EMPIRICAL VALIDATION AND ASSESSEMENT OF WEB METRICS

A Dissertation submitted in the partial fulfillment for the award of

MASTER OF TECHNOLOGY
IN
SOFTWARE ENGINEERING

By

RAMEEZ RAJA
(2K11/SWE/10)

Under the guidance of

Dr. RUCHIKA MALHOTRA
(Assistant Professor, Dept. of Software Engineering)

DEPARTMENT OF COMPUTER ENGINNERING
DELHI TECHNOLOGICAL UNIVERSITY
BAWANA ROAD, DELHI-110042
2013

DECLARATION

I hereby declare that the thesis entitled "**EMPIRICAL VALIDATION AND ASSESSEMENT OF WEB METRICS**" which is being submitted to the **Delhi Technological University**, in partial fulfillment of the requirements for the award of degree of **Master of Technology** in **Software Engineering** in the Department of Computer Engineering is an authentic work carried out by me. The material contained in this thesis has not been submitted to any university or institution for the award of any degree.

Rameez Raja
2k11/SWE/10
Email: rameez254@gmail.com

CERTIFICATE

DELHI TECHNOLOGICAL UNIVERSITY

(Govt. of National Capital Territory of Delhi)

BAWANA ROAD, DELHI-110042

Date: _____

This is to certify that the Major Project-2 Report entitled "**EMPIRICAL VALIDATION AND ASSESSEMENT OF WEB METRICS**" submitted by **RAMEEZ RAJA (Roll Number: 2K11/SWE/10),** in partial fulfillment of the requirements for the award of degree of Master of Technology in Software Engineering, is an authentic work carried out by her under my guidance. The content embodied in this thesis has not been submitted by her earlier to any institution or organization for any degree or diploma to the best of my knowledge and belief.

.

Guide:

Dr. Ruchika Malhotra
Assistant Professor
Department of Computer Engineering
DTU, Delhi

ACKNOWLEDGEMENT

I would like to express my gratitude to the many people who helped me in different ways with the development of this Major Project. Without their continuous support and guidance, the completion of my research leading to this Major Project would be impossible.

I wish to express my most sincere thanks to my supervisor **Dr. Ruchika Malhotra** for her invaluable advice, support and encouragement to proceed with my research at Delhi Technological University. I will carry out her guidance throughout my life.

I wish to convey my sincere gratitude to **Prof. Daya Gupta**, Head of Department, and all the faculties of Computer Engineering Department, Delhi Technological University who have enlightened me during my project.

Meanwhile, I would like to thank my colleagues at the Software Engineering Lab at Delhi Technological University for their support and feedback, and providing such a stimulating and friendly working atmosphere. It is fortunate for me to work and study with them.

This Project would not have been possible without the continuous support and guidance of my parents and all my friends who gave me constant support and strength to succeed.

RAMEEZ RAJA

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Websites have become an integral part of our day to day life. They act as a source of information and also as a medium of communication. With such important characteristic development of web sites should be done carefully. The quality of web sites is typically concerned with performance and usability and is measured using web metrics. Web metrics are the measure of attributes of a web page. Collecting, analyzing and interpreting web metrics is referred as web analytics.

In this study we categorize websites into three categories which are collected from pixel awards website and then analyze these categories using web metrics. For analysis we have created a web scrapper tool which evaluates web sites Using web page metrics and applied eight machine learning algorithm such as naïve bayes, bagging, random forest, random tree, multilayer perceptron, nnge and oner on these web page metrics to predict goodness of websites and also classified them within a particular category. The result of this paper will provide an empirical foundation for web site designing.

CHAPTER 1: INTRODUCTION

1.1 Introduction

Improving the effectiveness and quality of websites is a question of significant importance for the success of e-commerce, e-health, entertainment and other aspects of life that use the Internet technology as an effective medium to reach the target audience. Various website evaluation tools are now available to assess the effectiveness and quality of the websites associated to these industries. Normally most of these tools are designed for use by website users to provide their subjective opinions on the value and usefulness of the information for which they are targeted. Subsequently, user's feedback and comments form a basis of website improvement where changes are deemed necessary.

Specifically, there exists no established mechanism or relevant studies on assessing the quality and the effectiveness of online resources. However, few websites like webby awards, pixel awards etc. have made an attempt to provide a good rated websites which can be the basis of evaluating a website from quality perceptive. These websites provide the list of top rated websites of different categories like Entertainment, music, shopping websites etc. which are close to perfect in terms of their quality and popularity. Quality of any web site is represented in terms of how easily an individual navigate through web site i.e. user –friendliness, amount of information present on one page, visibility, web traffic. Website popularity is defined in terms of the number of hits on a web page generated by a website visitor. Web designing is the field of designing the web site. Web designing has been moved from the modest beginning of a text page to the high end of animated designs on the web pages. Web site designs have reached a great level of importance with the growing demand of internet marketing.

Information available on the internet is growing with tremendous amount of rate and usage of World Wide Web is also increasing, we need quality websites. Today we do not require any expert to create a web site, due to these web sites which are created are poor in quality and

making difficult for an individual to get desired outputs. Also there in no general web site design guidelines on which there is global agreement of web developers.

Metrics are important measures in analyzing any web site. Since, 1990s different set of metrics has been proposed for evaluating website quality attributes such as reliability, security, usability and maintainability. Quality assessment of websites made by experts are not cost-effective and cannot be fair and also regular quality assessment of websites have to carried out which cannot be possible manually. Therefore in this work we automate the website quality evaluation process.

There is no set of evaluation guidelines on which the answers should base, or a tool that would help users to best judge the quality of the relevant website. Furthermore, the approaches do not encourage users to provide feedback that would help website owners to subsequently improve the effectiveness of the websites.

When developing website assessment tools, the technical and content aspects of the website should be taken into consideration. While content aspect refers to those features related to the content of a website (e.g. accuracy, objectivity and relevancy), the technical aspect refers to those features related to the design and usability of a website (e.g. navigation, interactivity and accessibility). The two aspects make up website quality dimensions or simply website metrics. In this study, we define website metrics as sets of indicators to take into account when assessing the perceived quality of the website material.

1.2 Motivation of the Work

Web is becoming more and more important each day for conducting business, sharing information and for communication. Every passing day the number of companies, organizations and individuals publishing their web sites is increasing [25]. Considering all the information available on the web every individual should desire to find and access useful information. For example companies want to learn what their competitors do and what products they offer using the web. By the help of this information companies may learn from their rivals and improve their own web sites to increase their competitiveness. Considering their web sites, companies or

institutions have various methods to attract customers use their web sites to purchase goods or services or make customers take advantage of their web sites. The task of evaluating and improving the web sites can be intimidating considering the number of web sites available and the frequency of updates. As a result, automated support for web designers and web site owners become more important. Automated usability tools can help save time and money in design. It can improve consistency and quality of the web site [26].

Surprisingly, no studies have derived web design guidelines directly from web sites that have been assessed by human judges. In a recent paper from Ivory [15], it has been reported that the results of empirical analyses of the page-level elements on a large collection of expert reviewed web sites. These metrics concern page composition (e.g., word count, link count, graphic count), page formatting (e.g., emphasized text, text positioning, and text clusters), and overall page characteristics (e.g., page size and download speed). The results of this analysis allows one to predict with 65% accuracy if a web page will be assigned a very high or a very low rating by human judges. Even more interestingly, if we constrain predictions to be among pages within categories such as education, community, living, and finance, the prediction accuracy increases to 80% on average. This forms the basis of the work in this thesis.

1.3 Aim of the Work

It is not feasible to specify a checklist to evaluate a web site with constant controls which will ignore even daily changing web site contents. To provide a dynamic evaluation process, evaluation shall be separated into parts those will monitor and evaluate ordinarily stable and frequently changing characteristics. **To provide a solution to the problem this work presents an approach to evaluate the quality of web sites with the help of different web metrics which can assess the quality of website.** The main objective of this work is to implement a web site evaluation tool covering most important aspects of web site evaluation criteria in terms of quality of Website.

1.4 Organization of Thesis

The remainder part of thesis is organized in the following chapters:

Chapter 2 highlights the literature survey that has been done in the field of evaluation of a website. Chapter 3 describes the research background of the work done in detail, i.e. it gives the brief introduction of the various quantitative web interface measures and the independent and dependent variables selected for our study. In Chapter 4 the empirical datasets and their characteristics are discussed with the brief introduction of tool developed for the metrics estimation process. It also presents the various machine learning algorithms that are used for data analysis and validation of the results. In Chapter 5 we evaluate and judge the performance of our results. This section discusses the comparative analysis of results by applying various machine learning algorithms onto the collected data sets. Chapter 6 presents conclusion drawn from the research work. It also incorporates the scope of future integration.

CHAPTER 2: RELATED WORK

There are many criteria to evaluate a web site. Those may include: usability, authority, currency, objectivity, coverage, performance, traffic ranking, link popularity, accessibility, security, design patterns, HTML syntax analysis, and browser compatibility. Output data of traffic based and time-based analysis must be interpreted in order to identify usability problems. Server logs are problematic because they only track unique navigational events (e.g. do not capture use of back button).

In HTML syntax analysis it inspect the static HTML for pre-determined guidelines, such as number of words in link, links which are image links, all images contain an ALT attribute[5].These guidelines may cover universally accepted guidelines or guidelines accepted in a specific society. A list of Automatic Evaluation tools which depends on the characteristics of HTML.

1. Web XM

WebXM is used to automate inspection of some page defects. These defects include broken links, spelling errors, slow loading pages, poor search and navigation to help improve usability of the web site. WebXM automates more than170 accessibility checks, namely appropriate text and background color contrast or the presence of text equivalent alt tags on images. These accessibility checks ensure the accessibility of a web site for disabled people. The target of WebXM is to improve visitor experience. This target is obtained by exposing usability issues that may be causing visitors keep themselves away [29].

2. Booby

Bobby is a web accessibility testing tool. It is designed to help remove barriers on accessibility issues. It also encourages compliance with existing accessibility guidelines, including Section 508 of the US Rehabilitation Act and the W3C's Web Content Accessibility Guidelines (WCAG) [28]. Bobby examines every page of a website and tests every page of web site individually.

Then it checks the web site for several accessibility requirements. These requirements cover readability by screen readers, the provision of text equivalents for all images, animated elements, audio and video displays. It performs over 90 accessibility checks. In an evaluation session it examines HTML for compliance with specific guidelines for generating report for each page of the web site. A syntactic analysis is applied for HTML code. There are three priority levels according to WAI. These levels base on definition of guidelines, checkpoints and priorities. For each guideline appliance of content development scenarios are explained in checkpoint definitions [27]. Each checkpoint has one of the three priority levels according to the effect on accessibility issues. Eventually, the three conformance levels base on the satisfaction of all checkpoints of an increasing number of priority levels.

3. NIST Web Metrics

The US National Institute of Standards and Technology (NIST) have developed prototype tools. These tools aim to evaluate web site usability [19]. There tools are WebSAT, WebCAT.

3.1 WebSAT

The Web Static Analyzer Tool (WebSAT) is a prototype tool that inspects the HTML code of web pages for usability problems. WebSAT allows the webmaster to investigate these problems. Then webmaster can remove these problems from the web page design. WebSAT not only applies its own set of usability rules but also applies the IEEE Std.2000-1999 (NIST 2001b). [18] Likewise Bobby, accessibility is measured in accordance with the three priority levels suggested by WAI recommendations [28].

3.2 WebCAT

The Web Category Analysis Tool (WebCAT) allows webmaster to conduct a simple category analysis in the web quickly [12]. This is based on traditional card sorting techniques. The webmaster creates a set of categories and a number of items which are to be assigned by test subjects to the categories. Then the Webmaster can compare the real assignments with intended assignments which will meet user needs.

Limitations of existing Web Evaluation tools

1. A number of existing website evaluation methods generally requires the evaluator who has IT background to assess the qualities in a website. It is difficult to apply if the people do not have any IT skills.
2. Many new website software technologies and rules are not considered in existing website quality evaluation methods. The web developer is confused by the overall picture of the evaluation criteria. A new website evaluation methods need to involve the all identified new software technologies as the numbers of new criteria.
3. The specific quality criteria for a website's reputation are clarified in many existing website evaluation methods, however most creditable criteria are immeasurable.
4. The strengths and weaknesses of the web evaluation results should be applied to the user's expectations, and ease of understanding.

The most closely related work is done in Ivory et.al [15] [16] which provides preliminary analysis of collection of web pages and captures various web metrics associated with the rated websites, and predicts how the pair-wise correlations are manifested in the layout of the rated and unrated sites pages. This work does not apply various machine learning algorithms to predict the best suited model that can provide high accuracy.

According to K.M.Khan [6] quality of website can be defined in terms of functional and non-functional attributes. To derive quality metrics K.M.Khan adopts Goal-Question-Metric approach. According to GQM initially all the goals are defined that are to be measured, then for each goal, questions are derived that are required to determine if the goals are fulfilled, and finally the answers of these questions are known as metrics.

M. Zorman et.al [17] has proposed an algorithm to find the good or relevant websites for keywords provided by the user. The algorithm works on term frequency in a website using TFIDF heuristics search tool and evaluate website using decision tree machine learning algorithm.

CHAPTER 3: RESEARCH BACKGROUND

Our research moves in the way to identify the various characteristics of a web page i.e. web metrics and with the help of these metrics predict the class (TV and Movies websites belong to Entertainment Category) and goodness of website within a category.

ISO 9126 defines external quality, internal quality and quality in use. As we are going to use an automatic procedure, we are only concerned with the external quality. So our quality model will encompass the six well-known ISO9126 quality characteristics [3]: *Functionality, Reliability, Usability, Efficiency, Maintainability* and *Portability*.

Since we are concerned with the automatic collection of web sites quality metrics, the targets of that collection include the online artifacts that result from the server programming, which largely include html, style sheets and scripts. The server programming in C#, php, Perl, Java are out of our reach and so are the type of server used, the database used, and the hardware used. However, we can collect some server signatures automatically.

In this chapter, first we introduce the web site structure in section 3.1, and then we will see different web metrics proposed by different researches in section 3.2 and the metrics selected for study in section 3.2 and in section 3.4 machine learning algorithms for the analysis of metrics.

3.1 A web-site structure

A website is a collection of web pages, images, videos and other digital assets that are hosted on a Web server, usually accessible via the Internet or a LAN .It is a document, typically written in HTML or XHTML format, and may provide navigation to other web pages via hypertext links, that is almost always accessible via HTTP, a protocol that transfers information from the Web server to display in the user's Web browser. All Publicly accessible websites are seen collectively as constituting the "World Wide Web". Web pages may consist of files of static text stored within the web server's file system (static web pages), or the web server may construct the (X) HTML for each web page when it is requested by a browser (dynamic Web pages).

Web site structure is just like the blue print of the building. It should not be complicated nor need to be very fancy. The website should be organized in such a manner that visitors easily find what they want. The easier it's to use, the longer the users will stay with the website, and more they will see of it. Good web sites structure can easily grow logically. It will be very easy to add new contents without changing the graphical design of the web site that is build for a given customer.

3.2 Web Page Metrics

Metrics, as we know, refer to standards of measurement. Therefore, web metrics are standardized ways of measuring something that relates to the Web. Web metrics helps organizations to understand, manage and improve their web systems and hence enhance the quality of their online presence. Depending on the ISO 9126 quality model different metrics are defined in the literature.

3.2.1 Efficiency based web metrics

Efficiency metrics include related to size of a web page and the load time of a website/webpage [9, 20]. In Table 3.1, we summarize the website efficiency metrics.

Metric	**Meaning**
efficiency_css_size	Css size per page
efficiency_homepage_load_time	Homepage load time
Efficiency_image_size	Image size
efficiency_javascript	Script size per page
efficiency_page_load_time	page load time
efficiency_page_size	Page size

Table 3.1 Efficiency based web metrics list

3.2.2 Functionality based web metrics

It includes navigation, forms, identity and other aspects related to the functionality offered by the site. In Table 3.2, we summarize the website functionality metrics.

Metric	Meaning
forms_form_info_request [10, 23]	presence of contacts/info form
forms_labels[23]	number of label tags
Identity_auther[9]	Average presence of author
Identity_logo[9]	presence of site name in title
Identity_sitename_title[23]	Presence of navigation bar
Navigation_bar[20]	Presence of navigation bar
Navigation_bread_crums[20]	Presence of bread_crums(path metric)
Navigation_quality_of_links[9]	Presence of page title in links

Table 3.2 Functionality based web metrics list

3.2.2 Maintainability based web metrics

These metrics includes aspects related to the number of items to maintain (e.g. scripts, styles used and tables). In Table 3.2, we summarize the website maintainability metrics.

Metric	Meaning
Maintenance_num_script[23]	Script files no per page
Maintenance_num_styles[23]	Css file number per page
Maintenance_num_tables[1]	Tables number per page

Table 3.2 Maintainability based web metrics list

3.2.4 Portability based web metrics

These metrics includes aspects related to page layout, use of html standards, etc. In Table 3.4, we summarize the website portability metrics.

Metric	Meaning
Page_layout_device_specific[20]	Presence of specific css to device
Page_layout_html_standards[23]	Use of html notations in formatting
Pagelayout_num_divs[23]	Number of divs
Page_layout_num_frames[23]	Number of frames

Pagelayout_num-tables[23]	Number of tables
Pagelayout_num_table_inside_tables [1]	Presence of table inside table

Table 3.4 Portability based web metrics list

3.2.5 Reliability based web metrics

It includes aspects related to the validation and links status. In Table 3.5, we summarize the website reliability metrics.

Metric	**Meaning**
Links_avg_num_words[1]	Average number of words in links
Links_links_titles[1]	Links with title attributes
Links_num_broken_links[10, 20]	Number of broken links
Link_num_extern_links[20]	Number of broken link to another site
Link_num_image_link[1]	Number of link with images
Link_num_intern_broken_link[20]	Number of broken links in the same site
Link_num_intern_links[20]	Number of intern links
Links_num_links[10, 20]	Number of links
Links_num_non_implemented_links[10]	Number of non-implemented links
Link_page_withot_link[10, 20]	Pages without links in the site
Links_num_non_implemented_links[1]	Number of non-implemented links
Validation errors[20]	Html warning par page

Table 3.5 Reliability based web metrics list

3.2.6 Usability based web metrics

It includes aspects related to accessibility, multimedia and textual contents. In Table 3.6, we summarize the website usability metrics.

