المجلة الإلكترونية الشاملة متعددة التخصصات
العدد السادس والعشرون شهر(7) 2020

Performance Evaluation of Websites Using Machine Learning

Mr. Mohammad Ghattas

Ph.D. Badie Sartawi

College of Computer Science & Technology || Al-Quds University || Palestine

E-mail: m.ghattas@paluniv.edu.ps

Abstract: The web is playing a main role in various application domains such as business, education, engineering, and entertainment. As a result, there are increasing interests in designing and developing an efficient website to deliver a high degree of performance. Therefore, automated support for web designers is becoming more important to evaluate websites performance. Hence, many of the previous studies tried to measure websites performance by developing a static model and it is unless used for more domain.

The aims of this study are: (i) to explore the most effective metrics that almost all affect website performance; (ii) propose a dynamic model for performance evaluation of websites by using machine learning that called is PEML; and (iii) to support webmaster and decision-makers to understand what improvements are needed to enhance the performance and therefore the final relative weights of metrics in the level of the hierarchy.

This research proposes a dynamic model to performance evaluation of websites using machine learning method by applied two regression methods experiments namely, multiple linear regression and support vector machine regression on the same dataset that collected, to take the best performance of regression methods to generate weight for every metric and then developing a new dynamic model to evaluate websites performance.

Keywords: website performance, regression, machine learning, web metrics, support vector machine, multiple linear regression, evaluation, Rapid Miner.

1.Introduction

Lately, we have been become witness to an important alteration of our lives to a worldwide with the incipience of the web era. The web is an increasingly more vital asset in many sides of life: government, education, commerce and more [3]. Hence, Websites are a key element in obtaining the right information about the institutions. However, when it comes to a huge number of synchronous users these websites performance decreases considerably.

Utilizing the web devices many institutions become been able to raise their being customer-focused and their attributes of services and products. The analysis of the web site is currently thought to be an essential facet of attracting customers' attention [3]. In this study, it is logical to explore metrics into measure the performance of websites, whether to study the communication efficiency that they represent or in order to build useful appraisal metrics.

As result of the above requirements, it is important to provide a method to evaluate the performance quality of websites, which include various technological and logical factors. Each definition of performance quality from literature leads to lists of criteria about what constitutes a good quality website and how to measure the performance [8]. Therefore, it is important to build a model into evaluation websites performance, thus ensuring the development of modern websites and keeping abreast of modern technology.

This study employed machine learning to build a mathematical model approach to evaluating the performance quality of websites. In this study, we suggest a method based on appropriate metrics for evaluating websites performance.

This study proposed to build an understandable and applicable dynamic model for evaluating websites performance by using previous studies as a case study. By establishing a practical model, it is expected that organizations can better understand whether a given website can meet the expectations of its users, they serve in order to grow their satisfaction level.

2.Background

In this study, we covered basic concepts of performance evaluation of websites and more specifically regression techniques.

2.1 Study Terminologies

In this section will describe the web metrics used in performance evaluation of websites:

- Response Time: A Website server should respond to a browser request within certain parameters [24].
- Load Time: It is used to calculate the time required to load a page and its graphics [24].
- Markup Validation: It is utilized to assess and calculate the number of HTML errors, which exist on the website, such as orphan codes, coding errors, missing tags and etc [24].
- Broken Link: Broken links always reduces the quality of the website. Websites have internal or external links. A visitor expects the links to be valid, loads successfully to the clicked page [24].
- Design Optimization: The scripts, HTML or CSS codes optimized for quicker loading. The optimization also decreases the number of website parts such as images, scripts, HTML, CSS codes or video [24].
- Page Size: The size of the Web pages in the Website [25].
- No. of Request: The number of request/response between a client and a host [25].

2.2 Regression Techniques

In this section, we would like to explain the techniques employed in this study.

2.2.1 The Linear Regression

The linear regression type describes the output of website's performance y (a scalar) as an affine combination of the input metrics $x_1,x_2,...,x_p$ (each a scalar) plus a noise term ε, $y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p + \varepsilon$ [35]. We refer to the coefficients $\beta_0,\beta_1,...\beta_p$ as the weight for each metric in the model, and we refer to β_0 as the intercept term. The noising term ε for non-systematic, i.e., random, errors between the data and the model [35]. Hence, the linear regression model can namely be used for, at least, two several purposes: to describe relationships in the dataset by interpreting the weight to metrics $\beta= [\beta_0\ \beta_1\ ...\ \beta_p]^T$, and to predict future website performance by metrics [35].

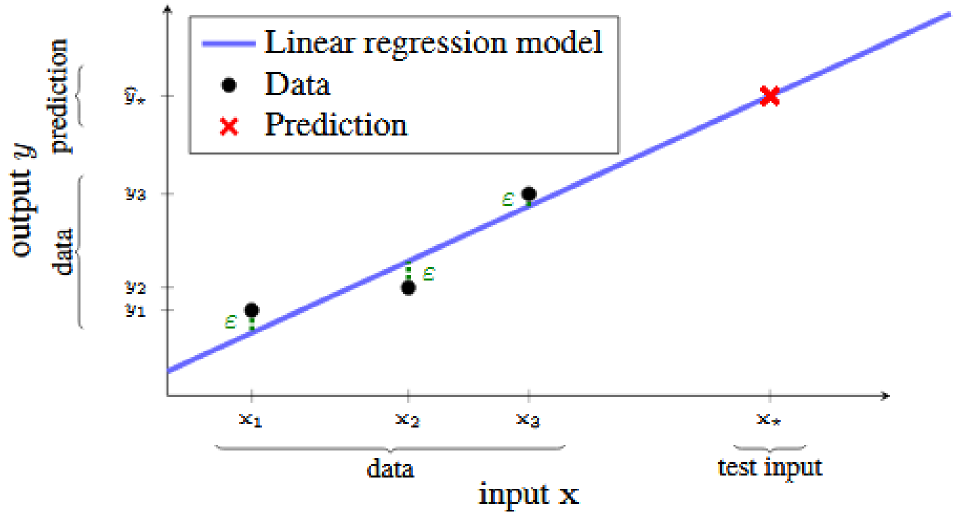


Figure 2.2 : linear regression model.[35]

To use the linear regression model, we first necessity to learn the unknown weight to each metric β0,β1,...,βp from a training dataset T. The training data consists of n samples of the output variable y, we call them yi (i= 1,...,n), and therefore the corresponding n samples xi(i= 1,...,n) (each a column vector). We write the dataset within the matrix form [35]:

$$\mathbf{X} = \begin{bmatrix} 1 & -\mathbf{x}_1^{\mathsf{T}}- \\ 1 & -\mathbf{x}_2^{\mathsf{T}}- \\ \vdots & \vdots \\ 1 & -\mathbf{x}_n^{\mathsf{T}}- \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \text{where each } \mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}.$$

[35]

Hence, X may be a ×(p+ 1) matrix, and website performance (y) an n dimensional vector. The first column of X, with only ones, corresponds to the intercept term β0 in the linear regression model. If we also stack the unknown weight to each metric β0,β1,...,βp into a (p+ 1) vector [35].

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix},$$

[35]

We can express the linear regression model by two equations:
Linear regression for single metric:

$$y = X\beta + \varepsilon, \text{ [35]}$$

Multiple Linear regression for multiple metrics:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_p x_{ip} + \varepsilon_i \textbf{ for } \boldsymbol{i} \textbf{ = 1,2, ... } \boldsymbol{n} \quad \text{[35]}$$

2.2.2 Support Vector Machine Regression

Support vector machine (SVM) could also be a standard machine learning tool for classification and regression, 1st known by Vladimir Vapnik and his colleagues in 1992 [18]. Support Vector Machine may be used as a regression method, preserve all the most features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses similar basics as the SVM for classification, with only a few minor differences because the output is an actual number it becomes very difficult to predict the information at hand, which has infinite possibilities. As shown in figure 2.3 in the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which might have already requested from the problem. But besides this fact, there's also a more complicated reason, the algorithm is more complicated therefore to be taken into consideration. However, the main idea is usually the same: to minimize error, individualizing the hyper plane which maximizes the margin, keeping in mind that a part of the error is tolerated [16].

