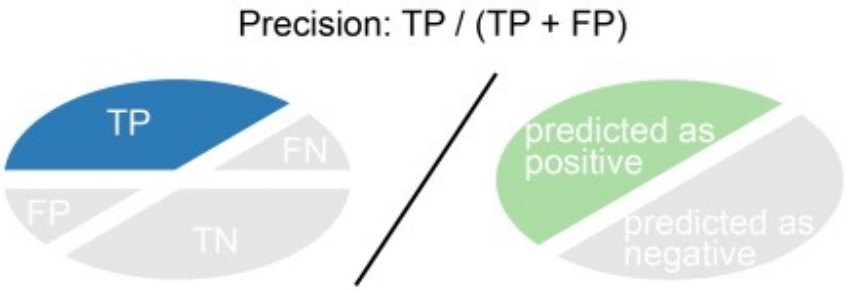
Figure 8. Error rate equation

Error rate is calculated as the total number of two incorrect predictions (FN + FP) divided by the total number of a dataset (P + N).

1.1.8 Precision

Precision or Positive predictive value is calculated as the number of correct positive predictions divided by the total number of positive predictions. It is also referred as positive predictive value or PPV where the best precision is 1.0 and the worst is 0.0.

$$\textbf{\textit{Precision}} = \frac{TP}{TP + FP}$$



$$PREC = \frac{TP}{TP + FP}$$

Figure 9. Precision equation

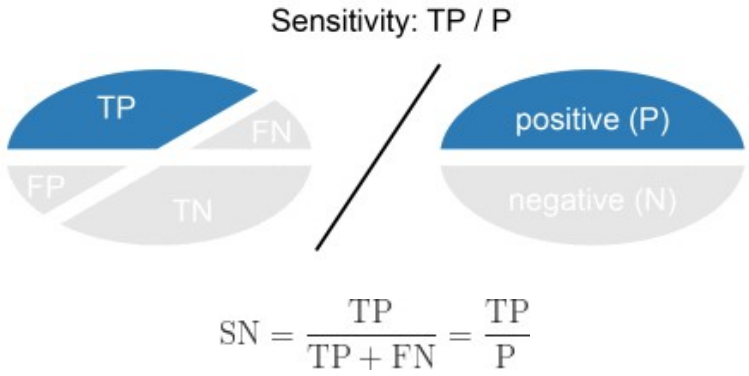
Precision is calculated as the number of correct positive predictions (TP) divided by the total number of positive predictions (TP + FP).

We divide the total number of correctly classified positive to get the value of precision where higher Precision indicates a greater number of positive or small number of FP.

1.1.9 Recall

Recall, sensitivity or True Positive rate is referred as the number of correct positive predictions divided by the total number of positives which also called recall (REC) or true positive rate (TPR). The worst is 0.0 whereas the best sensitivity is 1.0.

$$\textit{Recall} = \frac{TP}{TP + FN}$$



$$SN = \frac{TP}{TP + FN} = \frac{TP}{P}$$

Figure 10. Recall equation

Sensitivity is calculated as the number of correct positive predictions (TP) divided by the total number of positives (P).

The ratio of the total number of correctly classified positive examples divide to the total number of positive examples are refereed as Recall. High Recall indicates the class is correctly recognized or small number of FN. Out of all the positive classes, how much we predicted correctly should be high as possible.

High recall, low precision: Indicates most of the positive examples are correctly recognized with low FN but there are a lot of false positives.

Low recall, high precision: Indicates a lot of miss in positive examples or high FN but those we predict as positive are indeed positive or low FP.

1.1.10 F-measure

F-measure is a weighted average of the true positive rate (recall) and precision. To compare two models with low precision and high recall or vice versa is difficult to compare. Recall and Precision are measured at the same time with the help of F-score. By punishing the extreme values more, it uses Harmonic Mean in place of Arithmetic Mean.

F-measure helps to have a measurement that represents Recall and Precision. Harmonic Mean is being used in place of Arithmetic Mean as it punishes the extreme values more, therefore, F-Measure is always be closer to the smaller value of Precision or Recall.

$$F\text{-}measure = \frac{2*Recall*Precision}{Recall + Precision}$$

Figure 11. F-measure equation

1.2 Motivation of the research

Research on predictive analysis in medical section is growing ever more rapidly and opening new opportunities. USA federal government has several systems for collecting hospital and healthcare data. Forecasting standard health deficiency based on scope and severity, and health inspection cycle has a lot potentiality in this sector.

Unfortunately, we could not find many researches for forecasting the standard health deficiency, therefore, the aim of this research is to predict standard health deficiency based on scope and severity, and health inspection cycle with different combination of training and testing data set and measure the effectiveness using Confusion Matrix.

1.3 Problem statement

To point out the need for further understanding and investigation, we failed to find enough researches for predicting standard health deficiency. Lack of reliable source of data, not having enough data coverage, not considering effective factors as a mean of research are few of the many short comings that can be found in this limited research scope.

1.4 Research questions

The research questions are following to determine efficiency among different dataset combinations

- Are there any relationships between dependent (scope and severity) and independent variables (scope and severity, and health inspection cycle)?
- Is it possible to make future-oriented predictions considering the independent variables (scope and severity, and health inspection cycle)?
- Is Support Vector Machine an effective algorithm in this context?
- What is the set of variables that can define this forecasting in a more meaningful manner?

1.5 Research objective

As a result, the research objectives are

- To predict standard health deficiency based on scope and severity, and health inspection cycle.
- To find a suitable dataset combination for training and testing.

1.6 Research scope

We had identified few research gaps and based on these lacking we chose Support Vector Machine algorithm to predict the standard health deficiency. After applying the algorithm with the different data set both in training and testing phase, we had used Confusion Matrix to analyse the effective of these predictions. At the end, we compared among these different predictions to find an effective combination.

1.7 Thesis organization

In the following chapters, we discussed the literature including the research gap, described the research methodology along with data description, mentioned the results and discussions and finally discussed the conclusion with recommendations with our research findings, limitations and directions.

CHAPTER 2

LITERATURE REVIEW

CHAPTER 2

LITERATURE REVIEW

2.1 Previous literature

Research statement on health deficiencies prediction, Support Vector Machine (SVM) and Confusion Matrix are discussed in this literature review section.

2.1.1 Previous research on health deficiencies prediction

Severity and Scope is a system of rating the seriousness of deficiencies. The term Deficiency refers a regulatory requirement for failing to meet any facilities or put officially during the survey as "is found not met" (Centers for Medicare and Medicaid Service CMS, 2019).

Table 1
Health Inspection Score: Weights for Different Types of Deficiencies

Severity	Scope		
	Isolated	Pattern	Widespread
Immediate jeopardy to resident health or safety	**J** 50 points* (75 points)	**K** 100 points* (125 points)	**L** 150 points* (175 points)
Actual harm that is not immediate jeopardy	**G** 20 points	**H** 35 points (40 points)	**I** 45 points (50 points)
No actual harm with potential for more than minimal harm that is not immediate jeopardy	**D** 4 points	**E** 8 points	**F** 16 points (20 points)
No actual harm with potential for minimal harm	**A** 0 point	**B** 0 points	**C** 0 points

Figure 12. Health Inspection Score chart

According to their scope and severity, Health Inspection Results, points are assigned to individual health deficiencies, therefore, widespread and more serious deficiencies receive more points. For instance, a “past non-compliance” status deficiency with “immediate jeopardy” (i.e., J-, K- or L-level) is associated with a G-level deficiency are assigned. Here, Life Safety survey deficiencies are not included in the rating calculations (Centers for Medicare and Medicaid Service CMS, 2019).