Metric	**Meaning**
Accessibility_img_alt [20]	presence of alt attribute in images
accessibility_img_title [1]	presence of title attribute in images

accessibility_validate_access [20, 9, 23]	accessibility issues per page
multimedia_num_img [20]	image number per page
text_font_size_average_em [23]	average of font size in em (percentage) in css
text_font_size_average_px [23]	average font size in css in pixels
text_font_size_max_em [23]	maximum font size in em (percentage) in css
text_font_size_max_px [23]	max font size in pixels
text_font_size_min_em [23]	minimum fonts size in em (percentage) in css
text_font_size_min_px [23]	min font size in pixels
text_heading_len [20]	average heading length
text_heading_reverse_order [23]	number of headings in reverse order
text_italic_text [23]	number of italic text bigger than 20 chars
text_num_diferent_colors [23]	number of different text colors in css
text_num_diferent_fonts [20]	number of different text fonts in css
text_num_sentences_in_paragraph [20]	number of sentences per paragraph
text_num_subheading_heading [20]	number of sub headings per heading
text_num_syllables_in_word [20]	number of syllables per word
text_num_words_in_sentence [20]	number of words per sentence
text_num_words_meta_description[20]	number of words in metatag description
text_num_words_meta_keywords[20]	number of words in metatag keywords
text_paragraph_max_size [20]	maximum size of paragraph
text_paragraph_size [20]	paragraph size
text_subheading_len [20]	sun heading length
text_total_newlines [20]	total number of newlines
text_total_sentences [20]	total sentences
text_total_syllables [20]	total syllables
text_total_words [20]	total words
text_uppercase_text [23]	number of uppercase sentences

Table 3.6 Usability based web metrics list

3.3 Independent and Dependent Variables

The dataset comprises of 21 measures to be used for capturing various information related to web pages, one dependent and twenty one independent variables. These variables cover those attributes that can be computed automatically. Out of all above mentioned metrics, Section 3.4 describes the 21 metrics that we have selected as variables for our study. We developed a Web Scrapper tool developed in PYTHON technology to compute these metrics which has been explained in later chapter. We have used attribute selection technique for reducing data dimensionality provided in WEKA tool [32]. Table 3.4 summarizes the name and category of web metrics used in study.

Metric	Category
Meta tag [23]	Usability
Meta keywords [23]	Usability
Min keyword length [23]	Usability
Max keyword length [23]	Usability
Meta descriptor [23]	Usability
Total link [10, 20]	Reliability
Image link [1]	Reliability
Average number of words in a link [23]	Usability
Total images [23]	Usability
Alt images [20]	Usability
Words in alt images [20]	Usability
Division tag [23]	Portability

Paragraph [23]	Usability
Scripts [23]	Usability
Page size [9]	Efficiency
Body word count [23]	Usability
Title length [9]	Functionality
Tables[23]	Portability
Load time [9]	Efficiency
Total headings [23]	Usability
Link headings[23]	Reliability

Table 3.7 Web Metrics

The description of the attributes used in this study is given below:

1. *Meta tag*

Total number of Meta tag on a page. This attribute is calculated by counting total number of Meta tag on a page. The text in meta tags is not displayed by browser.

2. *Meta Keywords*

total number of meta keywords on a page. This attribute is calculated by counting total number of words in meta tag where name attribute is keywords. Meta Keywords are separated with comma (,) therefore words between two commas is considered as one keyword.

3. *Minimum Meta Keyword length*

Attribute is calculated by identifying a meta keyword from Meta Keywords which contain minimum number of characters. No spaces, commas, new line and tab are considered while counting characters in a keyword.

4. *Maximum Meta Keyword length*

Attribute is calculated by identifying a meta keyword from Meta Keywords which contain maximum number of characters No spaces, commas, new line and tab are considered while counting characters in a keyword.

5. *Meta Descriptor words*

Total number of words in meta tag where name attribute is descriptors. No spaces, commas,|,",\n,\t are considered while counting descriptor words.

6. *Total links*

Total number of links in a web page. The links which are in comments are not counted.

7. *Image links*

Total number of links which are images in a web page i.e. clicking on image leads to a new page.

8. *Average number of words in a link*

Attribute is calculated by dividing sum of words in all text links by total text links.

9. *Total images*

Attribute is calculated by counting number of images in a page.

10. *ALT Images*

Total number of images which contain an alt attribute. When a particular image is not loaded by a browser then the text present in alt attribute is displayed in place of the image.

11. *Words in alt images*

Attribute is calculated by summing all words present in all alt images.

12. *Division tag*

Total number of divisions of a web page i.e. in how many section a web page is divided.

13. *Paragraph*

This attribute is calculated by counting total paragraph in a web page.

14. *Scripts*

This attribute is calculated by counting total number of java scripts used in a web page.

15. *Size*

Attribute refers to number of bytes requires to store a web page on a system. We assume that only one server request is send at one time.

16. *Body Word Count*

this attribute is calculated by counting total words present in a body tag of a web page. We discard all words which are in comments and script tags.

17. *Title Length*

Total number of words present in web page title. Special characters are also considered.

18. *Tables*

Total number of tables in a web page. Tables are also used sometime for division of a web page.

19. *Load Time*

Attribute refers to time taken by a web page to load i.e. difference between the first request and first response time.

20. *Total Headings*

Total number of headings in a web page. We consider all six types of headings. Headings font size is bigger than rest of the word i.e. they are more visible than other word.

21. *Link Headings*

Total number of headings which are link i.e. on clicking that heading we move to a new page.

3.4 Empirical Data Collection

We analyze the web pages collected from pixel awards website. The pixel awards web site was established by Erick & Laubach in 2006.The Pixel Awards judges are proven innovators in their fields with broad web expertise and a knack for spotting extraordinary talent with fairness and accuracy [22]. Each website is evaluated on the basis of innovation, content, navigation, visual design, functionality and overall site experience.

The Pixel Awards judges are proven innovators in their respective fields with broad web expertise and a knack for spotting extraordinary talent with fairness and accuracy as described in Pixel Awards [22]. The websites placed in 24 categories are judged on the basis of creative and technical blend of impeccable graphic design, artistry, technological expertise, and a powerful, stimulating user experience [22]. These sites are the best of the web, thus each site for its respective category is evaluated for innovation, content, navigation, visual design, functionality and overall site experience.

For over study, we collected data from 7 categories of pixel awards for each year from 2006 to 2012. The categories we selected are TV, Movies, Blogs, Community, Food & Beverage, Travel and Commerce. We have merged the categories which have some properties in common and created three models.

MODEL 1: In this model we have collected data from Blogs and Community websites. Blogs also sometime referred as community and people often build a community over a blog. Data set is created from 119 web pages and some level-1 pages are also included.

MODEL 2: In this model we have collected data from TV and Movies websites as both have similar structure and purpose. Data set is created from 51 web pages

MODEL 3: In this model we have collected data for Food & Beverage, Travel and Commerce websites. All three behave as e-commerce websites. Data set is created from 129 web pages and some level-1 pages are also included.

The web pages used to predict good- bad class for each model is tabulated in Table 3.8.

Model	Good web pages	Bad web pages	Total web pages
Model 1	35	84	119
Model 2	12	39	51
Model 3	28	101	129

Table 3.8 Data

The web pages used to predict the class of each website within the particular model is tabulated in Table 3.9.

Model	Websites	Good web pages	Total web pages
Model 1	Blog	62	119
	Community	57	
Model 2	TV	25	51
	Movies	26	
Model 3	Food & Beverages	39	129
	Travel	41	
	Commerce	49	

Table 3.9 Classification of Data

CHAPTER 4: REASERCH METHODOLOGY

4.1 Methodology

We employ quantitative web-page attributes in our methodology to classify websites belongs to same domain (TV, movies website belong to entertainment domain) and in order to these we construct a model. Figure 4.1 shows the flowchart of methodology.

The flowchart of methodology is divided into three sections.

a. **Empirical Data Collection**: - In this section we first select the website from different nominated categories in 6 years (2006-2012) from Pixel awards for which metrics estimations are to be calculated. Secondly we download the source code of the website and then apply a web scrapper to calculate the different metrics for these websites.

b. **Web Scrapper**: - web scrapper will automate the process of web metric extraction from a web page. Firstly it will preprocess the source files to remove all unwanted things and then metrics are calculated.

c. **Result Analysis**: - In this section we use different machine learning techniques to analyze and compare data. Comparison shows which algorithm gives better results compared to others.

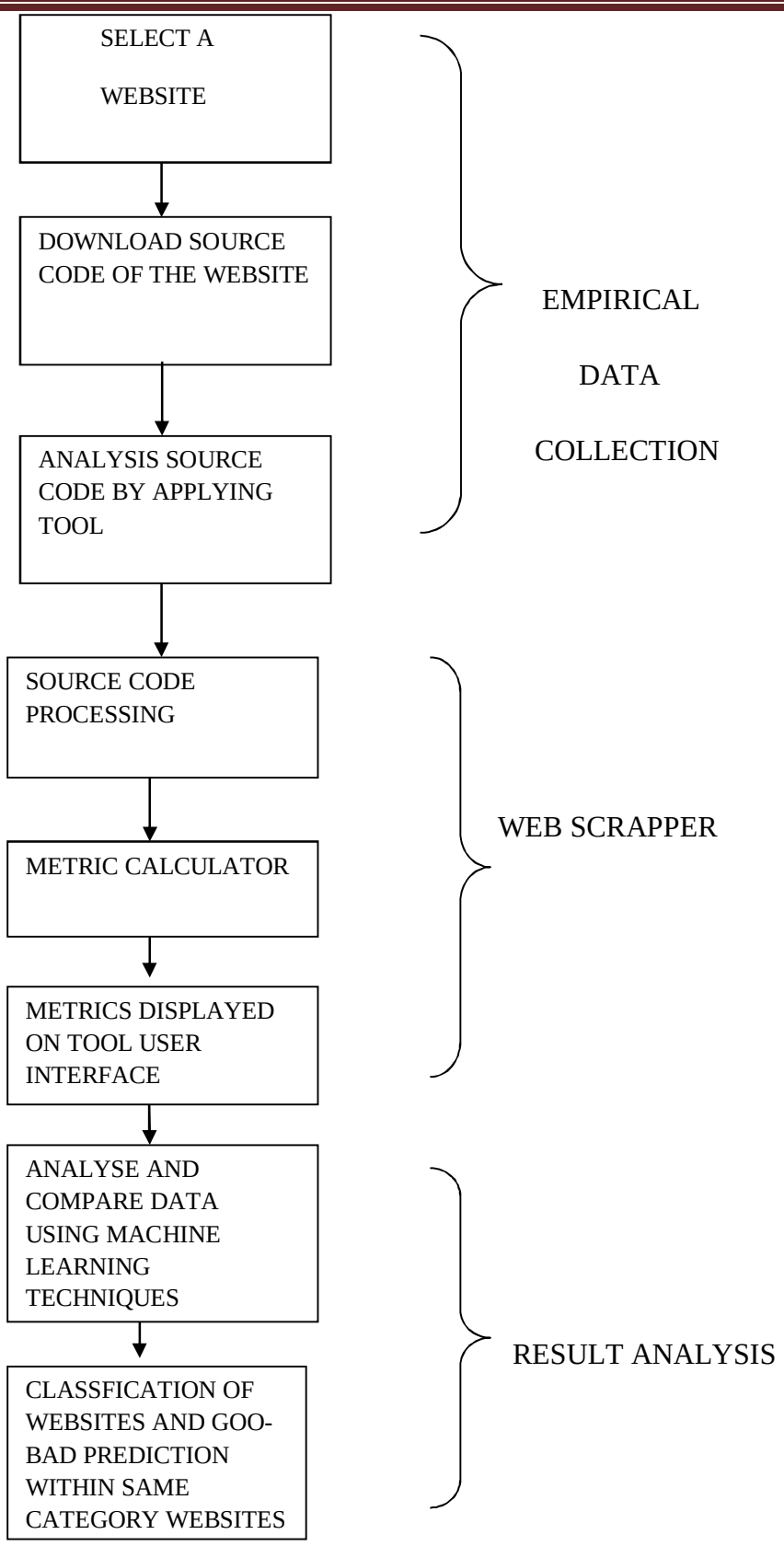

Figure 4.1 Flowchart of Methodology

4.2 Description of Tool

We have developed a web scrapper in python language that will automate the process of web metric extraction from a web page.

Web scraping (web harvesting or web data extraction) [30] is a computer software technique of extracting information from websites. Usually, such software programs simulate human exploration of the World Wide Web by either implementing low-level Hypertext Transfer Protocol (HTTP), or embedding a fully-fledged web browser, such as Internet Explorer or Mozilla Firefox [8prev].

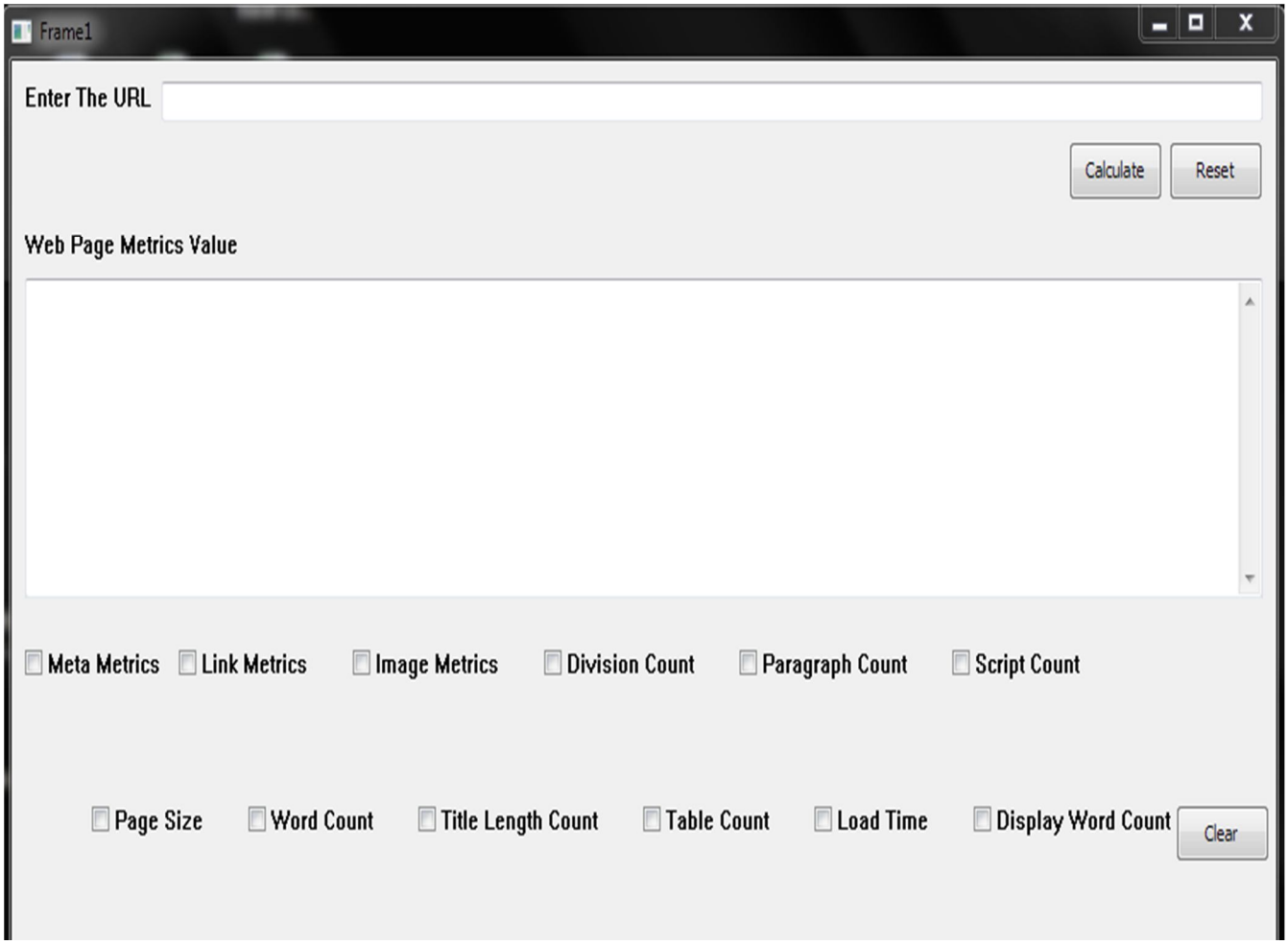

Figure 4.2 GUI of Tool

4.2.1 Algorithm of web scrapper

INPUT - URL of a Web Page

OUTPUT-Web page Metrics (only those whose checkbox are on)

STEPS:-

1. Enter the URL of the web page. The format of the URL
   http:// www.example.com
2. Store the Source Code of the page in a Variable called "soup" using method urllib2 and bs4(Beautiful Soup)
3. Variable "soup" is passed as a parameter to different methods in order to calculate the web page metrics.
4. The Page Metrics which we calculate are
   a. ***Metrics related to meta tag:***
      - Number of meta tag:- Calculated by counting all the meta tag in a page.
      - Meta keywords:- Calculate all the keywords which are present in a meta tag attribute keywords. Keywords are separated with comma (,) therefore words between two commas is considered as one keyword.
      - Maximum and Minimum length of a meta keyword:- The length of each meta keyword is find out . Length here represents number of words in a keyword. Out of these length minimum and maximum length are find out.
      - Meta Descriptors:- Calculate number of words in a meta tag attribute descriptor. While calculating number of words we also consider special symbols like '|', '"', ',' , ',' etc but neglect the spaces between the words.

   b. ***Metrics related to Links:***
      - Total links in a page:- Search all anchor tag in a page as we know anchor tag are used to
      - Image links :-
        If between <a> and an <img> tag is present
        Then it an image link

Else not an image link

- Average number of words in a link:- for all links which are not image link there is string between <a> and .

  Link name = sum of words between <a> and

  Average number of words in a link = sum of Link name of each non image link/total links.

c. Metrics related to Images:

- Total images: - total <img> tag which are present in a web page.
- Alt images: - total images which has an alt attribute.
- Words in alt images:- Sum of the words in a string of an alt attribute for all Alt images.

d. Number of division of a web page:

- Search all tag in a web page. This searching is done using the help of Beautiful Soup.

e. Total paragraph in a web page:

- Search all<p> tag in a web page.

f. Number of Scripts in a web page:

- Search all <script> tag with attribute type = “text/javascript”

g. Size of a web page

- Calculating web page size we retrieve the entire resource and calculate its length .This is done using the urlopen method of urrllib2 module.

h. Body Word Count:

- Word count represents total number of words on a web page when we load it. We calculate word count by converting an html page into a text page.
- HTML to TEXT conversion

- Remove all comments. In html comments are thing which are between <!- and -- >. Eg. <!—this is a comment -- >
- Remove all <script> tag as they contain definitions of function which are not displayed by browsers.
- The words which are in alt attribute of an image are not included in word count. Title length not included in word count.

i. Title length:

- Number of words which are present <title> ……</title>
  - Commas are not included in title length.
  - Semi-colons are not included in title length.
  - Newline and Tab are not counted in title length.

j. Load Time:

- Number of time took by urlopen module to read the web page. Time module of python is also used to calculate load.

