REPUBLIC OF TURKEY
FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND
APPLIED SCIENCE

FEATURE SELECTION FOR EFFICIENT CLASSIFICATION OF PHISHING WEBSITE DATASET

TWANA SAEED MUSTAFA

Master Thesis
Department: Software Engineering
Supervisor: Asst. Prof. Dr. Murat KARABATAK

JANUARY – 2017

REPUBLIC OF TURKEY
FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

FEATURE SELECTION FOR EFFICIENT CLASSIFICATION OF PHISHING WEBSITE DATASET

MASTER THESIS

TWANA SAEED MUSTAFA

(142137112)

Submission Date to the Institute: 28 December 2016

Thesis Presentation Date: 10 January 2017

Thesis Supervisor: Asst. Prof. Dr. Murat KARABATAK (F. U.)

Other Juries: Prof. Dr. Abdulkadir ŞENGÜR (F. U.)

: Assoc. Prof. Dr. Muhammed Fatih TALU (I. U.)

JANUARY – 2017

ACKNOWLEDGMENT

I thank all who in one way or another contributed in the completion of this thesis. First, I give thanks to God for protection and ability to do work.

I would like to express my sincere gratitude to my supervisor Asst. Prof. Dr. Murat KARABATAK for his patience, kind support, immense knowledge, motivation, directions and thorough guidance during my research work. His guidance helped me in all the time of research. At many stages of this project I benefited from his advice, particularly so when exploring new ideas. His positive outlook and confidence in my research inspired me and gave me confidence. His careful editing contributed enormously to the production of this thesis

I would like to thank my all friends, who have supported me throughout the entire process, both by keeping me harmonious and helping me putting pieces together. Your friendship makes my life a wonderful experience. I cannot list all of the name here, but you are always on my mind. I will be grateful forever for your kindness.

Last but not the least, I have to thank my parents for their love, encouraged me, prayed for me, and supported me throughout my life. Thank you both for giving me strength to reach for the stars and chase my dreams. My brothers, sister, auntie and cousins deserve my wholehearted thanks as well

Sincerely

TWANA SAEED MUSTAFA

Elazığ, 2017

TABLE OF CONTENTS

SUMMARY

Feature Selection for Efficient Classification of Phishing Website Dataset

The Internet is gradually becoming a necessary and important tool in everyday life. However, Internet users might have poor security for different kinds of web threats, which may lead to financial loss or clients lacking trust in online trading and banking. Phishing is described as a skill of impersonating a trusted website aiming to obtain private and secret information such as a user name and password or social security and credit card number. However, there is no single solution that can prevent most phishing attacks. For phishing attacks, various methods are required. In this thesis, a feature selection method and the Navie Bayes classifier are presented for the phishing Websites dataset. In this study, phishing dataset retrieved from UCI machine learning repository is used. This dataset consists of 11055 records and 31 features. The research presented in this thesis aims at reducing the number of features of the used dataset as well as obtaining the best classification performance. Feature selection algorithms are used to reduce the dataset features and to obtain a high system performance. In addition, the performance of feature selection algorithms is compared using the Naive Bayes classifier. Finally, a comparative performance in reducing dataset features using the common classification algorithms is given. The results show that an effective phishing detection can be made with feature selection that reduces the dataset features.

Keywords: Data Mining, Feature Selection, Naïve Bayes, Phishing Website

ÖZET

Phising Web Sitesi Veri Setinin Etkili Sınıflandırması için Özellik Seçimi

Internet, giderek insan hayatında önemli bir gerekli bir araç haline gelmiştir. Bununla birlikte, internet kullanıcılarının, farklı web tehditlerine karşı güvenlikleri oldukça yetersizdir. Çevrim için ticaret ve bankacılığa güvenmeleri, buradan doğabilecek web tehditlerine karşı daha riskli durumlar ortaya çıkarmaktadır. Phishing (Kimlik Avı), güvenli bir web sitesi görünümündeki bazı sitelerin, kişinin kullanıcı adı, şifre, sosyal güvenlik numarası ve kredi kartı numarası gibi gizli ve özel bilgilerini elde etmeyi amaçlayan bir yöntem olarak açıklanmaktadır. Çoğu kimlik avı saldırısını tespit etmek için genelde tek bir çözüm yolu yoktur ve çeşitli yöntemler gerekmektedir. Bu tezde, kimlik avı veri seti için özellik seçimi yöntemi ve sade bayes sınıflandırıcı tartışılmıştır. Bu çalışmada, UCI makine öğrenmesi deposundan alınan Phishing (Kimilik avı) veri seti kullanılmıştır. Bu veri seti 11055 adet kayıt ve 31 adet özellikten oluşmaktadır. Tezde, bu veri setinin özellik sayısını azaltılması ve en iyi sınıflandırma performansı elde edebilmesi hedeflenmiştir. Veri setini indirgemek için ve iyi bir sistem performansı elde etmek için özellik seçimi algoritmaları kullanılmıştır. Ayrıca, özellik seçimi algoritmalarının performansı naive bayes sınıflandırıcı kullanılarak karşılaştırılmıştır. Son olarak, indirgenmiş veri seti üzerinde diğer sınıflandırma algoritmalarının performansı karşılaştırmalı olarak verilmiştir. Elde edilen bulgular, özellik seçimi ile kimlik avı veri setinin indirgenmesi sayesinde etkili bir kimlik avı tespiti yapılabileceği sonucunu ortaya çıkarmıştır.

Anahtar Kelimeler: Veri Madenciliği, Özellik Seçimi, Sade Bayes, Phishing Web Sitesi

LIST OF FIGURES

Page No

LIST OF TABLES

Page No

ABBREVIATIONS

DM	**:** Data mining
KDD	**:** Knowledge Discovery from Data
CMAD	**:** Compliance Monitoring for Anomaly Detection
FS	**:** Feature Selection
SFS	**:** sequential forward Feature Selection
SBS	**:** Sequential Backward Feature Selection
FTC	**:** Elected Trade Commission
NN	**:** Neural Networks
MBA	**:** Market basket analysis
LVF	**:** Las Vegas Channel
CFS	**:** Correlation-based Feature Selection
PCA	**:** Principal Components Analysis
APWG	**:** Anti-Phishing Working Group
HTTP	**:** Hyper-Text Transfer Protocol
HTTPS	**:** Hyper Text Transfer Protocol Secure
IP	**:** Internet Protocol
URL	**:** Universal Resource Locator
PIN	**:** Personal Identification Number
ccTLD	**:** Country Code Top Level Domain
SFH	**:** Server Form Handler

1. INTRODUCTION

Phishing is a criminal exercise utilizing familiar (social) engineering techniques. Phishers try to fraudulently achieve sensitive personal information [1]. The Internet is not only significant for individual clients, but also for associations doing business online. These associations ordinarily offer transaction exchanging over the Internet [2]. Internet-clients might be powerless against various types of web-dangers that may bring about monetary harm, data fraud, loss of private data, brand notoriety harm, and loss of clients' trust in e-trade and online Web managing and account banking. Along these lines, Internet reasonableness for business exchanges gets to be dubious. Phishing has viewed a form of the web-dangers that is characterized as a special art of impersonating a reliable site aiming to get private data, for example usernames, secret words (passwords) and social security numbers, and credit card details [3]. When phishing pages obtain entrance, they can utilize your own data to confer wholesale fraud, charge credit cards, take advantage of unfilled ledgers, empty bank accounts, scan emails, and lock a person out of an online account by changing the needed password [4]. eBay and PayPal are two of the most targeted companies; online banks are also familiar targets. Phishing is regularly carried out utilizing email or an instant message, and generally directs users to send information and details to a Website, despite the fact that telephone contact has been utilized too [1].

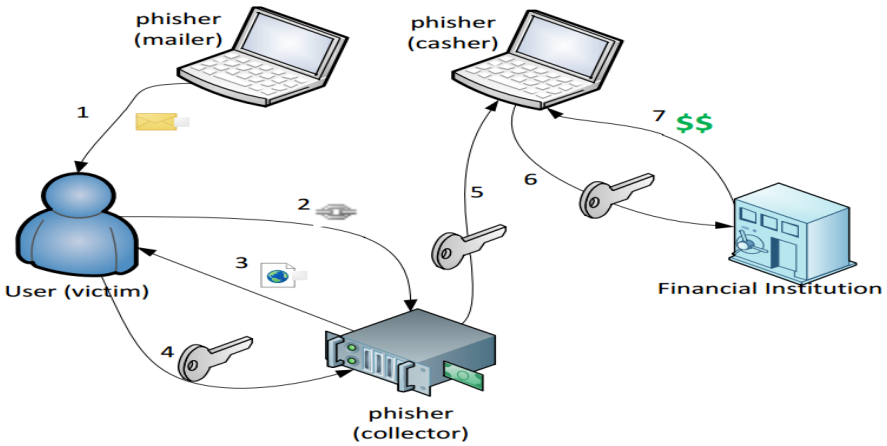

Figure 1.1. Phishing information [3]

In general, two methodologies are utilized in distinguishing phishing sites. The first one relies on blacklists [5]. The second way is known as a heuristic-based method. [6].

There are numerous approaches to battle phishing some of them:

- Legitimate arrangements: it is conducted by nations' hunting down exercise. The U.S. was the chief to work with the rules in contradicting phishing exercises, and numerous phishers have been captured and extorted [7].
- Training: The primary standard in battling phishing and information security dangers is purchaser's mindfulness. In the event that web clients can be fulfilled to check the security highlights inside the site page, later the issue is essentially left [8].
- Technical solution: weak points that seemed when depending on former stated resolutions led to the need of advanced resolutions. Many academic researches, business, and non-business resolutions are put forward to manage phishing. Further, some non-profitable organizations like "APWG", "Phish Tank" and "Miller Smiles" bring in meetings of ideas and distributing of the greatest exercise that could be systematized against phishing [8].
- Boycotts approach (blacklists approach): In this approach, the requested URL is contrasted and pre-characterized as phishing URLs. The shortcoming of this method is that the boycott ordinarily cannot shield all phishing site pages as a recently made fake site page takes a considerable amount of time before it is added to the rundown [5].
- Heuristic approach (experiential method): This method is recognized as experiential-founded method, where many structures are mined from web page to categorize it as phishy or legitimate [7].

Advance in mechanized information retrieval and storage innovation has brought about the improvement of gigantic records. This has happened in each normal of human endeavors, from the regular (like general store exchange information, charge card usage records, telephone call purposes of intrigue, and government bits of knowledge) to the more extraordinary (like pictures of galactic bodies, sub-nuclear databases, and restorative reports) [9].

Humans are overpowered by information - investigational data, mending data, demography information, and money related information, and promoting information.

Individuals have no opportunity not to use this information. Human thought has transformed into the acknowledged favorable position. Along these appearances, users have to decide techniques to definitely look at the information, to consequently group it, to naturally plot it, and to unavoidably find and characterize diagrams in it. This is a champion between the liveliest and strengthening assortments of the database contemplate bunch. Specialists in regions tallying figures, delineation, computerized reasoning, and machine learning are adding to this range [10]. Contemporary PC frameworks are gathering information at an inconceivable rate and from a broad assortment of establishments: from reason for offer machines in the remarkable path to apparatuses indexing each check opportunity, bank money removal, and charge postcard trade, to earth perception cables in space.

A few cases for enormous information stream:

- The present NASA earth surveillance satellites deliver a terabyte (i.e. 10^9bytes) of information consistently. This is more than the aggregate amount of information ever conveyed by all first reconnaissance satellites.

- The human genome obligation is securing many bytes for each of numerous billion hereditary bases.

- As long back as 1990, the US Enumeration gathered over a million bytes of data.

- Many firms protect awesome data warehouses of customer associations.

- There are gigantic measures of data confirmed ordinary on oblivious footage gadgets, similar to charge card exchange files and web logs, and furthermore non-run of the mill data, for instance CCTV recordings.

Feature selection, in view of application data field and the objective of the mining effort, is represented as the pick of human examiner of a subsection of the components found in the principal information set. The procedure of feature selection can be manual or automated by some robotized ways. In this regard, highlight determination strategies are connected in one of three hypothetical foundations: the channel show, the wrapper display, and inserted structure. These three principle families fluctuate in how the learning calculation is consolidating in evaluating and selecting highlights. Primary elements of both ways are given in Figure 1.2. Finally, the inserted methods coordinate component view and the learning calculation into a solitary improvement tricky beginning. At the point when the quantity of tests and measurements turns out to be huge, the channel approach is normally

chosen due to its computational proficiency and nonpartisan inclination towards any learning procedure [11].

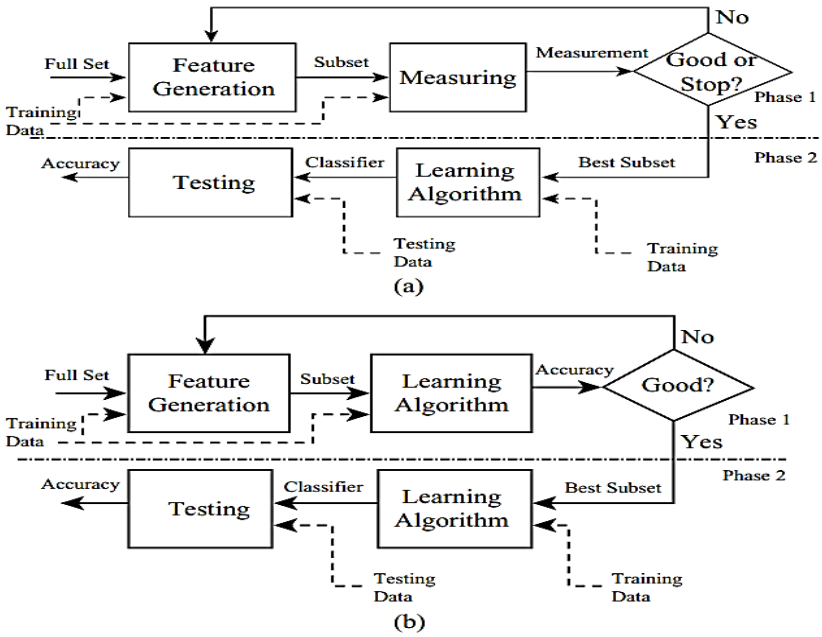

Figure 1.2. Feature-selection approaches. (a) filter model; (b) wrapper model [11].