For body development and preventing diseases, body requires several different vitamins and minerals that are crucial which are often referred as micronutrients and are not produced naturally in the body, therefore, need to get them from the diet. When the body doesn’t absorb or get from food the necessary amount of a nutrient, a nutritional deficiency occurs. A variety of health problems can result from these deficiencies including digestion problems, skin disorders, stunted or defective bone growth, and even dementia (Centers for Medicare and Medicaid Service CMS, 2019).

To rank deficiencies, nursing home surveyors use this matrix. The “scope” of the deficiency refers to the number of affected residents where the “severity” of the deficiency refers to the degree of harm, while (Allison Kite, 2016).

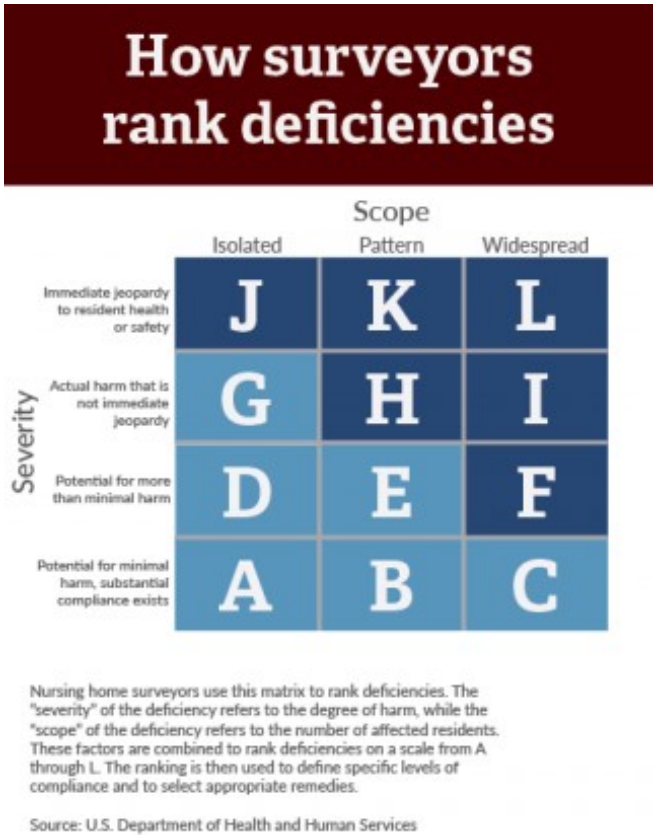


Figure 13. Deficiencies raking chart

2.1.2 Previous research on SVM

In 1992, Support Vector Machine (SVM) was first heard, and later introduced by Boser, Guyon, and Vapnik in COLT-92. This algorithm is a set of related supervised learning methods which are used for classification and regression and belong to a family of linear classifiers generalizations. To simplify, SVM is a regression and classification prediction tool to use in machine learning theory for maximizing accuracy prediction while avoiding automatically over-fit to the data. (Vikramaditya Jakkula, Washington State University).

Various ongoing researches have reported that in terms of classification accuracy the Support Vector Machine in general can deliver higher performance than the most other data classification algorithms. This algorithm is good for text categorization, hand-written digit recognition, tone recognition, image classification and object detection, micro-array gene expression data analysis, data classification prediction (DURGESH K. SRIVASTAVA, LEKHA BHAMBHU, 2009).

2.1.3 Previous research on confusion matrix

A confusion matrix is a tabular representation that often used to describe the classification model performance on a set of testing data. Where the related terminology can be confusing, the confusion matrix itself is relatively simple to understand. Where output can be two or more classes, it is a performance measurement for machine learning classification problem. A tabular representation of 4 different combinations of predicted and actual values, is extremely useful for measuring Recall, Precision, Specificity and Accuracy (A. K. Santra, C. Josephine Christy, 2012).

The Geometric Mean of Precision or Recall Gmeasure generally normalizes TP to the Geometric Mean of Real Positives and Predicted Positives, and content corresponds Information represented by Recall and Precision to the Arithmetic Mean (David Martin Ward Powers, 2011).

2.3 Research gap

Here are few research gaps that I have found and focusing on these lacking, I have conducted this research.

- Review of literature suggests that a few researches have been carried out to study the standard health deficiency.
- There is a need for better understanding of the data attributes which are essential to describe the data structure.

2.4 Summary

This research is motivated by the need for concentrating study on healthcare data to predict standard health deficiency with the help of data plotting visualizations, and Confusion Matrix analysis using Accuracy, Precision, Recall and F-Measures for finding a better dataset combination for training and testing to understand the effectiveness of the prediction model.

CHAPTER 3

RESEARCH METHODOLOGY

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Methodology

For research methodology, we have used Support Vector Machine (SVM) to predict the standard health deficiency. We developed the models with given dataset, trained the models and analyzed against both training and testing datasets along with data plotting visualizations. Finally, by analyzing the in-sample and out-of-sample datasets with Confusion Matrix, we have used Accuracy, Precision, Recall and F-Measures for finding a better dataset combination for training and testing to understand the effectiveness of the prediction model.

3.2 Data description

Health Deficiencies, a list of all health deficiencies currently listed on Nursing Home Compare, including the nursing home that received the deficiency, the associated inspection date, deficiency tag number, scope and severity, the status of the deficiency and the correction date. Data are presented as one deficiency per row.

Source: data.medicare.gov, a federal government website managed by the U.S. Department of Health & Human Services, 200 Independence Avenue, S.W. - Washington, D.C. 20201.

Web link: https://healthdata.gov/dataset/health-deficiencies?fbclid=IwAR3k9DbM1aLB0VPl_LNJes2Sjn_0Oawis8gdCZ-FPGpMh-PKyvBXE-QCMXE

Field	Value
Publisher	Centers for Medicare & Medicaid Services
Bureau Code	009:38 - Centers for Medicare and Medicaid Services
Modified	2019-04-24
Release Date	2019-04-24
Homepage URL	https://data.medicare.gov/d/r5ix-sfxw
Identifier	https://data.medicare.gov/api/views/r5ix-sfxw
License	http://opendefinition.org/licenses/odc-odbl/
Contact Name	Nursing Home Compare
Contact Email	BetterCare@cms.hhs.gov
Public Access Level	Nursing Home Compare
Program Code	009:000 - Department of Health and Human Services - (Primary Program Not Available)
License	http://opendefinition.org/licenses/odc-odbl/

Harvested from data.medicare.gov.

Harvest Source Title	data.medicare.gov
Harvest Source URI	https://data.medicare.gov/data.json
Last Harvest Performed	Mon, 05/06/2019 - 02:58

Table 14. Data authenticity and harvest information

Centers for Medicare & Medicaid Services, the Centers for Medicare & Medicaid Services (CMS), previously known as the Health Care Financing Administration (HCFA), is a federal agency within the United States Department of Health and Human Services (DHHS) that administers the Medicare program and works in partnership with state governments to administer Medicaid, the State Children's Health Insurance Program (SCHIP), and health insurance portability standards.