Start time = when reading of a page started

End time = when reading of a page ended.

Load time = End time – Start time (in sec)

k. Metrics related to display word count:

- Total headings in a web page:- There are 6 types of heading in html h1,h2,h3,h4,h5,h,6

Total headings = total <h1>tag + total<h2>tag + total<h3>tag

total<h4>tag + total<h5>tag + total<h6>tag.

- Link headings: - Heading which are also links.

IF <a> between <h1> and </h1>

Then it is link heading.

ELSE

Not a link heading.

5. Web metrics which are calculated are displayed on tool user interface. This estimation can be saved for the further references.

4.3 Machine Learning Algorithms for Data Analysis

4.3.1 Naive Bayes Classifier

Naive bayes classifier is a supervised machine learning algorithm (needs to be trained) based on the Bayesian theorem. Naive bayes is also called idiot bayes and it assumes that all features are conditionally independent given the class label [7].

To demonstrate the concept of naive bayes classifier, consider an example

Let there be a set of variables X ={x1, x2, x3……xn} and variables are classified as C1 and C2. Initially n1 belong to C1 class and n2 belong to C2 class. Consider a new variable z which is to be classified

According to Bayes theorem

$$\text{Posterior} = \frac{\text{Prior} * \text{likelihood}}{\text{Evidence}}$$

Therefore

Posterior probability of z being C1 = prior probability of C1 * likelihood of z in C1

Prior probability of C1 = n1/X

$$\text{Likelihood of z in C1} = \frac{\text{number of C1 in vicinity of z}}{\text{n1}}$$

Similarly cases for C2 class and the z belong to class which has higher posterior probability.

Advantages

It is quite accurate and very fast.

Out performs more sophisticated classifiers on many datasets, achieving impressive results [2].

4.3.2 Bagging

Bagging [bootstrap aggregating] was proposed by Leo Breiman [8] in 1966 to reduce the variance of predictor. It improves the classification by combining classifications of randomly generated training sets.

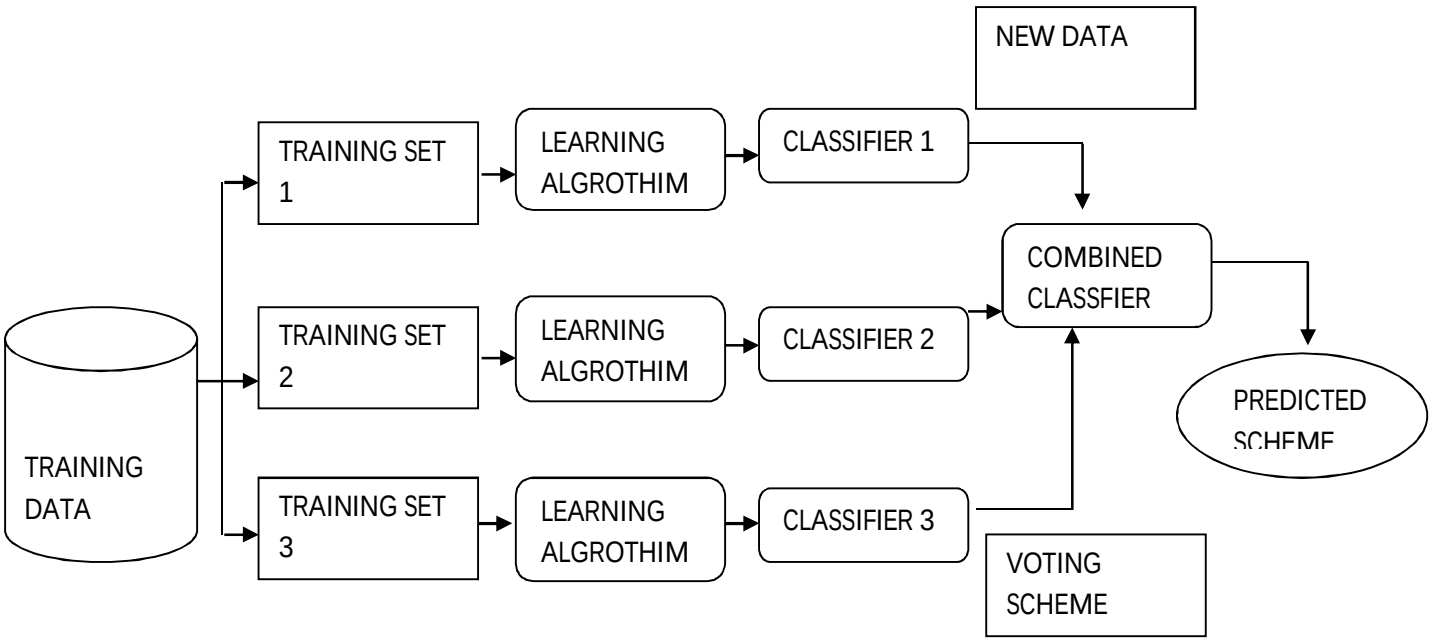

Figure 4.3 Flowchart of bagging classifier

Classification: - Voting scheme.

Prediction: - Averaging scheme.

The effect of combining different classifiers (hypotheses) can be explained with the theory of bias-variance decomposition

• Bias – an error due to a learning algorithm

• Variance – an error due to the learned model (data set related)

• The total expected error of a classifier = Bias + Variance

Bagging provides a substantial reduction in prediction error for regression as well as classification methods. Since the method employs averaging of several predictors, it is not useful for improving linear models.

4.3.3 Random Forest

The random forest (Breiman, 2001) is an ensemble approach that can also be thought of as a form of nearest neighbor predictor. An ensemble classifier uses many decision tree models. The main principle behind ensemble methods is that a group of "weak learners" can come together to form a "strong learner" .It can be used for both classification and regression

4.3.3.1 The Algorithm

Consider T number of trees

1. Sample N cases at random with replacement to create a subset of data. The subset should be about 66% of the total set.
2. At each node:
    1. For some number m, m predictor variables are selected at random from all the predictor variables.
    2. The predictor variable that provides the best split, according to some objective function is used to do a binary split on the node.
    3. At the next node, choose another m variable at random from all predictor variables and do the same.

Depending upon the value of m, there are three slightly different systems:

Random splitter selection: m =1
Breiman's bagger: m = total number of predictor variables.
Random Forest: m<<number of predictor variables. Brieman suggests three possible values form: $1/2\sqrt{m}$, $\sqrt{m}$ and $2\sqrt{m}$.

Running a Random Forest: When a new input is entered into the system, it is run down all of the trees. The result may either be an average or weighted average of all of the terminal nodes that are reached, or, in the case of categorical variables, a voting majority.

Strengths

Runtime is very fast.
Unbalanced and missing data can be handle.

Weaknesses

When used for regression they cannot predict beyond the range in the training data. They may over-fit data sets that are particularly noisy.

4.3.4 AdaBoostMI

AdaBoostMI, is an acronym for Adaptive Boosting, which is a machine learning algorithm, formulated by Yoav Freund and Robert Schapire in 1995. It is a meta-heuristic, and thus used in conjunction with many other learning algorithms to improve their performance [34]. It was the first algorithm that could adapt to the weak learners.

4.3.4.1 The Algorithm

Initialization

- All instances are equally weighted.
- A learning algorithm is applied.
- The weight of incorrectly classified examples is increased ("hard" instances), correctly decreased ("easy" instances).
- The algorithm concentrates on incorrectly classified "hard" instances. Some "had" instances become "harder" some "softer".
- A series of diverse experts (classifiers) is generated based on the reweighed data.

Steps

1. Set the weight value, w = 1, and assign it to each object in the training data set.
2. For each of t iterations, perform:
3. Apply a learning algorithm to the weighted training data set.

Compute classification error e for the weighted training data set.

If e = 0 or e>=.5, then terminate the classifier generation process and go to 4; otherwise multiple the weight w of each object by e/ (1 – e) and normalize the weights of all objects.

4. Classification

   Assign weight q = 0 to each decision (class) to be predicted.

5. For each of (or less) classifiers, add –log e/(1 – e) to the weight of the decision predicted by the classifier and output the decision with the highest weight.

For e = 0 all training examples (objects) are correctly classified (a perfect classifier) and therefore there is no reason to modify the object weights, i.e., for e/(1 – e) = 0 all new weights w become .

For e = .5, the expression – log e/(1 – e) = 0, and therefore the weights q = 0 are not be modified and therefore no decision is generated due to high classification error e.

4.3.5 Random Tree

Random tree is a single tree constructed in the Random Forest, or we can say that random forest is constructed by bagging ensembles of random trees. At each node of random tree, we select a given number of random features to find the best split and grow the tree to the maximum extent. There is no pruning.

4.3.6 Multilayer Perceptron

A Multilayer Perceptron is a feed forward (no recurrent connection) artificial neural network. Each MLP is composed of a minimum of three layers consisting of an input layer, one or more hidden layers and an output layer [13]. The input layer distributes the inputs to subsequent layers. Input nodes have liner activation functions and no thresholds. Each hidden unit node and each output node have thresholds associated with them in addition to the weights. The hidden unit nodes have nonlinear activation functions and the outputs have linear activation functions.

4.3.6.1 The Algorithm

1. Initially set all weights of network randomly between 1 and -1.
2. Obtain the output based on first training pattern.
3. Compare the network output with the target output.
4. Propagate the error backwards.

a. correct weights of output layer using the following formula.

$$W_{ho} = W_{ho} + (\eta \delta_o O_h)$$

W_{ho} – weight between hidden unit h and output unit o.

η - learning rate

O_h – output at hidden unit h.

$$\delta_o = O_o(1 - O_o)(t_o - O_o)$$

O_o– output at node o of output layer.

t_o - target output for that node.

b. correct the input weights using the following formula

$$W_{ih} = W_{ih} + (\eta \delta_h O_i)$$

W_{ih} – weight between input layer node i and hidden layer node h.

O_i- input a node I of input layer.

5. Calculate the error, by taking the average difference between the target and the output vector.
6. Repeat from 2 for each pattern in the training set to complete one epoch.

7. Shuffle the training set randomly. This is important so as to prevent the network being influenced by the order of the data.

8. Repeat from step 2 for a set number of epochs, or until the error ceases to change

4.3.7 NNGE

The third classifier used is Non-Nested Generalized Exemplars (NNGE), which is an algorithm introduced by Brent, 1995. It performs generalization by merging exemplars, forming hyper rectangles in attribute space that represent conjunctive rules with internal disjunction. The algorithm forms generalization each time a new example is added to the database, by joining it to its nearest neighbor of the same class.

The algorithm learns incrementally by first classifying, then generalizing each new example. When classifying an instance, one or more hyper rectangles may be found that the new instance is a member of, but which are of wrong class. The algorithm prunes these so that the new example is no longer a member. Once classified, the new instance is generalized by merging it with the nearest exemplar of the same class, which may be a single instance or a hyper rectangle.

The only thing that may pose a problem is that the algorithm tends to produce rules that test a large number of attributes. Because of this they are not very intelligible to people.

4.3.8 ONER

OneR stands for “One Rule”, classification algorithm in which for each prediction in data one rule is generated and out of all these rule there is “One Rule” with smallest total error. To create a rule for a prediction, we construct a frequency table for each predictor against the target.

4.3.8.1 OneR Algorithm

For each predictor,

For each value of that predictor, make a rule as follows;

Count how often each value of target (class) appears

Find the most frequent class

Make the rule assign that class to this value of the predictor

Calculate the total error of the rules of each predictor

Choose the predictor with the smallest total error.

CHAPTER 5: RESULT ANALYSIS

In this section we analyze the web metrics predictions calculated for finding the relationship between the web page metrics and goodness of websites and also classify websites of same category. To analyze the results we have employed various machines learning algorithm. The algorithms which we used are naïve bayes, random forest, random tree, OneR, AdaBoostMI, Bagging, NNgr, Multilayer perception. The following measures are used to evaluate the performance of each predicated model.

1. Sensitivity and Specificity: To predict the correctness of the model sensitivity and specificity of the model are computed [13]. The percentage of websites correctly predicted to be good among all the websites is known as sensitivity (true positive rate) of the model. The percentage of websites correctly predicted to be not good among all websites is known as specificity (1-false positive value) of the model. Both the sensitivity and specificity should be high in order to predict as well as classify the websites.

2. Receiver Operating Characteristics (ROC) Curves: The performance of the outputs of the predicted models is evaluated using ROC analysis [24]. X coordinate specifies 1-specificity and Y –coordinate specify sensitivity on ROC curve. We select many cutoff points between 0 and 1 after the construction of ROC curve to calculate specificity and sensitivity at each cut off point. The cutoff point that maximizes both sensitivity and specificity is called optimal cutoff point and is selected for ROC curve. Accuracy of a model is computed by area under curve (AUC). AUC is a combined measure of sensitivity and specificity.

To predict the accuracy of the model we apply it to different data sets. We therefore, performed a 10-cross validation of the models [14].In this each dataset is randomly divided into 10 equal subsets. Out of these 10 subsets each time one is selected as test set and rest of them as training set.

Section 5.1 shows the descriptive statistics; section 5.2 gives the analysis of logistic regression

statistical technique and section 5.3-5.10 describe the analyses of machine learning techniques and finally section 5.11 discusses the evaluated results.

5.1 Descriptive Statistics

In this section we show min, max, mean and standard deviation for all 3 models. Table 5.1 presents the descriptive statistics of Data set 1(Model 1).

Metric	Min	Max	Mean	SD
Meta Tag	2	36	8.85	6.27
Meta Keywords	0	217	6.01	23.85
Min Keyword length	0	23	2.58	3.94
Max Keyword length	0	28	5.59	7.58
Meta Descriptor	0	241	17.9	24.34
Total Link	15	1805	156.57	203.35
Image Link	1	152	18.99	22.64
Average no. of words in a link	1.23	14.17	2.79	1.58
Total Images	1	273	32.722	42.22
Alt Images	0	89	21.15	19.46
Words in alt images	0	418	52.29	73.16
Division tag	6	1513	150.44	209.62
Paragraph	0	203	27.02	38.522
Scripts	0	63	16.6	14.07
Page Size	6775	474723	6886.752	77609.1
Body Word Count	17	18158	1147.05	1986.07
Title Length	0	21	6.73	3.83
Tables	0	6	0.21	0.699
Load Time	.034	7.328	1.67	1.349
Total Headings	0	145	15.84	21.21
Link Headings	0	61	6.1	11.059

Table 5.1: Descriptive Statistics of Model 1

Similarly, Table 5.2 shows the descriptive statistics of Data set 2(Model 2).

Metric	Min	Max	Mean	SD
Meta Tag	0	24	8.09	5.55
Meta Keywords	0	37	8.019	8.87
Min Keyword length	0	25	3.13	4.52
Max Keyword length	0	28	10.7	9.59
Meta Descriptor	0	51	20.01	13.2
Total Link	1	392	97.27	91.06
Image Link	0	91	19.35	24.26
Average no. of words in a link	0	13.28	2.34	1.97
Total Images	0	230	34.33	48.39
Alt Images	0	210	29.35	45.39
Words in alt images	0	255	50.05	72
Division tag	1	596	101.01	127.43
Paragraph	0	108	17.66	24.01
Scripts	1	55	16.49	12.17
Page Size	3657	169755	50542.67	46794.08
Body Word Count	0	7864	796.08	1242.61
Title Length	1	20	7.82	5.13
Tables	0	4	0.27	0.93
Load Time	0.012	4.429	1.09	0.84
Total Headings	0	95	19.74	24.99
Link Headings	0	36	4.27	8.16

Table 5.2: Descriptive Statistics of Model 2

Similarly, Table 5.3 shows the descriptive statistics of Data set 3(Model 3).

Metric	Min	Max	Mean	SD
Meta Tag	2	23	9.03	5.08
Meta Keywords	0	140	11.17	27.79
Min Keyword length	0	67	4.75	7.33
Max Keyword length	0	67	11.45	12.04
Meta Descriptors	0	136	19.44	20.23
Total Link	1	493	118.605	91.3
Image Link	0	117	14.55	22.18
Average no. of words in a link	0	3.91	1.93	0.605
Total Images	0	182	33.82	44.49
Alt Images	0	182	24.48	40.69
Words in alt images	0	243	33.93	51.23
Division tag	1	344	105.93	86.25
Paragraph	0	116	19.02	22.704
Scripts	1	40	17.57	11.0452
Page Size	3161	358927	59212.61	52557.98
Body Word Count	12	2176	736.22	562.09
Title Length	1	18	6.68	3.72
Tables	0	96	2.21	10.95
Load Time	0.001	4.444	1.22	0.944
Total Headings	0	133	19.5	22.71
Link Headings	0	52	4.67	11.62

Table 5.3: Descriptive Statistics of Model 3

5.2 Naïve Bayes Analysis

Tables 5.4 and 5.5 represent website prediction and 10-cross validation results for goodness of a webpage for all the 3 models by naïve bayes classifier.

The observations which are made from Tables 5.4 and 5.5 are summarized below:

- In model 1, 32 website out of 35 are correctly predicted as good and 39 websites out of 84 are correctly predicted as bad.
- In model 2, 7 websites out of 12 are correctly predicted as good and 35 websites out of 39 are correctly predicted as bad.
- In model 3, 19 websites out of 28 are correctly predicted as good and 74 websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	32	7	19
Number of bad websites correctly predicted	39	35	74

Table 5.4 Goodness of Websites Using Naïve Bayes Classifier for Model 1, 2 And 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.714	0.714	0.834	0.776
Model 2	0.833	0.795	0.248	0.793
Model 3	0.714	0.723	0.409	0.805

Table 5.5 10-Cross Validation Results for Models Using Naïve Bayes Classifier

Figures 5.1-5.3 shows the ROC Curves for Model 1, Model 2 and Model 3 using Naïve Bayes Classifier.