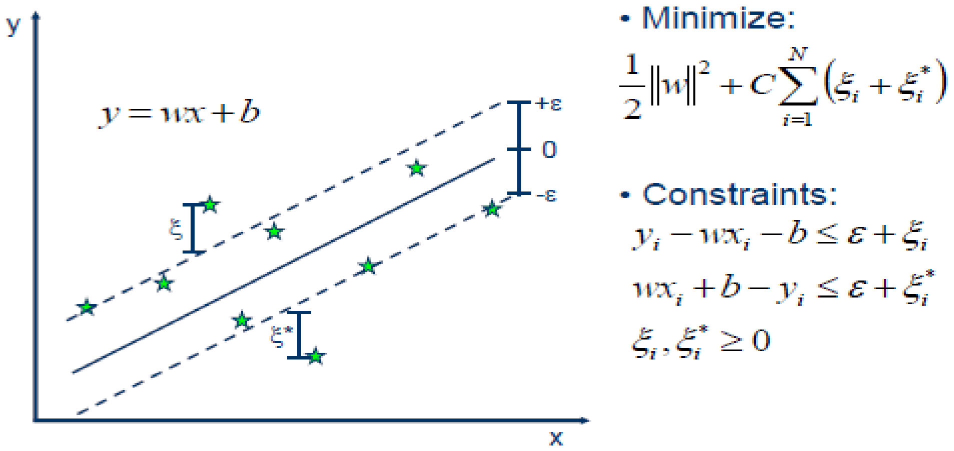


Figure 2.3 : Linear SVR [16]

We can express the linear SVR:

$$y = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) \cdot \langle x_i, x \rangle + b$$

[16]

3.Literature Review

3.1 Website Evaluation Studies

Lately, there is no model for evaluating airline websites, and also the existing methods do not enough understanding for airlines' proprietors to ascertain whether their websites meet the recognized guidelines from the metric of website performance. In this study, researchers have suggested a hybrid model to combine Entropy Weight Method and Grey Relational Analysis for determining and evaluating the performance of airline websites with a sample of eleven airline websites. and they have assessed many metrics of performance and each metric include design optimization, load time, response time, mark up and broken links ..etc and these metrics were measured by using on-line diagnostic tools. Vatansever et al. (2017) [3]

Kaur et al. (2016) present an empirical performance analysis of universities website that usability is currently important by website developers who will develop websites and also the performance of a website are often an important issue to its success. In this study focused methodology has been made to find all possible metrics in the website design. The researchers evaluated and compared the automated testing tools to determine their performance, speed, number of requests, load time, page size, SEO, mobile and security for university websites of Punjab [1].

Harshan et al. (2016) the active presence of library websites on the internet is becoming a hallmark of academic networks obligation to facilitate the community to access the knowledge depositories about the world. In this research the model was developed on the base of a conceptual framework, which consisted of eight quantitative performance attributes identified from an extensive literature review also as discussions with experts which include the design optimization, load time, page size, number of items, page speed, broken links, response time and mark-up validation . This study suggested a model by using AHP approach to gauge the performance of library websites. Finally, the model can be used as a gauge website design guideline that helps to develop usable websites across library domains [2].

Devi et al. (2016) the main aim of this paper is to design the website evaluation framework for academic websites. The quality of an internet site makes an internet site profitable, easy and accessible, and it conjointly offers helpful and reliable information, providing good design and visual look to satisfy the user's needs and expectations. The researchers design new evaluation framework based on the main quality determinants of the chosen base model (ISO 9126-1) and rearranged to group factors with an equivalent semantic meaning in one category by removing existing repetitions and different factor names. thus, This model to evaluate the quality of websites using different quality assessment techniques starting in the earlier stages of the website design, during the intermediate design stages and the deployment stages [5].

Khan et al. (2013) this study aimed to check the Asian airline's website quality via online web diagnostic tools. The researchers used the analytical hierarchy process which generates the weights for each metrics and makes it easy to judge the better results to evaluate the website performance of each airline in Malaysia. The researchers used the metrics include Load time page size, no. of items, response time, page speed, availability, broken links, response time, mark up validation, design optimization, page rank and traffic to make the better performance website and to provide a future approach for customer satisfaction with the websites [7].

There an enormous growth of web applications and also the web applications are not simply static, document-oriented but dynamic applications with several technologies to form complex, heterogeneous web systems and applications. Many of the current website evaluation techniques and criteria for evaluating web application are unable to assess the performance and quality of web application, and most of them focus purely on usability and accessibility. And therefore, the researches presented an analysis methodology consistent with measurement approaches used in the performance evaluation domain and guideline review approaches used in the quality evaluation domain and they propose an automatic tool to calculate the quality and aesthetic factors of web application. Kulkarni et al. (2012) [6].

Dominic et al. (2011) the researchers suggested a methodology for choosing and evaluating the best e-government website based on many metrics of website performance. they used a group of metrics namely load time, response time, page rank, the frequency of update, traffic, design optimization, page size, number of the item, accessibility error, markup validation, and broken link. Thus, they proposed some methodologies for determining and measuring the best e-government sites based on many metrics of website performance, consisting of analytical hierarchy process (AHP), fuzzy analytical hierarchy process (FAHP), linear weightage model (LWM) and also one new hybrid model (NHM). This NHM has been implemented using LWM and FAHP to generate the weights for the metric which are much better and guaranteed more fairly preference of metric. and then they employ a hybrid model among linear weightage model and fuzzy analytical hierarchy process approach for the website. Then the results of this study confirmed that most Asian websites fail in performance and quality criteria. By applying the hybrid model approach [8].

Jati et al. (2011) this study applies the test to evaluate the e-government of website performance for some Asian countries by using web diagnostic tools online. they suggested a methodology for choosing and evaluating the better e-government website supported several metrics of website performance. They used the PROMETHEE II technique to get the perfect ranking of the e-government websites. Analytical hierarchy process (AHP) has been proposed for determining the better website to support researcher into the decision-making activity, that aims to determine the better website between a grouping of e-government website. The final score obtains for each website across each metric is calculated by using multiplying the weight of each metric with the weight of each website. The website which has got the highest score is suggested as the best website and decision maker may consider that one as the best decision choice. Results of the e government websites performance based on load time, response time, page rank, the frequency of update, traffic, design optimization, size, number of items, accessibility error, mark-up validation, and broken link [9].

Islam et al. (2011) the presented study concentrate both the user's point of view and applied automated tools to evaluate the performance of some academic websites in Bangladesh by using two on-line automated tools, such as web page analyser and HTML toolbox were used along with a questionnaire directed to users of that websites. They used Webpage Analyzer to test the internal metrics of the websites including the total no of images, HTML page sizes, the total no of HTML files and other relevant items of websites. The researchers recommended that these websites ought to be designed supported further content; incorporate many academic data, and priority ought to run for coming up with easy websites [4].

3.2 Performance Standard

Every webpage design has its own features and these features have disadvantage and benefits. There is a mechanism for measuring the effects of the webpage component towards the performance and quality of the website. This mechanism measuring time and the size, component needed by the user in order to downloading a website. The main factors that will affect download time are page size (bytes), number and types of component, number of a server from the accessed web. Research makes by IBM may be used as a regular for measuring performance (Amerson et al., 2001) [33].