Provider Name	Provider State	Deficiency Tag Number	Scope Severity Code	Inspection Cycle	Standard Deficiency
MEADOWS OF CENTRAL MASSACHUSETTS (THE)	MA	157	3	2	1
LIFE CARE CENTER OF AUBURN	MA	333	3	2	1
BENJAMIN HEALTHCARE CENTER	MA	600	6	1	0
GROSVENOR PARK HEALTH CENTER	MA	226	3	2	1
CARE ONE AT CONCORD	MA	441	3	3	1
BENJAMIN HEALTHCARE CENTER	MA	282	3	3	1
ROYAL MEGANSETT NURSING & REHABILITATION	MA	813	3	1	1
WINGATE AT NEEDHAM	MA	689	3	1	0
BENJAMIN HEALTHCARE CENTER	MA	657	4	1	1
BENJAMIN HEALTHCARE CENTER	MA	865	3	1	1
COMMONS RESIDENCE AT ORCHARD COVE	MA	759	3	1	1
ALLIANCE HEALTH AT MARINA BAY	MA	244	4	3	1
HARBOR HOUSE NURSING & REHABILITATION CENTER	MA	757	3	1	1
ROYAL MEGANSETT NURSING & REHABILITATION	MA	625	1	1	1
HILLCREST COMMONS NURSING & REHABILITATION CENTER	MA	812	3	1	1

Table 15. Data sample with Scope Severity Code, Inspection Cycle, Standard Deficiency

Here, we have used Scope Severity Code and Inspection Cycle columns as independent variance to predict Standard Deficiency

3.3 Data preprocessing

Healthdata.gov has opened Health Deficiencies project where we found 354,271 rows of dataset along with 19 columns. We have considered first 10,000 rows and 3 columns as Scope Severity Code, Inspection Cycle, Standard Deficiency for this analysis which has been selected from Python codebase.

For Scope Severity Code, the alphabetic (A-L) have been converted to (0-11) weight to make to analysis more effective.

Scope Severity Code	Weighted numeric value
A	0
B	1
C	2
D	3
E	4
F	5
G	6
H	7
I	8
J	9
K	10
L	11

Table 16. Scope Severity Code weight chart

3.4 Summary

The predictive analysis for standard health deficiency using Support Vector Machine with the help of Confusion Matrix can be an imperative way for predicting healthcare data. Defining research scope with SVM algorithm, preparing data, analyzing through different data plotting visualizations and concluding effectiveness with confusion matrix are the methodologies that we had followed in this research.

CHAPTER 4

RESULTS AND DISCUSSIONS

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Data analysis technique

We applied the analysis directly into the download source file which had 354,271 rows with 19 columns. Among these, we had selected first 10,000 rows and 3 columns (Scope Severity Code, Inspection Cycle and Standard Deficiency) for this research scope.

After selecting the research data, we generated 3 different data visualizations with 100, 1000 and 10,000 data rows to understand the data variations. Then we had created 2 Support Vector Machine (SVM) analyses based on the test data provisioning ratio, analysis A with less training (20%) but most testing dataset (80%) and the opposite for analysis B with most training (80%) but less testing set (20%). And for both these analysis A and B, we had measured the performance with in-sample (training dataset) and out-of-sample (testing dataset) scope by using Confusion Matrix and with the help of Accuracy, Precision, Recall and F-Measurement to compare efficiency among these analyses.

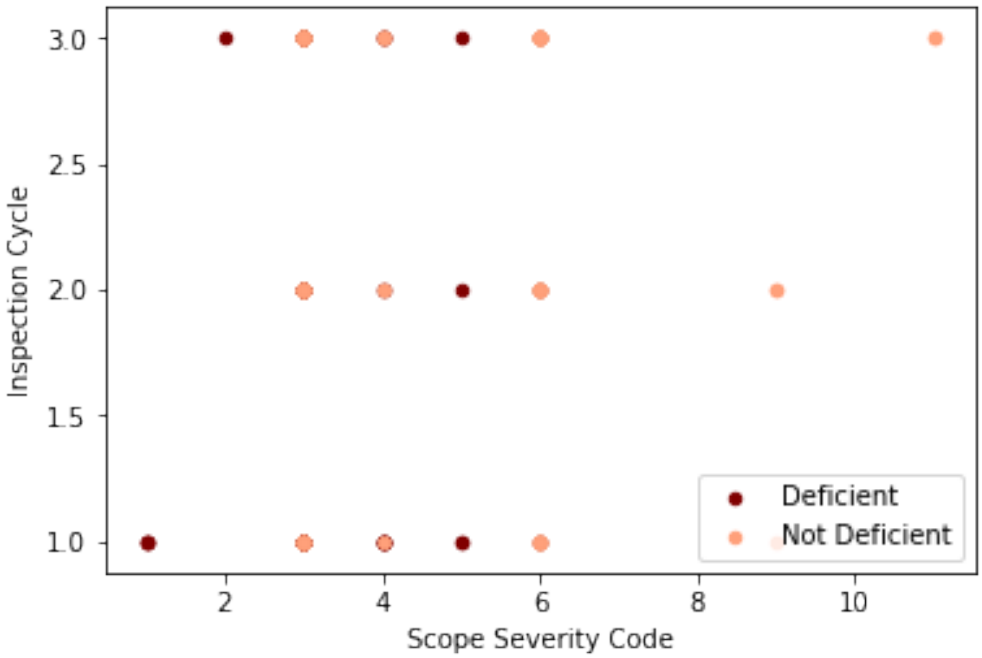
4.2 Data visualization

By using the selected dataset, we have plotted 3 data visualizations with 100, 1000 and 10,000 data rows to understand the data variations. In these visualizations, we have used Scope Severity Code in X-axis, Inspection Cycle in Y-axis and Standard Deficiency as the size and color deepness of the point.

For Scope Severity Code in X-axis, we have 12 data points (A-L with 0-11 range) and Inspection Cycle in Y-axis with 3 data points (1-3 inspection cycles). And the deeper color represents more deficiency data points on Standard Deficiency as Deficient (1 as Deficient and 0 as Not Deficient in the graph legend) .

4.2.1 Data visualization of 100 rows

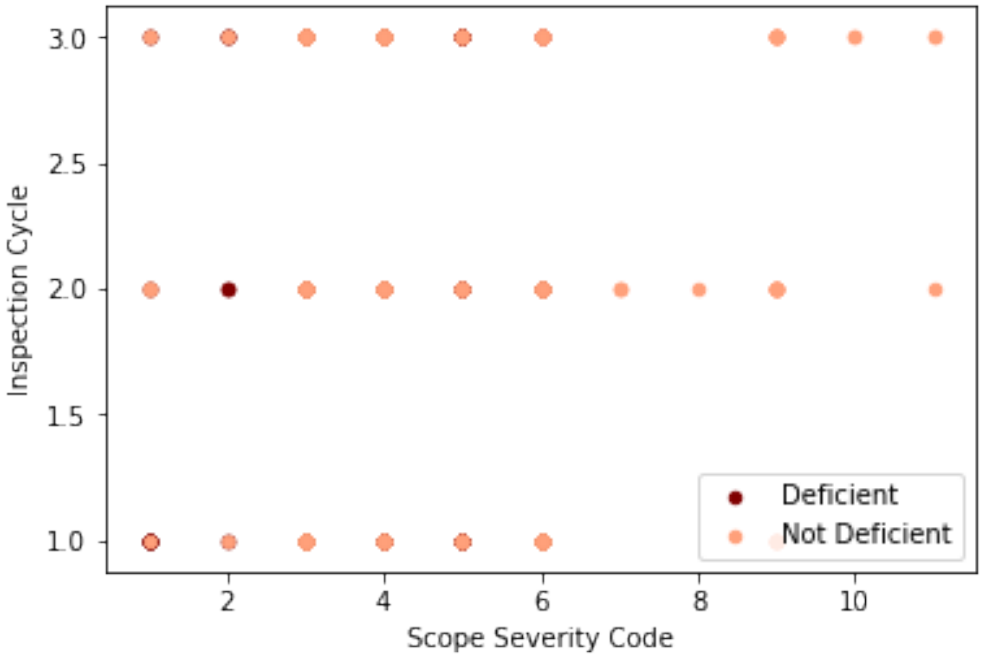
As the data points are only 100, we can't find that much coverage in this visualization. Moreover, within this limited data scope, there are many prominent (deeper) points as data plots have overwritten the same points in many cases.