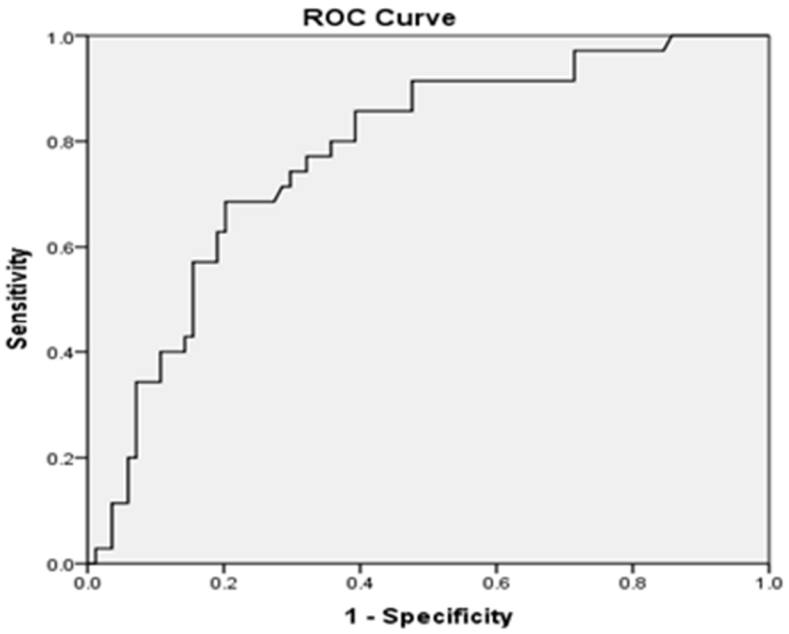

Figure 5.1 ROC Curve for Model 1 Using Naïve Bayes Classifier

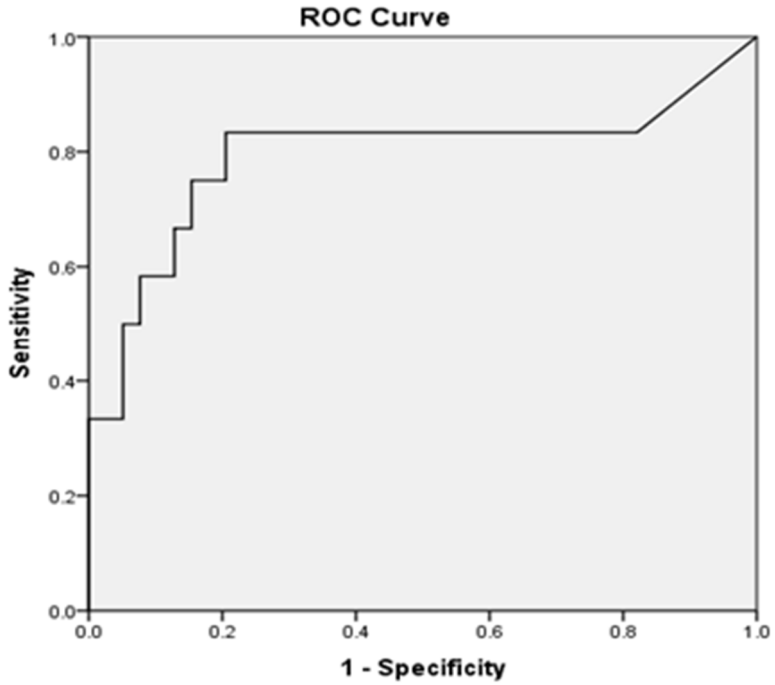

Figure 5.2 ROC Curve for Model 2 Using Naïve Bayes Classifier

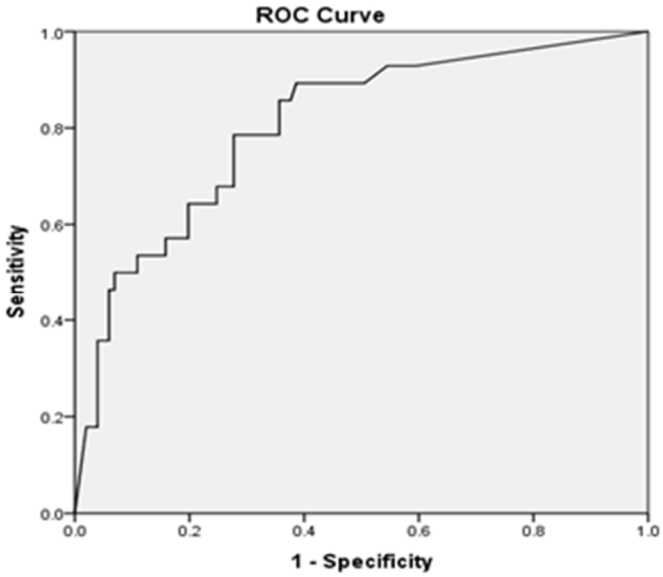

Figure 5.3 ROC Curve for Model 3 Using Naïve Bayes Classifier

Tables 5.6 and 5.7 represent web page prediction and 10-cross validation results for classification of web page of same category of all the 3 models by naïve bayes classifier.
The observations which are made from Tables 5.6 and 5.7are summarized below:

- In model 1, 20 websites out of 62 are correctly predicted as Blog and 47 websites out of 57 are correctly predicted as Community.
- In model 2, 13 websites out of 25 are correctly predicted as TV and 20 websites out of are correctly predicted as 26 Movies.
- In model 3, 22 websites out of 39 are correctly predicted as Food & Beverages, 38 websites out of 49 are correctly predicted as Commerce and 22 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	20
	Community	47
Model 2	TV	13
	Movies	20
Model 3	Food & Beverages	22
	Commerce	38

	Travel	22

Table 5.6 Class Prediction of Websites Using Naïve Bayes Classifier for Model 1, 2 and 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.661	0.614	0.190	0.657
Model 2		0.750	0.652	0.767	0.674
Model 3	Food & Beverages	0.61	0.817		0.855
	Commerce	0.58	0.828		
	Travel	0.78	0.81		

Table 5.7 10-Cross Validation Results for Models Using Naïve Bayes Classifier

Figures 5.4-5.5 shows the ROC Curves for Model 1, Model 2, and Model 3 using Naïve Bayes Classifier.

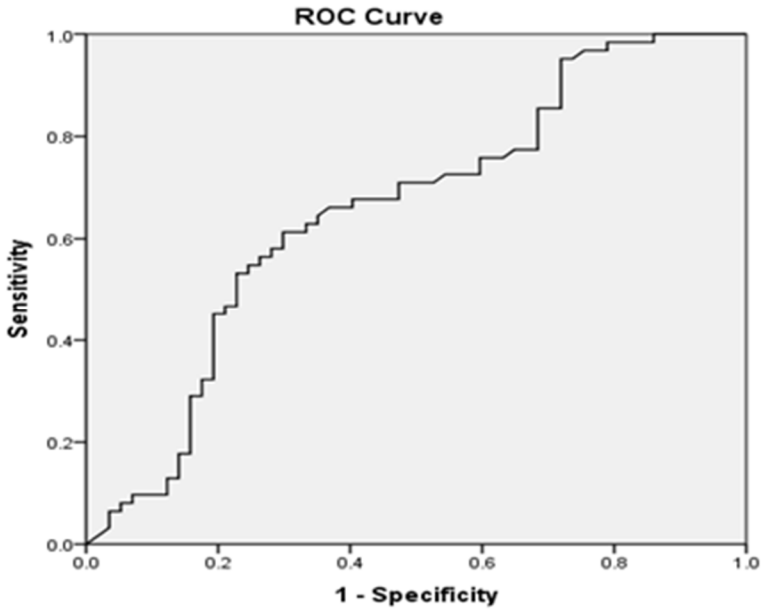

Figure 5.4 ROC Curve for Model 1 Using Naïve Bayes Classifier

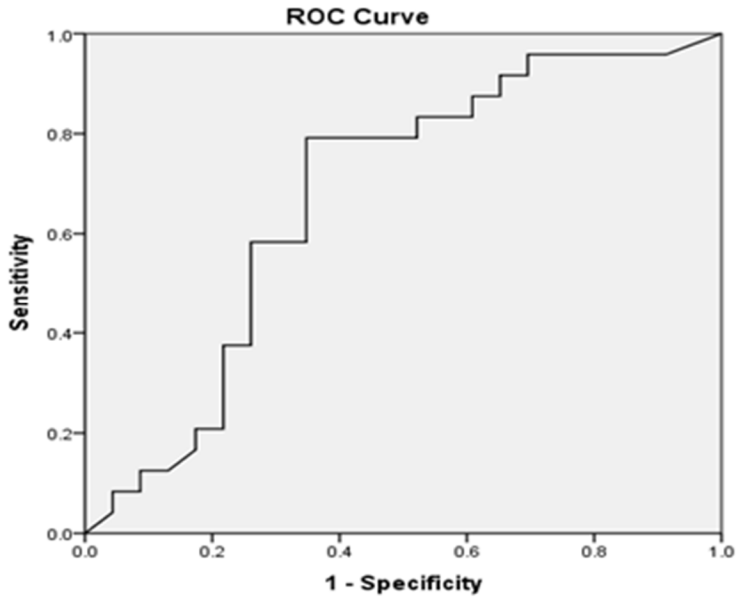

Figure 5.5 ROC Curve for Model 2 Using Naïve Bayes Classifier

5.3 Multilayer Perceptron

Tables 5.8 and 5.9 represent web page prediction and 10-cross validation results for all the 3 models by Multilayer Perceptron.

The observations which are made from Table 5.8 and 5.9 are summarized below:

- In model 1, 23websites out of 35 are correctly predicted as good and 71 websites out of 84 are correctly predicted as bad.
- In model 2, 8websites out of 12 are correctly predicted as good and 32websites out of 39 are correctly predicted as bad.
- In model 3, 10websites out of 28 are correctly predicted as good and 91websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	23	8	10
Number of bad websites correctly predicted	71	32	91

Table 5.8 Goodness of Websites Using Multilayer Perceptron Classifier for Model 1, 2 and 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.743	0.738	0.351	0.805
Model 2	0.750	.744	0.270	0.755
Model 3	0.714	0.713	0.4095	0.753

Table 5.9 10-Cross Validation Results for Models Using Multilayer Perceptron Classifier

Figures 5.6-5.8 shows the ROC Curves for Model 1, Model 2, and Model 3 using Multilayer Perceptron Classifier.

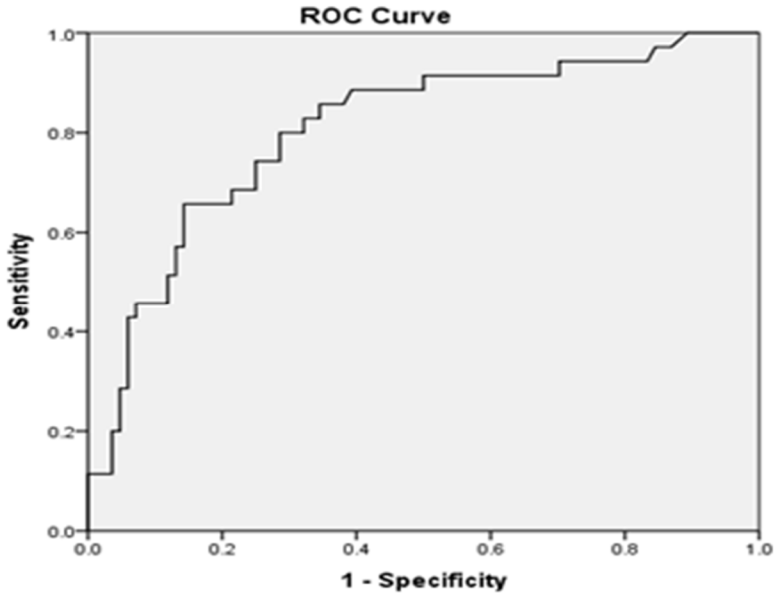

Figure 5.6 ROC Curve for Model 1 Using Multilayer Perceptron Classifier

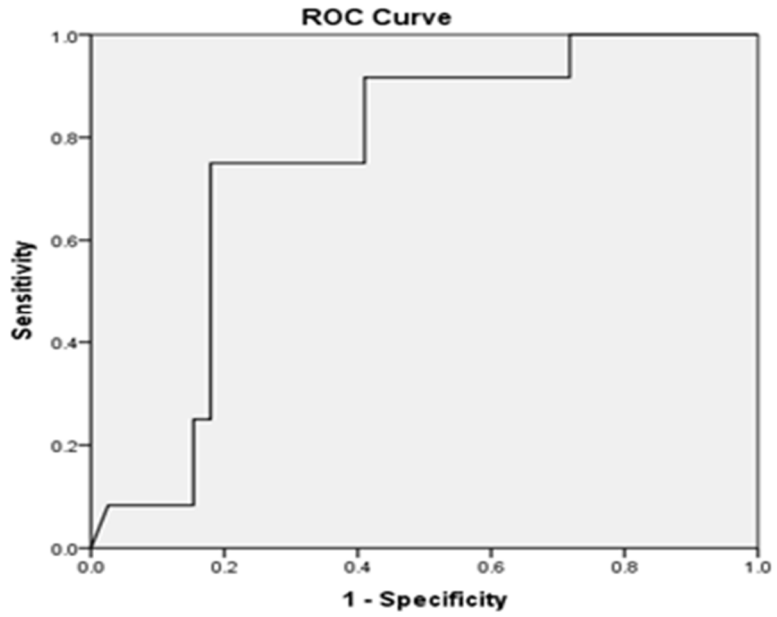

Figure 5.7 ROC Curve for Model 2 Using Multilayer Perceptron Classifier

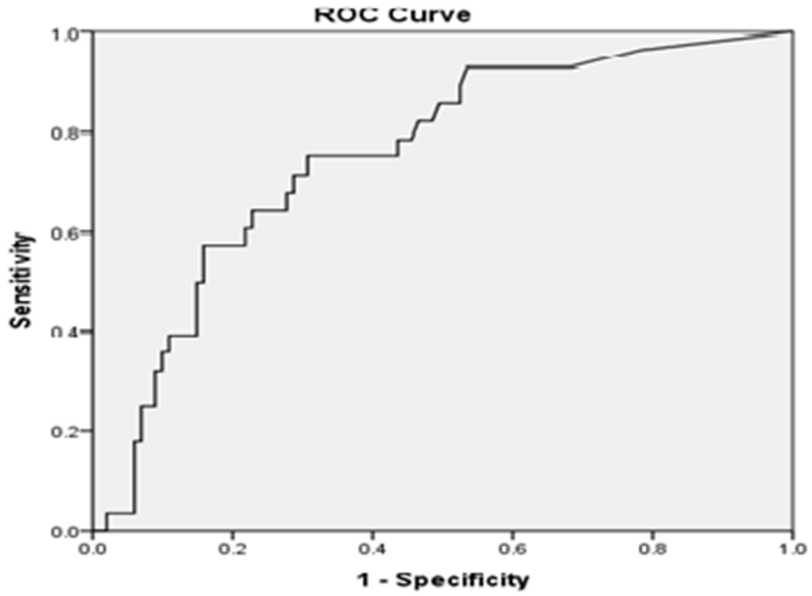

Figure 5.8 ROC Curve for Model 3 Using Multilayer Perceptron Classifier

Tables 5.10 and 5.11 represent web sites prediction and 10-cross validation results for classification of web sites of same category of all the 3 models by MLP classifier.
The observations which are made from Table 5.10 and 5.11 are summarized below:

- In model 1, 38 websites out of 62 are correctly predicted as Blog and 43 websites out of 57 are correctly predicted as Community.
- In model 2, 14 websites out of 25 are correctly predicted as TV and 19 websites out of are 26 correctly predicted as Movies.
- In model 3, 23 websites out of 39 are correctly predicted as Food & Beverages, 39 websites out of 49 are correctly predicted as Commerce and 31 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	38
	Community	43
Model 2	TV	14
	Movies	19
Model 3	Food & Beverages	23
	Travel	31
	Commerce	39

Table 5.10 Class Prediction of Websites Using Multilayer Perceptron Classifier for Model 1, 2 and 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.742	0.737	0.446	0.758
Model 2		0.640	0.654	0.360	0.702
Model 3	Food & Beverages	0.79	0.82		0.853
	Commerce	0.88	0.86		
	Travel	0.67	0.87		

Table 5.11 10-cross Validation Results for Models Using Multilayer Perceptron Classifier

Figure 5.9-5.10 shows the ROC Curves for Model 1, Model 2, and Model 3 using Multilayer Perceptron Classifier.

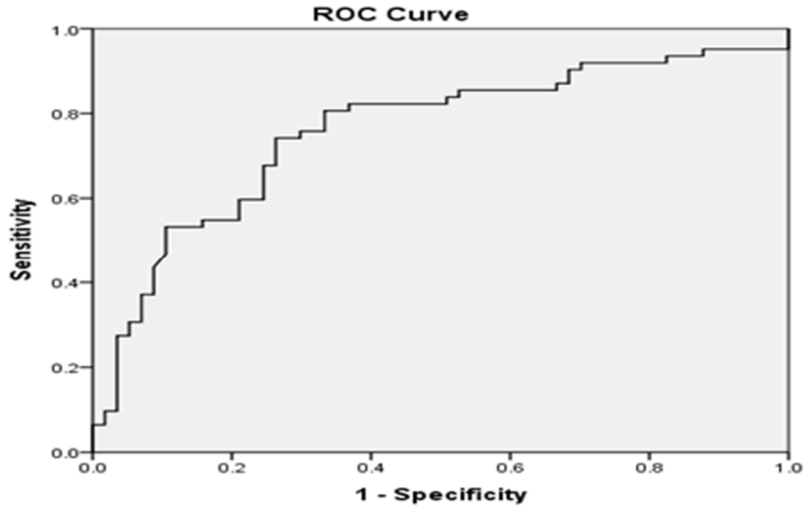

Figure 5.9 ROC Curve for Model 1 Using Multilayer Perceptron Classifier

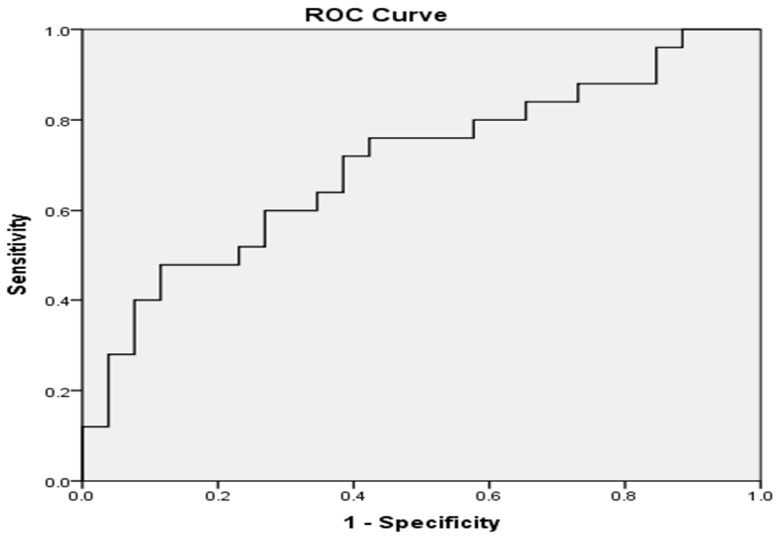

Figure 5.10 ROC Curve for Model 2 Using Multilayer Perceptron Classifier

5.4 ADABoostMI

Tables 5.12 and 5.13 represent web page prediction and 10-cross validation results for all the 3 models by AdaBoostMI.