Table 3.2 describes all of the metric and performance standards that should be fulfilled by a website to be a good quality website. Tested metrics consist of: webpage loading time, average server response time, number of item per page and webpage size in bytes.

Standard international download time in order to this performance can be used as a ref to categories the tested webpage. Automation in testing for website performance is a new opportunity and a new method, and should be applied for evaluating the performance of the website. For leveraging the effectiveness of continuous performance enhancement, the developer community has been aggressive in attaining TQM strategies by implementing ISO 9001:2000 kind (Sakthivel et al., 2007) [34].

Evaluate Metric	**Performance standard**
Average server response time	< 0.5 second
Number of item per page	< 20 item
Webpage loading time	< 30 second
Webpage size in byte	< 64 Kbytes

Source: Amerson et al. (2001)

Broken links can give a bad effect for the truthfulness of a website. truthfulness is very important in the World Wide Web, because transaction between customer and seller is not on the spot and the risk of fraud is several times higher. The customer would truthfulness choose to buy from a website that looks professional.

4.Proposed Method

In this study proposes a new approach for evaluating the websites performance using machine learning. As shown in figure 4.1 to implement this research.

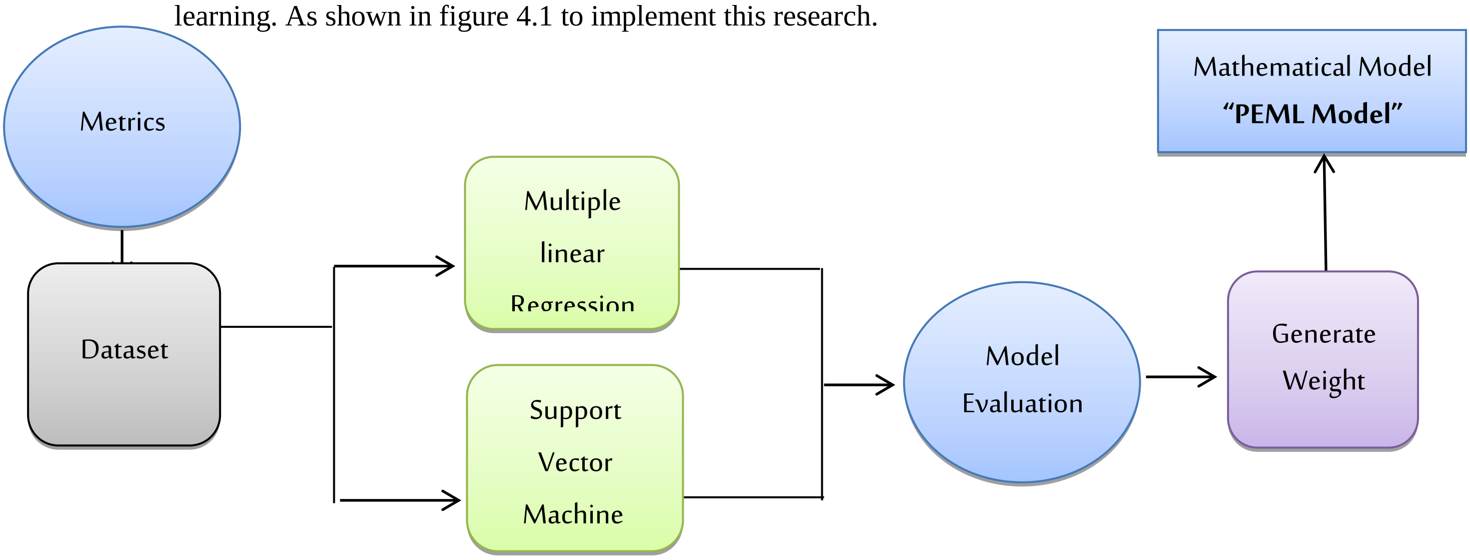


Figure 4.1: The steps of implement the methodology model

4.1 Identification of metrics that affect the performance of the website

There is a large number of metrics that affect websites performance; in our study, we have selected all metrics from the previous study and make Online Questionnaire to find out the local experts opinion for asking them "What are the best metrics that affect websites performance?". Thus, we take the metrics selected was good and excellent from the online questionnaire .

4.2 Experiments Setup

In this section, we have a description of the experimental environment of the experiments on a machine with properties that is Intel (R) Core (TM) i5-4210U CPU @ 1.70 GHz (4CPU), 4.00 GB

RAM, 500 GB hard disk drive and Windows 7, the 64-bit operating system installed and determined the experimental tools that are used in the experiments, finally determine the setting of the experiments in the research.

4.3 Collection of data and creating of the dataset

In the study, 26 metrics were identified for evaluating the performance of the website primarily. The number of metrics was reduced to 11 metrics by 4 experts. The experts were computer engineers and experienced in software, web design, and web masters; as shown in Table 4.3 the metrics were used in this study.

Table 4.3: Website Performance Measurement Metric

Web Metric
Response Time
Load Time
Broken Links
No. of Request
page Size
mark-up validation
design optimization
Page Speed
Start time render
Connection time
DNS lookup

Using website diagnostic tools for collecting data for all metrics, and creating the dataset will take place. All of the data for this research was taken using PC with specification: Intel(R) Core(TM) i5-4210 CPU @ 1.70GHz, using Local Area Network internet connection with 8 Mb/s internet speeds; Table 4.4 shows website diagnostic tools .

We collected data for 174 random websites in different domains, such as: (Education, health, government, and business) to the dataset considered for analysis.

After that, We used SPSS statistical tool to find the most influence metric to enhancing the website performance among all the collected metrics and rule out every metric unless has no affect website performance.

4.4 Determine machine learning method

Machine learning methods are the backbone of our approach in the research where used to generate the weight of the metric. Hence, the task of regression and classification is to predict website performance (y) based on metrics (X), based on the dataset :

If Y is numerical, the task is called regression.

If Y is nominal, the task is called classification.[17]

There are various algorithms for regression methods. Hence, we applied linear regression and support vector machine regression that depends on the volume and structure of the dataset. In this study, we have two different algorithms for conducting the experiments on the same dataset, namely, linear regression and support vector machine to explains the comparison of the models that give the best results in terms of the Correlation coefficient in the performance evaluation metric.

4.4.1 Linear regression model

The technique is a statistical approach to construct a linear model predicting the value of the metric while knowing the values of the other metrics. It employs the least mean square method in order to adjust the parameters of the linear model/function [12]. Figure 4.5 shows the main process of linear regression method that we applied on the experiment; this method is implemented via Rapid Miner tools:

