Iraqi Journal of Intelligent Computing and Informatics (IJICI)
Vol. 2, 1, June 2023, pp. 11~25
ISSN: 2791-2868, DOI: 10.52940/ijici.v2i1.27

Webpage Classification Using Ensemble Machine Learning

Hind Sabah Rahim[1], Aliea Salman sabir[2], Nahla Abbas Flayh [3]
[1,2,3] Department of Computer information systems, College of Computer Science and Information Technology, University of Basrah, Iraq

Article Info

Article history:

Received February 22, 2023
Revised March 08, 2023
Accepted March 24, 2023

Keywords:

Webpage Classification
Natural language Processing
Machine Learning
Ensemble Machine Learning
Stacking

ABSTRACT

These days, it is not easy to get the correct information after typing a keyword into a search engine because so many results are returned. Classification of Web pages is a technique that helps us locate the wanted information quickly and effectively. In addition, website categorization is crucial for businesses that provide marketing and analytical solutions because it enables them to create a well-balanced mix of search engine and directory listings. This will give marketers a better idea of where their local company listings appear online, allowing them to have more judgment about initiative and strategy.

Therefore, the research aimed to construct a classification system based on a dataset of English web pages. This information has been acquired from the Kaggle website and consisted of *14*08 distinct rows organized into 16 categories.

The research has employed mixed strategies to determine which strategy for Web page categorization would yield the best results. The first strategy puts into practice a collection of machine-learning algorithms. It assesses how well they accomplish the given classification task. Ensemble stacking is the second strategy, and it is employed to enhance the classification of websites.

Comparing the results of the two strategies reveals that Ensemble stacking, the second strategy, was the more influential architecture for classifying web pages this approach had 0.95 F1-score, 0.95 accuracy, 0.95 precision, and 0.95 recall achieved by this method. The first approach, which made use of machine learning techniques, on the other hand, received an F1-score of 0.93, 0.94 for precision, 0.93 for recall, and 0.93 for accuracy.

Corresponding Author:

Hind Sabah Rahim
Department of Computer information systems, College of Computer Science and Information Technology, University of Basrah, Iraq
Email: itpg.hind.sabah@uobasrah.edu.iq

1. INTRODUCTION

During the past decade, social networking sites have grown exponentially. Currently, social networking services deal with vast amounts of data collected and shared by the public. Many people share their thoughts and feelings on various topics through social networking sites.

Getting information from many sources and extracting it for later use is the process of text analysis. It is crucial to analyze social network data to look at people's viewpoints and ideas on a particular issue to predict and improve the future. Identifying potentially damaging information is crucial to preventing the exploitation of a website or social network blog [1].

The number of web pages on the Internet is rising rapidly. In 2022, there were over 1.9 billion web pages on the world wide web, and as time passes, the number is increasing [2]. See Figure 1.

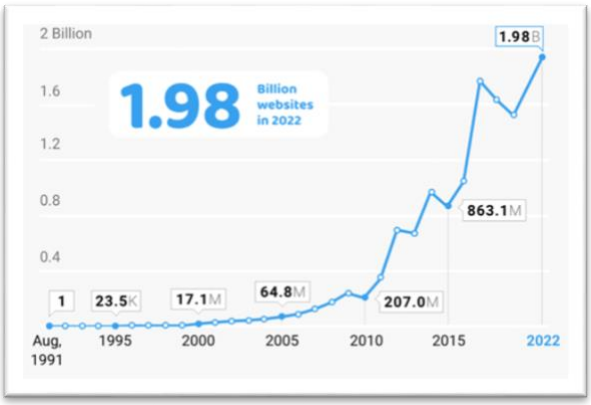

Figure 1. Total number of web pages across time

Classification of Web pages is a method of data retrieval that delivers usable data that is a foundation for numerous application fields. Organizing web pages into categories gives valuable information for practical Internet usage, filtering spam, and numerous other applications. Obtaining relevant results fast among billions of websites is an intricate problem that search engines tackle. Several search engines need a topic-based categorization of web pages to offer better consumer results [3].

Classifying web pages manually is impractical due to the vast volume of information accessible via the Internet. The web presents a dynamic setting that frequently changes, making it challenging to create a categorization model that can categorize numerous web pages [4].

The classification of web pages is necessary for extracting knowledge and retrieving tasks, including the creation, development, and keeping up of Web directories; improving search for better results; improving the quality of question-and-answer platforms; developing; specialized web crawling; filtering web content; assisting web browsing; and contextual advertising.

2. WEBPAGES CLASSIFICATION TYPES

Classifying Websites may be divided into subfields, including subject classification, functional classification, sentiment classification, and other classification methods.