The observations which are made from Table 5.12 and 5.13 are summarized below:

- In model 1, 23websites out of 35 are correctly predicted as good and 68 websites out of 84 are correctly predicted as bad.

- In model 2, 7websites out of 12 are correctly predicted as good and 32websites out of 39 are correctly predicted as bad.
- In model 3, 15websites out of 28 are correctly predicted as good and 97websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	23	7	15
Number of bad websites correctly predicted	68	34	97

Table 5.12 Goodness of Websites Using AdaBoostMI Classifier for Model 1, 2 And 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.743	0.750	0.399	0.828
Model 2	0.750	0.744	0.230	0.846
Model 3	0.679	0.644	0.137	0.767

Table 5.13 10-Cross Validation Results for Models Using AdaBoostMI Classifier

Figures 5.11-5.13 shows the ROC Curves for Model 1, Model 2, and Model 3 using AdaBoostMI Classifier.

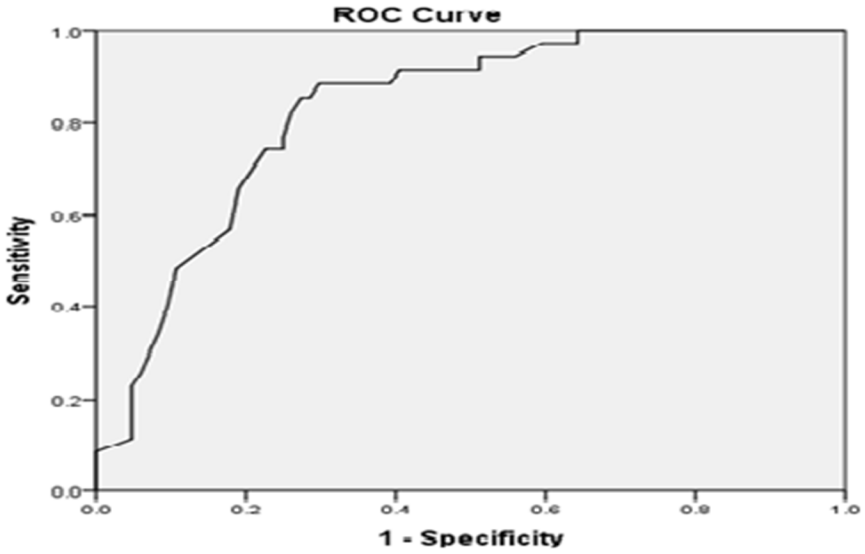

Figure 5.11 ROC Curve for Model 1 Using ADABoostMI Classifier

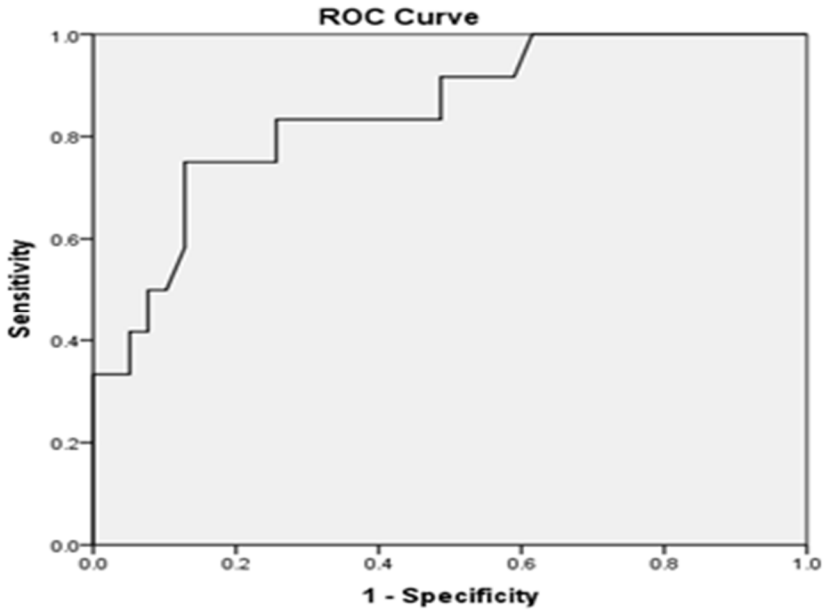

Figure 5.12 ROC Curve for Model 2 Using ADABoostMI Classifier

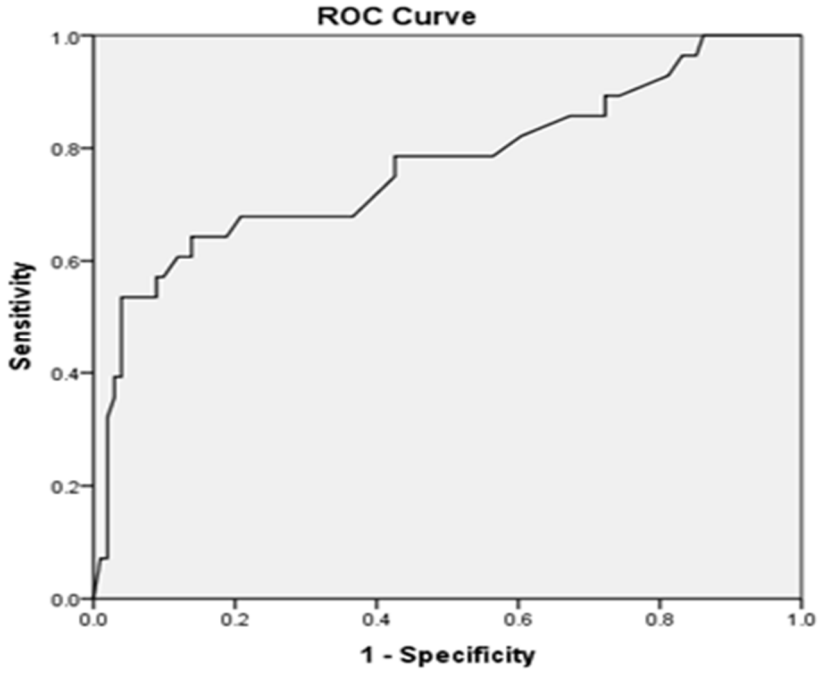

Figure 5.13 ROC Curve for Model 3 Using ADABoostMI Classifier

Tables 5.14 and 5.15 represent website prediction and 10-cross validation results for classification of website of same category of all the 3 models by AdaBoostMI classifier.
The observations which are made from Table 5.14 and 5.15 are summarized below:

- In model 1, 49 websites out of 62 are correctly predicted as Blog and 48 websites out of 57 are correctly predicted as Community.
- In model 2, 20 websites out of 25 are correctly predicted as TV and 14 websites out of 26 are correctly predicted as Movies.
- In model 3, 36 websites out of 39 are correctly predicted as Food & Beverages, 26 websites out of 49 are correctly predicted as Commerce and 5 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	49
	Community	48
Model 2	TV	20
	Movies	14
Model 3	Food & Beverages	36
	Travel	5
	Commerce	26

Table 5.14 Class Prediction of Website Using AdaBoostMI Classifier for Model 1, 2 And 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.790	0.789	0.405	0.858
Model 2		0.667	0.652	0.465	0.755
Model 3	Food & Beverages	0.45	0.938		0.682
	Commerce	0.65	0.74		
	Travel	0.55	0.70		

Table 5.15 10-Cross Validation Results for Models Using AdaBoostMI Classifier

Figures 5.14-5.15 shows the ROC Curves for Model 1, Model 2, and Model 3 using AdaBoostMI Classifier.

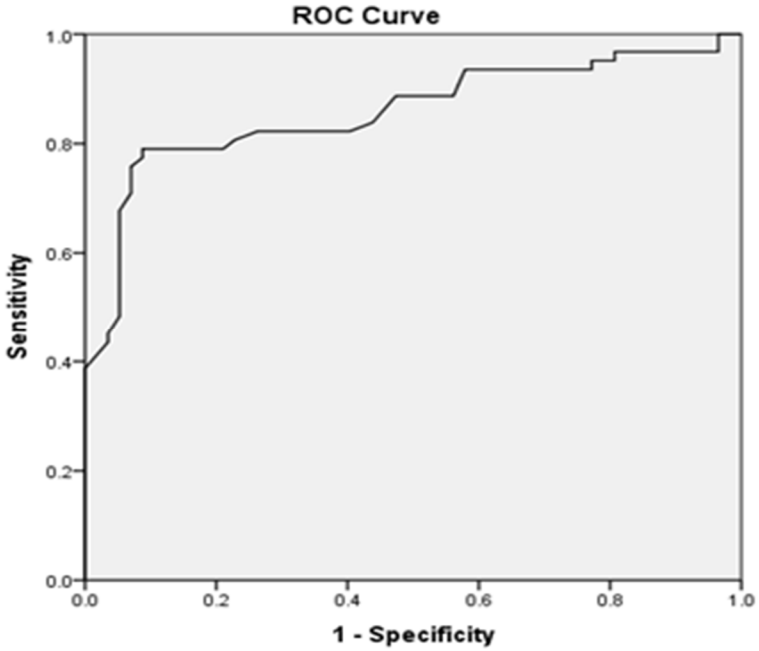

Figure 5.14 ROC Curve for Model 1 using ADABoostMI Classifier

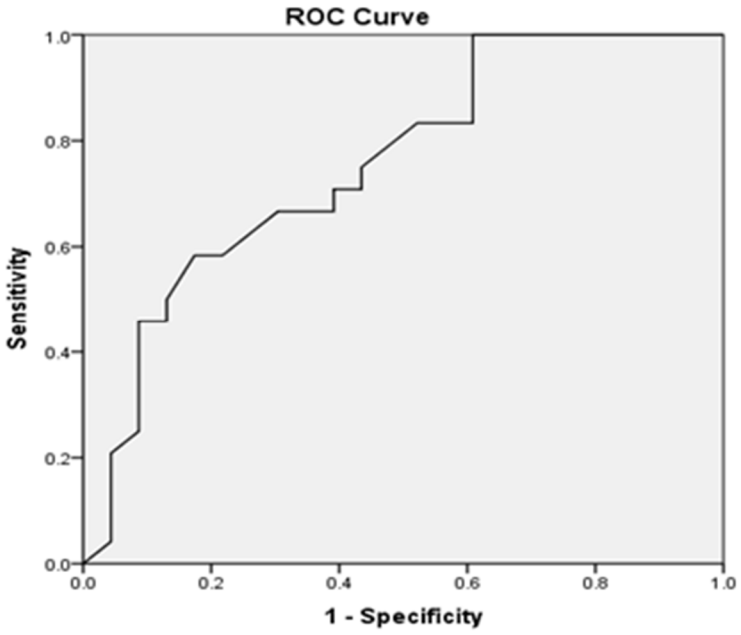

Figure 5.15 ROC Curve for Model 2 using ADABoostMI Classifier

5.5 Bagging

Tables 5.16 and 5.17 represent web page prediction and 10-cross validation results for all the 3 models by bagging classifier.

The observations which are made from Table 5.16 and 5.17 are summarized below:

- In model 1, 24websites out of 35 are correctly predicted as good and 78 websites out of 84 are correctly predicted as bad.
- In model 2, 4websites out of 12 are correctly predicted as good and 36 websites out of 39 are correctly predicted as bad.
- In model 3, 14websites out of 28 are correctly predicted as good and 100 websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	24	4	12
Number of bad websites correctly predicted	78	36	100

Table 5.16 Goodness of Website Using Bagging Classifier for Model 1, 2 And 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.829	0.833	0.332	0.900
Model 2	0.833	0.846	0.315	0.847
Model 3	0.750	0.743	0.161	0.800

Table 5.17 10-Cross Validation Results for Models Using Bagging Classifier

Figures 5.16-5.18 shows the ROC Curves for Model 1, Model 2, and Model 3 using Bagging Classifier.

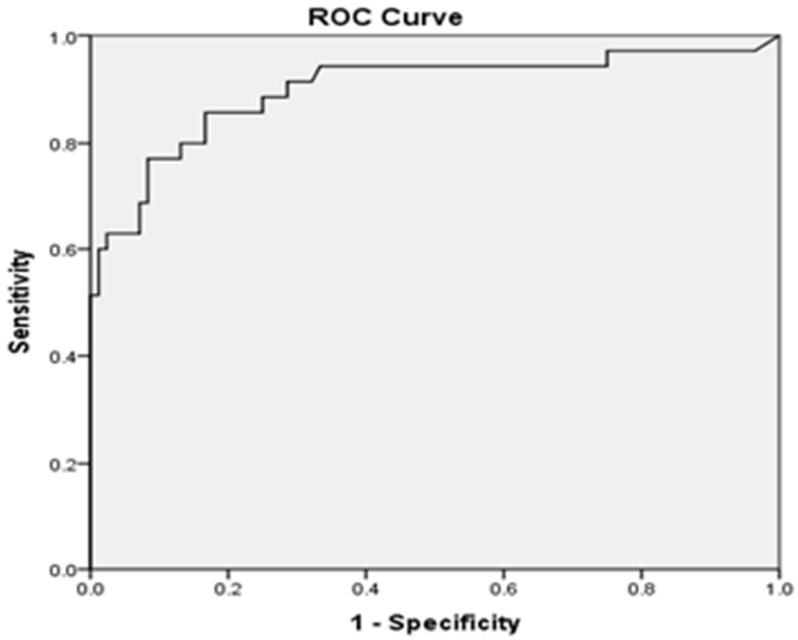

Figure 5.16 ROC Curve for Model 1 Using Bagging Classifier

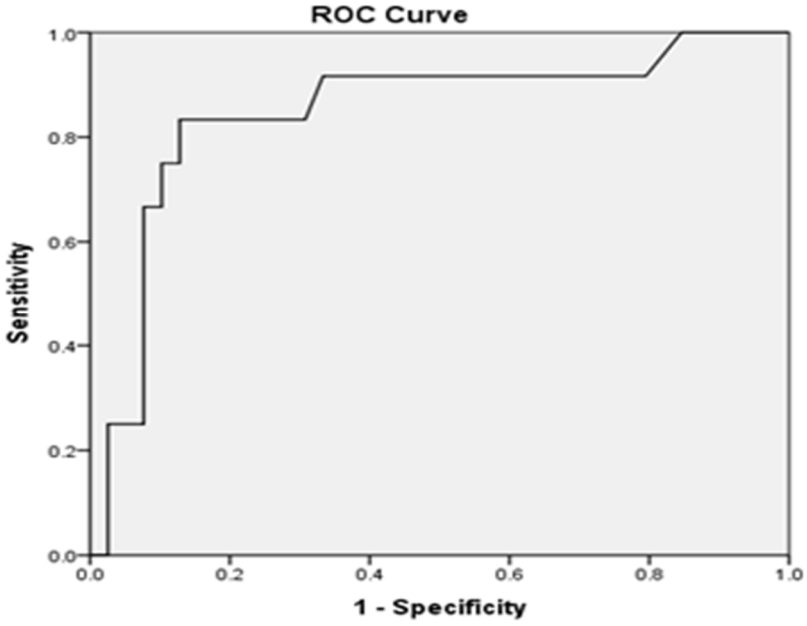

Figure 5.17 ROC Curve for Model 2 Using Bagging Classifier

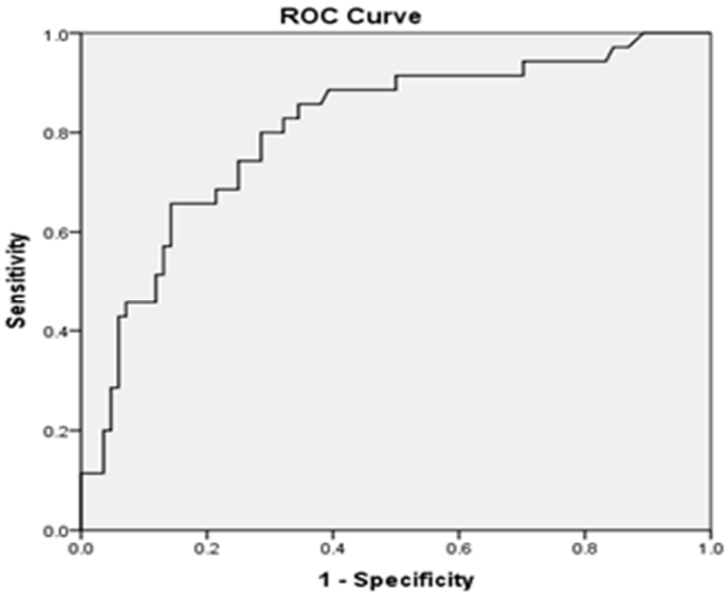

Figure 5.18 ROC Curve for Model 3 Using Bagging Classifier

Tables 5.18 and 5.19 represent web site prediction and 10-cross validation results for classification of web sites of same category of all the 3 models by bagging classifier.
The observations which are made from Table 5.18 and 5.19 are summarized below:

- In model 1, 48 websites out of 62 are correctly predicted as Blog and 48 websites out of 57 are correctly predicted as Community.
- In model 2, 17 websites out of 25 are correctly predicted as TV and 20 websites out of 26 are correctly predicted as Movies.
- In model 3, 36 websites out of 39 are correctly predicted as Food & Beverages, 42 websites out of 49 are correctly predicted as Commerce and 32 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	48
	Community	48
Model 2	TV	17
	Movies	20
Model 3	Food & Beverages	36
	Travel	32

	Commerce	42

Table 5.18 Class Prediction of Website Using Bagging Classifier for Model 1, 2 And 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.823	0.825	0.441	0.885
Model 2		0.720	0.692	0.472	0.808
Model 3	Food & Beverages	0.87	0.96		0.938
	Commerce	0.84	0.911		
	Travel	0.84	0.90		

Table 5.19 10-Cross Validation Results for Models Using Bagging Classifier

Figures 5.19-5.20 shows the ROC Curves for Model 1, Model 2, and Model 3 using Bagging Perceptron Classifier.