1.1 Data Mining

Simply speaking, data mining means separating or “mining” learning from a lot of information. Recollect that the withdrawal of gold-plated from pillars or sand is alluded to as gold mining as opposed to shake or sand mining. In this way, data mining ought to have been more fittingly named “learning withdrawal from information,” which is tragically truly long. “Data mining,” a shorter-term, will not mirror the accentuation on mining from a lot of information. Things being what they are, mining is an unmistakable term depicting the procedure that finds a little arrangement of valuable pieces from a lot of crude material as illustrated in Figure 1.3. Various distinctive terms pass on a practically identical or hardly extraordinary intending to data mining, for example, information mining from information, learning extraction, data/plan examination, information prehistoric studies, and information digging. Numerous people regard data mining as a corresponding word for additional broadly used period, Information Detection from Statistics, or KDD [10].

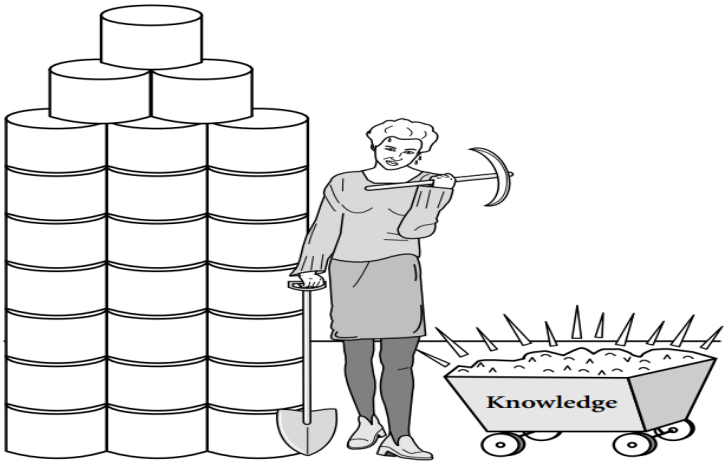

Figure 1.3. Data mining—searching for knowledge (interesting patterns) in your data [10].

1.2 Phishing

Phishing is a type of social activities in which an assailant, called a phisher, endeavors to deceitfully recover honest to goodness clients' privacy or sensitive information by copying electronic interchanges from a reliable or open association in a computerized design [12]. The term "phishing" was coined around 1995, when Internet scam artists were utilizing email baits to "fish" for passwords and money related data from the ocean of Internet clients. Here, "ph" is a typical programmer substitution of "f", which originates from the first type of hacking, "phreaking", on phone switches amid 1960s [13]. Early phishers duplicated the code from the AOL site and created pages that seemed as though they were a piece of AOL, and sent parodied messages or texts with a connection to this fake site page, requesting that potential casualties to uncover their passwords [14]. The process of phishing attach is illustrated in Figure 1.4.

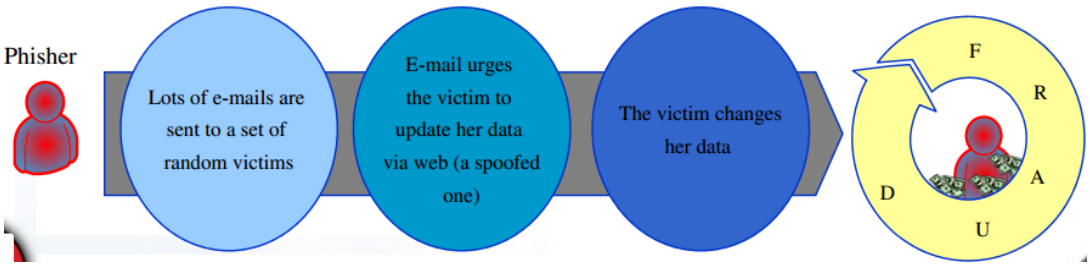

Figure 1.4. A process of phishing attacks [14]

An entire phishing assault includes three parts of phishing process. Firstly, mailers convey countless messages (for the most part through botnets), which guide clients to false sites. In addition, authorities set up false sites (generally facilitated on bargained machines), which effectively incite clients to provide private data. Finally, cashers utilize the classified data to accomplish payments out. Fiscal trades are conducted between those phishers on regular basis.

The most recent insights uncover that banks and monetary organizations alongside the online networking destinations keep on being the fundamental concentration of phishers. Some committed projects are likewise getting to be prevalent among phishers in light of the fact that with them phishers can breach the money related data of casualty as well as utilize existing prize focuses as cash. U.S. remains the biggest target of phishing, representing 61% of phishing destinations described in June 2016 [15]. An investigation of demographic variables proposes that women are more powerless to phishing than men are and clients between the ages of 18 and 25 are more vulnerable to phishing than other age categories [16]. Phishing assaults that at first target general purchasers are currently developing to incorporate prominent targets, intending to take licensed innovation, corporate privileged insights, and delicate data concerning national security.

1.3 Feature Selection

A procedure chooses a subset of unique components. The optimality of an element subset is measured by an assessment basis, as the dimensionality of an area extends the quantity of elements N increments. Finding an ideal component subset is typically, recalcitrant [17] and numerous issues identified with feature selection have been appeared to be NP-hard [18]. A run of the mill feature selection process comprises of four fundamental strides, as depicted in Figure 1.5, to be specific, subset stage, subset assessment, ceasing standard, and result approval [19]. The subset stage is an inquiry methodology [20] that produces competitor includes subsets for assessment in light of a specific hunt procedure. Every subset is assessed and contrasted and the past best one is indicated by a specific assessment rule. If the new subset ends up being better, it replaces the past best subset. The procedure of subset stage and assessment is rehashed until a given ceasing paradigm is fulfilled. At that point, choosing the best subset typically should be

approved by earlier learning or diverse tests by means of engineered and additionally genuine information sets.

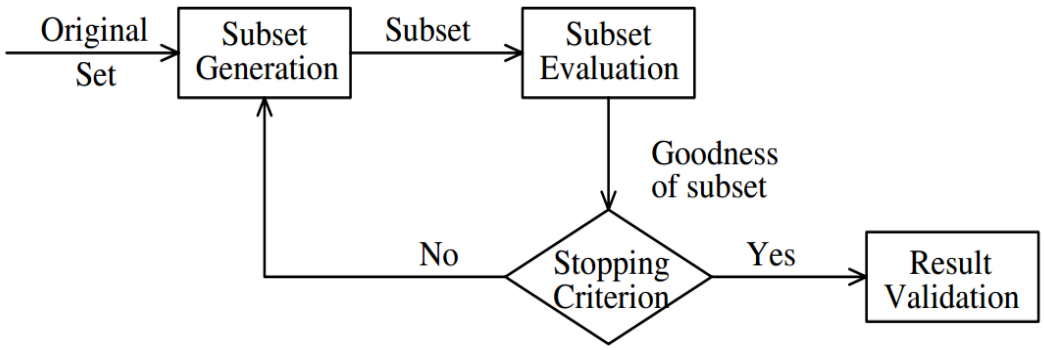

Figure 1.5. Four key steps of feature selection [19]

Feature selection can be found in numerous areas of information mining, for example, characterization, grouping, affiliation rules, and relapse. For instance, the "include choice" is called subset or variable choice in statistics [21].

2. DATA MINING

Data mining is an interdisciplinary subfield of computer science. It is the computational procedure of finding patterns in extensive datasets including strategies at the crossing point of artificial intelligence.

2.1 Data Mining Process

Without a doubt, some people may think of data mining as just a principal venture during the time spent learning a specific feature or task. Knowledge discovery as a procedure is delineated in Figure 2.1 and it comprises of an iterative arrangement of the following accompanying strides:

- Data cleaning
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern evaluation
- Knowledge presentation

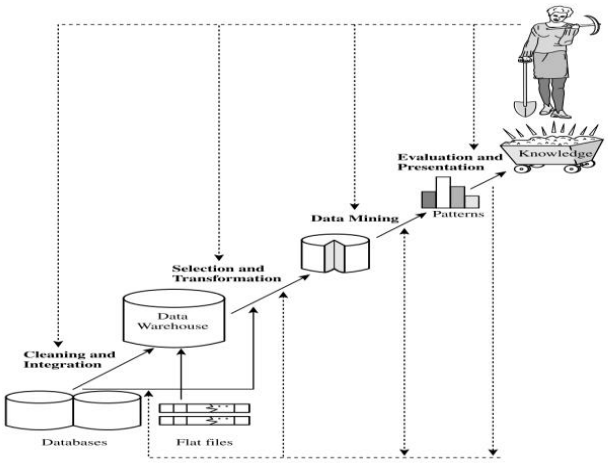

Figure 2.1. Data mining as a step in the process of knowledge discovery[10]

2.1.1 Data Cleaning

Evacuating commotion and right conflicting information is basically called information cleaning [10]; it is a level where clamor information and insignificant information are expelled from the accumulation [11].

2.1.2 Data Integration

Combining information with numerous sources into a rational information store, for instance, an information distribution center [10]. At this stage, various information sources, regularly heterogeneous, might be joined in a typical source [11].

2.1.3 Data Selection

Combining information with numerous sources into a rational information store, for example, an information distribution center [10]. At this stage, various information sources, regularly heterogeneous, might be joined in a typical source [11].

2.1.4 Data Transformation

This stage is where information is changed or merged into structures fitting for withdrawal by accomplishment rundown or collection processes [10]. Otherwise, it is called information combination; the chosen information is changed into structures proper for the mining method in a stage [11]. By applying a change strategy to standardize or institutionalize the factors would be the proper way.

2.1.4.1 2.1.4.1. Min-Max Normalization

Plays out an immediate change on the initial information, which assumes that *minA* and *maxA* are the base and most extreme estimations of a quality, A. Min-max standardization maps an esteem, *vi*, of *A* to v_i' in the variety [new *minA*, new *maxA*] as presented in equation 2.1.

$$v_i' = \frac{vi - min_A}{max_A - min_A}\ (new\ \ max A\ \ - new\ \ min A) + \ new\ \ min A. \qquad (2.1)$$

Min-max standardization secures the associations in the middle of the first information values. It will experience a "beyond the field of play" mistake if a future data circumstance for institutionalization decreases outdoor of the first information goes for A.

Equation 2.1 can be used for min-max standardization for different cases; for example, assume that the base and most extreme quantity for the characteristic wage are \$11,000 and \$97,000 respectively. It might want to guide the wage to the range [0.0, 1.0]. By min-max standardization, an estimation of \$72,400 for money is changed to the following:

$$\frac{72{,}400 - 11{,}000}{97{,}000 - 11{,}000}\ (1.0 - 0) + 0 = 0.714$$

2.1.4.2 2.1.4.2. Z-Score Normalization

Zero-mean standardization, the qualities for a property *an*, are standardized in light of the mean and standard deviation of F. An esteem, v**i** , of *an* is standardized to $V_i' i$ by registering where *an* and F^- are the mean and standard deviation, respectively, of trait F; where $F^- = \frac{1}{n}\ (v1 + v2 + \cdots Vn)$ and F^σ is processed as the square base of the fluctuation of:

$$\text{F}\ (\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \ x^-)^2 = \frac{1}{n}\left[\sum x_i^2 - \frac{1}{n}\ (\sum x_i)^2\right]) \qquad (2.2)$$

This technique for standardization is valuable when the very least and most extreme of trademark are dark, or when there are anomalies that rule the min-max standardization.

For instance, if the mean and standard deviation of the qualities for the ascribed pay are \$53,000 and \$15,500, respectively, with Equation 2.2 for z-score standardization, an estimation of \$72,400 for money is changed to $\frac{72{,}400 - 53{,}000}{15{,}000} = 1.251$. A variety of this z-score standardization replaces the standard deviation of the equation ($v_i' = \frac{\text{vi} - \overline{\text{v}}}{\sigma_{\text{F}}}$) by the

mean total deviation of F. The mean outright deviation of F, denoted S_F, is $S_F = (|v_1 - F^- | + |v_2 - F^-| + \cdots + |v_n - F^-|)$.

Thus, z-score normalization using the mean total deviation will be given in Equation 2.3 below:

$$v_i' = \frac{vi - \bar{v}}{S_A} \quad (2.3)$$

The mean outright deviation, S_F, is stronger to exceptions than the standard deviation, SF. While registering the mean aggregate deviation, the deviations from (| Fi − F^- |) are not squared; thus, the effect of anomalies is to some degree lessened.

2.1.4.3 2.1.4.3. Normalization by Decimal

Scaling is standardized by moving the fraction purpose of characteristics of quality A. The quantity of fraction focuses upon the extreme supreme estimation of A. An esteem, v_i, of *an* is standardized to v_i' by Equation 2.4:

$$v_i' = \frac{v^i}{10^j} \quad (2.4)$$

Where *j* is the smallest integer such that Max ($|v_i'|$) < 1

2.1.5 Data Mining

Data mining is an essential procedure where keen strategies are linked to concentrate information designs [10]. It is the essential stride in which sharp strategies are linked to concentrate designs possibly valuable [12].

2.1.6 Pattern Evaluation

Pattern emulation is employed to perceive the truly intriguing examples related to learning in the view of some intriguing quality measures [10]. Entirely intriguing examples related to learning are distinguished in view of given measures [11].

2.1.7 Knowledge Presentation

Knowledge presentation is applied where observation and learning illustration systems are used to display the extracted information to the client [10]. Further, the last stage in which the discovered information is outwardly spoken to the client. This key stride utilizes representation strategies to help customers comprehend and translate the information mining comes about [11].

2.2 Naive Bayes Classifier

The Naive Bayes classifier is executed for learning the duty where every occurrence *x* is defined by a partnership of feature values and the goal function *f*(*x*) can catch on any value from some constrained set *V*. *B* is the preparation case set of the aim function which is given, and a recent example is displayed and labelled by the tuple of characteristic values < *a1*, *a2*, . . , *an*>. The trainee is requested, that is foresees the aim quality, or classification, for this new example.

The Bayesian method is used for classifying the new item to assign the most probable target value, *vMAP*, given the characteristic values < *a1*, *a2*, . . ., *an* > that describe the item as follows:

$$UMAP = argmax_{vj \in V} \left(P\left(V_j \,|a_1, a_2, \ldots, a_n\right)\right) \tag{2.5}$$

Utilizing Bayes theorem:

$$UMAP = argmax_{vj \in V} \left(\frac{P\left(a_1, a_2, \ldots, a_n | v_j\right) p(V_j)}{P\left(a_1, a_2, \ldots, a_n\right)}\right) \tag{2.6}$$

Where $argmax_{vj \in V} \left(P\left(a_1, a_2, \ldots, a_n \,|v_j\right)\ p(v_j)\right)$

The Naive Bayes classifier prepares the farther improving supposition that the item values are restrictively independent given the aim value. Accordingly,

$$UNB = argmax_{vj \in V}\left(p(v_j) \prod_j \quad p(a_i \backslash v_j)\right) \tag{2.7}$$

Where *vNB* indicates the goal value yield by the Naive Bayes classifier. The restrictive feasibility $P(a_i \backslash v_j)$ should be evaluated from the training set. The previous expectation $P(v_j)$ should also be altered in a few designs (regularly by just checking the frequencies from the training set). The possibility for contrasting hypotheses can be measured by balancing the values obtained for every hypothesis. Naive Bayes is a simple but very successful classifier [22].