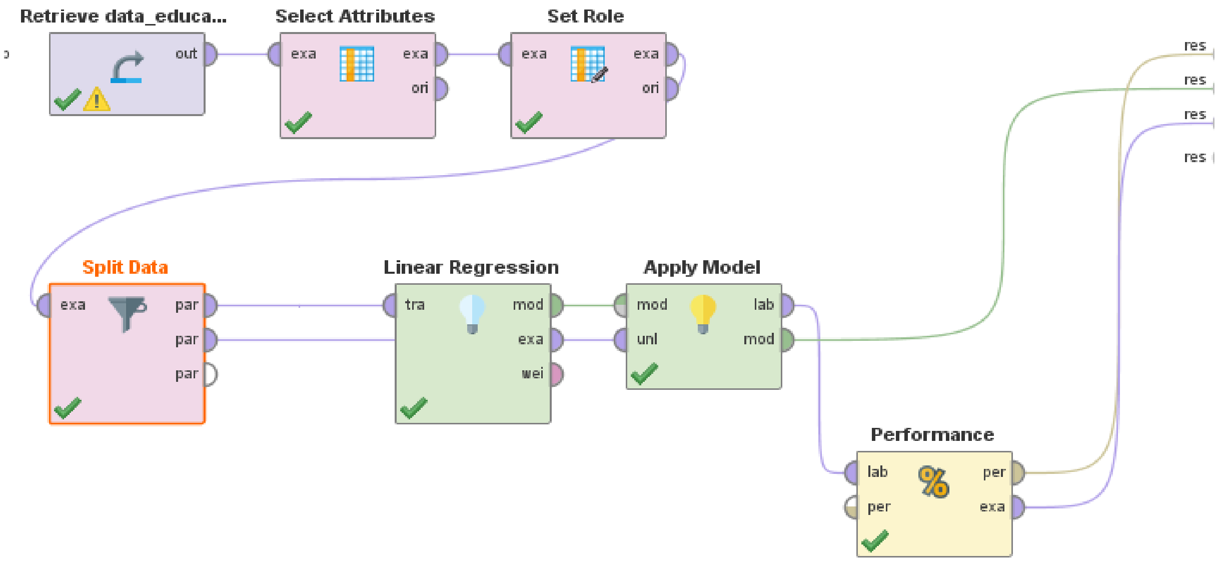


Figure 4.5: The main process of linear regression method in Rapid Miner tool

4.4.2 Support vector machine regression model

The algorithm builds support vectors in a high-dimensional feature area. Then, hyperplane with the maximal margin is constructed. The kernel function is used to transform the data, whose augments the dimensionality of the data. This augmentation stimulates that the data can be separated with a hyperplane with much higher probability, and establish a minimal prediction probability error measure [12]. Figure 4.6 shows the main process of support vector machine method that we applied on the experiment; this method is implemented via RapidMiner tools:

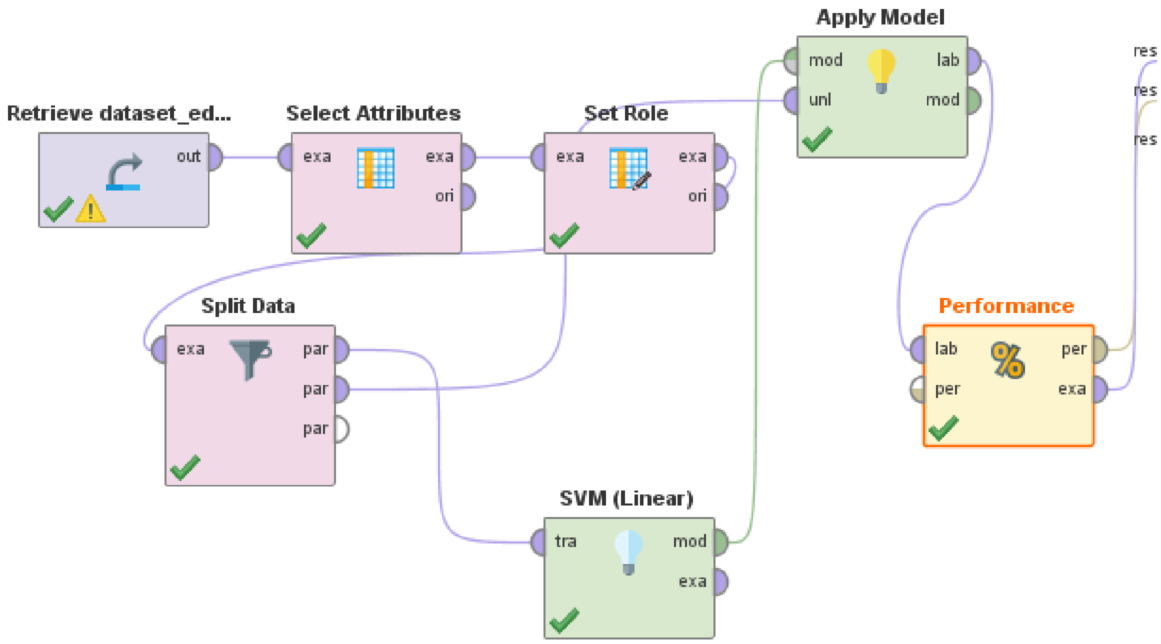


Figure 4.6: The main process of support vector machine method in RapidMiner tool

4.5 Calculating weights for every metric

In this step, we generated a weight for every metric by using regression methods. Moreover, after generated weight to every metric, we can arrange the most affect metrics on the website's performance on the level of the hierarchy as shown in figure 4.7.

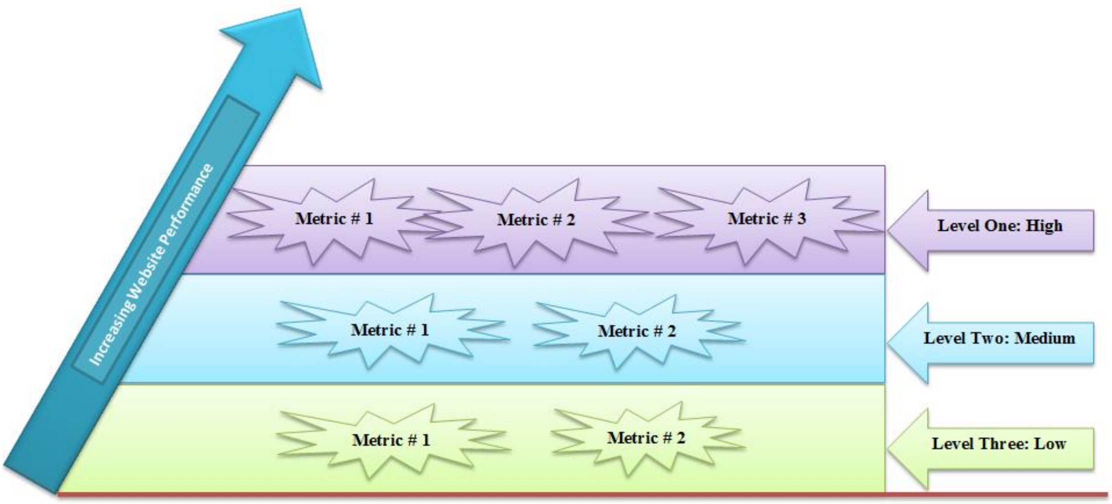


Figure 4.7: The level of the hierarchy of web metrics

4.6 Model evaluation

The weights of metrics were calculated by using the regression Methods and then evaluate the performance of the websites using mathematical model. Hence, after building different regression models namely, linear regression model and support vector machine regression model. There are criteria whereby they can be evaluated and compared to take the best performance among the models.

- Average absolute error: it represents the average absolute deviation of the prediction from the actual value (it is expressed in website performance) [11].

- Average relative error: it is calculated as the average of the prediction that sees in the numerator the error in absolute value among the predicted values and the respective real values and the denominator the real value (it is expressed in percentage) [11].

- Correlation: it provides a percentage correlation value among predicted and actual values in a range between 0 and 100 where 100 represents the perfect forecast of data by the model (it is expressed in percentage) [11].

5.Model Analysis and Evaluation

In this section, we present the results of research experiments that presented in the previous section and finally we discuss these results.

5.1 Model Analysis

In this section are discussed experimental analysis by using SPSS tools and RapidMiner, in order to get most affected metrics and to take the best algorithm performance.

5.1.1 Model Analysis Using SPSS Tool

In order to determine the most influential metric on the performance of websites from the dataset collected, as mentioned in section 4.3, we run SPSS on the same dataset. Thus, the number of metrics was reduced to 7 metrics were the most affect website performance based on significant in coefficient table. Table 5.1 shows the coefficient table after performing the statistical analysis into the SPSS tool.