Graph 1. Data visualization of 100 rows

4.2.2 Data visualization of 1000 rows

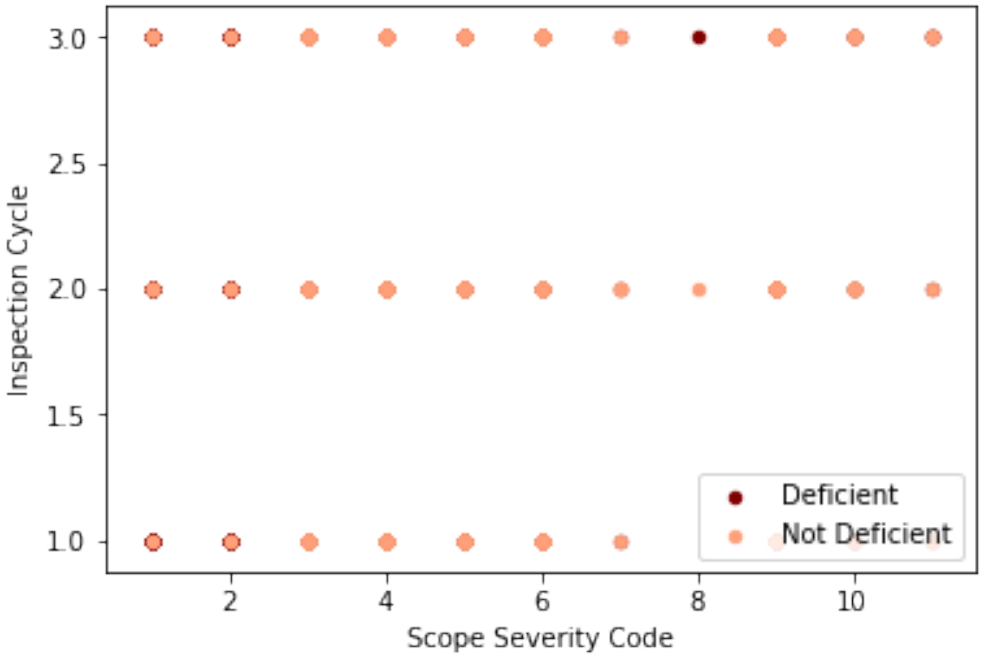
Unlike 100 rows visualization, we can find more coverage in this graph for 1000 rows. Moreover, within this data scope, there are not many prominent (deeper) points as data plots have overwritten evenly in most cases.



Graph 2. Data visualization of 1000 rows

4.2.3 Data visualization of 10000 rows

Unlike pervious 2 visualizations, we have maximum coverage on data points here for 10,000 rows visualization. Moreover, the data plots are evenly distributed into different points which indicates a healthy data variation.

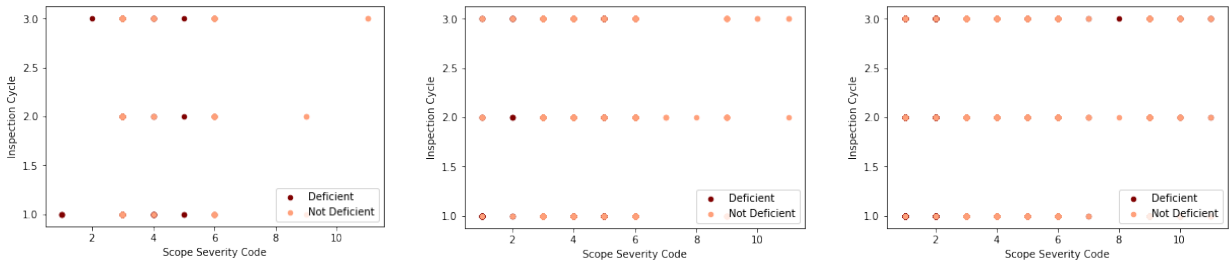


Graph 3. Data visualization of 10000 rows

4.2.4 Data visualization comparison

By comparing these 3 graphs, we can find that the visualization with 10,000 data rows has the best data point coverage where most of the data points are with good shape and balanced color that indicates even data distribution for Scope Severity Code, Inspection Cycle, Standard Deficiency.

Therefore, we can conclude that first 10,000 data rows can represent a healthy data set to consider for this research for predicting standard health deficiency.



100 rows visualization 1000 rows visualization 10,000 rows visualization

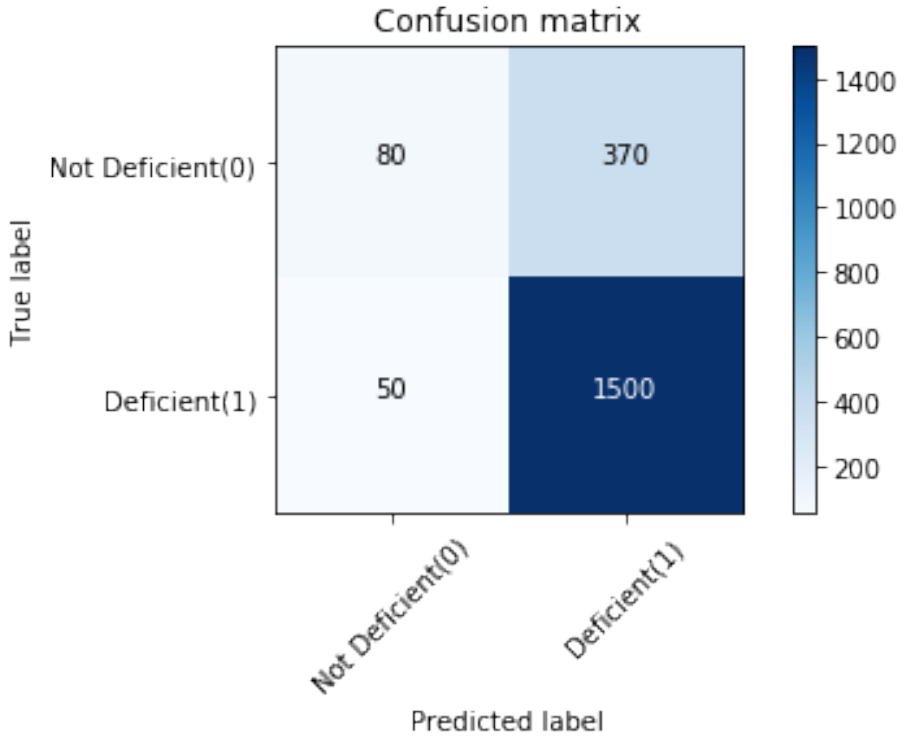
Graph 4. Data visualization comparison

4.3 Analysis A – training (20%) and testing (80%) model

In analysis A, we are considering less training dataset (20% of 10,000 is 2000 data rows) and more testing dataset (80% of 10,000 is 8000 data rows) to understand how Support Vector Machine (SVM) preforms with this ratio. We are using Confusion Matrix to visualize the prediction data for standard health deficiency and with the help of Accuracy, Precision, Recall, F-Measures to measure the efficiency.

4.3.1 In-sample confusion matrix

After training the model with (20%) 2000 training dataset, we are applying the same dataset (in-sample dataset) again to understand the SVM prediction efficiency.