2.3 Application Field of DM

Data mining takes place to be high stage to action as professional techniques to fix problems and would not need presumptions to be made about data. These times, minimal enthusiasm is operating out of the technicians connected with method but needs knowledge of data and business problem to expect the designs and actions in a way that is automated. However, data mining is advantageous to deal with currently unknown habits in data which is certainly wide, although mainly utilized for data dredging on behave of employing data mining techniques incorrectly to show inaccurate or untrue findings. Features clearly mention the much further element of data mining usually elevated in facts development applications plus in inclusion developed a range of techniques to keep away from issues being such data mining strategies [23]. It is actually beneficial to mention that dredging can be used as exploratory resource when developing and speculations can be obviously made.

- Forecast with what usually takes destination when you look at the near upcoming.
- Categorizing things into groups based on practices.
- Connecting activities which may tend to be similar and possible to happen concurrently.
- Obtaining the individuals into groups based on their particular certain characteristics.

It can be found in a predictive method for a number of programs to experience objective that is business. Standard uses of data mining tend to be given just under:

a) Fraud or non-compliance anomaly detection: Data mining isolates their components and the prompt to fraudster, excess and manhandle. for example, MasterCard fraud recognition checking [24].

b) **Intrusion detection**: This technique checking and investigates the occasions occurring in a PC framework with an end that is specific to distinguish indications of safety problems [25].

c) **Lie detection (SAS Text Miner):** SAS text miner uses the devices intelligence to detect and identify lies which will assist superiors in automatically detecting anomalies within the Internet or e-mail data [26].

d) **Market Basket Analysis (MBA)**: MBA is fundamentally applicable data mining method in understanding what things tend to be purchased collectively based on connection rules, mostly because of the goal of acknowledging options that are cross-selling [27].

e) **Aid to marketing or retailing:** Via data mining entrepreneurs can directly get important and precise habits on buying behavior of the customers that is useful to them in predicting which things their customers are thinking about purchasing [28].

f) **Phenomena of "beer and baby diapers":** This story of utilizing data mining to find a Web connect beer is certainly intermediary diapers is informed, retold and place into like other legend [29].

g) **Financial, banking and credit or risk scoring:** Data mining can really help cash relevant foundations in a variety of paths, by way of example credit scoring, credit evaluation [30].

h) **Satellite sensing:** There is an endless number of satellites far and wide: some are geo-stationary over a district, and some are circling around the Earth, yet all are sending a constant stream of information to the surface. NASA, which controls a substantial number of satellites, gets more information consistently than what all NASA specialists and architects can adapt to. Numerous satellite pictures and information are made open when they are gotten in the trusts that different analysts can break down them [30].

2.4 Labelled and Unlabelled Data

In like way have a dataset of cases (known as circumstances), each of which includes the values of a true quantity of factors, which in data mining tend to be frequently known as features (a couple of kinds of information) which are dealt with by significantly

different means for the kind that is the first specially designated feature. Also, the main point is to work with the data provided to anticipate the estimation of the quality, for examples that have not yet been seen. This kind of information is called labelled. Data mining that makes use of labelled data is known as supervised learning in the sense that the assigned high quality is absolute. In other words, it must take one of various values that are particular, for example, “great” or “poor” or (in an item acknowledgment application) “auto”, “bike”, “individual”, “transfer” or “taxi”; the task is known as classification. The task is known as relapse on the off chance that the assigned quality is numerical; for example, the normal cost price of a house or the opening cost of an offer on tomorrow's securities trade. Information that doesn't have any uncommon quality that is assigned is known as unlabelled. Data mining of unlabelled data is recognized as unsupervised learning. Here the genuine point is essentially to remove the most absolute data it can from the information available [31].

2.4.1 Classification

Classification pertains to an assignment that happens much associated with the correct amount of time in ordinary life. An opinion surveying business might wish to classify individuals in terms of satisfaction and whether they are most likely to vote in favour of all of numerous governmental policies or tend to be undecided. Similarly, it might wish to classify a student task in terms of difference, quality, pass or fail for example. Or a hospital may choose to classify therapeutic patients into those who are at high, moderate or low risk of suffering an ailment with certainty. This example reveals a scenario that is regular, as shown in Table 2.1. It now has a dataset as a table of understudies, which include evaluations on five subjects (the estimations of characteristics Eng. this is certainly soft, HCI, CSA and Project) by general degree classifications [31]. There are many ways in which it is able to do that, including the following.

2.4.1.1 Nearest Neighbour Matching

Nearest neighbour matching depends on distinguishing say the five situations that are “nearest” in an understanding that a few are unclassified ones. If the five nearest next-door

neighbours have amounts next, First, Second, 2nd and Second, then it might correctly deduce that in the brand name instance that is a brand new be classified as 'Second' [31].

Table 2.1. Degree classification data [31]

SoftEn	ARIN	HCI	CSA	Project	Class
A	B	A	B	B	Second
A	B	B	B	B	Second
B	A	A	B	A	Second
A	A	A	A	B	First
A	A	B	B	A	First
B	A	A	B	B	Second
.....	.....	.....	.....	.....	.....
A	A	B	A	B	First

2.4.1.2 Classification Rules

The instructions can be chosen that individuals may use to anticipate the classification of circumstances that are unseen for the situation, as follows:

THEN Class = First IF SoftEng = A AND Project = A

IF SoftEng = A AND Project = B AND ARIN = B THEN Class =Second

THEN Class = Second IF SoftEng = B

2.4.1.3 2.4.1.3. Classification Tree

A proven way of producing classification principles is via an advanced construction that is tree-like which is called a classification tree or a choice tree, as illustrated in Figure 2.2 [31].

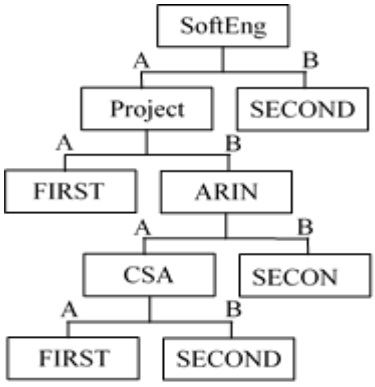

Figure 2.2. Possible decision tree corresponding to the degree classification data [31]

2.4.2 Numerical Prediction

Classification is a kind of forecast, in which the high standard to be anticipated is a level. Numerical anticipation (frequently known as relapse) is an addition. For this circumstance, it desires to anticipate a numerical high standard, as a case, a business's profits or a provided expense. A very common means of carrying this out is to utilize a Neural Network as displayed in Figure 2.3, which is also known by a simplified name as Neural Net. This really is a confusing proving strategy considering a design of a neuron that is human. A neural net is to

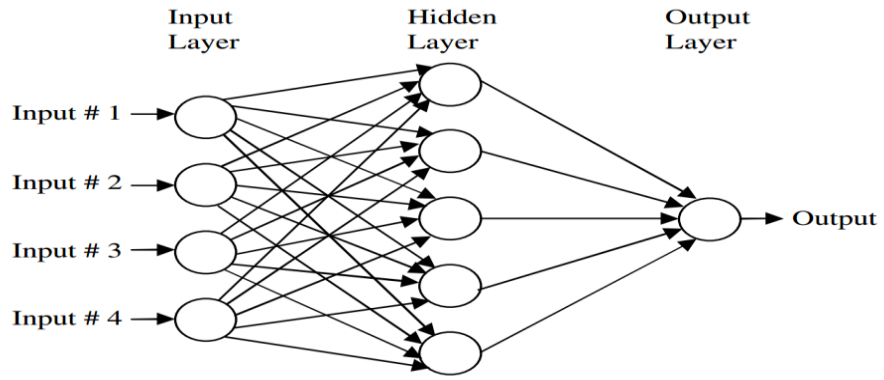

Figure 2.3. Neural network [31]

Provide an order of inputs and it is utilized to anticipate several outcomes. Neural networks are considered important for data mining [31].

2.4.3 Association Rules

Occasionally, it may be desired to use a planning set to find any association that continues among the estimations of variables. When it comes to a part that is mostly instructions it is referred to as association standards. Generally, there are several possible rational association guidelines from any offered dataset. Most of these are not top quality; therefore, association concepts frequently become indicated with a few additional data.

2.4.4 Clustering

Clustering algorithms analyse the information to find gatherings of things that are similar. For example, a guarantee business might cluster clients relating to earnings, age, types of plans bought or earlier needed knowledge. In an analysis that is defective, utility problems could be grouped in keeping with the values of certain key issues as shown in Figure 2.4 [31].

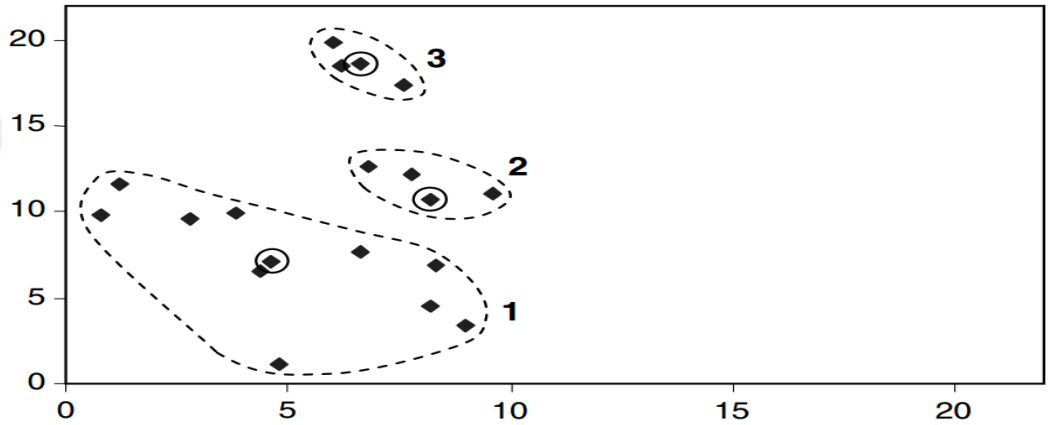

Figure 2.4. Clustering of data [31]

2.5 Feature Selection Techniques

Here occur two practical strategies to look at for feature selection: search for the subset that is best in terms of prescient features (for building productive expectation models) or find all the important features for the class feature. The last is efficient by employing a positioning of the qualities as indicated by their particular prescient vitality, ascertained by means of various strategies: (i) register the performance of a classifier developed with every single customizable, (ii) process measurement, for example a relationship coefficient or even the edge and (iii) utilize data theory activities, similar to the data [32]. In any case, this approach doesn't decide repetitive features that have been demonstrated to impair the order procedure of the Naïve Bayes classifier [33]. Thus, most feature selection systems focus on searching for the subset that is most helpful in terms of prescient features. They vary in two viewpoints that are vital – the search methodology utilized in addition to the feature subset assessment procedure [30].

Feature selection algorithms are often divided in device discovering writing into filter strategy (or filters), wrapper technique (or wrappers) and embedded methods (in other words, methods embedded inside the learning procedure of particular classifiers) [34].

2.5.1 Search Strategies

For the feature selection issue, your request of the search space is $O(2^{|F|})$. Thus, doing a comprehensive search is unfeasible beside spaces with just a couple of features. Full search procedures carry out an aggregate search for the ideal subset, predictable (agreeing with) towards the evaluation reason made utilization of their specific intricacy is smaller than $O(2^{|F|})$, on the grounds that not all subsets are analysed. The optimality with respect to the choice would be ensured in full. Partners for this class have a tendency to be branch and bound with backtracking or expansive first search.

A less inspected procedure is search; this is unquestionably discretionary which limits the sheer number of analysed subsets by setting a most extreme number of practical emphases. The optimality connected with choice differs as indicated by the sources offered and values that are adequate to particular parameters. Delegated with this gathering is the Las Vegas search algorithm. Furthermore, a specific level of haphazardness can be found in innate algorithms and mimicked strengthening; ravenous slant climbing can be infused with arbitrariness by beginning from a preparatory subset that is irregular [35].

2.5.2 Filtering Methods

A filter feature that is executed independently of a particular classifier is propelled by the properties connected with data division itself. There are distinctive algorithms that are strong in abstraction which utilize a channel strategy. Among the most reported are Alleviation [36], LVF [37], Center [38], Connection based channel – CFS [39], or factual works on considering hypothesis tests. LVC (Las Vegas Channel) [37] utilizes a probabilistic-lead and the way it searches is surely discretionary investigating the characteristic subspace, and a consistency appraisal measure more distinctive than the fundamental one used by core interest. The technique is proficient, and has now the fundamental preferred standpoint of having the capacity to discover subsets that are useful

for datasets with sound. Also, an awesome estimation for the answer; this is absolutely last promptly accessible amid the execution connected with the algorithm. One detriment could be the truth, as it won't make utilization of past information so it might take more time to acquire the reply than algorithms making utilization of heuristic era techniques.

2.5.3 Wrapper Method

Since filters disregard catching the predispositions inborn in learning algorithms, for the genuine motivation behind enhancing the class performance, channel techniques won't achieve upgrades that are significant. Rather, wrapper strategy ought to be contemplated. Exploratory results, which approve this assumption, are accessible in [33, 40]. Wrappers [41], rather than channel rehearses, scan for the subset and this is absolutely ideal using an experiential risk estimation for a specific classifier (they perform exact danger minimization). Subsequently, they have been changed to your specific relations between your classification algorithm and the instruction that can be found. One downside is that they are normally rather lazy.

As a rule, a wrapper system comprises three fundamental steps:

- an era technique
- an assessment technique
- an approval technique

In this way, a wrapper is a 3-tuple with respect to the kind <*generation, evaluation, validation*>. The feature selection technique chooses the insignificant subset of features, considering the estimate execution as investigation capacity: reducing the mistake that is approximated or comparably boosting the normal exactness.

3. FEATURE SELECTION FOR DATA MINING

3.1 Introduction

The flood of a vast assortment of organized and semi-organized data has a prompt effect on the various ways that are initiated to gather information. Still, it is a customary feeling that the rate of development of the data available is not coordinated by the improvement of techniques that suitably utilize this data. Thus, the field of Data Mining (DM) has seen rising consideration both in mainstream researchers and in the market, and diverse techniques are being produced to attempt to mine astounding data concealed in information. Normally, the problems handled by data mining are (a) the identification of association rules, that is, rules that express specific blends of elements that are present in the information with high recurrence or likelihood and (b) classification, where one is stated “objects” in association with distinctive classes and is to discover a tenet ready to distinguish components of one class from components of another class.