Table 5.1: Coefficients of used metrics

Coefficients[a]

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	t	Sig.
(Constant)	79.006	1.965		40.201	.000
Broken_link	-.109	.014	-.445	-7.555	.002
page_size	-.419	.121	-.194	-3.464	.001
response_time	-.563	.318	-.099	-1.769	.048
No_of_Request	-.056	.016	-.209	-3.545	.001
Optimization	.054	.022	.132	2.398	.018
load_time	-.175	.075	-.130	-2.319	.022
Markup_validation	-.012	.006	-.104	-2.022	.045

Result of above coefficient table :

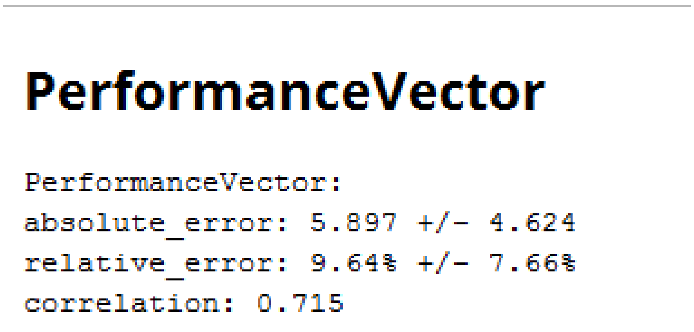
Multiple regression were run to predict performance from metrics. These metrics statistically significantly predicted performance, $p < .05$. Hence, we retain to those metrics whose significant level is < 0.05 and remove those metrics whose significance level is > 0.05 from the model.

5.1.2 Model Analysis Using Machine Learning

After determining the metrics that significantly impact the performance of websites from the dataset as mentioned in section 5.1.1. Therefore, we have used various regression methods namely linear regression and support vector machine regression on the same dataset as mentioned in section 4.4. The experiments aimed to compare machine learning algorithms to take the best algorithm to create a model for the evaluation of the website performance.

5.1.2.1 Linear Regression Results and Analysis

In order to evaluate the performance of the linear regression model by using Rapidminer tool, we run an experiment on the dataset. As Shown in figure 5.1 to understand how the prediction is successful, correlation, average absolute error, and average relative error as mentioned in section 4.4.1.



PerformanceVector

```
PerformanceVector:
absolute_error: 5.897 +/- 4.624
relative_error: 9.64% +/- 7.66%
correlation: 0.715
```

Figure 5.1: Performance of model by LR

Figure 5.2 the plot of prediction of performance of the websites versus the linear line using the linear regression method, The straight line in blue represents the real values of the performance of websites, and the red line indicates the deviation in the prediction of linear regression.

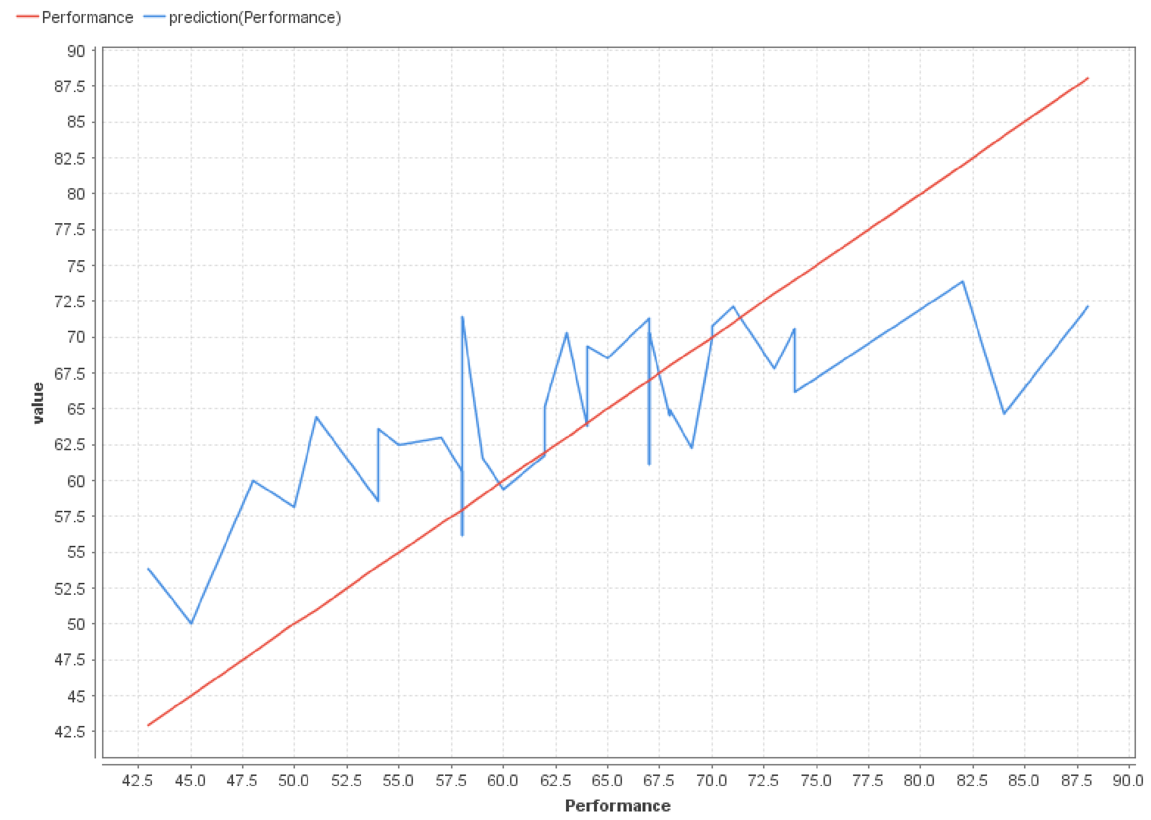


Figure 5.2 the plot of prediction of performance of the websites versus the linear line using the linear regression method

5.1.2.2 Support Vector Machine Results and Analysis

In order to evaluate the performance of the support vector machine model by using Rapidminer tool, we run an experiment on the same dataset. As Shown in figure 5.3 to understand how the prediction is successful, correlation, average absolute error, and average relative error as mentioned in section 4.4.2.

PerformanceVector

```
PerformanceVector:
absolute_error: 6.993 +/- 5.277
relative_error: 11.72% +/- 9.75%
correlation: 0.652
```

Figure 5.3: Performance of model by SVM

5.2 Model Evaluation

The experiments aimed to compare machine learning algorithms to create a model for the evaluation of the website's performance. In order to evaluate the performance of our model. We take the best algorithm based on correlation, average absolute error, and average relative error as mentioned in section 4.6.

Our approach aims to achieve the best performance results in comparison to the state between the two models. We evaluated our approach on the same dataset. Table 5.3 shows the comparison results of Models. The correlation in linear regression model shows a good prediction is 71.5 % compared with the correlation support vector machine 65.2%. However, the linear regression provides the best result with the minimal average absolute error is 5.897 +/- 4.624 and the minimal average relative error 9.64% +/- 7.66% with the other model.

Table 5.3: Results comparison results of models

Model	Correlation (Min/Max %)	Average Absolute Error	Average Relative Error	Time To Build Model
Multiple linear Regression	71.5 %	5.897 +/- 4.624	9.64% +/- 7.66%	1 Sec
Support Vector Machine	65.2 %	6.993 +/- 5.277	11.72% +/- 9.75%	3 Sec

In this research, we have used measurement metrics namely: correlation, average absolute error, and average relative error. After the analysis, we concluded that the different between linear regression and support vector machine is that the linear regression model gives the best performance result and it has the lowest error rate. It also takes less time to build the model. Hence, we concluded that linear regression gives the highest accurate model to generate weights for metrics.

Figure 5.4 and Figure 5.5 show some output results concerning the comparison with real websites performance data and predictive ones using linear regression and support vector machine according to the cases of Table 5.3. The results must be read as follow:

If the prediction is similar to the real data concerning website performance will follow the same trend, otherwise will occur a trend variation.