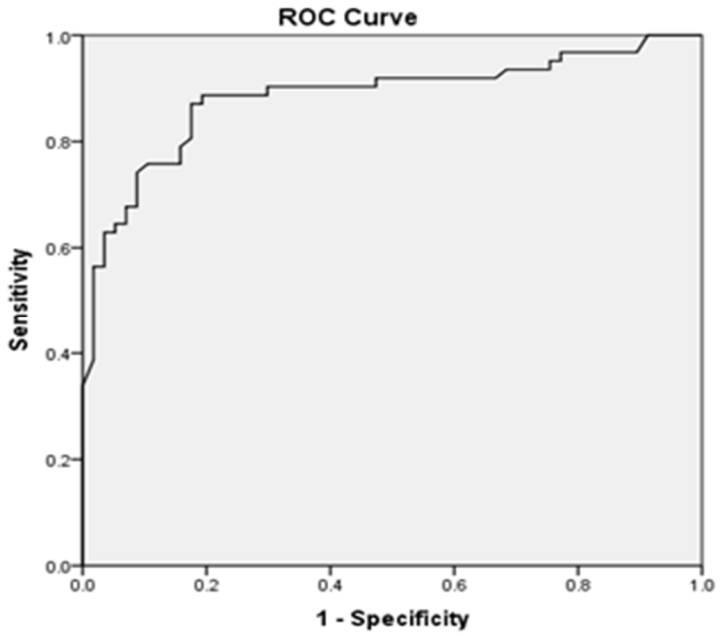

Figure 5.19 ROC Curve for Model 1 using Bagging Classifier

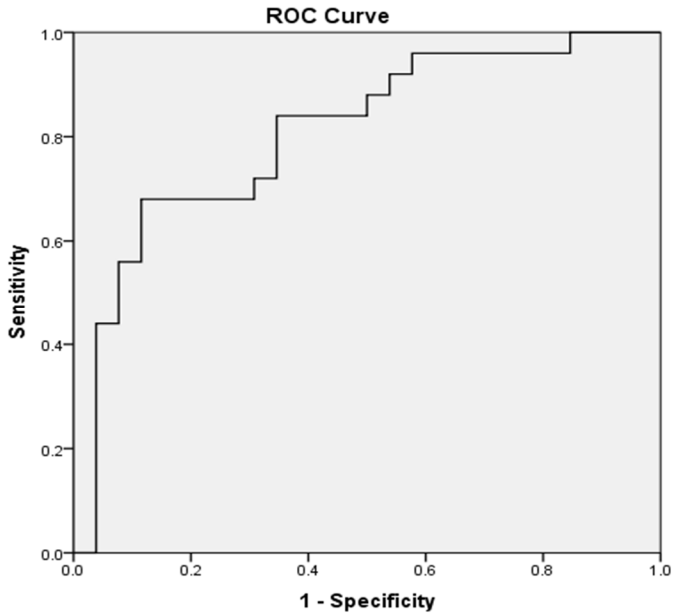

Figure 5.20 ROC Curve for Model 2 using Bagging Classifier

5.6 NNGR

Tables 5.20 and 5.21 represent web site prediction and 10-cross validation results for all the 3 models by nngr.

The observations which are made from Table 5.20 and 5.21are summarized below:

- In model 1, 22websites out of 35 are correctly predicted as good and 76 websites out of 84 are correctly predicted as bad.
- In model 2, 6websites out of 12 are correctly predicted as good and 35 websites out of 39 are correctly predicted as bad.
- In model 3, 13websites out of 28 are correctly predicted as good and 100websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	22	6	13
Number of bad websites correctly predicted	76	35	100

Table 5.20 Goodness of Web Site Using Nngr Classifier for Model 1, 2 And 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.629	0.905	0.50	0.767
Model 2	0.500	0.897	0.50	0.699
Model 3	0.464	0.99	0.50	0.727

Table 5.21 10-Cross Validation Results for Models Using Nngr Classifier

Figures 5.21-5.23 shows the ROC Curves for Model 1, Model 2, and Model 3 using Nngr Classifier.

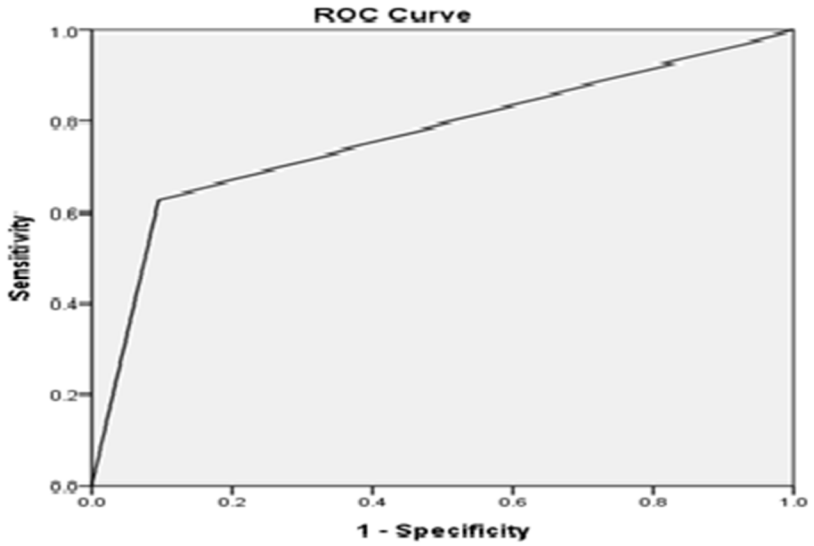

Figure 5.21 ROC Curve for Model 1 using Nngr Classifier

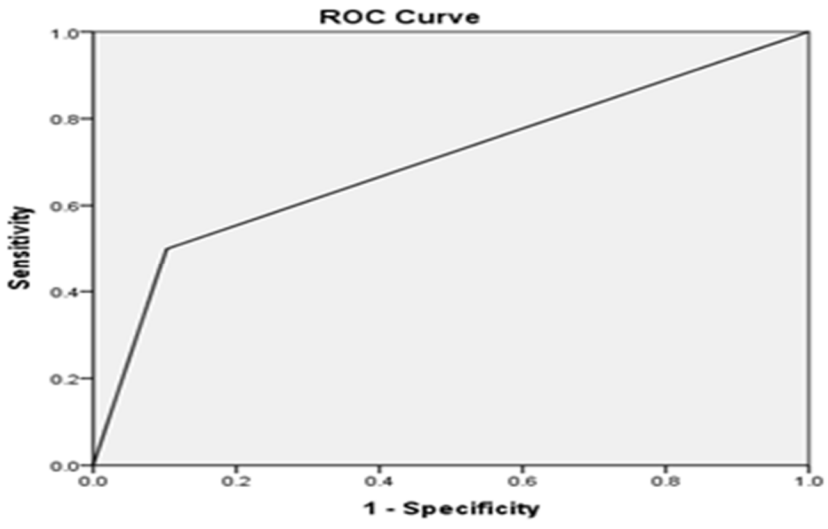

Figure 5.22 ROC Curve for Model 2 using Nngr Classifier

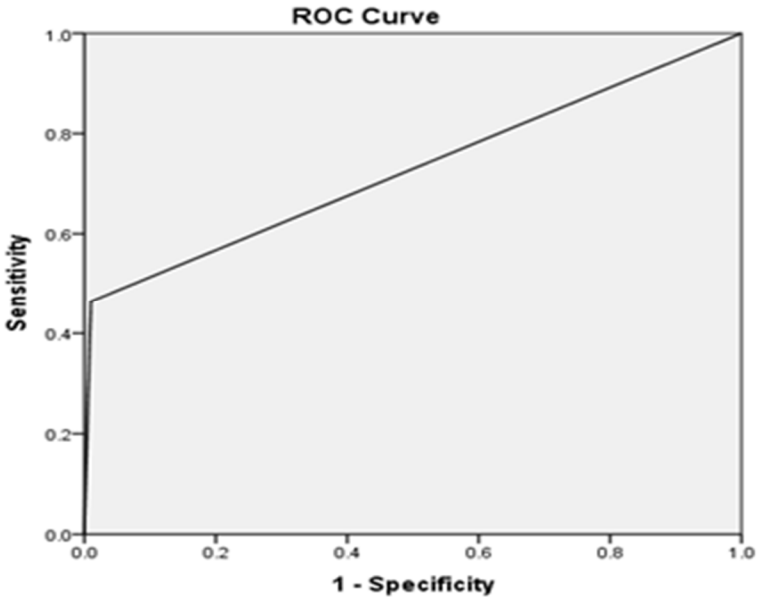

Figure 5.23 ROC Curve for Model 3 using Nngr Classifier

Tables 5.22 and 5.23 represent web page prediction and 10-cross validation results for classification of web page of same category of all the 3 models by nngr classifier.

The observations which are made from Table 5.6 and 5.7 are summarized below:

- In model 1, 54 websites out of 62 are correctly predicted as Blog and 48 websites out of 57 are correctly predicted as Community.
- In model 2, 15 websites out of 25 are correctly predicted as TV and 18 websites out of 26 are correctly predicted as Movies.
- In model 3, 30 websites out of 39 are correctly predicted as Food & Beverages, 47 websites out of 49 are correctly predicted as Commerce and 30 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	54
	Community	48
Model 2	TV	15
	Movies	18
Model 3	Food & Beverages	30
	Travel	30

	Commerce	47

Table 5.22 Class Prediction of Website Using Nngr Classifier for Model 1, 2 And 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.871	0.842	0.50	0.857
Model 2		0.600	0.692	0.50	0.646
Model 3	Food & Beverages	0.857	0.90		0.87
	Commerce	0.82	0.97		
	Travel	0.810	0.87		

Table 5.23 10-Cross Validation Results for Models Using Nngr Classifier

Figures 5.24-5.25 shows the ROC Curves for Model 1, Model 2 and Model 3 using Nngr Classifier.

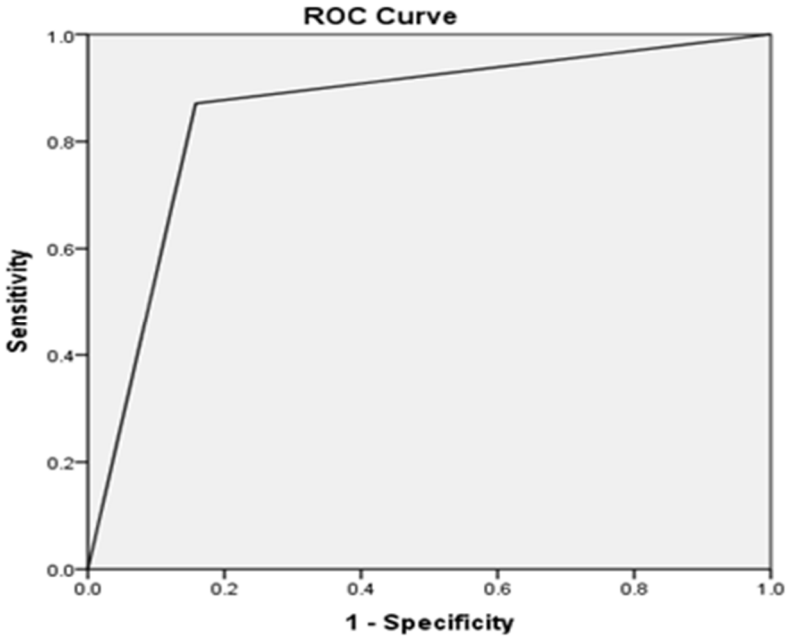

Figure 5.24 ROC Curve for Model 1 Using Nngr Classifier

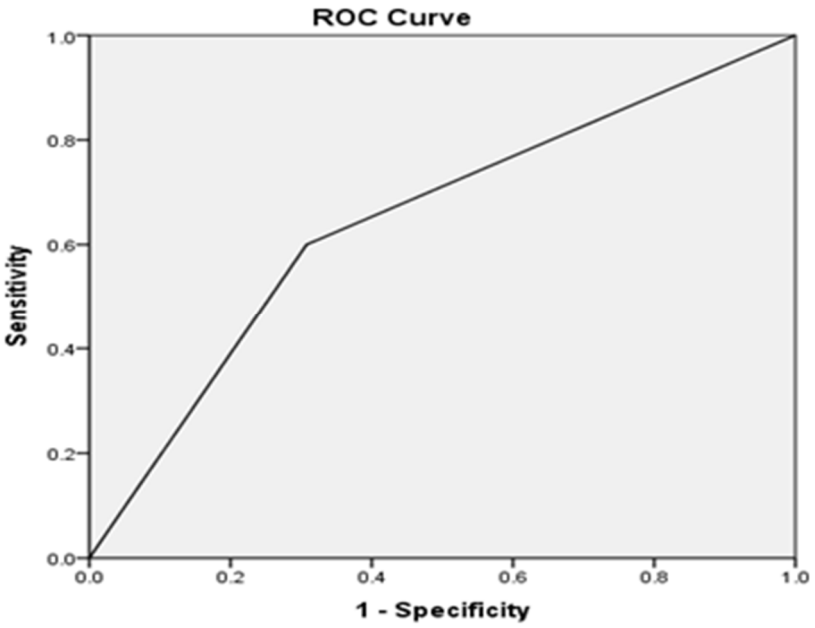

Figure 5.25 ROC Curve for Model 2 Using Nngr Classifier

5.7 OneR

Tables 5.24 and 5.25 represent web page prediction and 10-cross validation results for all the 3 models by oner.

The observations which are made from Table 5.24 and 5.25 are summarized below:

- In model 1, 15websites out of 35 are correctly predicted as good and 66 websites out of 84 are correctly predicted as bad.
- In model 2, 9websites out of 12 are correctly predicted as good and 34 websites out of 39 are correctly predicted as bad.
- In model 3, 14websites out of 28 are correctly predicted as good and 96websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	15	9	14
Number of bad websites correctly predicted	66	34	96

Table 5.24 Goodness of Website Using OneR Classifier for Model 1, 2 And 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.429	0.786	0.50	0.607
Model 2	0.750	0.872	0.50	0.811
Model 3	0.50	0.95	0.50	0.725

Table 5.25 10-Cross Validation Results for Models Using OneR Classifier

Figures 5.26-5.28 shows the ROC Curves for Model 1, Model 2, and Model 3 using OneR Classifier.

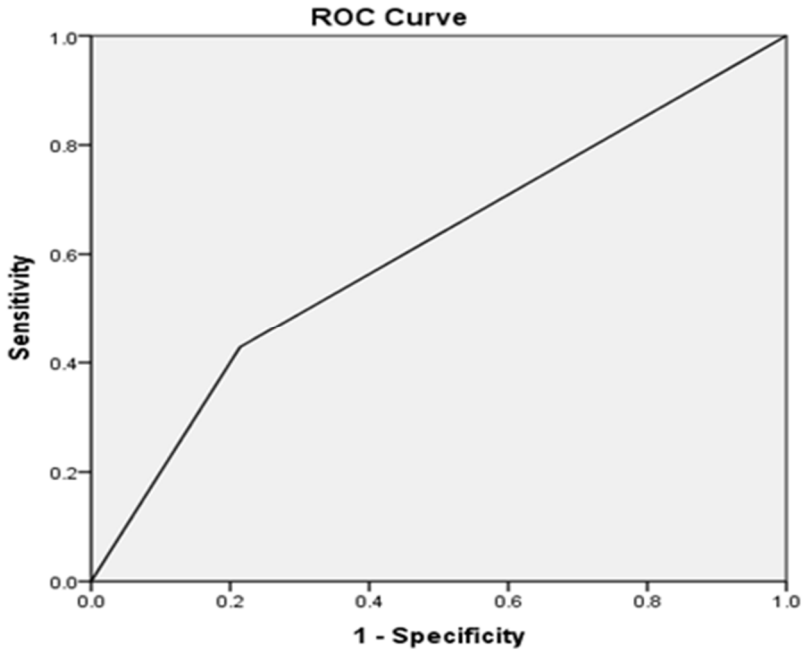

Figure 5.26 ROC Curve for Model 1 Using OneR Classifier

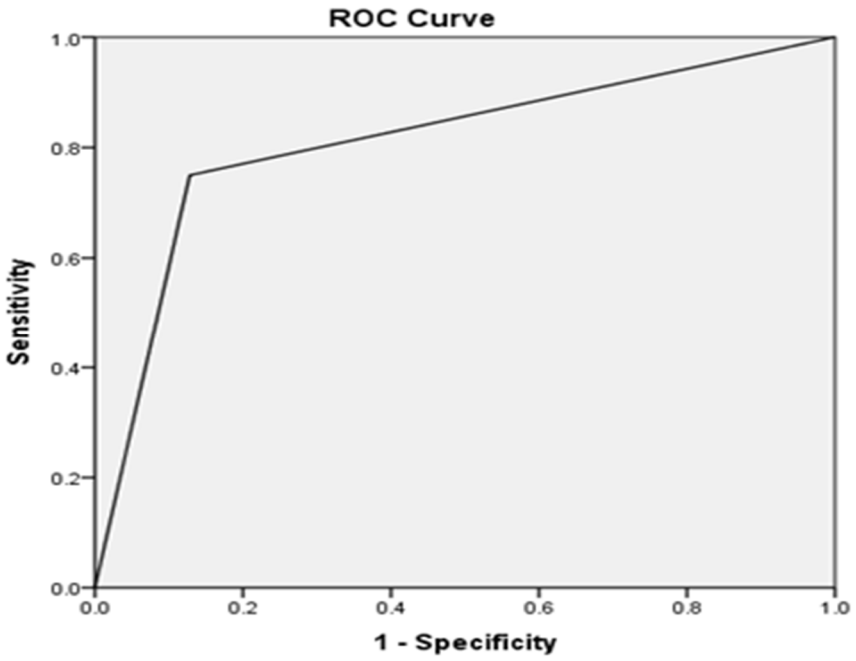

Figure 5.27 ROC Curve for Model 2 Using OneR Classifier

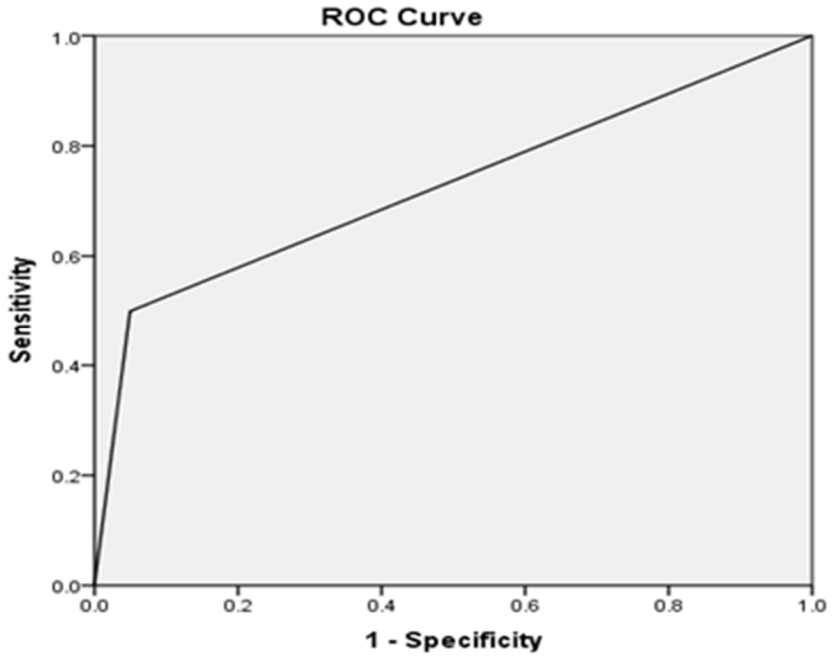

Figure 5.28 ROC Curve for Model 3 Using OneR Classifier

Tables 5.26 and 5.27 represent website prediction and 10-cross validation results for classification of web page of same category of all the 3 models by oner.