In various cases, FS can be seen as an autonomous assignment in the DM procedure that pre-forms the data before they are managed by a DM technique that consistently may crash and burn or have colossal computational issues in treating a specific dataset with countless information.

The standard points of interest in utilizing FS as a part of DM may in this manner be illustrated as follows:

- Diminishment in the aggregate of data expected to set up a DM algorithm.
- Better characteristics of the principles gained from information.
- Simpler collection and capable of gaining the data recognized to a smaller whole number of “helpful” components.
- Less expense for obtaining the data (much of the time FS centres on determining a pleasing subset of the accessible elements by minimizing the strong practical cost of getting that component in this present reality) [42].

3.2 Role of Feature Selection in Data Mining

Highlighting the choice is one of the prime considerations in the field of information mining. Expressing a dataset highlighting the choice could be open as the strategy of selecting a subset of elements for creating further information examination. This choice of subset of elements is anticipated by gaining the most extreme data currently in the dataset; that is the choice includes a subset that ought to contain the most evident components related to the model development. Highlight determination is particularly imperative in high dimensional datasets since it reduces dimensionality and thereby refutes the effects of the scourge of dimensionality. Assisting in various real life frameworks, including the choice is vital in arranging the conduct and execution of the framework. In particular, in biomedical applications, highlight choice could assume an essential part in order biomarkers. In an infection characterization issue in genomic consideration, for instance, highlights choice procedures could arrange the qualities that differentiate the unhealthy and sound cells. This is not only helping the information expert in diminishing information measurement, but on the other hand is a gigantic achievement for scientists to grasp the organic framework and order the malady activating qualities [43].

3.3 Feature Selection Algorithms

3.3.1 Forward Feature Selection

A forward selection exploration begins with a single assessment of every feature. For every element, a feature selection standard, *J* feature, is estimated. In addition, the feature by the best record (highest estimation of the performance rule) is chosen for the following step of the exploration (a "victor" – a predecessor of the sub-tree). At that point, in the following stage, one extra element is added to the choice of "victory" feature (having the best estimation of the measure) after past stage, starting with all conceivable, as illustrated in Figure 3.1.

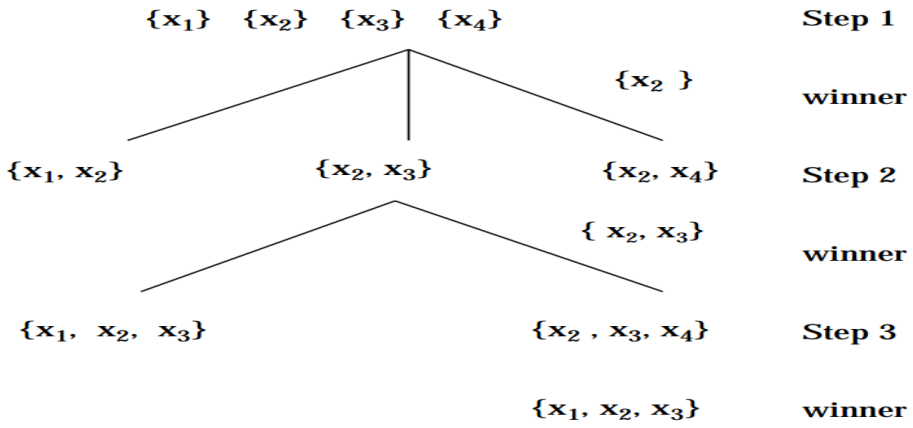

Figure 3.1. Sequential forward feature selection search [44].

Two-feature subsets have a "victor" in every subsection by combining elements that are routinely assessed. Besides, those showing the highest increment of the performance basis are chosen as a champ and replacement of the following stage. The technique proceeds up to the greatest *m*-feature subsection where the "champ" of the m^{th} stage has been prepared [44].

Algorithm: Feature selection by stepwise forward search.

Assumed: A data set $Tall$ with $Nall$ named designs comprising *n* elements *X* = {*x*1, *x*2, ⋯ , *x*n}; a number *m* of features in the resultant subset of best elements; and a feature subsection assessment foundation $Pfeature$ with a characterized method for its estimation in view of a constrained size data set $TXfeature$.

1. Set an underlying "victory" feature subsection as a vacant set $Xvictory, 0 = \{\ \}$.
2. Set a stage number $p = 1$.
3. Procedure conceivable $n - p + 1$ subsections, with an aggregate of j components, that include a triumphant "victory" $p - 1$ feature subset $Xvictory, p - 1$ from the past step, by one new element included.
4. Assess the feature selection paradigm for every element subset shaped in stage *p*. Select as a victory a subsection $Xvictory, p$ with a bigger increment *A* of the performance paradigm $Pfeature$ as compared with the highest rule numerical quantity (the victory subset $Xvictory, j - 1$) after the past step.

5. If $P = m$, in that point of break off. The victor $Xvictory, p$ subsection in stage P is the last chosen subsection of m features. Alternatively, set $p = p + 1$ and proceed from stage 3.

The forward selection algorithm gives a suboptimal solution, since it doesn't inspect every single conceivable subset of components. The basic forward selection strategy accepts that the number of features m in a subsequent subsection is recognized. This method will need precisely m stages. At this time, the best possible number of features m must be found. This circumstance characterizes an additional exploration procedure by ceasing the foundation $Pfeature, length$. Here, a conceivable halting measure for discovering the correct number m of elements in a last feature subsection choice can be, for instance, a characterized $\in length$ of maximum performance adds for two continuous steps. Moreover, the terminating point is reached when the expansion in the feature selection basis for the P^{th} stage victory element $Xvictory, j$, as contrasted and the comparing execution or a victory feature subsection after the past step $p - 1$, is a smaller extent than the characterized limit $\in length$ [44] as follows:

$$Pfeature, length = Pfeature\,(Xvictory, j) - Pfeature\,(Xvictory, j - 1) < \in length$$

3.3.2 Backward Feature Selection

Backward selection is similar to forward selection; however, it applies a turned around method of feature selection, beginning with the whole list of capabilities and removing features each one in turn. In backward selection, accepting a known number m of latest elements, the examining begins through the assessment of the whole arrangement of n features. For the whole list of capabilities, a selection rule $Jfeature$ is assessed. At that point, in the following step, every conceivable subset containing features from the past step with one component disposed of is organized and their performance standards are assessed. At every progression, one feature, which gives the least reduction in the esteem of feature selection determination incorporated into the past step, is disposed of. The methodology proceeds pending the best m-feature subsection is found in [44, 45].

Algorithm: Feature selection by stepwise backward search.

Assumed: A dataset $Tall$ with $Nall$ marked examples comprising *n* components $X = \{x1, x2, \cdots, xn\}$; a number m of features in the subsequent subsection of finest elements and an element subsection assessment measure $Jfeature$ with a characterized strategy for its computation is dependent upon a limited-size of data set $TXfeature$.

1. Assess a feature selection foundation $Jfeature$ (X) for a set X of all n elements.
2. Set a stage number $j = 1$ with a rundown X of all n elements.
3. Procedure all $n - j + 1$ conceivable subset with $n - j$ features by disposing of one component at once from the rundown of features of the past step.
4. Assess a feature selection standard for every element subsection organized in step j. Select as a "victory" a subset $Xvictory, j$ with the least abatement of a performance basis $Jfeature\ (Xvictory, j)$ as contrasted and the foundation esteem from the past step (which compares to its greatest esteem for this progression from a pool of all subsections). The throwaway feature from the past step, which brought on the making of the victory subset $Xvictory, j$ is then disposed of from a pool of elements utilized as a part of the following step (next step), and winning subset turns into a progenitor of a more profound sub-tree.
5. If $j = m$, then halt: the victory subsection in step j is the last chosen subsection of m features. Else, set $j = j + 1$ and proceed from stage 3.

The forward selection algorithm gives an imperfect "suboptimal" answer (arrangement), since it does not test all conceivable subsections of features. The backward selection algorithm imposes more extra escalated computations than the forward selection. Regardless of likenesses, both algorithms may give distinctive results for the similar circumstances.

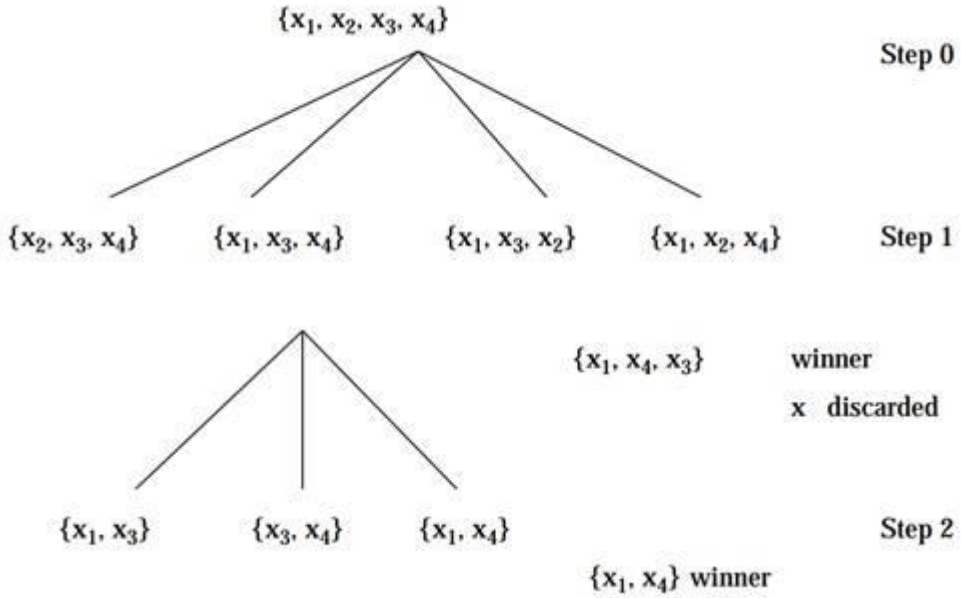

Figure 3.2. Sequential backward feature selection search [45]

On the off chance that the number *m* of last features is an obscure one from the earlier, then additional best search ought to be utilized. Discovery of the best possible number *m* of features in the last chosen feature subsection can be acknowledged in a way like the technique portrayed before for forward selection.

The forward and backward search techniques can be consolidated in a few techniques, permitting them to cover more component subsets through expanded computations, and along these lines to discover better problematic capabilities. For example, in the so-called full stepwise search, processes at every progression begin as in the backward search. All subsections are made by expelling one variable after the past step pools are assessed. On the off chance that the feature selection rule abatement is underneath a characterized edge, then a variable is removed. In the event that none of the variables give an abatement beneath the limit, then a variable is included, as in the forward search technique.

3.3.3 Individual Feature Selection

The easiest technique, and maybe the one giving the lowest performance, for picking the finest *N* features is to relegate a separation control gauge to each of the features in the first set, *X*, individually. In this manner, the features are requested in the following way:

$$J\,(x\_1)\ \geq\ J\,(x\_2)\ \geq \ldots \geq\ J\,(X_P) \tag{3.1}$$

Furthermore, the choice as our best set of *N* arrangements for the *N* features with the finest individual record:

$$\{X_i \backslash\, i\ \leq\ N\} \tag{3.2}$$

Occasionally, this strategy can create sensible feature sets, particularly if the features in the first set are unrelated; meanwhile the technique disregards multivariate connections. Notwithstanding, if the features of the initial set are greatly connected, the picked feature set will be problematic, as a portion of the features will include minimal biased authority. There are situations when the *N* best features are not the best *N* features notwithstanding when the variables are independent [46,47].

3.3.4 Plus-l Take Away-r Feature Selection

This is a technique that permits some backtracking in the feature selection procedure. If $l\ >\ r$, it is a base "bottom-up" method. l features are put into the current set utilizing SFS; after that the nastiest r features are detached utilizing SBS. This algorithm removes the difficulty of nesting because the set of features obtained at a given stage is certainly not inevitably a subset of the features at the following stage of the methodology. If $l\ <\ r$ then the strategy is "top-down", beginning with the total set of features, expelling *r*, then adding l progressively pending the prerequisite number is accomplished [47, 48].

Generalized plus l – take away r selection

The version that is generalized of $l-r$ algorithm utilizes the algorithms at every stage as opposed to the SFS and SBS processes. Generalizing the strategy more by permitting

the true numbers l and r to be made from several elements l_i, i = 1, ... , n_l, and r_j, j= 1, ... , n_r (where n_l and n_r will be the number of elements), satisfying the following rules:

$0 \leq l_i \leq l \qquad 0 \leq r_j \leq r$

$\sum_{i=l}^{n_l} li = l \qquad \sum_{j=1}^{nr} rj = r$

In this generalization (speculation), rather than applying the summed up consecutive forward selection in a single stage of l variables (indicated SFS(l)), the feature set is increased in nl stages by including li features ($i = 1,\ldots,nl$) at every addition; that is, applying SFS(li) progressively for $i = 1,\ldots,nl$. This reduces the computational multifaceted nature. Likewise, SBS(r) is swapped by applying SBS(r_j), $j = 1,\ldots,nr$, progressively. The algorithm is alluded to as the (z_l , z_r) algorithm, wherever Z_l and Z_r indicate the grouping of whole numbers l_i and l_j as follows:

$$z_l = (l_1 , l_2 , \ldots l_{n1})$$

$$z_r = (r_1 , r_2 , \ldots r_{n1})$$

The suboptimal quest algorithms examined in this subsection and the thorough search methodology might be thought to be unique instances associated with the (Z_l , Z_r) algorithm.

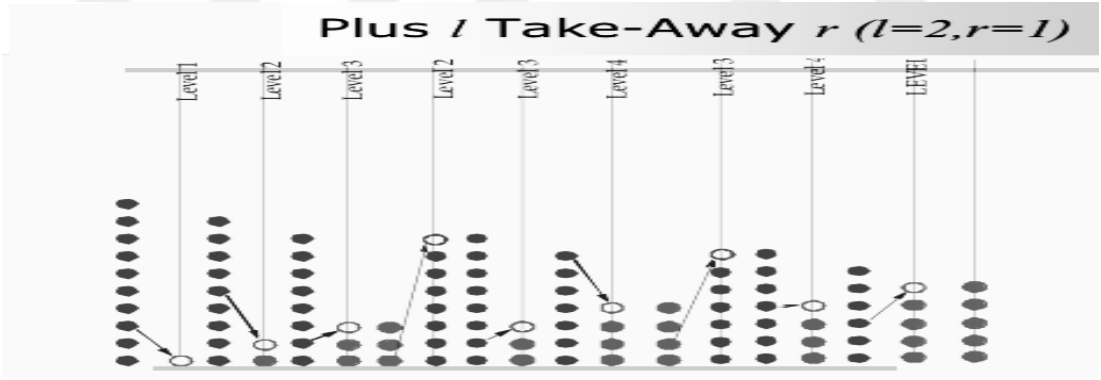

Figure 3.3. Plus-l take Away-r feature selection process [48].