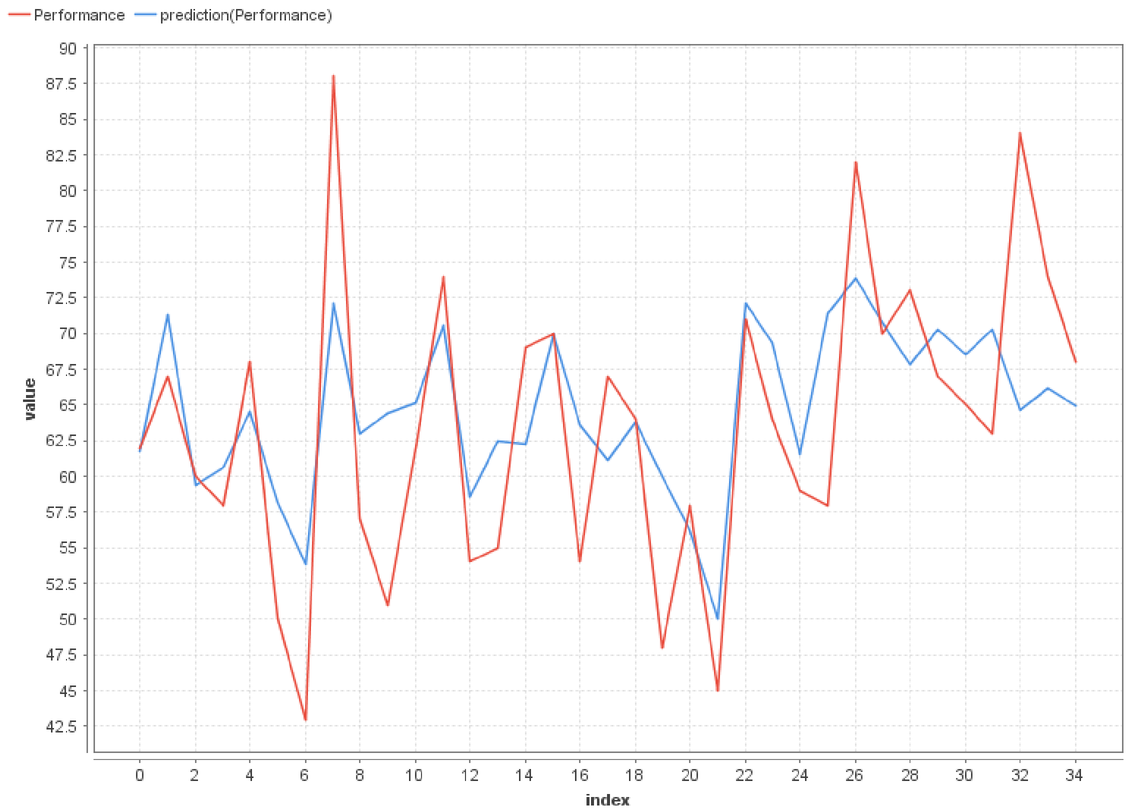


Figure 5.4: comparison with real websites performance data and predictive ones by linear regression model

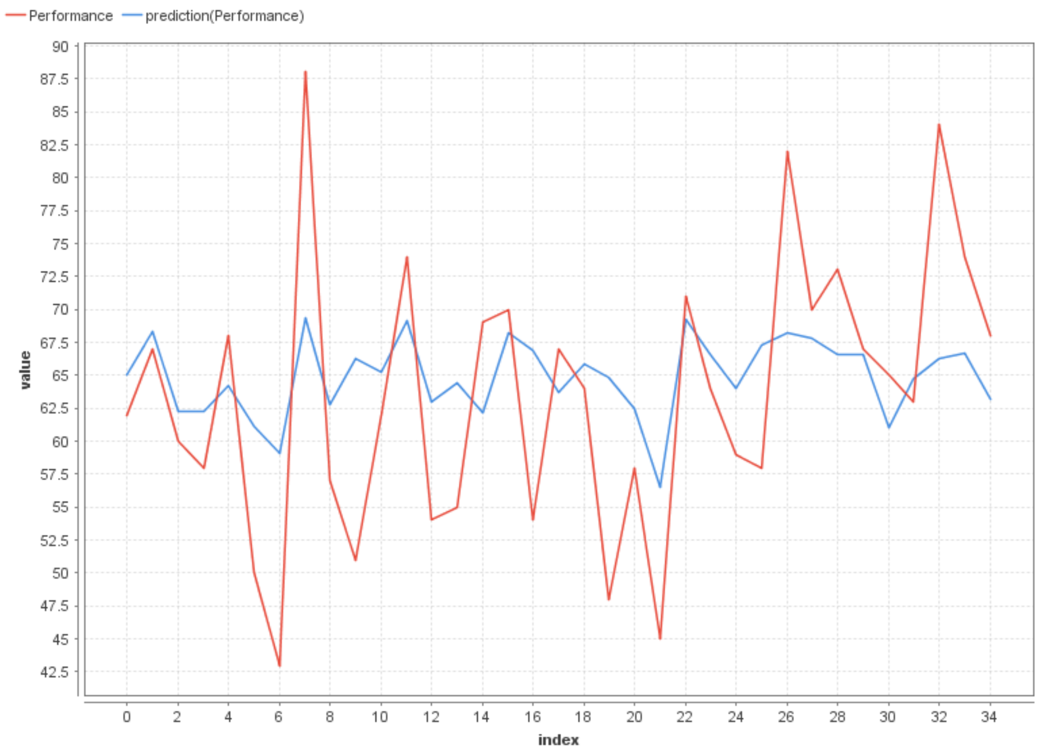


Figure 5.5: comparison with real websites performance data and predictive ones by Support Vector Machine model

Result of below correlations matrix :

Figure 5.6 shows the correlation between all metrics and it can produce a weights vector based on these correlations. And also correlation is a statistical mechanism in order to can show whether and how strongly pairs of metrics are related.

Attributes	Response_time	Load_time	Broken_link	Page_Size	Markup_validation	Optimization	No_of_Request	Performance
Response_time	1	0.283	0.158	-0.005	0.111	-0.315	0.145	-0.289
Load_time	0.283	1	0.294	0.293	0.095	-0.124	0.133	-0.400
Broken_link	0.158	0.294	1	0.351	-0.016	0.029	0.422	-0.650
Page_Size	-0.005	0.293	0.351	1	-0.017	0.064	0.288	-0.438
Markup_validation	0.111	0.095	-0.016	-0.017	1	-0.098	0.117	-0.154
Optimization	-0.315	-0.124	0.029	0.064	-0.098	1	0.210	0.121
No_of_Request	0.145	0.133	0.422	0.288	0.117	0.210	1	-0.469
Performance	-0.289	-0.400	-0.650	-0.438	-0.154	0.121	-0.469	1

Figure 5.6: The correlation matrix among metrics

A correlation is a number between -1 and +1 that measures the degree of association between two metrics (call them X and Y). A positive value for the correlation implies a positive association like the association between website performance and design optimization, where the optimal design can lead to the best website performance. And also a negative value for the correlation implies a negative or inverse association like the association between website performance and response time, where any decrease in response time can result to the best performance.

5.3 Identifying most important metrics

Figure 5.7 shows calculate the relevance of the metrics by computing the value of correlation for each metric with respect to website performance as mentioned in the section 4.5. Thus, we arranged the metrics from a high correlation to low correlation based on the weight to every metric.