The observations which are made from Table 5.26 and 5.27 are summarized below:

- In model 1, 46 websites out of 62 are correctly predicted as Blog and 36 websites out of 57 are correctly predicted as Community.
- In model 2, 18 websites out of 25 are correctly predicted as TV and 15 websites out of 26 are correctly predicted as Movies.
- In model 3, 20 websites out of 39 are correctly predicted as Food & Beverages, 30 websites out of 49 are correctly predicted as Commerce and 20 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	46
	Community	36
Model 2	TV	18
	Movies	15
Model 3	Food & Beverages	20
	Travel	20
	Commerce	30

Table 5.26 Class Prediction of Website Using OneR Classifier for Model 1, 2 And 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.742	0.632	0.50	0.687
Model 2		0.720	0.577	0.50	0.648
Model 3	Food & Beverages	0.60	0.80		0.653
	Travel	0.55	0.74		
	Commerce	0.47	0.75		

Table 5.27 10-Cross Validation Results for Models Using OneR Classifier

Figure 5.29-5.30 shows the ROC Curves for Model 1, Model 2, and Model 3 using OneR Classifier.

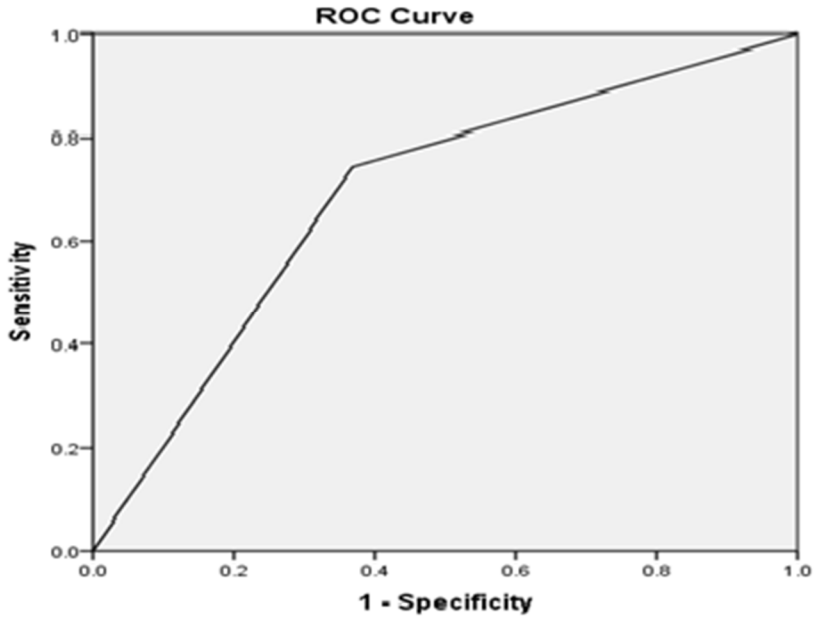

Figure 5.29 ROC Curve for Model 1 Using OneR Classifier

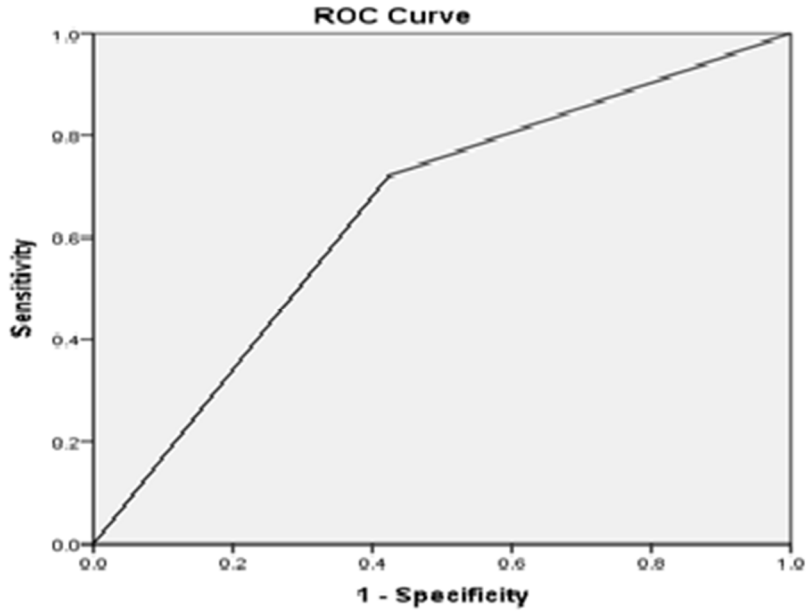

Figure 5.30 ROC Curve for Model 2 Using OneR Classifier

5.8 Random Forest

Tables 5.28 and 5.29 represent web site prediction and 10-cross validation results for all the 3 models by Random Forest.

The observations which are made from Table 5.28 and 5.29 are summarized below:

- In model 1, 26websites out of 35 are correctly predicted as good and 76 websites out of 84 are correctly predicted as bad.
- In model 2, 9websites out of 12 are correctly predicted as good and 34 websites out of 39 are correctly predicted as bad.
- In model 3, 16websites out of 28 are correctly predicted as good and 95websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	26	9	16
Number of bad websites correctly predicted	76	34	95

Table 5.28 Goodness of Website Prediction Using Random Forest Classifier for Model 1, 2 and 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.743	0.857	0.250	0.89
Model 2	0.917	0.846	0.250	0.922
Model 3	0.821	0.752	0.150	0.857

Table 5.29 10-Cross Validation Results for Models Using Random Forest Classifier

Figure 5.31-5.32 shows the ROC Curves for Model 1, Model 2, and Model 3 using Random Forest Classifier.

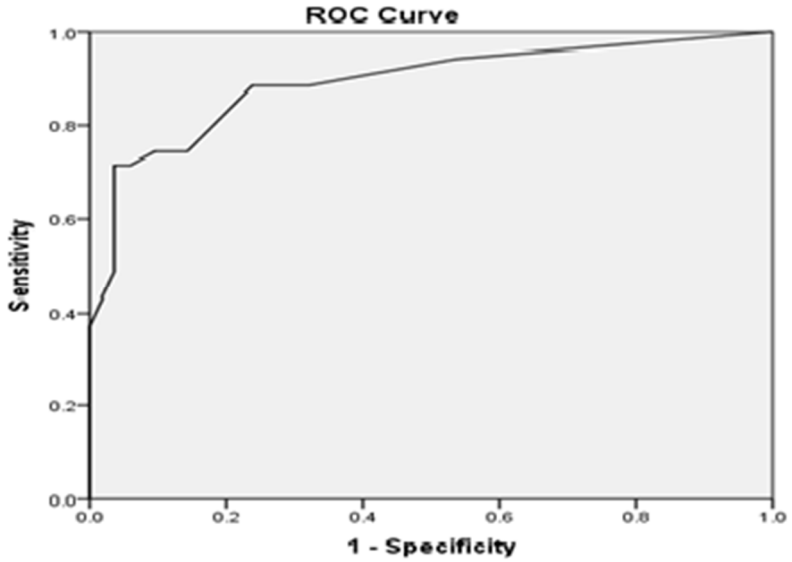

Figure 5.31 ROC Curve for Model 3 Using Random Forest Classifier

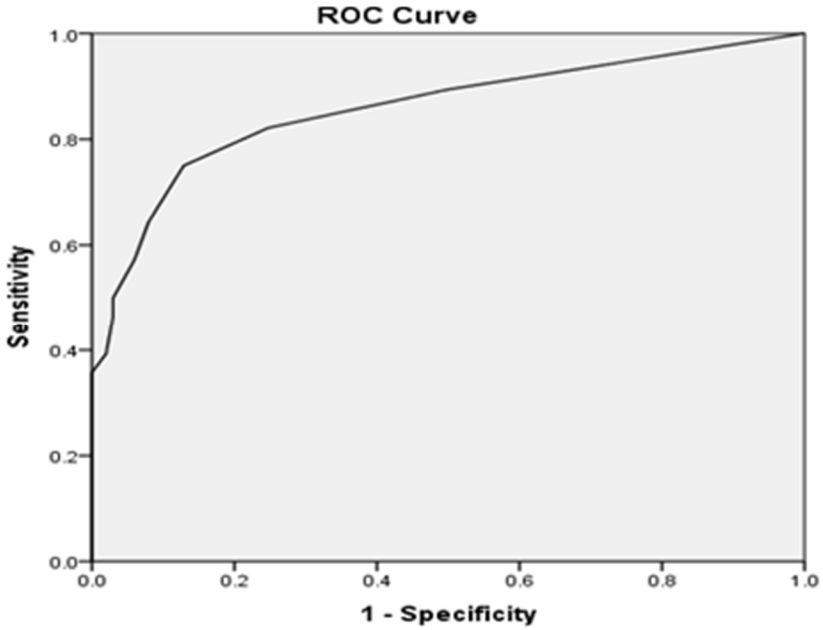

Figure 5.32 ROC Curve for Model 3 Using Random Forest Classifier

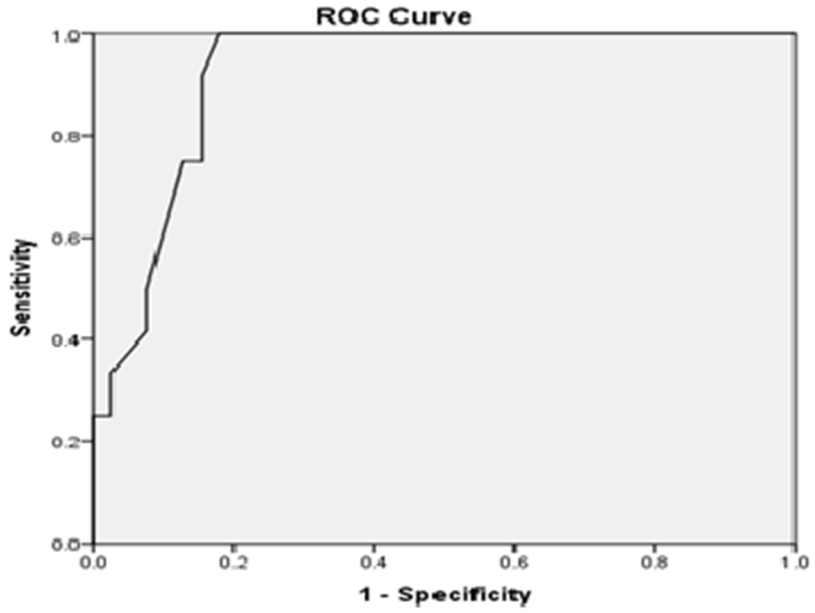

Figure 5.33 ROC Curve for Model 3 Using Random Forest Classifier

Tables 5.30 and 5.31 represent web sites prediction and 10-cross validation results for classification of web sites of same category of all the 3 models by Random Forest classifier. The observations which are made from Table 5.30 and 5.31 are summarized below:

- In model 1, 58 websites out of 62 are correctly predicted as Blog and 49 websites out of 57 are correctly predicted as Community.
- In model 2, 21 websites out of 25 are correctly predicted as TV and 19 websites out of 26 are correctly predicted as Movies.
- In model 3, 37 websites out of 39 are correctly predicted as Food & Beverages, 44 websites out of 49 are correctly predicted as Commerce and 32 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	58
	Community	49
Model 2	TV	21
	Movies	19
Model 3	Food & Beverages	37

	Travel	32
	Commerce	44

Table 5.30 Class Prediction of Website Using Random Forest Classifier for Model 1, 2 and 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.903	0.895	0.550	0.958
Model 2		0.750	0.739	0.550	0.822
Model 3	Food & Beverages	0.90	0.97		0.969
	Commerce	0.88	0.93		
	Travel	0.84	0.90		

Table 5.31 10-Cross Validation Results for Models Using Random Forest Classifier

Figures 5.34-5.35 shows the ROC Curves for Model 1, Model 2, and Model 3 using Random Forest Classifier.

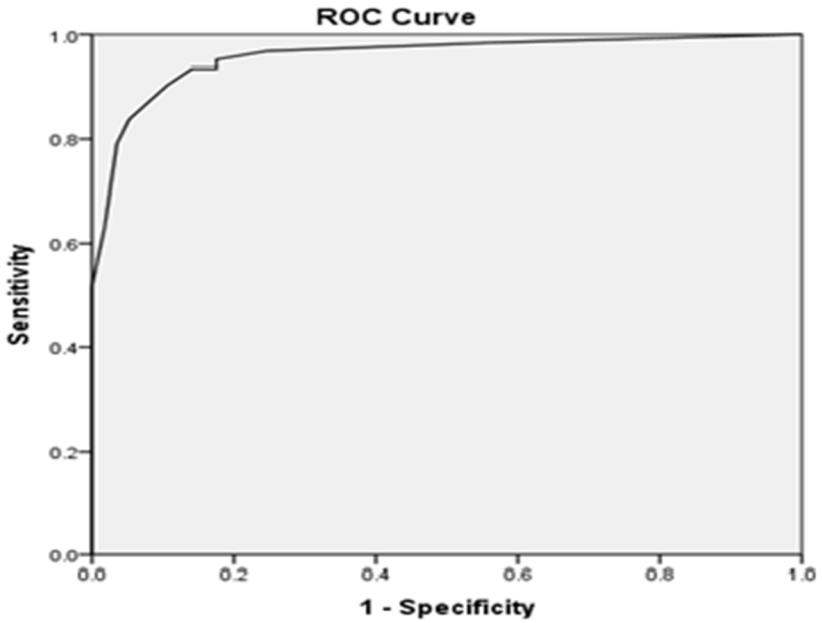

Figure 5.34 ROC Curve for Model 1 Using Random Forest Classifier

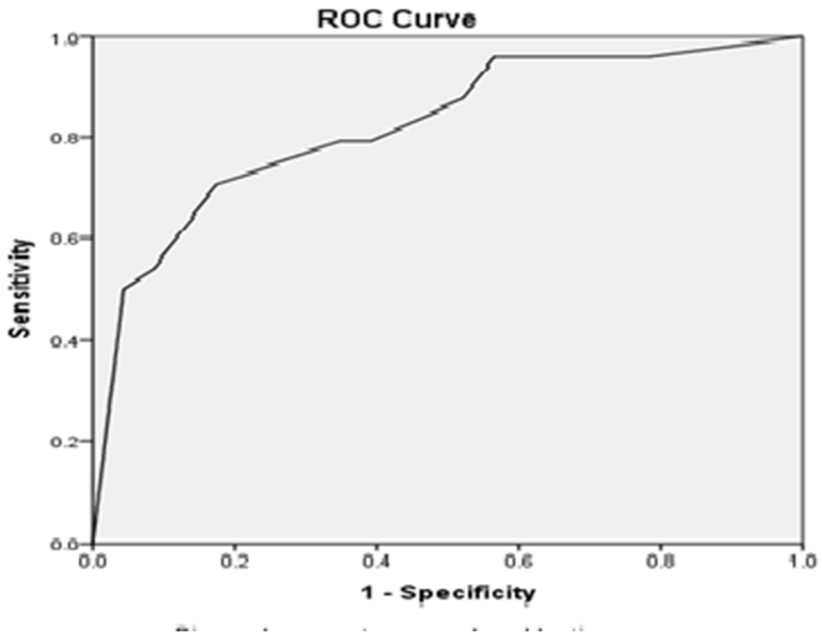

Figure 5.35 ROC Curve for Model 2 Using Random Forest Classifier

5.9 Random Tree

Tables 5.32 and 5.33 represent web sites prediction and 10-cross validation results for all the 3 models by Random Tree.

The observations which are made from Table 5.32 and 5.33 are summarized below:

- In model 1, 26websites out of 35 are correctly predicted as good and 78 websites out of 84 are correctly predicted as bad.
- In model 2, 7websites out of 12 are correctly predicted as good and 37 websites out of 39 are correctly predicted as bad.
- In model 3, 16websites out of 28 are correctly predicted as good and 91websites out of 101 are correctly predicted as bad.

Parameter	Model 1	Model 2	Model 3
Number of good websites correctly predicted	26	7	16
Number of bad websites correctly predicted	78	37	91

Figure 5.32 Goodness of Website Prediction Using Random Tree Classifier for Model 1, 2 and 3

Model	Sensitivity	Specificity	Cutoff	AUC
Model 1	0.743	0.929	0.50	0.836
Model 2	0.583	0.949	0.50	0.766
Model 3	0.571	0.901	0.50	0.736

Table 5.33: 10-Cross Validation Results for Models Using Random Tree Classifier

Figures 5.36-5.38 shows the ROC Curves for Model 1, Model 2, and Model 3 using Random Forest Classifier.