3.3.5 Association Rules Feature Selection

Association Rules (AR) is a method used for finding the associations and/or relationships among items in large databases. Therefore, it can be used for detecting relations among inputs of any system and later eliminating some unnecessary inputs. There is more than one technique for AR algorithm; however, AR1 has been used in this thesis.

AR1 is an AR technique that uses all input parameters and all their records to find relations among the input parameters. If rules that have enough support and high confidence values can be found, then some inputs can be eliminated relying on these rules [49].

3.4 Phishing Techniques

Phishing is the strategy used to take individual data through spamming or other deceptive means. There are various diverse phishing procedures used to obtain individual data from clients. As innovation turns out to be more exceptional, the phishing procedures being utilized are likewise more progressed. To anticipate Internet phishing, clients ought to know about different sorts of phishing systems and they ought to likewise know about combatting phishing methods to shield themselves from being compromised. In the following subsections, user take a look at some of these phishing procedures [50].

3.4.1 Email / Spam

Phishers may send a similar email to many clients, asking them to fill in individual points of interest. These points of interest will be utilized by the phishers for their unlawful exercises [50].

3.4.2 Instant Messaging

Texting is the strategy in which the client receives a message with a connection guiding them to a fake phishing website, which has an indistinguishable look and feel from the genuine website. In the event that the client doesn’t take a look at the URL, it might be difficult to differentiate between the fake and real websites. At that point, the client is requested to give individual data on the page [50].

3.4.3 Trojan Hosts

Trojan hosts are imperceptible programmers attempting to sign into your client record to gather certifications through the local machine. The procured data is then transmitted to phishers [50].

3.4.4 Key Loggers

Key lumberjacks allude to the malware used to distinguish contributions from the keyboard. The data is sent to the programmers who will interpret passwords and different sorts of data [50].

3.4.5 Content Injection

Content injection is the procedure where the phisher changes a part of the substance on the page of a solid website [50].

3.4.6 Phishing through Search Engines

Some phishing tricks include web indexes where the client is coordinated to item destinations, which may offer minimal effort items or administrations. Here, when the client tries to purchase the item by entering the credit card security elements, the information is gathered by the collecting website. There are numerous fake bank sites offering Visas or advances to clients at a low rate; however, they are really phishing websites [50].

3.4.7 Phone Phishing

In telephone phishing, the phisher makes telephone calls to the client and requests that the client dial a number. The target is to gain individual data of the financial balance through the telephone. Telephone phishing is for the most part carried out with a fake guest ID [50].

3.4.8 Malware Phishing

Phishing tricks including malware oblige it to keep running on the client's PC [50]. Malware is a piece of programming developed either with the end goal of attacking a figuring device or for gaining profit by the disadvantage of its client [51]. The malware is generally attached to the email sent to the client by the phishers. When you tap on the connection, the malware will begin working.

3.5 Definition of Phishing Website

There are numerous definitions of a phishing site; it need to be exceptionally watchful how users characterize the term since it is always advancing. One of these definitions comes from the Anti-Phishing Working Group (APWG): “Phishing assaults utilize both social designing and specialized subterfuge to take buyers' close to home character information and money related record certifications” [52]. Normally, a phishing assault is a mix of deceitful messages, fake sites and wholesale fraud. Web clients or clients of numerous banks and budgetary organizations are the objectives of phishing assaults [53].

Phishing is a specific sort of junk mail, which replicates common structures. Phishing fakes are described as endeavours to take on the appearance of a reliable individual or copy a set up and reputed business in an automated correspondence; for example, email or site [54]. The goal is to trap beneficiaries into revealing security data, for example, financial balance numbers, passwords, and credit card points of interest. A person occupied with phishing exercises is known as a phisher [55]. Phishing site assaults utilize sites intended to look as though they originate from a known and true association, keeping in mind the end goal to swindle clients into revealing individual, money related or PC account data. The aggressor can then utilize this data for criminal purposes; for example, wholesale fraud, theft or misrepresentation. Clients are deceived into revealing their data such as ledgers, Mastercards and so forth. Moreover, by transferring and introducing antagonistic programming [56].

3.6 Evolution of Phishing

Around the beginning of phishing history, phishers were regularly acting alone or in small, unsophisticated social groups. Composing consistently delineates early phishers as youngsters obtaining account data to realize mischief and to make long-distance phone calls, generally speaking with a low level of affiliation or perniciousness [57]. As money related associations have extended their online availability and interest, the monetary advantage of trading online account information has extended greatly. Phishing ambushes are ultimately becoming progressively capable, organized and methodical.

From the 1990s, after the prominence of the web, America Online (AOL) transformed into the primary concentration of the phishing assaults. The principal attempts at hacking into AOL were aimed at genuine AOL accounts, and the phishing attacks were connected with the mass of items, which were exchanged through criminal programs. There were projects (such as AOHell) that automated the strategy of phishing for records and credit information. In those days, phishing wasn't utilized as much as a piece of email appeared differently in relation to Web Hand-off Talk (IRC) or the advice prepared framework that AOL used. Phishers regularly presented themselves as AOL staff and sent messages to the customers. They sent messages, for instance, "check your record" or "assert charging information" to draw victims into revealing passwords or other sensitive information. The information they procured would be used to trade the items in mass.

With the growing advancement of online monetary services and e-business, the convergence of phishing strikes swung to purchasers of online banks, online retailers and other online organization providers, for instance, eBay or PayPal. The dubious media of phishing are seen in online social affairs of e-banks, web transfer talking (IRC), texting (IM) and email. Customarily, the phishers present themselves as a specialist of an online affiliation, they gain trust from the customers of the affiliation, and a short time later misdirect the buyers into passing on their sensitive information.

Phishing started as messages with the goal of gaining a reply with the data asked for. This is still the most generally perceived methodology for beginning phishing attacks (assaults); however, today phishers use a couple of differing ways to deal with accumulating the data they require. Repeated locales, Trojans, key-lumberjacks and screen shots are just different unmistakable strategies they are in the blink of an eye using [58].

Phishers began to make fake locales to increase the worrying rate of phishing. For instance, phishers enrol a large number of domain names that look like a renowned brand, for instance, “www.cit1bank.com” or “www.citi-bank.com”. Victims who enter one of these destinations by making mistakes in typing or by falling for the phisher's ploy, may assume that the site is the honest to goodness one, and work their record on the site. Phishers embed site plans into the email messages, complementing them with stolen logos and trademarks from the association concerned, and create the entry address so that the delivery appears to originate from the genuine affiliation [59].

Various new attacks fuse a connection to a bona fide money saving site beyond anyone’s ability to see; however, a fake “login” box is put front of the genuine site. Unmistakably it is, moreover, convincing in light of the fact that the genuine site and the scam appear, in every way, to be from a comparative source. In the wake of surrendering individual financial information data on a phishing site, the casualty is occupied with the veritable landing page of the association being focused on. Thus, the loser won’t question if the site is false [60].

Phishing site attacks are created at a rapid pace. The number of phishing attacks and reported phishing destinations are extending every year, even every month. Harm realized by phishing is outrageous. The APWG is an industry association focused on wiping out discount misrepresentation and information theft that results from the development of phishing and fake emails. This international based affiliation provides a platform to discuss phishing issues, trials and appraisals of potential innovation arrangements, and gives access to a collaborative source of reports on phishing attacks [61].

The number of special phishing sites identified by this association demonstrated that there has been a tremendous increment in one of a kind phishing sites everywhere throughout the world. From October 2015 to March 2016, according to the APWG worldwide phishing overview reported there were no less than 447,945 phishing attacks as shown in Figure 3.4 [62].

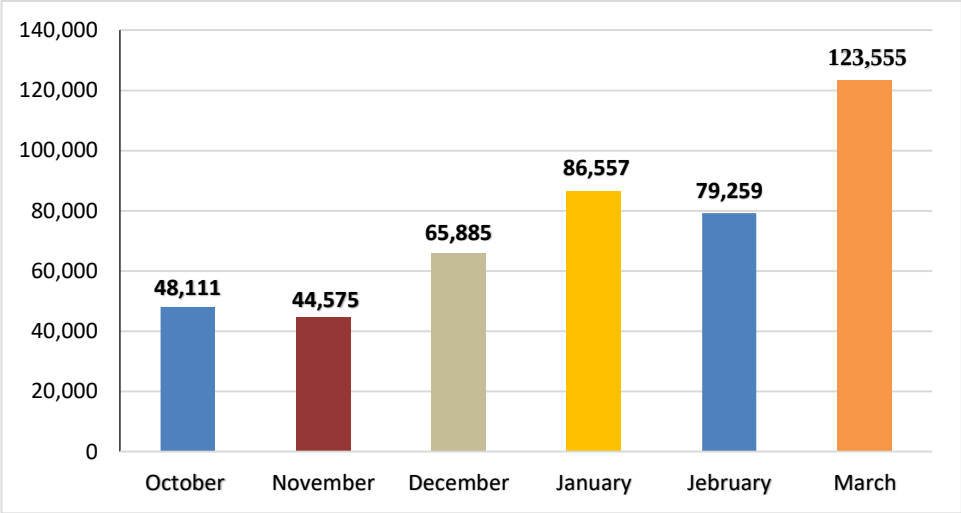

Figure 3.4. Unique phishing sites detected october 2015- march 2016 [62].

In the first half of 2016, a report distributed by APWG demonstrated that the aggregate number of unique phishing sites rose and there were 755,436 phishing attacks as shown in Figure 3.5.

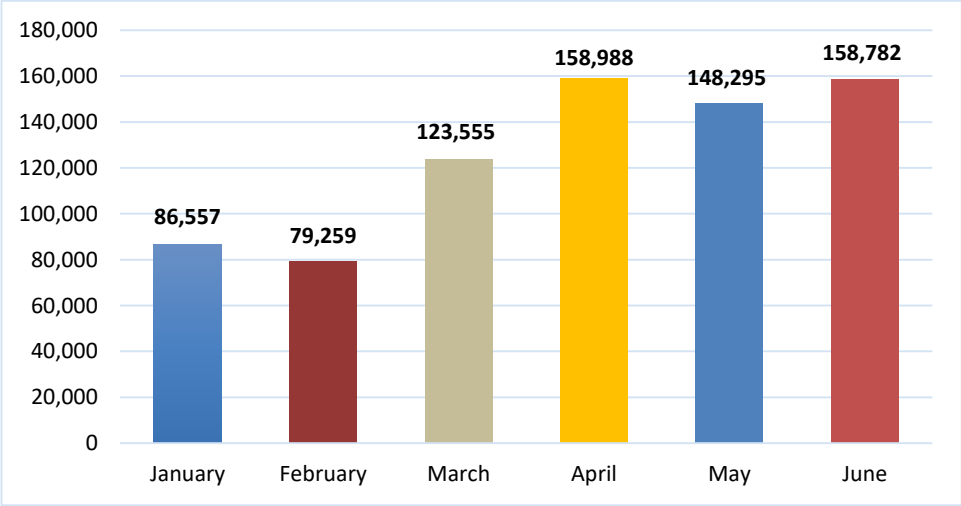

Figure 3.5. Unique phishing sites detected january - june 2016 [62].

In any case, the aggregate number of remarkable phishing websites seen in the second quarter of 2016 was 466,065. This was 61% higher than the past quarterly record in Q4, 2015 [62].

3.7 Types of Phishing

Phishing has spread beyond email to include VoIP, SMS, texting, long-range informal communication destinations, and essentially multiplayer recreational games. The accompanying are some real classes of phishing.

Clone Phishing: In this type of phishing, a phisher makes a cloned email. He does this by gaining information, for instance substance and recipient addresses from an authentic email, which was passed on officially [63].

Spear Phishing: Skewer phishing focuses on a specific group. So rather than sending out a considerable number of messages, stick phishers centre on selected groups of people with something in common, for example people from a comparable affiliation [64].

3.8 Phishing Websites Dataset

A dataset is an accumulation of data information coordinated as a stream of bytes in the logical report. The physical structure of every record is about the same and homogeneous through a dataset. This means the data might be of any kind (binary, floating point or characters) without utilizing a false end-of-record condition [47].

In this study, a phishing Website Dataset that includes 31 features taking one of two, a binary or a ternary value, was used. This dataset was taken from UCl Machine Learning Repository [65]. There are 11,055 records in this dataset and each record has 31 features. The imperative elements that have turned out to be sound and powerful in foreseeing phishing sites will be presented [66,67].

Table 3.1. Features of Phishing Website Dataset

ID	Feature Name	ID	Feature Name	ID	Feature Name
1	Having_Ip_Address	12	HTTPS_Token	23	Iframe
2	URL_Length	13	Request_URL	24	Age_Of_Domain
3	Shortinig_Service	14	URL_Of_Anchor	25	Dnsrecord
4	Having_At_Symbol	15	Links_In_Tags	26	Web_Traffic
5	Double_Slash_Redirecting	16	SFH	27	Page_Rank
6	Prefix_Suffix	17	Submitting_To_Email	28	Google_Index
7	Hsving_Sub_Domain	18	Abnormail_URL	29	Links_Poiniting_To_Page
8	Sslfinal_State	19	Redirect	30	Statistical_Report
9	Domain_Registeration_Length	20	On_Mouseover	31	Result
10	Favicon	21	Right Click		
11	Port	22	Popupwidnow		

4. APPLICATION AND RESULT

The main goal in this part is to reduce the number of features, in order to reduce the running time of the feature for the feature selection algorithm and find the highest classification performance for the best set of features. For this issue MATLAB software has been used for coding and designing my application. The dataset was taken from UCl Machine Learning Repository constituting 11,055 records with 31 features and the 31st feature is output of this dataset as mentioned earlier in Chapter Three [67]; moreover, the minimal number of feature selections that could be accustomed to phishing sites. This study was performed on a phishing website dataset by using the classification process. The Naive Bayes classifier was used for the classification process. The FS algorithm process was used to reduce the number of features. Some of the FS algorithms such as individual, forward selection, backward selection, association rules, and plus-l take away-*r* were used. The plus-l take away-*r* algorithm was performed with *l*=3 and *r*=1. For the first step, the dataset was input to our software, then feature selection algorithm and Naive Bayes classification was applied. The steps of application are explained in Figure 4.1.