Graph 5. In-sample confusion matrix for analysis A

Total 2000	Not Deficient (0)	Deficient (1)	
Not Deficient (0)	80 [TN]	370 [FP]	450
Deficient (1)	50 [FN]	1500 [TP]	1550
	130	1870	

	precision	recall	f1-score	support
0	0.62	0.18	0.28	450
1	0.80	0.97	0.88	1550
micro avg	0.79	0.79	0.79	2000
macro avg	0.71	0.57	0.58	2000
weighted avg	0.76	0.79	0.74	2000

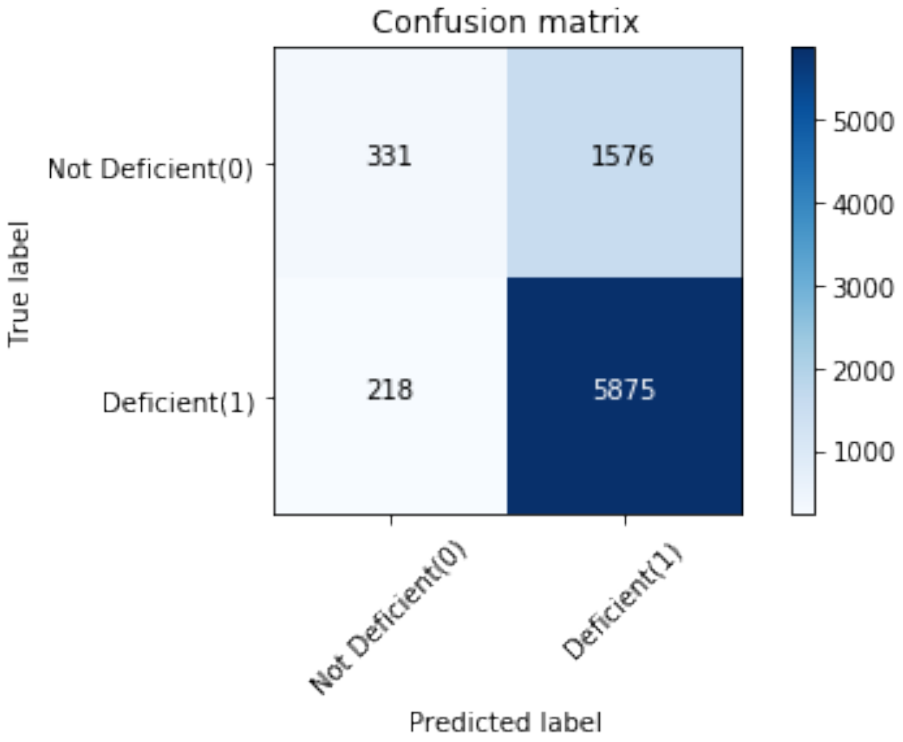
Table 6. In-sample confusion matrix for analysis A calculation

Here are the Jaccard score and F1-score with in-sample scope for analysis A.

- Jaccard score: 0.79
- F1-score: 0.7418935269207501

4.3.2 Out-of-sample confusion matrix

Here, we considered (80%) 8000 rows as testing dataset (out-of-sample dataset) to understand how well the SVM prediction efficiency.



Graph 7. Out-of-sample confusion matrix for analysis A

Total 8000	Not Deficient (0)	Deficient (1)	
Not Deficient (0)	331 [TN]	1576 [FP]	1907
Deficient (1)	218 [FN]	5875 [TP]	6083
	549	7451	

	precision	recall	f1-score	support
0	0.60	0.17	0.27	1907
1	0.79	0.96	0.87	6093
micro avg	0.78	0.78	0.78	8000
macro avg	0.70	0.57	0.57	8000
weighted avg	0.74	0.78	0.72	8000

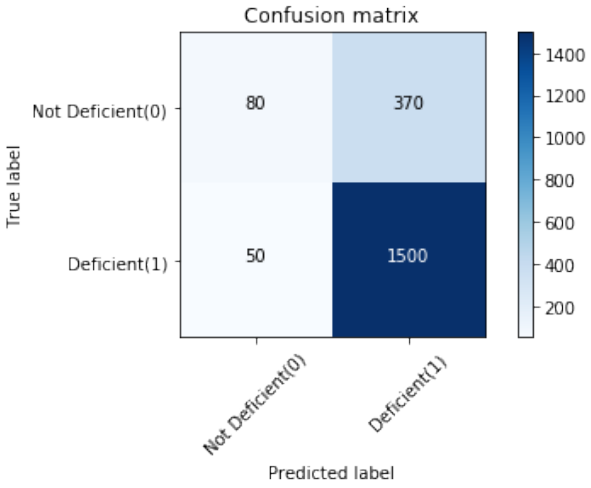
Table 8. Out-of-sample confusion matrix for analysis A calculation

Here are the Jaccard score and F1-score with out-of-sample scope for analysis A.

- Jaccard score: 0.77575
- F1-score: 0.7249948476530109

4.3.3 In-sample vs out-of-sample confusion matrix comparison

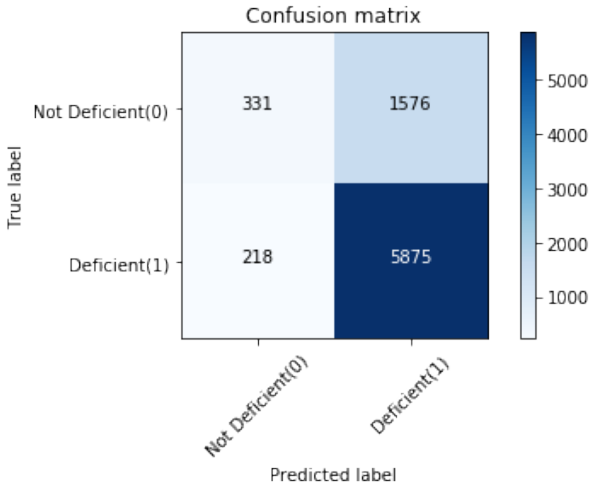
For analysis A, by comparing in-sample and out-of-sample confusion matrix measurements, we can conclude that there is a visible difference with the F1-score along with differences in Jaccard score, Precision and Recall.



	precision	recall	f1-score	support
0	0.62	0.18	0.28	450
1	0.80	0.97	0.88	1550
micro avg	0.79	0.79	0.79	2000
macro avg	0.71	0.57	0.58	2000
weighted avg	0.76	0.79	0.74	2000

- Jaccard score: 0.79
- F1-score: 0.7418935269207501

In-sample confusion matric for Analysis A



	precision	recall	f1-score	support
0	0.60	0.17	0.27	1907
1	0.79	0.96	0.87	6093
micro avg	0.78	0.78	0.78	8000
macro avg	0.70	0.57	0.57	8000
weighted avg	0.74	0.78	0.72	8000