- **Subject classification** is interested in the topic or subject of a web page. For instance, subject classification would be determining whether a page is about arts, business, or sports. [5].
- **Functional classification** takes into consideration the function that the website page serves. For example, functional classification determines whether a page is to be a personal homepage, course page,or admission page [6].
- **Sentiment classification** emphasizes the point of view presented on a webpage, often known as the author's perspective toward any topic [7].

Genre classification and search engine spam classification are two other types of classification. Another type of classification depends on the number of categories on the page; binary and multi-class classification are the two parts of the classification problem. Multi-class classification divides the dataset into many classes based on a classification rule. In contrast, binary classification divides the dataset into only two classes. The has explored multi-class subject classification. See Figure 2 for that Clearfield binary and multi-classification. [8]

Because of the tremendous amount of web pages that increase every minute, and with the variety of content and amount of information that webpages contain, most of it is not structured, so classifying webpages manually is impossible for many reasons. That required building a system classifying webpages to their category in an automatic way that can serve a wide range of domains like businesses and families, text analysis, search engines, and web mining.

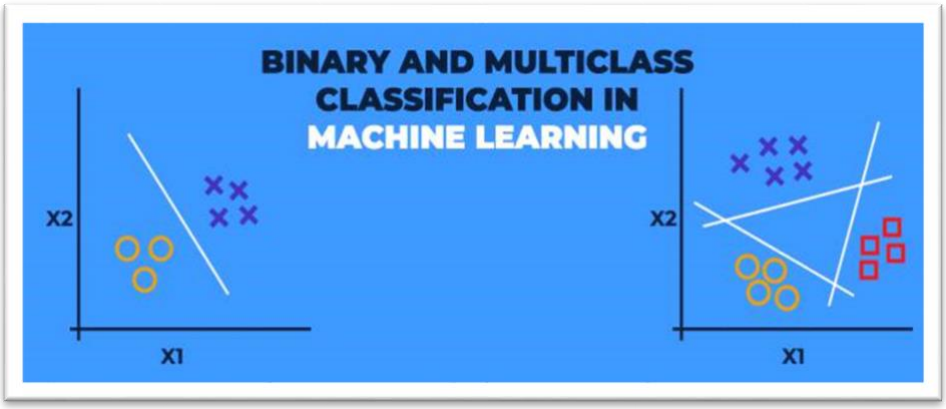

Figure 2. Binary and Multi-classification.

3. RELATED WORK:

Many studies have been undertaken in recent years utilizing Natural Language Processing (NLP) approaches to categorize web pages into different categories. Even though most traditional Machine Learning (ML) techniques prioritize the required feature qualities of web pages and categorize them into predetermined categories using ML algorithms.

- In 2016, [9] compared the abilities of classifier ensembles to identify text documents using keywords accurately. The keyword extraction techniques deal with high-dimensional feature spaces by simply collecting essential keywords from the text documents. To evaluate the effectiveness of statistical keyword extraction methods in combination with ensemble learning algorithms, a comparison of base learning algorithms (Naive Bayes, Support Vector Machines, Logistic Regression, and Random Forest) with five popular ensemble methods (AdaBoost, Bagging, Dagging, Random Subspace, and Majority Voting) is made. The practical analysis demonstrates that combining ensemble learning with keyword-based reconstructions of textual information can make text classification schemes more accurate and able to be used on a larger scale.

- In 2017, [10] suggested a system for classifying Turkish text that uses a mix of multinomial NB, SVM, multivariate Bernoulli NB, and RF. To connect the base learners, they use stacking and majority voting. When stacked classifiers are compared to single classifiers for different datasets, the results show that the success rate ranges from 2% to 13%.

- In 2019, [11] Utilizing HTML and URL, a stacking model has been suggested. Features to detect phishing sites. Light gradient boosting machine (Light GBM), gradient boosted decision tree, and XGradientBoost has been merged to create a stacking model that enables various models to work in harmony, enhancing the effectiveness of phishing webpage recognition. They outperformed several using a variety of measures, and machine learning models produced results with an accuracy of 97.30%.

- By combining the Nave Bayes, Support Vector Machine, and Random Forest algorithms, [12] in 2021 presented a stacking ensemble approach. By obtaining Web sites of Indian academics from international university websites, the Stacking approach improves the Stacking ensemble's two-stage learning process. The advantages of individual base classifiers are combined to improve the classification system's performance. The outcome demonstrates that an ensemble stacking method outperforms the individual classifiers.

- In 2022, [13] presented a classification approach based on the categories Ham and Spam for email text. Importing the dataset, pre-processing (removing stop words and vectorizing), and feature choosing (weighing and selecting), The steps involved in developing a classification model included separating the data into a train set (80%) and test set (20%), importing classifiers, and training classifiers. The model was evaluated before being deployed along with a spam filtering application on a server (Heroku) using the Flask framework. The testing of the system indicated that its performance was satisfactory.