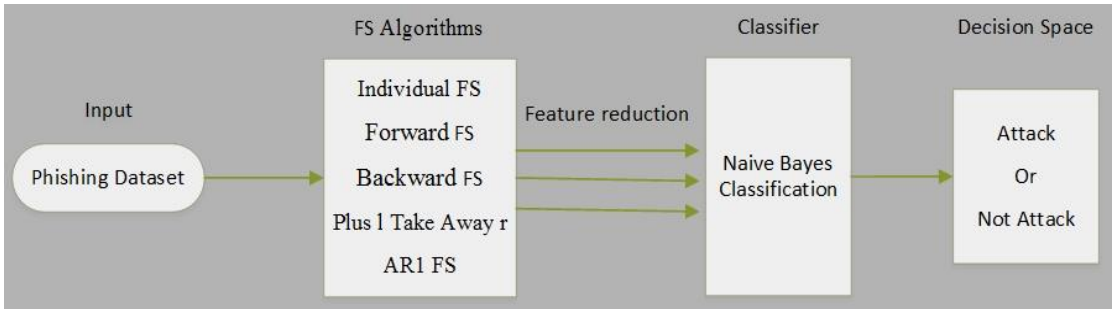

Figure 4.1. Flow diagram of application

The dataset was separated into training data and testing data. It facilitates the display of training data and testing data, which were used to find the best performance accuracy for Plus-l take away-*r* which includes 27 features.

In the classification process, the 5-fold cross-validation technique was used. The dataset is divided into five parts X1, X2, X3, X4, X5, which was applied for calculating average performance values automatically. Each part of cross-validation consists of 2,211 records. Table 4.1 shows the 5-fold cross-validation performance accuracy.

Table 4.1. 5-Fold cross validation performance accuracy

5-Fold	Training data	Test data	Classification rate (%)
Fold 1	X2,X3,X4,X5	X1	93.6228
Fold 2	X1,X3,X4,X5	X2	93.5323
Fold 3	X1,X2,X4,X5	X3	93.6228
Fold 4	X1,X2,X3,X5	X4	93.3062
Fold 5	X1,X2,X3,X4	X5	92.8539
		Average	**93.3876**

The outcome performance in the NB classification technique is based on the evaluation of the confusion matrix. A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier"). Confusion matrix includes four cases True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Confusion matrix that comprises three cases: Sensitivity, Specificity and Accuracy. True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values are utilized to calculate the Sensitivity, Specificity and Accuracy. TP, is a properly reported true positive test result in the data. FP, false positive is any error reported in the data indicating improper test results, which are negative as positive. TN is a properly reported true negative test result in the data. FN, false negative is any error reported in the data indicating improper test results, which are positive as negative.

Table 4.2. Confusion matrix

	Predicted condition	
True Condition	**Predicted Condition positive**	**Predicted Condition negative**
	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

The confusion matrix is a dimension of NxN (mostly 2x2) table containing TP, TN, FP and FN.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad \text{Spesitivity} = \frac{TN}{TN+FP} \quad \text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP}$$

4.1 Feature Selection for Phishing Dataset with Naïve Bayes Classifier

This section shows the results of classifications and removing features. For classifications the Naïve Bays classifier was used. Removing features used a feature selection algorithm such as individual, forward, backward, AR1, and plus-l take away-r.

4.1.1 Individual Feature Selection (IFS)

IFS was explained in Chapter Three. The algorithm performance was compared for every feature category in our model. Different performance results were obtained for each individual algorithm. A performance of 88.88% was obtained using one feature. Various experiments were tested to obtain the accuracy range. Table 4.3 presents performance accuracies based on the conducted experiments. As highlighted in Table 4.3, the best performance is obtained with an accuracy of 88.8919 % where 1 feature was included.

Table 4.3. The results of individual feature selection algorithms with Naïve Bayes classifier

Selected Feature	Accuracy	Selected Feature	Accuracy	Selected Feature	Accuracy
8.th	88.8819	1.th	56.1800	22.th	55.6900
14.th	84.7200	2.th	55.9700	23.th	55.6900
26.th	68.5800	3.th	55.6900	27.th	55.6900
7.th	66.2500	5.th	55.6900	29.th	55.6900
15.th	63.0800	10.th	55.6900	24.th	55.6400
9.th	62.4700	12.th	55.6900	4.th	55.4800
13.th	61.4200	17.th	55.6900	16.th	55.4600
28.th	58.5300	18.th	55.6900	20.th	55.3900
6.th	57.5500	19.th	55.6900	11.th	55.3600
30.th	56.8400	21.th	55.6900	25.th	55.2800

Table 4.4 shows the confusion matrix for IFS and NB classification. The predicted number of True Positive (TP) attacks is 4,197 and the predicted number of False Negative (FN) attacks is 701, the predicted number of False Positive (FP) attacks is 527, and the predicted number of True Negative (TN) attacks is 5630.

Table 4.4. Confusion matrix for IFS by use NB classifier

		Classified	
		Attacked	Not attacked
Actual	Attacked	4197	701
	Not Attacked	527	5630

Also, the sensitivity and specificity are found.

$$\text{Sensitivity} = \frac{4197}{4197 + 701} = 85.68\ \%$$

$$\text{Spesitivity} = \frac{5630}{5630 + 527} = 91.44\ \%$$

Following, the individual feature was tested together to achieve better performance. The best performance, in this case, is found at an accuracy of 93.0077% where 27 features were included. For the least number of features, performance was found at an accuracy of 92.7454% where 12 features were used; all the results are shown in detail in Table 4.5.

Table 4.5. The results of individual feature selection algorithms together with Naïve Bayes classifier

Num. of Features	Individual Features Selection Process	Accuracy
1	8	88.8919
2	8,14	91.2619
3	8,14,26	91.0629
4	8,14,26,7	90.7282
5	8,14,26,7,15	91.3252
6	8,14,26,7,15,9	91.6237
7	8,14,26,7,15,9,13	91.9674
8	8,14,26,7,15,9,13,28	92.1212
9	8,14,26,7,15,9,13,28,6	92.6368
10	8,14,26,7,15,9,13,28,6,30	92.6368
11	8,14,26,7,15,9,13,28,6,30,1	92.673
12	**8,14,26,7,15,9,13,28,6,30,1,2**	**92.7454**
13	8,14,26,7,15,9,13,28,6,30,1,2,3	92.7092
14	8,14,26,7,15,9,13,28,6,30,1,2,3,5	92.7273
15	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10	92.7001
16	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12	92.673
17	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17	92.7273
18	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18	92.7001
19	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19	92.7182
20	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21	92.7454
21	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22	92.7363
22	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23	92.7273
23	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27	92.7092
24	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27,29	92.673
25	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27,29,24	92.664
26	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27,29,24,4	92.6368
27	**8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27,29,24,4,16**	**93.0077**
28	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27,29,24,4,16,20	92.9534
29	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27,29,24,4,16,20,11	92.9715
30	8,14,26,7,15,9,13,28,6,30,1,2,3,5,10,12,17,18,19,21,22,23,27,29,24,4,16,20,11,25	92.9806

Table 4.6 shows the confusion matrix for IFS and NB classification obtained for features together. Individual features were tested together to achieve better performance. The predicted number of TP attacks is 4,432 and the predicted number of FN attacks is 466, the predicted number of FP attacks is 313, and the predicted number of TN attacks is 5,844.

Table 4.6. Confusion matrix for IFS by NB classifier

		Classified	
		Attacked	Not attacked
Actual	Attacked	4432	466
	Not Attacked	313	5844

Also, the sensitivity and specificity are found.

$$\text{Sensitivity} = \frac{4432}{4432 + 466} = 90.48\ \%$$

$$\text{Spesitivity} = \frac{5844}{5844 + 313} = 94.91\ \%$$

4.1.2 Forward Feature Selection (FFS)

FFS was explained in Chapter Three. For FFS, several experiments were conducted; the best performance was found with an accuracy of 93.3514%, which was obtained when 24 features were used. For the second least number of features, performance is found with an accuracy of 93.34% when 27 features are used. Another good performance with fewer features is found at an accuracy of 93.3333% when 19 features are used. The detailed experiment results for FFS are tabulated in Table 4.7.

Table 4.7. The results of forward feature selection algorithms with Naive Bayes classifier

Num. of Features	Forward Feature Selection Process	Accuracy
1	8	88.8919
2	8,14	91.2619
3	8,14,6	91.7322
4	8,14,6,15	92.3293
5	8,14,6,15,16	92.4921
6	8,14,6,15,16,28	92.6097
7	8,14,6,15,16,28,26	92.7635
8	8,14,6,15,16,28,26,9	92.9353
9	8,14,6,15,16,28,26,9,24	93.1796
10	8,14,6,15,16,28,26,9,24,2	93.1705
11	8,14,6,15,16,28,26,9,24,2,20	93.1976
12	8,14,6,15,16,28,26,9,24,2,20,27	93.2519
13	8,14,6,15,16,28,26,9,24,2,20,27,23	93.2519
14	8,14,6,15,16,28,26,9,24,2,20,27,23,21	93.2519
15	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11	93.2610
16	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11.3	93.2791
17	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19	93.2791
18	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17	93.3062
19	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12	93.3333
20	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10	93.3062
21	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22	93.3152
22	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5	93.3333
23	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4	93.2881
24	**8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4,18**	**93.3514**
25	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4,18,29	93.2429
26	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4,18,29,1	93.2881
27	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4,18,29,1,25	93.3424
28	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4,18,29,1,25,30	93.3243
29	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4,18,29,1,25,30,13	92.9806
30	8,14,6,15,16,28,26,9,24,2,20,27,23,21,11,3,19,17,12,10,22,5,4,18,29,1,25,30,13,7	92.9806

Table 4.8 shows the confusion matrix for FFS and NB classification. Several experiments were conducted; the predicted number of TP attacks is 4,445 and the predicted number of FN attacks is 453. The predicted number of FP attacks is 282 and the predicted number of TN attacks is 5,875.

Table 4.8. Confusion matrix for FFS by NB classifier

		Classified	
		Attacked	Not attacked
Actual	Attacked	4445	453
	Not Attacked	282	5875

Also, the sensitivity and specificity are found.

$$\text{Sensitivity} = \frac{4445}{4445 + 453} = 90.75\ \%$$

$$\text{Spesitivity} = \frac{5875}{5875 + 282} = 95.41\ \%$$

4.1.3 Backward Feature Selection (BFS)

BFS was explained in Chapter Three. Similarly, for backward selection, several experiments are conducted. The best performance is obtained at an accuracy of 93.1886% when 25 features are used. For the second least number of features, performance is found at an accuracy of 93.1705% when 27 features are included. The arrow symbol (→) shows "remove one feature in each iteration". The complete series of the experiment results is listed in Table 4.9.

Table 4.9. The results of backward feature selection algorithms with Naïve Bayes classifier

Num. of Features	Backward Feature Selection Process	Accuracy
30	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30	93.0077
29	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,25,26,27,28,29,30 → **24**	93.0891
28	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,25,26,28,29,30 → **27**	93.1705
27	1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18,19,20,21,22,23,25,26,28,29,30 → **11**	93.1796
26	1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,19,20,21,22,23,25,26,28,29,30 → **18**	93.1705
25	**1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,19,20,21,23,25,26,28,29,30 → 22**	**93.1886**
24	1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,19,20,21,25,26,28,29,30 → **23**	93.1796
23	1,2,3,4,6,7,8,9,10,12,13,14,15,16,17,19,20,21,25,26,28,29,30 → **5**	93.1524
22	1,2,3,4,6,7,8,9,10,12,13,14,15,16,17,19,20,25,26,28,29,30 → **21**	93.1615
21	1,2,3,4,6,7,8,9,10,12,13,14,15,16,17,20,25,26,28,29,30 →**19**	93.1524
20	1,2,3,4,6,7,8,9,12,13,14,15,16,17,20,25,26,28,29,30 → **10**	93.1253
19	1,2,4,6,7,8,9,12,13,14,15,16,17,20,25,26,28,29,30 → **3**	93.1253
18	1,2,4,6,7,8,9,12,13,14,15,16,17,25,26,28,29,30 → **20**	93.1343
17	1,2,4,6,7,8,9,12,13,14,15,16,17,25,26,28,30 → **29**	93.0981
16	1,4,6,7,8,9,12,13,14,15,16,17,25,26,28,30 → **2**	93.0710
15	1,4,6,7,8,9,12,13,14,15,16,25,26,28,30 → **17**	93.0529
14	1,4,6,7,8,9,13,14,15,16,25,26,28,30 → **12**	93.0348
13	1,6,7,8,9,13,14,15,16,25,26,28,30 → **4**	93.0077
12	1,6,7,8,9,13,14,15,16,25,26,28 → **30**	92.9715
11	6,7,8,9,13,14,15,16,25,26,28 → **1**	92.9444
10	6,7,8,9,13,14,15,16,26,28 → **25**	92.8901
9	6,8,9,13,14,15,16,26,28 → **7**	92.8087
8	6,8,9,14,15,16,26,28 → **13**	92.9353
7	6,8,9,14,15,16,26 → **28**	92.7725
6	6,8,14,15,16,26 → **9**	92.5825
5	6,8,14,15,16 → **26**	92.4921
4	6,8,14,15 → **16**	92.3293
3	6,8,14 → **15**	91.7322
2	8,14 → **6**	91.2619
1	8 →**14**	88.8919

Table 4.10 shows the confusion matrix for BFS and NB classification. Several experiments were conducted, and the predicted number of TP attacks is 4,428, the

predicted number of FN attacks is 470, the predicted number of FP attacks is 283 and the predicted number of TN attacks is 5,874.

Table 4.10. Confusion matrix for BFS by NB classifier

		Classified	
		Attacked	Not attacked
Actual	Attacked	4428	470
	Not Attacked	283	5874

Also, the sensitivity and specificity are found.

$$\text{Sensitivity} = \frac{4428}{4428 + 470} = 90.40\ \%$$

$$\text{Spesitivity} = \frac{5874}{5874 + 283} = 95.40\ \%$$

4.1.4 Plus-*l* Take Away-*r* Feature Selection

The theory of Plus-*l* Take Away-*r* was explained in Chapter Three. For Plus-*l* Take Away-*r* (*l*=3, *r*=1), several experiments were conducted and the best experiment results were chosen. The best performance is found with an accuracy of 93.39% when 27 features are used. The second least number of features with the highest performance is obtained at an accuracy of 93.3424% when 18 features are used. As shown in the table, two arrow symbols are presented, the first is the right arrow (→), which shows "remove one feature in each iteration", and the second is the left arrow (←) which shows "add three features in each iteration". The detailed experiment results are listed in Table 4.11.