- Jaccard score: 0.77575
- F1-score: 0.7249948476530109

Out-of-sample confusion matric for Analysis A

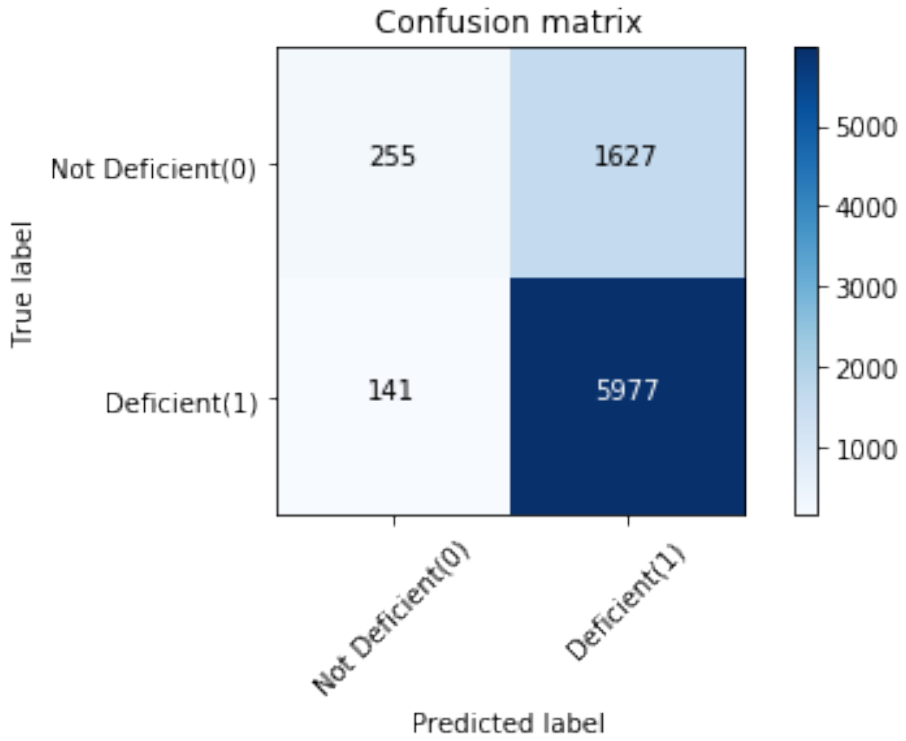
Graph 9. In-sample vs out-of-sample confusion matrix comparison for analysis A

4.4 Analysis B – training (80%) and testing (20%) model

In analysis B, we are considering less training dataset (80% of 10,000 is 8000 data rows) and more testing dataset (20% of 10,000 is 2000 data rows) to understand how Support Vector Machine (SVM) preforms with this ratio. We are using Confusion Matrix to visualize the prediction data for standard health deficiency and with the help of Accuracy, Precision, Recall, F-Measures to measure the efficiency.

4.4.1 In-sample confusion matrix

After training the model with (80%) 8000 training dataset, we are applying the same dataset (in-sample dataset) again to understand the SVM prediction efficiency.



Graph 10. In-sample confusion matrix for analysis B

Total 8000	Not Deficient (0)	Deficient (1)	
Not Deficient (0)	255 [TN]	1627 [FP]	1882
Deficient (1)	141 [FN]	5977 [TP]	6118
	396	7604	

	precision	recall	f1-score	support
0	0.64	0.14	0.22	1882
1	0.79	0.98	0.87	6118
micro avg	0.78	0.78	0.78	8000
macro avg	0.71	0.56	0.55	8000
weighted avg	0.75	0.78	0.72	8000

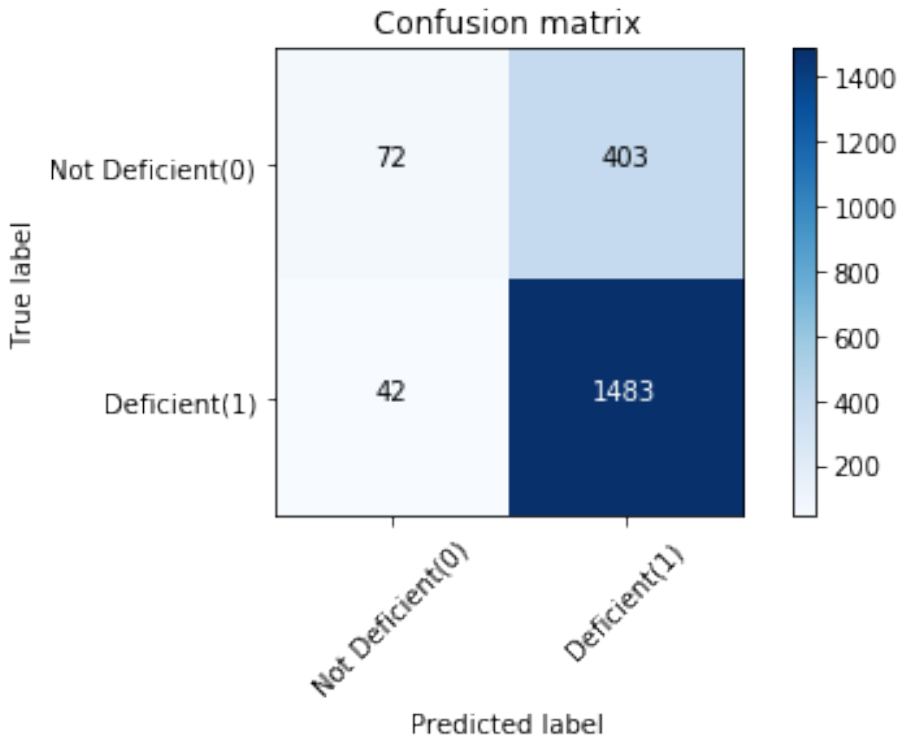
Table 11. In-sample confusion matrix for analysis B calculation

Here are the Jaccard score and F1-score with in-sample scope for analysis B.

- Jaccard score: 0.779
- F1-score: 0.7188843147620011

4.4.2 Out-of-sample confusion matrix

Here, we considered (20%) 2000 rows as testing dataset (out-of-sample dataset) to understand how well the SVM prediction efficiency.



Graph 12. Out-of-sample confusion matrix for analysis B

Total 2000	Not Deficient (0)	Deficient (1)	
Not Deficient (0)	72 [TN]	403 [FP]	475
Deficient (1)	42 [FN]	1483 [TP]	1525
	114	1886	

	precision	recall	f1-score	support
0	0.63	0.15	0.24	475
1	0.79	0.97	0.87	1525
micro avg	0.78	0.78	0.78	2000
macro avg	0.71	0.56	0.56	2000
weighted avg	0.75	0.78	0.72	2000

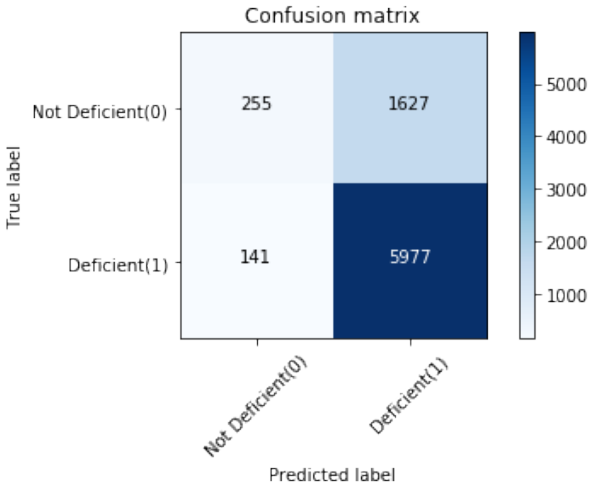
Table 13. Out-of-sample confusion matrix for analysis B calculation

Here are the Jaccard score and F1-score with out-of-sample scope for analysis B.