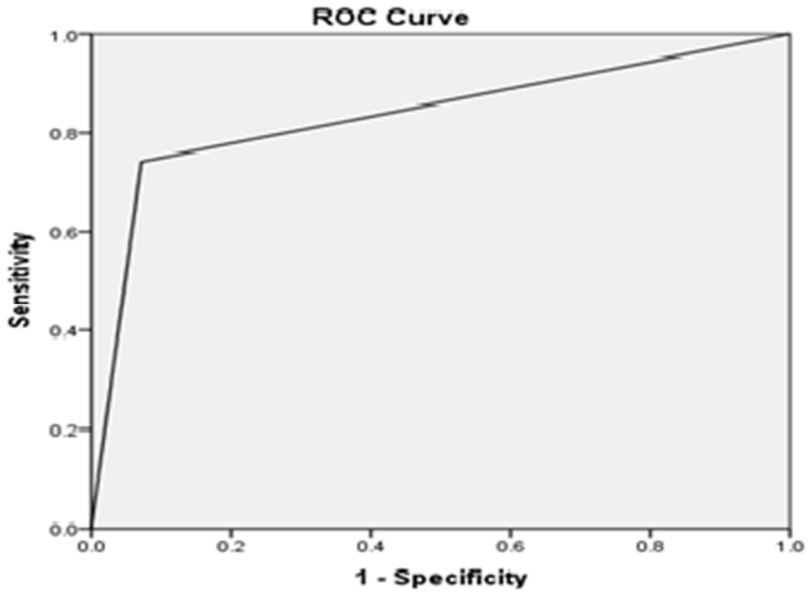

Figure 5.36 ROC Curve for Model 3 Using Random Tree Classifier for goodness

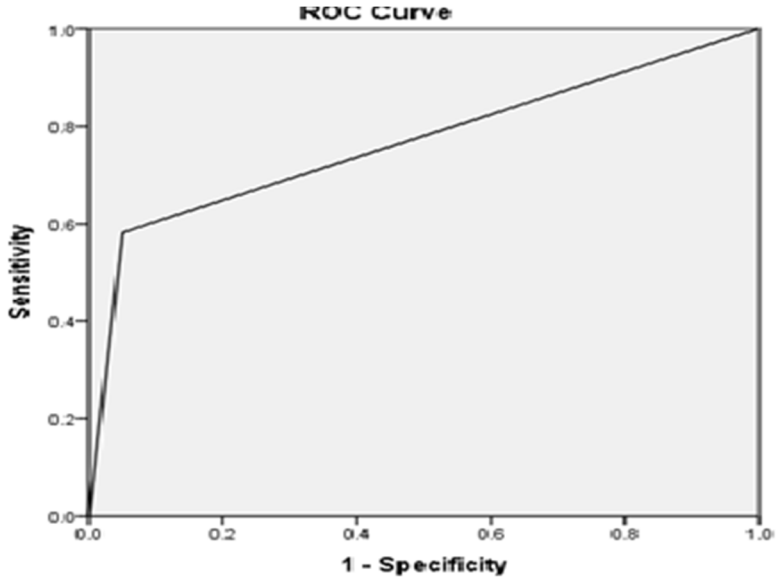

Figure 5.37 ROC Curve for Model 2 Using Random Tree Classifier

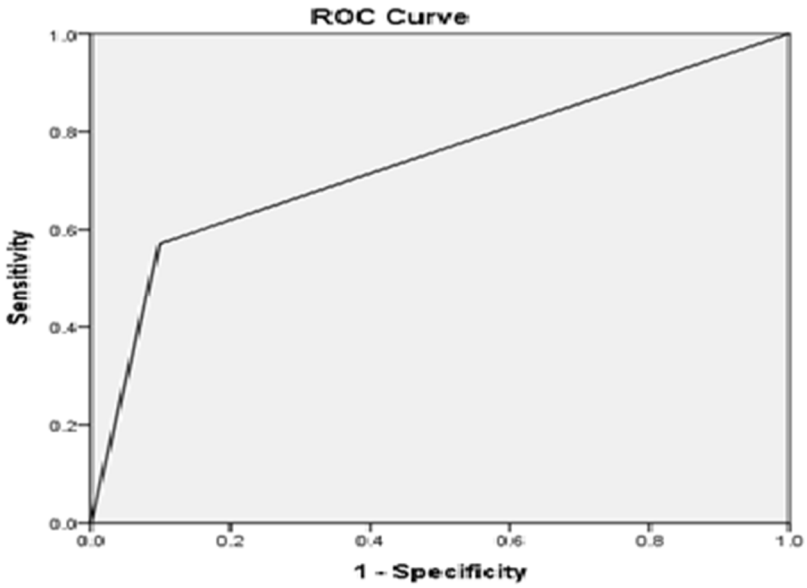

Figure 5.38 ROC Curve for Model 3 Using Random Tree Classifier

Tables 5.34 and 5.35 represent web site prediction and 10-cross validation results for classification of web site of same category of all the 3 models by Random Tree.
The observations which are made from Table 5.24 and 5.25 are summarized below:

- In model 1, 53 websites out of 62 are correctly predicted as Blog and 51 websites out of 57 are correctly predicted as Community.
- In model 2, 16 websites out of 25 are correctly predicted as TV and 17 websites out of 26 are correctly predicted as Movies.
- In model 3, 36 websites out of 39 are correctly predicted as Food & Beverages, 42 websites out of 49 are correctly predicted as Commerce and 30 websites out of 41 are correctly predicted as Travel.

Model	Websites	Data Points
Model 1	Blog	53
	Community	51
Model 2	TV	16
	Movies	17
Model 3	Food & Beverages	36
	Travel	30
	Commerce	42

Figure 5.34 Class Prediction of Website Using Random Tree Classifier for Model 1, 2 and 3

Model		Sensitivity	Specificity	Cutoff	AUC
Model 1		0.855	0.895	0.50	0.875
Model 2		0.640	0.654	0.50	0.647
Model 3	Food & Beverages	0.85	0.96		0.877
	Commerce	0.85	0.91		
	Travel	0.78	0.87		

Table 5.35 10-Cross Validation Results for Models Using Random Tree Classifier

Figures 5.39-5.40 shows the ROC Curves for Model 1, Model 2, and Model 3 using Random Tree Classifier.

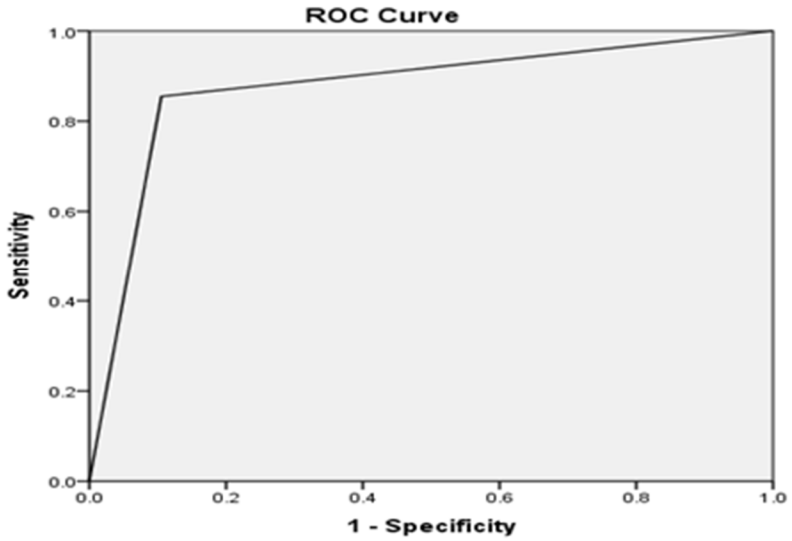

Figure 5.39 ROC Curve for Model 1 Using Random Tree Classifier

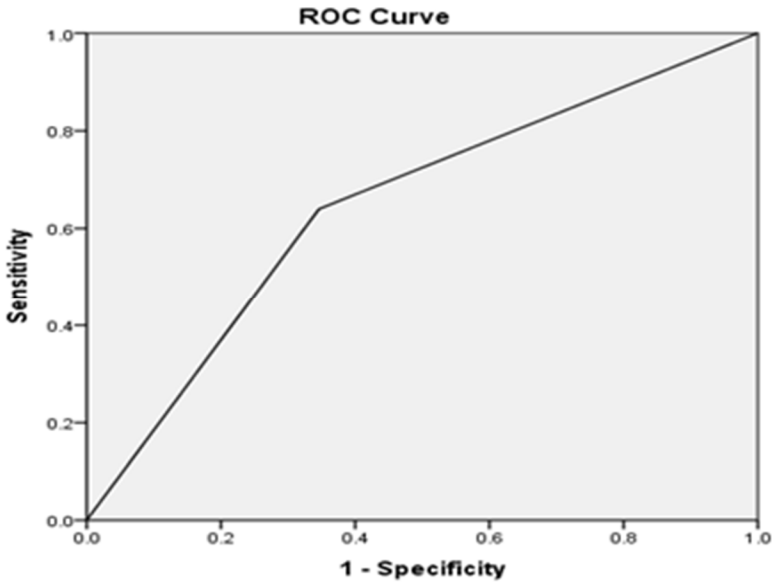

Figure 5.40 ROC Curve for Model 2 Using Random Tree Classifier

5.10 Model Evaluation

To reduce the data dimensionality we have used CFS technique [11] and the select a subset of attributes.

Model	Attribute List
Model 1	Meta Keywords, Meta Descriptor, Division, Script, Title Length.
Model 2	Paragraph, Script, Load Time.
Model 3	Meta Keywords, Meta Descriptor, Total Link, ALT words, Paragraph , Body Word Count, Title Length.

Table 5.36 Attributes Selected for Prediction

Model	Attribute List
Model 1	Meta tag, Meta Descriptor, Total Link, Image Link, Script.
Model 2	Meta tag, Total Link, Avg. word in Link, Division
Model 3	Meta tag, max keyword, Meta descriptor, Avg. word in link, ALT images, script, size, Title length.

Table 5.37 Attributes Selected for Classification

From table 5.36 we can predict that:

Meta Keywords, Meta Descriptor, Paragraph, Script, Title length are common in at most 2 models and can be consider as important factors to enhance the quality.

The cut-off points of predicted model are computed using ROC analysis and also measure sensitivity and specificity using it. Thus, accuracy of the predicted models is computed using ROC curve.

We have employed 8 machine learning techniques to evaluate their performance for predicting the quality of the websites. The AUC values of random forest for three models is greater than the AUC values of other machine learning techniques (Naïve bayes, Multilayer perception, AdaBoostMI, Bagging, Nngr, OneR, Random Tree). 0.89 is the AOC value for Model 1using

random forest which is greater than that using other techniques and similar trends for Model 2 having AOC value 0.922 and Model 3 having AOC .857. All models performed best with random forest classifier.

Similarly in case of evaluating performance of 8 machine learning techniques for predicting the class of a website within a category/model. The AUC values of random forest for three models is greater than the AUC values of other machine learning techniques (naïvebayes, Multilayer perception, AdaboostMI, bagging, Nngr, oneR, Random Tree). 0.958 is the AOC value for Model 1using random forest which is greater than that using other techniques and similar trends for Model 2 having AOC value 0.822 and Model 3 having AOC 0.969. All models performed best with random forest classifier.

Thus, we can say that that for both prediction of web site quality and class of a web site Random forest is the best model on the basis of Sensitivity, Specificity and AOC values.

CHAPTER 6: CONCLUSION AND FUTURE WORK

The goal of this research is to find the effect of web page measures on the categorization of web sites into good or bad and also their effect on classifying websites of same category/model. Different machine learning techniques have been applied for classifying and categorization of websites and also analyzed their performance.

The main contribution of this thesis is summarized as follows: First, we collected 3 sets of data of Pixel Awards for each category we created from 2006 to 2012, considering 0-level and some 1-level web pages for each website. Second, we computed 21 web page metrics for these web pages using a PYTHON based tool. Third, we applied machine learning methods such as Naïve Bayes, Random Forest, NNGR, OnerR, Bagging, AdaBoostMI, Random Tree, MLP to predict the effect of web page metrics on the classification of web pages into good or bad classes and predict the class of web site of same category. Although, this research is conducted for three categories only, in which two categories have two web sites and one have three websites, this study can be repeated for more categories. Our main results are summarized as follows:

1. The most significant metrics for categorization of web sites into good or bad for Model 1 are Meta Keywords, Meta Descriptor, Division, Script, and Title Length. Paragraph, Script, Load Time for Model 2 and Meta Keywords, Meta Descriptor, Total Link, ALT words, Paragraph , Body Word Count, Title Length for Model 3. This signifies that for different categories, the various attributes were included as important metrics for web site development.
2. The most significant metrics for categorization of web sites for Model 1 are Meta tag, Meta Descriptor, Total Link, Image Link, Script. Meta tag, Total Link, Avg. word in Link, Division for Model 2 and Meta tag, max keyword, Meta descriptor, Avg. word in link, ALT images, script, size, and Title length for Model 3. This signifies that for different categories, the various attributes were included as important metrics for web site development.
3. Random Forest outperformed the other models although all models predicted good area under ROC analysis.

6.1 Application of Work

In this work we established two relationships. Firstly, web metrics and quality of the website. Secondly, web metrics and class of the web site. In order to establish these relationships we have collected data sets from Pixel awards website which honors web site on different criteria's. This work will provide web designers with important metrics that can be used in web site design and also the model for verifying the quality of website. Website quality can easily be estimated by computing the values of values of web metrics and then applying the Random Forest model which is more effective than all the models. The websites which are classified as bad need more attention.

We also identify the class of website within same categories of websites. This will help web designers and researchers to be careful with web metrics values which are helpful in predicting the class so that websites can be distinguishable.

6.2 Future Work

This study confirms that web metrics can be helpful in predicting the goodness and class of the websites of same category with the help of machine learning methods. In future we can do similar study on different data set and also consider more web page metrics. We plan to carry our research for all the levels of web pages in the website.

References

[1] C. Calero, J. Ruiz, and M. Piattini, "Classifying web metrics using the web quality model", Online Information Review, vol. 29, no. 3, pp. 227- 248, Emerald Group Publishing, 2005.

[2] I. H. Witten, E. Frank, and M.A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. San Francisco: Morgan Kaufmann Publishers, 2011.

[3] ISO 9126-1, "Software Engineering-Product Quality - Part 1: Quality Model",October 2001.

[4] J. Jugini, S.Laskowski, "Design of a File Format for Logging Website Interaction", NIST Special Publication, 2001.

[5] J. Scholtz, S. Laskowski and L. Downey, "Developing usability tools and techniques for designing and testing web sites," In Proceedings of the 4th Conference on Human Factors & the Web, 1998.

[6] K. M. Khan, "Assessing Quality of Web Based System," IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, 2008.

[7] K. P. Murphy, "Naive Bayes Classifiers," Technical Report, October 2006.

[8] L. Breiman, "Bagging Predictors," In Machine Learning Journal, vol. 26, no. 2, pp. 123-140, ACM Digital Library, 1996.

[9] L. Mich, M. Franch, and L. Gaio, "Evaluating and Designing the Quality of Web Sites", IEEE MultiMedia, vol. 10, no. 1, pp. 34-43, IEEE Computer Society, 2003.

[10] L. Olsina and G. Rossi, "Measuring Web Application Quality with WebQEM", IEEE MultiMedia, vol. 9, no. 4, pp. 20-29, IEEE Computer Society, 2002.

[11] M. A. Hall, "Correlation-based Feature Subset Selection for Machine Learning," PhD dissertation, University of Waikato, Computer Science Dept., Waikato, N.Z., 1999.

[12] M. Marsico, S. Levialdi "Evaluating web sites: exploiting users expectations", International Journal of Human-Computer Studies, pp. 381-416, ELSEVIER, 2003.

[13] M. Nazzal Jamal, M. El-Emary Ibrahim, A. Najim Salam, "Multilayer Perceptron Neural Network for Analyzing the Properties of Jordan oil Shale", World Applied Sciences Journal 5, vol. 5, no. 5, pp. 546-552, IDOSI Publications, 2008.

[14] M. Stone, “Cross-validatory choice and assessment of statistical predictions,” In Journal of the Royal Statistical Society. Series B (Methodological), vol. 36, no. 2, pp. 111–147, Royal Statistical Society, 1974.

[15] M. Y. Ivory, R.Sinha and A. Marti Hearst, “Empirically Validated Web Page Design Metrics”, vol. 15, no. 5, ACM SIGCHI, 2001.

[16] M. Y. Ivory, R. Sinha and A. Marti Hearst, “Preliminary findings on quantitative measures for distinguishing highly rated information-centric web pages,” In Proceedings of the 6th Conference on Human Factors and the Web, IEEE Internet Computing, 2002.

[17] M. Zorman, V. Podgorelec, P. Kokol, and S. H. Babic, “Using machine learning techniques for automatic evaluation of Websites,” In Proceedings of the Third International Conference on Computational Intelligence and Multimedia Applications ICCIMA, pp. 169-173, IEEE Computer Society Press, 1999.

[18] National institute of standards and technology, IEEE Std 2001-1999, from Web Site: http://zing.ncsl.nist.gov/WebTools/WebSAT/ieee guide.html.

[19] National institute of standards and technology, Web Metrics Test bed: Technical Overview, from Web Site: http://zing.ncsl.nist.gov WebTools/tech.html.

[20] O. Signore, "A Comprehensive Model for Web Sites Quality", in proceedings of the Seventh IEEE International Symposium on Web Site Evolution, pp. 30-38, IEEE Society, 2005.

[21] P. Warren, C. Gaskell, and C. Boldyreff, "Preparing the Ground for Website Metrics Research," Proceedings of the 3rd International Workshop on Web Site Evolution, IEEE Computer Press, 2001.

[22] Pixel Awards, Web Awards Competition (2006), from Web Site, http://www.pixelawards.com/

[23] Rio. Americo “Websites Quality: Does it depend on the application Domain?” International Conference on the quality of Information and Communications Technology, 2011.

[24] S. Koukoulas and G. A. Blackburn, "Introducing new indices for accuracy evaluation of classified images representing semi-natural woodland environments," Photogrammetric Engineering & Remote Sensing , vol. 67, no. 4, pp. 499-510, American society for photogrammetry and remote sensing 2001.

[25] Sukhpal Kaur ,"An Automated Tool For Web site Evaluation", International Journal of Computer Science and Information Technologies, vol. 3, no. 3, 2012.

[26] V. M. R Penichet, C.Calero, M.D.Lozano, M.Piattini," Using WQM for classifying usability metrics", IADIS International Conference, 2006

[27] WAI, Checklist of Checkpoints for Web Content Accessibility Guidelines 1.0, from Web Site: http://www.w3.org/TR/WCAG10/full-checklist.html.

[28] WAI, Web Content Accessibility Guidelines, W3C Recommendation, from Web Site: http://www.w3.org/TR/WCAG10/.

[29] Watchfire, "Site Quality and Accessibility, from Web Site: http://www.watchfire.com products/webxm/siteusability.aspx.

[30] Web Scraping, from Web Site, http://en.wikipedia.org/wiki/Web_scraping

[31] WebTango, Automating Web Site Usability Evaluation, from Web Site http://webtango.berkeley.edu/.

[32] Weka3: Data Mining Software in Java. Available from http://www.cs.waikato.ac.nz/ml/ weka/.

[33] Wen Zhu, Nancy Zeng, NingWang , "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations", Health Care and Life Sciences, Northeast SAS users group , 2011.

[34] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," In Journal of computer and system sciences, pp. 119-139, 1997.

[35] Yogesh Singh, Ruchika Malhotra, Poonam Gupta,"Empirical Validation of Web Metrics for Improving the Quality of Web Page", International Journal of Advanced Computer Science and Applications, vol. 2, no. 5, 2011.