Table 4.11. The results of plus l take away r (l=3, r=1) feature selection algorithms with a Naive Bayes classifier

Num. of Features	Process	Plus-*l*-Takeaway-*r* Feature Selection	Accuracy
3	*l3*	8,14,6	91.7322
2	*r1*	8,14 → **6**	91.2619
5	*l3*	8,14 ← **6,15,16**	92.4921
4	*r1*	8,14,6,15 → **16**	92.3293
7	*l3*	8,14,6,15 ← **16,28,26**	92.7635
6	*r1*	8,14,6,15,16,28 → **26**	92.6097
9	*l3*	8,14,6,15,16,28 ← **26,9,24**	93.1796
8	*r1*	8,14,6,15,16,26,9,24 → **28**	92.9715
11	*l3*	8,14,6,15,16,26,9,24 ← **28,2,20**	93.1976
10	*r1*	8,14,6,15,16,26,9,24,28,2 → **20**	93.1705
13	*l3*	8,14,6,15,16,26,9,24,28,2 ← **20,27,21**	93.2429
12	*r1*	8,14,6,15,16,26,9,24,28,2,20,27 → **21**	93.2519
15	*l3*	8,14,6,15,16,26,9,24,28,2,20,27 ← **21,11,3**	93.2881
14	*r1*	8,14,6,15,16,26,9,24,28,2,20,27,21,11 → **3**	93.2700
17	*l3*	8,14,6,15,16,26,9,24,28,2,20,27,21,11 → **3,17,19**	93.3062
16	*r1*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19 → **11**	93.3062
19	*l3*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19 ← **10,5,11**	93.3152
18	***r1***	**8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,11 → 10**	93.3424
21	*l3*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,1 ← **22,10,12**	93.3152
20	*r1*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,11,22,10 → **12**	93.3152
23	*l3*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,11,22,10 ← **12,23,4**	93.2881
22	*r1*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,10,12,23,4 → **11**	93.3424
25	*l3*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,10,12,23,4 ← **30,18,1**	93.2610
24	*r1*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,10,12,23,4,30,18 → 1	93.2700
27	***l3***	**8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,10,12,23,4,30,18, ← 1,25,29**	**93.3905**
26	*r1*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,12,23,4,30,18,1,25,29 → **10**	93.3152
29	*l3*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,12,23,4,30,18,1,25,29, ← **10,11,13**	92.9806
28	*r1*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,12,23,4,30,18,1,25,29,10,11 →**13**	93.3876
30	*l3*	8,14,6,15,16,26,9,24,28,2,20,27,21,3,17,19,5,22,12,23,4,30,18,1,25,29,10,11, ← **13, 7**	92.9806

Table 4.12 shows the confusion matrix for the Plus-*l* Take Away-*r* (*l*=3, *r*=1) and NB classifier. Several experiments were conducted. The predicted number of TP attacks is 4,438 and the predicted number of FN attacks is 459. The predicted number of FP attacks is 271 and the predicted number of TN attacks is 5,887. Here, the correctly classified instances are 10,325 and the incorrectly classified instances are 730.

Table 4.12. Confusion matrix for Plus-l Take Away-r (l=3, r=1) by NB classifier

		Classified	
		Attacked	Not Attacked
Actual	Attacked	4438	459
	Not Attacked	271	5887

Also, the sensitivity and specificity are found.

$$\text{Sensitivity} = \frac{4438}{4438 + 459} = 90.62\ \%$$

$$\text{Spesitivity} = \frac{5887}{5887 + 271} = 95.59\ \%$$

4.1.5 Association Rules Feature Selection

The last and final algorithm is Association Rules (AR1), which provides the best performance at an accuracy of 93.28% when 26 features are used, as seen in Table 4.13.

Table 4.13. The result of Association Rules features selection algorithms with Naïve Bayes classifier

Num. of Features	**Association Rules**	**Accuracy**
26	1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18,19,22,24,25,26,27,28,29,30	93.28

Table 4.14 shows the confusion matrix for AR1 and NB classification. Several experiments were conducted. The predicted number of TP attacks is 4,448 and the predicted number of FN attacks is 464. The predicted number of FP attacks is 289 and the predicted number of TN attacks is 5854. Here, the correctly classified instances are 10,305, and the incorrectly classified instances are 753.

Table 4.14. Confusion matrix for AR1 by NB classifier

		Classified	
		Attacked	Not attacked
Actual	Attacked	4448	464
	Not Attacked	289	5854

Also, the sensitivity and specificity are found.

$$\text{Sensitivity} = \frac{4432}{4432 + 466} = 90.48\ \%$$

$$\text{Spesitivity} = \frac{5844}{5844 + 313} = 94.91\ \%$$

To summarize, different performances are obtained for each algorithm. For the individual, the best performance is 93.0077% using 27 features. For FFS, the best performance is 93.35% using 24 features. For backward selection, the best performance is 93.19% using 25 features. Finally, for Plus *l* Take Away *r* (*l*=3, *r*=1), the best performance is 93.39% using 27 features. The last algorithm is AR1, which provided a performance of 93.28% when 26 features are included. This means the ultimate best performance and classification are obtained by Plus *l* Take Away *r* (*l*=3, *r*=1).

To further illustrate the outcomes, a comparison between the classification rates, the algorithm performance as well as the accuracy for every feature is illustrated in Figure 4.2.

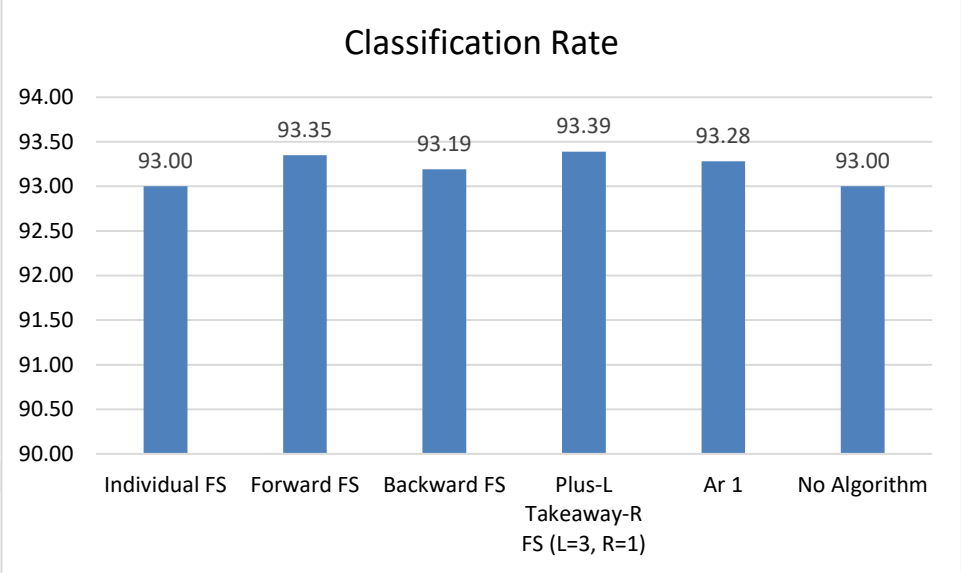

Figure 4.2. The results of classification rate for phishing website dataset using feature selection algorithms by Naive Bayes classifier

Table 4.15 below shows a number of features with performance accuracy and the name of the feature selection algorithm used by Naive Bayes classification.

Table 4.15. The results and comparison of feature selection algorithms with Naïve Bayes classifier rate

Feature Selection Algorithm	**Selected Features**	**Naïve Bayes Classification Rate %**	**Number of Features**
Individual FS	1,2,3,4,5,6,7,8, 9,10,12,13,14, 15,16, 17,18,19,21,22,23,24,26,27,28,30,29	93.00	27
Forward FS	2,3,4,5,6,8, 9, 10,11,12,14,15,16, 17, 18,19,20,21, 22,23,24,26, 27,28	93.35	24
Backward FS	1,2,3,4,5,6,7,8,9,10,12,13, 14,15,16,17,19,20,21,23,25,26,28,29,30	93.19	25
Plus-L Takeaway-R FS (L=3, R=1)	1,2,3,4,5,6,8,9,10,12,14,15,16,17,18,19, 20,21,22,23,24,25,26,27,28,29,30	93.39	27
AR-1	1,2,3,4,5,6,7,8,9,10,12,13,14,15,16,17,18, 19,22,24,25,26,27,28,29,30	93.28	26
No FS Algorithm	All features	93.00	30

4.2 Comparing other Classifier with FS Algorithms

The aim of this section is to compare various classifiers with reduced feature sets. For this issue WEKA software was used. WEKA is a well-designed software for a simple route and it has numerous data mining algorithms for classification. WEKA software is basically a Java application that is used for data mining.

During this study, many experiments were conducted. Table 4.16 shows various results for different feature selection algorithms. The classifiers namely BayesNet, SGD, Lazy. KStar, Randomizable-Filtered-Classifier (R. FClassifier), LMT and ID3 were used in the experiments. The classification performance was provided based on the accuracy as shown in Table 4.16.

The BayesNet algorithm provided the highest accuracy with 93.2519% for the Plus-l take away-r FS (l=3, r=1) features selection algorithm. The SGD algorithm provided the highest accuracy at 94.0027% for AR1. The Lazy.KStar algorithm produced the highest accuracy at 97.223% for AR1. The R. FClassifier algorithm found the highest accuracy at 94.6721% for AR. The LMT algorithm found the highest accuracy at 96.8883% for AR1. The last algorithm ID3 found the highest accuracy at 96.5084% for AR1.

Table 4.16. The results of feature selection algorithms with percentage of the classifier algorithms' accuracies

Feature Selection Algorithm	**Num of Features**	**Bayesnet**	**SGD**	**Lazy.Kstar**	**R.1F. Classifier**	**LMT**	**ID3**
Individual FS	13	92.6911	93.2338	94.0208	94.3284	94.8168	94.7716
forward FS	24	93.2067	93.4600	95.2420	94.0027	95.5767	95.1696
Backward FS	25	93.0167	93.8399	96.3908	94.6359	96.5355	96.1284
Plus-l Takeaway-*r* FS (*l*=3, *r*=1)	26	93.2519	93.4962	96.2189	94.4098	96.246	95.9294
AR 1	27	92.9353	94.0027	97.223	94.6721	96.8883	96.5084
Don't used	30	92.9896	93.9575	97.1958	94.2108	96.8792	96.4993

Table 4.17 shows the confusion matrix for Association Rules and Lazy.Kstar classifier. Several experiments were conducted, the predicted number of TP attacks is 4,676 and the predicted number of FN attacks is 222. The predicted number of FP attacks

is 85 and the predicted number of TN attacks is 6,072. Here, the correctly classified instances are 10,748 and the incorrectly classified instances are 307.

Table 4.17. Confusion matrix for Association rules by Lazy.Kstar classifier

		Classified	
		Attacked	Not Attacked
Actual	Attacked	4676	222
	Not Attacked	85	6072

Also, the sensitivity and specificity are found:

$$\text{Sensitivity} = \frac{4676}{4676 + 222} = 95.46\ \%$$

$$\text{Spesitivity} = \frac{6072}{6072 + 85} = 98.61\ \%$$

In the second part of the comparison, WEKA is used to find the worst algorithm for the used dataset analysis. During the second part, many experiments were conducted and the details are listed in Table 4.18.

Table 4.18. The results of feature selection algorithms with percentage of the worst classifier algorithms' accuracies

Feature Selection Algorithm	**Number Features**	**Multilayer Perceptron**	**JRip**	**PART**	**J48**	**Random Forest**	**RandomTree**
Individual FS	13	94.7987	93.268	94.6269	94.446	95.1515	94.7173
forward FS	24	95.0882	93.3786	95.1787	95.1606	95.6219	95.1967
Backward FS	25	96.038	94.3736	96.0742	95.8209	96.7707	96.0923
Plus-l takeaway-r FS (l=3, r=1)	26	95.8028	94.1384	95.64	95.6128	96.3546	95.4862
AR 1	27	96.6893	94.1384	95.64	95.8299	97.3134	96.3455
Don't used FS	30	96.9064	95.0158	96.7616	95.8752	97.3406	96.3727

Table 4.18 above shows the worst classification accuracies executed by WEKA application software. The first algorithm MultilayerPerceptron found the highest accuracy

at 96.6893% for AR1 when 27 features were included but at the same time if feature selection algorithms and features removal were not used, then the highest accuracy is found at 96.9064% of classification rate. The J48 algorithm obtained the highest accuracy at 95.8299% for Backward AR1 when 27 features were used; however, at the same time, if the feature selection algorithms and features removal were not used, then the highest accuracy is found at 95.8752% of classification rate. Similarly, for this algorithm, WEKA provided the same classification rate, which means it is the worst algorithm. The algorithm JRip found the highest accuracy at 94.3786% for Backward FS when 25 features were included but at the same time, if feature selection algorithms and features removal were not used, then the highest accuracy is found at 95.0158% of classification rate. The PART algorithm found the highest accuracy at 96.0742% for Backward FS when 25 features were included, but at the same time, if feature selection algorithms and features removal were not used, the highest accuracy is found at 96.7616% of classification rate. The RandomForest algorithm found the highest accuracy at 97.3134% for AR when 27 features were included but at the same time, if feature selection algorithms and features removal were not used, a better highest accuracy is found at 97.3406% of classification rate. The RandomTree algorithm found the highest accuracy at 96.3455% for AR when 27 features were included but at the same time, if feature selection algorithms and features removal were not used, a better highest accuracy is obtained at 97.3406% of classification rate.

5. CONCLUSION

In this study, the detection of phishing attacks was studied using feature selection and the Naive Bayes classification method. The phishing website dataset was taken from the UCI machine learning repository website. The phishing website dataset consists of thirty-one features. Some feature selection algorithms such as individual, forward, backward, pulse l – take away -r, and AR1 were used for the dataset. In the classification process, the Naive Bayes classifier was applied and in order to reduce phishing dataset features, feature selection algorithms were used. The datasets were separated into two parts. The first part was used for the training stage, and the second part was used for the testing stage. In the test stage, a 5-fold cross-validation method was applied to the dataset to ensure the correct performance accuracy. Several experiments were conducted during the classification process. The correct classification rate of the phishing website dataset has obtained a rate of 93.39% with 26 features; also, another good classification rate performance with fewer features was 93.35% with 24 features.