- Jaccard score: 0.7775
- F1-score: 0.7210885560000379

4.4.3 In-sample vs out-of-sample confusion matrix comparison

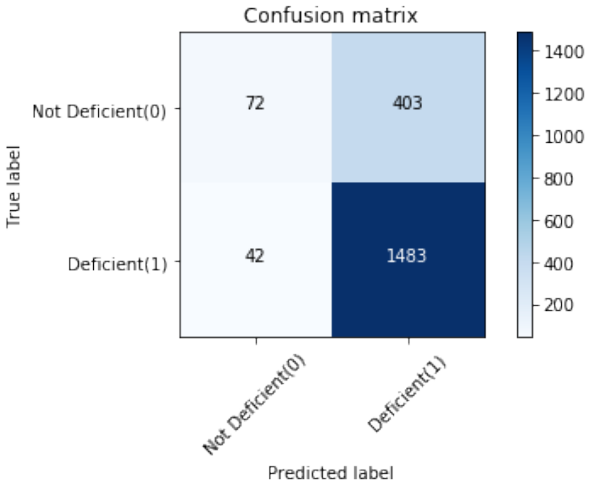
For analysis B, by comparing in-sample and out-of-sample confusion matrix measurements, we can conclude that there is a visible difference with the F1-score along with differences in Jaccard score, Precision and Recall.



	precision	recall	f1-score	support
0	0.64	0.14	0.22	1882
1	0.79	0.98	0.87	6118
micro avg	0.78	0.78	0.78	8000
macro avg	0.71	0.56	0.55	8000
weighted avg	0.75	0.78	0.72	8000

- Jaccard score: 0.779
- F1-score: 0.7188843147620011

In-sample confusion matric for Analysis B



	precision	recall	f1-score	support
0	0.63	0.15	0.24	475
1	0.79	0.97	0.87	1525
micro avg	0.78	0.78	0.78	2000
macro avg	0.71	0.56	0.56	2000
weighted avg	0.75	0.78	0.72	2000

- Jaccard score: 0.7775
- F1-score: 0.7210885560000379

Out-of-sample confusion matric for Analysis B

Graph 14. In-sample vs out-of-sample confusion matrix comparison for analysis B

4.5 Analysis comparison

Here we compared the analysis between A and B to understand the training and testing data distribution ratio along with comparison for in-sample and out-of-sample to prove the effectiveness of Support Vector Matrix in the standard health deficiency prediction context.

4.5.1 Comparison of analysis A and analysis B

Here are the confusion matrix measurements by analysis A, B and sample selection.

	precision	recall	f1-score	support
0	0.62	0.18	0.28	450
1	0.80	0.97	0.88	1550
micro avg	0.79	0.79	0.79	2000
macro avg	0.71	0.57	0.58	2000
weighted avg	0.76	0.79	0.74	2000

- Jaccard score: 0.79
- F1-score: 0.7418935269207501

In-sample confusion matric for Analysis A

	precision	recall	f1-score	support
0	0.60	0.17	0.27	1907
1	0.79	0.96	0.87	6093
micro avg	0.78	0.78	0.78	8000
macro avg	0.70	0.57	0.57	8000
weighted avg	0.74	0.78	0.72	8000

- Jaccard score: 0.77575
- F1-score: 0.7249948476530109

Out-of-sample confusion matric for Analysis A

	precision	recall	f1-score	support
0	0.64	0.14	0.22	1882
1	0.79	0.98	0.87	6118
micro avg	0.78	0.78	0.78	8000
macro avg	0.71	0.56	0.55	8000
weighted avg	0.75	0.78	0.72	8000

- Jaccard score: 0.779
- F1-score: 0.7188843147620011

In-sample confusion matric for Analysis B

	precision	recall	f1-score	support
0	0.63	0.15	0.24	475
1	0.79	0.97	0.87	1525
micro avg	0.78	0.78	0.78	2000
macro avg	0.71	0.56	0.56	2000
weighted avg	0.75	0.78	0.72	2000

- Jaccard score: 0.7775
- F1-score: 0.7210885560000379

Out-of-sample confusion matric for Analysis B

Table 15. Comparison of analysis A and analysis B calculation

For analysis A comparison, we can find higher values of F1-score for in-sample than out-of-sample calculation. As we have trained with only 20% dataset and again testing the model with same dataset, we have this score which is abnormal in comparison to other F1-scores. However, for out-of-sample dataset which has 80% of the dataset and showing lower score with compare to in-sample dataset.

For analysis B comparison, we can find very close values of F1-score both for in-sample and out-of-sample calculation. As we have trained with 80% dataset and again testing the model with same dataset, we have this score which seems quite reasonable in comparison to other F1-scores. However, for out-of-sample dataset which has 20% of the dataset and showing a little lower score with compare to in-sample dataset in the same analysis context.

4.6 Summary

In conclusion, it is proven by this research that more dataset considered for training can produce much better prediction in comparison to less dataset. Overall, SVM had produced a decent prediction for standard health deficiency in 10,000 rows dataset consideration. Confusion matrix measurements like accuracy, precision, recall and f1-score successfully could measure the system's efficiency and helped to come up with a comparison in conclusion.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion and Findings

Standard health deficiency prediction is an essential study not only for people involved in the sector but also for researchers to find new prospect of applications. This research will particularly help preventing deficiency cases before hand with the help of proven methods and measures.

Data visualizations helped to select appropriate dataset for this research by comparing different visual plots for Deficiency deepness and plotting variations where 10,000 data rows had created much balanced data distribution.

Considering 2 different analyses (A & B) with different ratio of training and testing datasets helped to understand the impact of training on the SVM algorithm where we found that more training data can produce much better predictions. Confusion matrix with different measurements (accuracy, precision, recall & f1-score) had proved to be every effective to measure the performance of the SVM algorithm. Finally, for this health deficiency prediction context, Support Vector Matrix is proven to a decent predictor.

5.2 Limitations

Limited access of dataset is the major limitation that we faced during this research. We had data with no specific timestamp, therefore, had visualizations based on so specific time information and eventually we could not confirm whether these confusion matrix measurements are still applicable with ongoing circumstances or not.

Moreover, the dataset we had used for this analysis did not have definition for all attributes. If we had proper description for more attributes, we could try creating other visualizations and eventually could include more attributes with in the analysis for predicting standard health deficiency more effectively.

Finally, we have analyzed with Support Vector Machine (SVM) regression techniques whereas considering other machine learning algorithms like Logistic tree, Random forest, Decision tree would help us to compare between variety of models to find a better fit predicting standard health deficiencies.

5.3 Recommendations for future works

Further research considering more detailed dataset with timestamp information and other machine learning algorithms with different visualizations and confusion matrix measurement calculations are imperative to find a suitable model for analyzing standard health deficiency predictions.

REFERENCES

Theodoros Evgeniou, Massimiliano Pontil (2001). Support Vector Machines: Theory and Applications. https://www.researchgate.net/publication/221621494_Support_Vector_Machines_Theory_and_Applications

Simon Tong, Daphne Koller (2001). Support Vector Machine Active Learning with Applications to Text Classification. http://www.jmlr.org/papers/volume2/tong01a/tong01a.pdf

Vikramaditya Jakkula. Tutorial on Support Vector Machine (School of EECS, Washington State University, Pullman 99164). https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf

Andrew Ng. CS229 Lecture notes. http://cs229.stanford.edu/notes/cs229-notes3.pdf

DURGESH K. SRIVASTAVA, LEKHA BHAMBHU (2009). DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE. http://www.jatit.org/volumes/research-papers/Vol12No1/1Vol12No1.pdf

Tristan Fletcher (2008). Support Vector Machines Explained. https://cling.csd.uwo.ca/cs860/papers/SVM_Explained.pdf

Yongli Zhang (2012). Support Vector Machine Classification Algorithm and Its Application (ICICA 2012: Information Computing and Applications pp 179-186). https://link.springer.com/chapter/10.1007/978-3-642-34041-3_27

Lei ShiQiguo DuanXinming MaMei Weng. The Research of Support Vector Machine in Agricultural Data Classification (CCTA 2011: Computer and Computing Technologies in Agriculture). https://link.springer.com/chapter/10.1007/978-3-642-27275-2_29