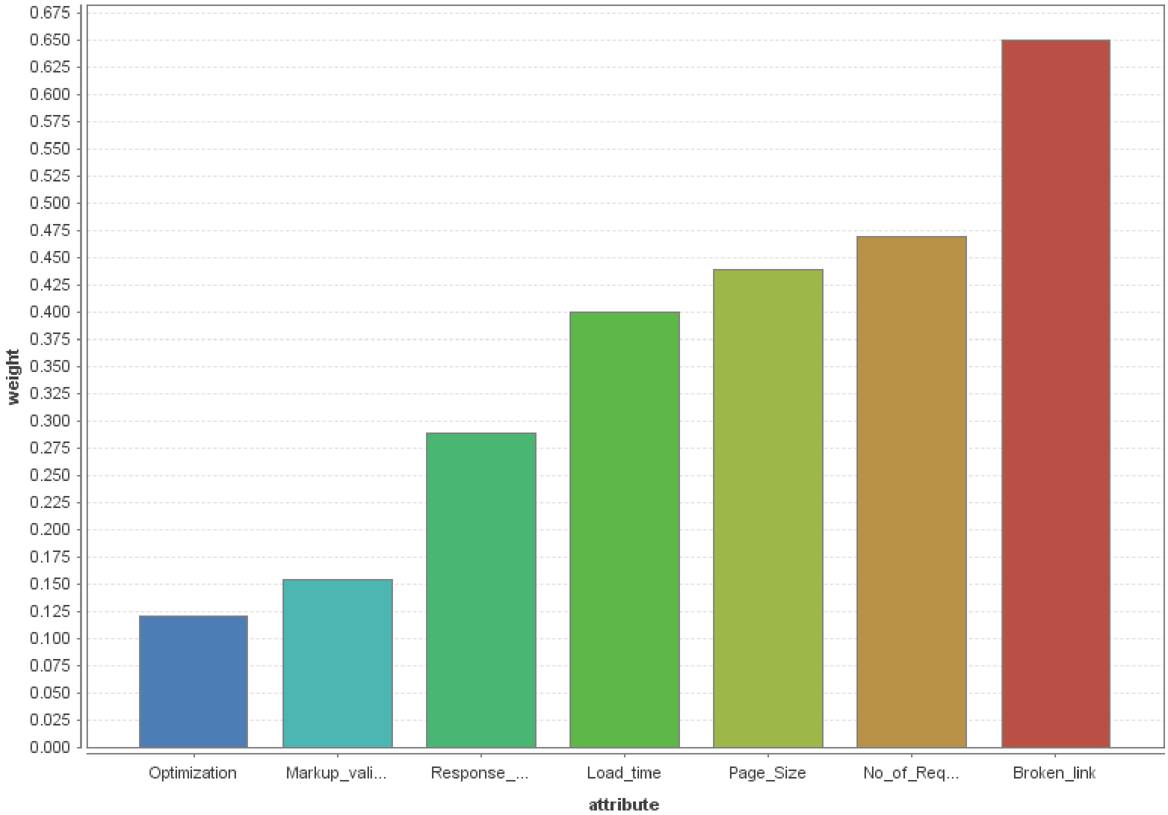


Figure 5.7 shows correlation the relevance of the metric

Therefore, Figure 5.8 shows arranged metrics in the level of the hierarchy help webmasters and decision-makers to know what improvements are needed to enhance the performance as shown in figure 5.7 above.

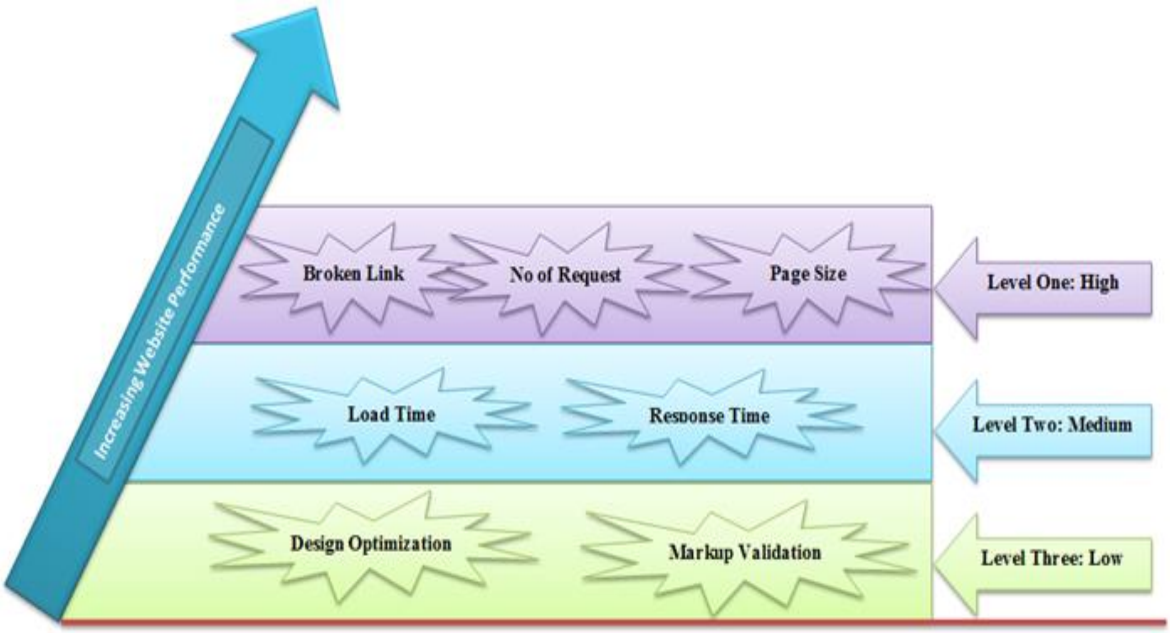


Figure 5.8: The level of the hierarchy of web metrics

5.4 Building Model

After determining the best performance between the two models as mentioned in section 5.2. we developed a new dynamic model to evaluate websites performance based on the proposed mathematical model that we called is PEML. Figure 5.9 shows the linear regression model using machine learning.

LinearRegression

```
- 0.596 * Response_time
- 0.154 * Load_time
- 0.105 * Broken_link
- 0.415 * Page_Size
- 0.013 * Markup_validation
+ 0.070 * Optimization
- 0.051 * No_of_Request
+ 77.610
```

Figure 5.9: The linear regression model

Finally, we extracted equation that used to evaluate websites performance by using the best performance among models. After that, we want to evaluate website performance based on the model in our study by the mathematical model:

Final website performance (%) = + 77.610 + - 0.596 * Response Time + - 0.154 * Load Time + - 0.105 * Broken Link + - 0.415 * Page Size + - 0.013 * Markup Validation + 0.070 * Design Optimization + - 0.051 * No of Request

5.5 Benchmarking

In order to validate a new model in this thesis that called is PEML, we want to compare with the previous studies by using the same dataset in the previous studies [8] [9]. The researchers in the previous studies measured sample data as shown in table 5.3 from national e-government portals of a chosen number of countries in Asia: Singapore, Korean, Japan, Hong Kong, and Malaysia based on many metrics of website performance, consisting of eleven metric: load time, response time, page rank, frequency of update, traffic, design optimization, page size, number of components, accessibility error, mark-up validation, and broken link. There are five models used in the previous studies [8] [9]: analytical hierarchy process model (AHP), fuzzy analytical hierarchy process model (FAHP), linear weightage model (LWM), hybrid model (combination among LWM and FAHP), and PROMETHEE II model.

As a result, we want to test our new model in this thesis on a new dataset from the previous studies [8] [9] as shown in table 5.4. Table 5.5 the final ranking of e-government websites based on five specific methods from the previous studies and the proposed a new model in this thesis. In accordance with the results generated by the suggested model, Korea website has the highest ranking in comparison with the rest of the e-government websites. The first column in Table 5.4 shows the metrics of the quality website. The metric elaborate in the website selection process using the proposed model are load time (A), response time (B), design optimization (C), page size (D), number of requests (E), markup validation (F), and broken link (G). The second column shows the measurement unit, and the rest of the columns represent the e-government website performance value.