In the second part of this study, other classification methods were compared with Naive Bayes such as BayesNe, SGD, lazy.KStar, R.F.Classifier, LMT and ID3. These classifiers provided different classification performance rates. In addition, during the comparison, some algorithms that showed bad performance such as Multilayer Perceptron, AdaBoostM1, JRip, PART, J48, RandomForest and RandomTree were classified and tabulated, and the results are discussed in Chapter four.

Finally, the experiments showed that feature selection can be used to increase the accuracy rate and achieve faster classification.

5.1 Further Work

The study can and should be further developed using various datasets on different algorithms. Classification, accuracy and reducing the dataset can be organized to enhance the overall performance of the different algorithms and it should be presented as a new direction of research. Moreover, the feature selection algorithm can also be implemented by a different programming language to save time and compare classification accuracy. Finally, the programming model can be developed, performing with minimal user value, and can also be used to investigate several datasets such as breast cancer, skin cancer, bank marketing and car evaluation.

REFERENCES

[1] http://www.anti-abuse.org/phishing-general-information Phishing: General Information. Accessed on 8 Feb. 2016.

[2] **Liu, J. and Ye, Y.,** 2001. E-commerce agents: *marketplace solutions, security issues, and supply and demand* (Vol. 2033). On Apr 18.Springer Science & Business Media. Germany.

[3] **APWG, G. Aaron and R. Manning.,** 2015. APWG Phishing Reports, APWG, [Online]. Available: http://www.antiphishing.org/resources/apwg-reports/. [October– December 2015, Published March 22, 2016] (Accessed on 10 Feb. 2016)

[4] https://safety.yahoo.com/Security/PHISHING-SITE.html Yahoo Safety Center Security. How Can I Identify a Phishing Website or Email? Accessed on 22 Jul. 2016.

[5] **Sanglerdsinlapachai, N., Rungsawang, A.,** 2010. Using domain top-page similarity feature in machine Learning-Based Web Phishing Detection. *Third Int. Conf on 2010 Jan 9. Knowledge Discovery and Data Mining,* Phuket, pp. 187–190.

[6] **Sophie, G.P., Gustavo, G.G., Maryline, L.** 2011. Decisive heuristics to differentiate legitimate from phishing sites. *In Network and Information Systems Security (SAR-SSI). Conf on 2011 May 18, IEEE,* pp. 1–9.

[7] **Dhamija, R, Tygar, J. D. and Hearst, M.,** 2006. Why Phishing Works, in Proceedings of the SIGCHI conference on 2006 Apr 22. Human Factors in computing systems, Cosmopolitan Montréal, Canada.

[8] **Mohammad, R., Thabtah, Fadi Abdeljaber and McCluskey, T.L**. 2014. Predicting phishing websites based on self-structuring neural network. *Neural Computing and Applications,* 2014 Aug. 25 (2). pp.443-458. ISSN 0941-0643.

[9] **David, H., Heike, M. and Padhraic, S.,** 2001**,** Principles of Data Mining. MIT press, Cambridge, Massachusetts, USA.

[10] **Jiawei, H., Micheline, K and Jian, P.,** 2012, Data Mining: Concepts and Techniques. Elsevier, Amsterdam, Netherlands.

[11] **Han, J, and Kamber, M.**, 2000, Data Mining: Concepts and Techniques. Morgan Kaufmann. Elsevier, Amsterdam, Netherlands.

[12] **Markus, J, and Steven, M.**, 2007, Phishing and countermeasures: understanding the increasing problem of electronic identity theft. John Wiley & Sons, Inc. New York City, USA.

[13] http://www.antiphishing.org/word_phish.html Anti-Phishing Working Group. Origin of the word "phishing". Accessed on 13 Aug. 2016.

[14] http://www.wordspy.com/words/phishing.asp Phishing - word spy. Accessed on 17 Aug. 2016.

[15] http://www.phishtank.com/stats/2016/01/?y=2016&m=01 Phish tank. Phish tank stats-Jan 2016. Accessed on 17 Aug. 2016

[16] **Steve, S., Mandy, H., Ponnurangam, K., and Lorrie, C., and Julie, D., 2010.** Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI Conference on Apr 10 Human Factors in Computing Systems*. ACM, New York, USA.

[17] **Kohavi, R. and John, G.H.,** 1997.Wrappers for feature subset selection. Artificial Intelligence, 97(1-2):273–324. Elsevier Science Publishers Ltd. Essex, UK

[18] **Blum, A.L., and Rivest, R.L.,** 1992.Training a 3-node neural networks is NP-complete. Neural Networks, USA. 5:117 – 127.

[19] **Dash, M., and Liu, H.,** 1997. Feature selection for classification. Intelligent Data Analysis: *An International Journal*, 1(3):131–156.

[20] **Langley, P.,** 1994. Selection of relevant features in machine learning. *In Proceedings of the AAAI, Fall Symposium on Relevance*, Nov 4, vol. 184. PP 140–144.

[21] **Miller. A.,** 2002. Subset Selection in Regression second editions. Chapman & Hall/CRC, Boca Raton, Florida.

[22] **Chen, M.S., Han, J., and Yu, P. S.**, 1996. Data mining: An overview from a database perspective *IEEE Transactions on Knowledge and data Engineering*, 8(6), pp.866-883.

[23] **Jensen, D.,** 2000. Data snooping, dredging and fishing: The dark side of data mining a SIGKDD99 panel report. *ACM SIGKDD Explorations Newsletter*, 1(2), pp.52-54.

[24] **Goldschmidt, P.S., Alert-Km Pty Ltd.** 2006. *Compliance monitoring for anomaly detection.* U.S. Patent 6,983,266.

[25] **Smyth, P.**, 2001, Breaking out of the Black-Box: research challenges in data mining, Paper presented at the Sixth Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD-2001), Santa Barbara, California, USA.

[26] **SAS Institute Inc,** 2002, Lie detector software: SAS Text Miner (product announcement), *Information Age Magazine*, London, UK.

[27] **Berry, M J A., and Linoff, G S.,** 2004, Data mining techniques: for marketing, sales, and relationship management. John Wiley & Sons, Inc. New York City, USA.

[28] **Delmater, R., and Hancock, M.**, 2002, Data mining explained: a manager's guide to customer-centric business intelligence, Digital press, New York.

[29] **Fuchs, G.,** 2004, Data Mining: if only it really were about Beer and Diapers, *Information Management Online*. July, 1, pp.1006133-1

[30] **Langdell, S.**, 2010, Use of data mining in financial applications, (Data Analysis and Visualization Group at NAG Ltd). Accessed on 09 Aug. 2016. https://www.nag.co.uk/IndustryArticles/UnleashingDMpotential.pdf

[31] **Max, B.,** 2007, Principles of Data Mining, (Vol. 180). Springer, UK – London.

[32] **Guyon, I., and Elisseeff, A.,** 2003, An introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3(Mar), pp. 1157-1182.

[33] **Langley, P., and Sage, S.,** 1994, Induction of selective Bayesian classifiers. *In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, CA, USA. pp. 399-406.

[34] **Nilsson, R.,** 2007, Statistical Feature Selection, with Applications in Life Science. *PhD Thesis*. Linkoping University, Sweden.

[35] **Devijver, P.A., and Kittler, J.,** 1982, Pattern Recognition – A Statistical Approach. Prentice Hall, London, UK.

[36] **Kira, K., and Rendell, L. A.,** 1992, the feature selection problem - Traditional methods and a new algorithm. *In Proceedings of tenth national Conference on AI.* AAAI-92. AAAI Press. pp. 129- 134.

[37] **Liu H., and Setiono, R.,** 1996, a probabilistic approach to feature selection–a filter solution. *In Proceedings of the 13th International Conference on Machine Learning,* Vol. 96. pp. 319-327.

[38] **Almuallim, H., and Dietterich, T.G.,** 1997, learning with many irrelevant features. *In Proceedings of Ninth National Conference on AI.* Anaheim, California. Vol. 2. AAAI Press. pp. 547-552.

[39] **Hall, M.,** 2000, Correlation-based feature selection for machine learning. *PhD Thesis,* Department of Computer Science, Waikato University, New Zealand.

[40] **Kohavi, R.,** 1995, Wrappers for Performance Enhancement and Oblivious Decision Graphs. *PhD thesis,* Stanford University, Computer Science Department.

[41] **John, G.H., Kohavi, R. and Pfleger, P.,** 1994, Irrelevant features and the subset selection problem. In *Machine Learning: Proceedings of the Eleventh International Conference*. Morgan Kaufmann. pp. 121-129.

[42] **Cherkauer, K.J., and Shavlik, J. W.,** 1996, Growing simpler decision trees to facilitate knowledge discovery. *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. AAAI Press. Vol. 96, pp. 315-318.

[43] **Huan, Liu., and Hiroshi, M, da., Rudy S., and Zheng, Z.,** 2000, Feature Selection: An Ever Evolving Frontier in Data Mining. FSDM. 2010 Jun, 10. pp.4-13.

[44] **Krzysztof, J, C., Witold, P., Roman, W, S., and Lukasz, A, K.,** 2007. Data Mining a Knowledge Discovery Approach, Springer Science+Business Media, LLC. pp 224- 227.

[45] **Marill, T., and Green, D.,** 1963. On the effectiveness of receptors in recognition systems. *IEEE Transactions on Information Theory,* 9(1):11–17.

[46] **Hand, D.J.** 1981a. Discrimination and Classification. John Wiley, Chichester, New York.

[47] **Andrew, R, W., John, W., and Sons, L.,** 2002. Statistical Pattern Recognition., Second Edition. West Sussex PO19 8SQ, England. pp 314-316.

[48] **Kittler, J.,** 1986. Feature selection and extraction. In Young, T.Y., and Fu, K.S. (Eds.), *Handbook of Pattern Recognition and Image Processing*, Academic Press, 59–83

[49] **Karabatak, M., and Ince, M. C.,** 2009. An expert system for detection of breast cancer based on association rules and neural network. *Expert systems with Applications.* 36(2), 3465-3469

[50] http://www.phishing.org/phishing-techniques Phishing Techniques. Accessed on 25 Aug. 2016.

[51] **Markus, J., and Steven, M., 2007.** Phishing and countermeasures: understanding the increasing problem of electronic identity theft. John Wiley & Sons, USA.

[52] APWG, (2005) Phishing Activity Trends Report, Access date [03/07/2016]. http://www.antiphishing.org/download/document/245/APWG_Global_Phishing_Report_2H_2014.pdf

[53] **Ding, B., and Li, R.,** 2006. Phishing and Anti-phishing. *Master of Science Thesis,* Department of Computer and Systems Sciences, University/Royal Institute of Technology. Stockholm, Sweden.

[54] **James, L.,** 2006. Phishing Exposed, Tech Target Article sponsored by: Sunbelt software.

[55] **Chhabra, S.,** 2005. Fighting Spam, Phishing and Email Fraud, *Doctoral dissertation*, University of California Riverside, California, USA.

[56] **Jakobsson, M., and Young, A.** 2005. Distributed Phishing Attacks, *IACR Cryptology ePrint Archive*, 91.

[57] **Watson, D., Holz, T., and Mueller, S.** 2005. Know your enemy: Phishing, behind the scenes of Phishing attacks, *The Honeynet Project & Research Alliance*.

[58] **Jagatic, T.N., Johnson, N.A., Jakobsson, M. and Menczer, F.,** 2007. Social phishing. *Communications of the ACM.* 50(10). pp.94-100.

[59] **Jagatic, T., Johnson, N., Jakobsson, M., and Menczer, F.** 2005. Social Phishing, School of Informatics Indiana University, Bloomington, ACM

[60] **Hernandez, I, and Leggio, J.** 2006. Combating Phishing Websites, *Project Proposal*.

[61] **Zin, A., and Yunos, Z.** 2005. How to Make Online Banking Secure, article published in The Star InTech.21.

[62] **APWG.**, 2016. Anti-Phishing Best Practices for ISPs and Mailbox Providers, a document jointly produced by the MAAWG and APWG, All Reports Available: http://www.antiphishing.org/apwg-news-center/APWG-News/

[63] **Yiru X.** 67-year-old man swindled 700k from online banking account. http://finance.
sina.com.cn/money/bank/guangjiao/20110326/16149598471.sh%tml.

[64] **Jason. H.,** 2011. Why have there been so many security breaches recently? http://cacm.acm.org/blogs/blog-cacm/107800-why-have-there-been-so-many-security-breaches-recently/fulltext Accessed on 23 Aug. 2016.

[65] http://www.icbtollfree.com/article_free.cfm? articleId=5926 I dentity thieves take advantage of VoIP. Accessed on 29 Aug. 2016.

[66] https://archive.ics.uci.edu/ml/datasets/Phishing+Websites **UCI Machine Learning Repository:** Phishing Websites Data Set . Accessed on 8 Feb. 2016.

[67] **Mohammad, R., Thabtah, F., and McCluskey, L.** 2014. Intelligent Rule based Phishing Websites Classification. *IET Information Security*, 8 (3) 153-160.

CURRICULUM VITA

Twana SHWANY
twanacsd@gmail.com

Nationality:	Iraq
Place of birth:	Erbil
Date of birth:	02 / 01 / 1990
Marital status:	Single

EDUCATION

03/ 2015 – Present	MCE (Master of Software Engineering), Firat University, Elazig, Turkey.
2013 – 2015	SABIS University, in the College of Civil Engineering, Erbil, Iraq.
09/ 2008 – 07/ 2012	Bachelor of Computer Science, Salahaddin University, Iraq.
09/ 2005 – 05/ 2008	High School, Rizgary High School Erbil, Iraq.
09/ 2002 – 05/ 2005	Intermediate School, Wali Dewana intermediate School Erbil, Iraq
09/ 1995 – 05/ 2002	Primary School, Halkawt Primary School, Erbil, Iraq

WORK EXPERIENCES

1. Women's Union Institute for Social Affairs, Erbil, Iraq, IT Manager.
2. Garay Private High School, Erbil, Iraq, Computer Teacher.
3. Arean Steel Company Erbil, Iraq, IT Manager.
4. Al-Yasra Company, Erbil, Iraq, Data Analyst.

PUBLICATIONS

Published a Paper under the title of: "Feature Selection for Phishing Website by Using Naive Bayes Classifier" 1-3- December- 2016 in International Engineering, Science and Education Conference (INESEC) – Page No: 1360 / Diyarbakir-Turkey