Dmitriy Fradkin, Ilya Muchnik. Support Vector Machines for Classification (DIMACS Series in Discrete Mathematics and Theoretical Computer Science). https://pdfs.semanticscholar.org/b944/a74a79c701d6c2a0ff2d28a0d16b9d8b1cd0.pdf

Sofia Visa, Brian Ramsay, Anca Ralescu, Esther van der Knaap (2011). Confusion Matrix-based Feature Selection. http://ceur-ws.org/Vol-710/paper37.pdf

A. K. Santra, C. Josephine Christy (2012). Genetic Algorithm and Confusion Matrix for Document Clustering. https://www.ijcsi.org/papers/IJCSI-9-1-2-322-328.pdf

Robert Susmaga. Confusion Matrix Visualization (Intelligent Information Processing and Web Mining pp 107-116). https://link.springer.com/chapter/10.1007/978-3-540-39985-8_12

Kai Ming Ting. Encyclopedia of Machine Learning and Data Mining (2017 Edition). https://link.springer.com/referenceworkentry/10.1007%2F978-1-4899-7687-1_50

Pierre Machart, Liva Ralaivola (2012). Confusion Matrix Stability Bounds for Multiclass Classification (arXiv.org > cs > arXiv:1202.6221). https://arxiv.org/abs/1202.6221

Cinthia O. A. Freitas João M. de Carvalho José Josemar Oliveira Jr. Simone B. K. Aires Robert Sabourin. Confusion Matrix Disagreement for Multiple Classifiers (CIARP 2007: Progress in Pattern Recognition, Image Analysis and Applications pp 387-396). https://link.springer.com/chapter/10.1007/978-3-540-76725-1_41

Marina Sokolova, Guy Lapalme (2009). A systematic analysis of performance measures for classification tasks. http://atour.iro.umontreal.ca/rali/sites/default/files/publis/SokolovaLapalme-JIPM09.pdf

Maria Navin J R, Pankaja R (2016). Performance Analysis of Text Classification Algorithms using Confusion Matrix. https://www.erpublication.org/published_paper/IJETR042741.pdf

Ronnie Merin George, Dr. Jose Alex Mathew (2016). Emotion Classification Using Machine Learning and Data Preprocessing Approach on Tulu Speech Data. https://www.ijcsmc.com/docs/papers/June2016/V5I6201617.pdf

Shantanu Godbole (2002). Exploiting confusion matrices for automatic generation of topic hierarchies and scaling up multi-way classifiers. http://www.godbole.net/shantanu/pubs/autoconfmat.pdf

Anderson, Harvey; Peterson, Dewey; H.Beaton, George (1982). Estimating nutrient deficiencies in a population from dietary records: The use of probability analyses. https://www.sciencedirect.com/science/article/pii/S0271531782800493

Dugdale, Alan E (2006). Predicting iron and folate deficiency anaemias from standard blood testing: the mechanism and implications for clinical medicine and public health in developing countries (Dugdale; licensee BioMed Central Ltd. 2006). https://tbiomed.biomedcentral.com/articles/10.1186/1742-4682-3-34

R. Goodman, D. Renfrew, M. Mullick (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka (European Child & Adolescent Psychiatry, June 2000, Volume 9, Issue 2, pp 129–134). https://link.springer.com/article/10.1007/s007870050008

Mar´ıa Alejandra Culman, Jairo Alejandro G´omez, Jes´us Talavera, Luis Alfredo Quiroz,Luis Eduardo Tob´on, Juan Manuel Aranda, Luis Ernesto Garreta, and Cristihian Jarri Bayona (2017). A Novel Application for Identification of NutrientDeficiencies in Oil Palm Using the Internet ofThings. https://www.researchgate.net/publication/316342740_A_Novel_Application_for_Identification_of_Nutrient_Deficiencies_in_Oil_Palm_Using_the_Internet_of_Things

Bich Tran Bruce K. Armstrong Kevin McGeechan Peter R. Ebeling Dallas R. English Michael G. Kimlin Robyn Lucas Jolieke C. van der Pols Alison Venn Val Gebski David C. Whiteman Penelope M. Webb Rachel E. Neale (2013). Predicting vitamin D deficiency in older Australian adults (Clinical Endocrinology, Volume 79, Issue 5). https://onlinelibrary.wiley.com/doi/abs/10.1111/cen.12203

Kite, Allison (2016). Lots of data, little agreement on nursing home quality. https://www.khi.org/news/article/lots-of-data-little-agreement-on-nursing-home-quality

Loran F. Nordgren, Mary-Hunter Morris McDonnell (2010). The Scope-Severity Paradox: Why Doing More Harm Is Judged to Be Less Harmful. https://journals.sagepub.com/doi/abs/10.1177/1948550610382308?journalCode=sppa

Minnesota Department of Health (2014). Annual Quality Improvement Report: The Nursing Home Survey Process. https://www.health.state.mn.us/facilities/regulation/docs/legrpts/2014nhqifylegrpt.pdf

Department of Health & Human Services (DHHS), Centers for Medicare & Medicaid Services (CMS) (2014). CMS Manual System, Pub. 100-07 State Operations. https://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/Downloads/R113SOMA.pdf

John P. Schomberg, Oliver L. Haimson, Gillian R. Hayes, Hoda Anton-Culver (2016). Supplementing Public Health Inspection via Social Media. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152117

Marina Sokolova, Nathalie Japkowicz, Stan Szpakowicz (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. https://www.researchgate.net/publication/225215404_Beyond_Accuracy_F-Score_and_ROC_A_Family_of_Discriminant_Measures_for_Performance_Evaluation

David Martin Ward Powers (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. https://www.researchgate.net/publication/276412348_Evaluation_From_precision_recall_and_F-measure_to_ROC_informedness_markedness_correlation; https://bioinfopublication.org/files/articles/2_1_1_JMLT.pdf; https://csem.flinders.edu.au/research/techreps/SIE07001.pdf

Zachary C. Lipton, Charles Elkan, and Balakrishnan Naryanaswamy (2014). Thresholding Classifiers to Maximize F1 Score. https://arxiv.org/pdf/1402.1892.pdf

Nan Ye, Kian Ming A. Chai, Wee Sun Lee, Hai Leong Chieu (2012). Optimizing F-Measures: A Tale of Two Approaches. https://arxiv.org/ftp/arxiv/papers/1206/1206.4625.pdf

Willem Waegeman, Krzysztof Dembczy´nski, Arkadiusz Jachnik, Weiwei Cheng, Eyke H¨ullermeier (2014). On the Bayes-Optimality of F-Measure Maximizers. http://jmlr.org/papers/volume15/waegeman14a/waegeman14a.pdf

Centers for Medicare and Medicaid Service CMS (2019). Design for Nursing Home Compare Five-Star Quality Rating System (Technical Users' Guide). https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandComplianc/downloads/usersguide.pdf

Ming-Jie Zhao, Narayanan Edakunni, Adam Pocock, Gavin Brown (2013). Beyond Fano's Inequality: Bounds on the Optimal F-Score, BER, and Cost-Sensitive Risk and Their Implications. http://jmlr.csail.mit.edu/papers/volume14/zhao13a/zhao13a.pdf

Weiwei Cheng, Krzysztof Dembczy´nski, Eyke H¨ullermeier, Adrian Jaroszewicz, and Willem Waegeman (2012). F-Measure Maximization in Topical Classification. https://weiweicheng.com/research/papers/cheng-jrs12.pdf