Table 5.4 Original data

Metric	Measurement unit	Singapore	Korea	Japan	Hong Kong	Malaysia
A	**Seconds**	30.77	0.30	68.93	41.94	77.51
B	**Seconds**	1.94	1.17	1.73	1.03	4.84
C	**Percentage**	37.50	57.00	36.50	33.00	22.00
D	**Number**	128,305.00	511.00	285,645.00	195,384.00	366,825.00
E	**Number**	26.00	1.00	60.00	15.00	22.00
F	**Number**	79.00	5.00	21.00	3.00	80.00
G	**Number**	4.00	0.00	1.00	1.00	9.00

Table 5.4 Final result for e-government websites performance

Method	*Singapore*	*Korea*	*Japan*	*Hong Kong*	*Malaysia*
LWM	0.499(3)	0.766(1)	0.456(4)	0.672(2)	0.252(5)
AHP	0.183(3)	0.313(1)	0.115(4)	0.305(2)	0.085(5)
FAHP	0.222(3)	0.390(1)	0.007(4)	0.380(2)	0.001(5)
Hybrid	0.620(3)	0.771(1)	0.431(4)	0.683(2)	0.162(5)
PROMETHEE II	0.019912(3)	0.298043(1)	-0.10962(4)	0.185212(2)	-0.39355(5)
PEML	71.5(3)	80.5(1)	64.9(4)	71.8(2)	61.0(5)

6. Conclusion and Future work

This section concludes our study. We represent a brief conclusion and future work.

6.1 Conclusion

This study proposed a dynamic model is namely PEML to evaluate the performance of the websites. The proposed approach was using the mathematical model and machine learning. We applied experiments on two algorithms namely, linear regression and support vector machine regression, we applied the experiments on the same dataset that collected to take the best performance of regression methods to generate weight to every the metric for developing a new dynamic model to evaluate websites performance.

6.2 Future Work

Future studies can adopt multi-attribute approaches to evaluate the effectiveness of websites and includes adding more metrics to evaluate website performance. The results of future studies then can be compared with those results presented in this study.

References

[1] Kaur, S., Kaur, K., & Kaur, P. (2016). An empirical performance evaluation of universities website. *Int J Comput Appl*, *146*, 15.

[2] Harshan, R. K., Chen, X., & Shi, B. (2017). Research Article Analytic Hierarchy Process (AHP) Based Model for Assessing Performance Quality of Library Websites.

[3] Vatansever, K., & Akgűl, Y. (2018). Performance evaluation of websites using entropy and grey relational analysis methods: The case of airline companies. *Decision Science Letters*, *7*(2), 119-130.

[4] Islam, A., & Tsuji, K. (2011). Evaluation of usage of university websites in Bangladesh. *DESIDOC Journal of Library & Information Technology*, *31*(6).

[5] Devi, K., & Sharma, A. (2016). Framework for evaluation of academic website. *Journal of International Journal of Computer Techniques*, *3*(2), 234-9.

[6] Kulkarni, R. B., & Dixit, S. K. (2012). Empirical and Automated Analysis of Web Applications. *International Journal of Computer Applications*, *38*(9), 1-8.

[7] Khan, H., & Dominic, P. D. D. (2013). Measuring Quality of Asian Airline Websites Using Analytical Hierarchy Process: A Future Customer Satisfaction Approach. *ISICO 2013*, *2013*.

[8] Dominic, P. D. D., et al. "A comparison of Asian e-government websites quality: using a non-parametric test." *International Journal of Business Information Systems* 7.2 (2011): 220-246.

[9] Jati, Handaru. "Quality Ranking of E-Government Websites: PROMETHEE II Approach." *International Conference for Informatics for Development, Yogyakarta*. 2011.

[10] Zahran, D. I., Al-Nuaim, H. A., Rutter, M. J., & Benyon, D. (2014). A comparative approach to web evaluation and website evaluation methods. *International Journal of Public Information Systems*, *10*(1).63

[11] Massaro, A., Maritati, V., & Galiano, A. (2018). Data Mining model performance of sales predictive algorithms based on RapidMiner workflows. *Int. J. Comp. Sci. Inf. Technol*, *10*, 39-56.

[12] Graczyk, M., Lasota, T., & Trawiński, B. (2009, October). Comparative analysis of premises valuation models using KEEL, RapidMiner, and WEKA. In *International conference on computational collective intelligence* (pp. 800-812). Springer, Berlin, Heidelberg.

[13] Duncan, R. J., Van Dongen, N., Missimer, C. A., Liu, S., & Jabrane, K. (2011). *U.S. Patent No. 7,966,558*. Washington, DC: U.S. Patent and Trademark Office.

[14] IBM SPSS Software = https://www.ibm.com/analytics/spss-statistics-software , note = Accessed: 2019-03-01

[15] Tutorial: Understanding Regression Error Metrics in Python = https://www.dataquest.io/blog/understanding-regression-error-metrics/, note = Accessed: 2018-12-05

[16] Support Vector Machine – Regression (SVR) = http://www.saedsayad.com/support_vector_machine_reg.htm, note = Accessed: 2018-09-01

[17] Vresk, T., & Čavrak, I. (2016, May). Architecture of an interoperable IoT platform based on microservices. In *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 1196-1201). IEEE.

[18] Vapnik, V. The Nature of Statistical Learning Theory. Springer, New York, 1995

[19] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006, August). Yale: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 935-940). ACM.

[20] Souders, Steve. "High Performance Web Sites: Essential Knowledge for Front-End Engineers. sl." (2007): 978-0596529307.

[21] Moustakis, V., Litos, C., Dalivigas, A., & Tsironis, L. (2004, November). Website Quality Assessment Criteria. In *ICIQ* (pp. 59-73).

[22] Palmer, J. W. (2002). Web site usability, design, and performance metrics. *Information systems research*, *13*(2), 151-167.64

[23] Hong, I. B. (2007). A survey of web site success metrics used by Internet-dependent organizations in Korea. *Internet research*, *17*(3), 272-290.

[24] Vatansever, K., & Akgűl, Y. (2018). Performance evaluation of websites using entropy and grey relational analysis methods: The case of airline companies. *Decision Science Letters*, *7*(2), 119-130.

[25] Albhaishi, A., Wahsheh, H., & Alghamdi, T. (2014, May). Evaluating Web Ranking Metrics for Saudi Universities. In *Zaytoonah University International Engineering Conference on Design and Innovation in Sustainability 2014* (No. 156988). Zaytoonah University, Amman, Jordan.

[26] Gollapudi, S. (2016). *Practical machine learning*. Packt Publishing Ltd.

[27] Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

[28] Riihimäki, T. (2014). Evaluating the value of web metrics.

[29] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.

[30] Maimon, O., & Rokach, L. (2009). Introduction to knowledge discovery and data mining. In *Data Mining and Knowledge Discovery Handbook* (pp. 1-15). Springer, Boston, MA.

[31] Tan, P.N., Steinbach, M., Kumar, V.: Introduction to data mining. 1st (2005)

[32] Brownlee, J. (2016). *Master Machine Learning Algorithms: discover how they work and implement them from scratch*. Jason Brownlee.

[33] Amerson, M., Fisher, G., Hsiung, L., Krueger, L., & Mills, N. (2001). Design for performance: analysis of download times for page elements suggests ways to optimize. *online] http://www. ibm. com/developerworks/websphere/library/techarticles/hipods/perform. html*.

[34] Sakthivel, M., Devadasan, S. R., Vinodh, S., Ramesh, A., & Shyamsundar, S. (2007). ISO 9001: 2000 based quality information management responsibility system. *International Journal of Business Information Systems*, *2*(2), 217.

[35] Lindholm, A., Wahlström, N., Lindsten, F., & Schön, T. B. Supervised Machine Learning.