4. THEORETICAL BACKGROUND

The main components of the classification process are introduced in this section. It also discusses modern procedures and approaches that are thought to be the foundation of the categorization process, like machine learning algorithms, the concept of Natural language processing, and the criteria for evaluating the model's final performance.

1- Machine Learning

Machine learning is one of the most important areas of Artificial Intelligence (AI). The machine's primary goal is to access data and use it to learn and find patterns in it. Predictions and data clustering may then be done using these patterns [14].

Classification models are used to classify input data, with output values or the goal (Y) being categorical, an example of classification is used to determine whether or not a patient is sick [15].

A machine learning model's ability to predict is based on the given data in many domains like the huge growth in social media, and the massive number of users has lured attackers to distribute harmful content through fake accounts [16], and governments started to use webpages for better government services delivered to citizens because of these e-services. [17]

algorithms in machine learning [18] split into four categories according to the kind of input data and predicted outputs:

- Supervised Learning.
- Unsupervised Learning.
- Semi-supervised Learning.
- Reinforcement Learning.

5. SUPERVISED MACHINE LEARNING ALGORITHMS

In the following section, the most common supervised approaches will be clarified:

1- Logistic Regression (LG)

The main function in LR is the sigmoid function converts each real number into a value between 0 and 1. The logistic/sigmoid function is essentially on top of a linear regression model [19]. This indicates that the output of this model is always between 0 and 1, giving us the likelihood of an observation being either 1 or 0. Logistic Regression computes the probability of a binary outcome. It classifies the data points into either outcome by establishing a threshold [20].

2- K-Nearest Neighbour (KNN)

A supervised machine-learning technique that can be used to solve regression and classification issues is the K-nearest neighbors (KNN) method. KNN classification employs majority voting over the k-nearest neighbors to predict the results of a new dataset. The testing phase is slow and expensive regarding computer resources because it is a slow learning model where computations only happen at run-time [21]. It uses distance functions like Euclidean distance to get the k-nearest sites.The performance of the KNN algorithm is mainly reliant on the selection of the number of nearest points.

3- Random Forest (RF)

A supervised machine learning classification approach and constructing a decision tree that produces many decision trees during model training. It is a form of additive model that makes predictions by combining the conclusions of base models [22]. The classification outcomes are given for each tree. The number of classes these trees create determines the highest classification (majority class).

4- Support Vector Machine (SVM)

It is the first of the most powerful and reliable statistical machine-learning techniques. The most important objective of the SVM classifier is to create a functional hyperplane to segregate trained data. The optimal hyper method is selected from a collection of hyper techniques with a large margin of safety [23]. SVM is a discriminative classifier technically defined by a separating hyperplane. In other words, the algorithm builds an ideal hyperplane for classifying new cases given labeled training data. If there are N features, the dimensions of the hyperplane will be N-1. In two dimensions, a hyperplane is a line that divides a plane into two sections, one for each class. The SVM method repeatedly generates the optimal hyperplane from among all possibilities [24]. The mathematical margin is the distance between each grouping's closest points, and in the hyperplane, a larger margin is preferred over a smaller one [25].

5 - Naive Bayes (NB)

The classifier method is based on the Bayes Theorem and also the idea of feature independence for conditions. Condition independence is used to learn the joint likelihood distribution of both input and output for each trained data set. foundation, the frequency after a given set of inputs x is then calculated. The strategy is easy and highly successful for learning and prediction [26].

5.1 ENSEMBLE MACHINE LEARNING TECHNIQUE

Ensemble approaches combine the results of numerous algorithms to get more accurate results and improve the model's overall performance. It may produce outcomes that are superior to those of any individual algorithm. Ensemble Methods include Stacking, Bagging, Boosting, Adaboost, etc., and are generally used for improving classification accuracy by aggregating the predictions of multiple classifiers [27].

The term Crowd's Wisdom refers to a decision-making process in which human beings make lower-level judgments.

5.2 STACKING CONCEPT

Stacked Generalization, or stacking for short, is an ensemble method used in machine learning that employs various learning algorithms as a base- model to enhance prediction performance. Diverse members are sought by changing the model types fitted to the training data. [28]
A different machine learning model is utilized to discover the most effective method for combining the forecasts provided by the base models.
The predictions that are made by the base models using data that is not in the sample are used to train the meta-model. To accomplish k-fold, cross-validation will need to perform on each base model, and all out-of-fold predictions will need to be saved. After that, the base models are trained on the entirety of the training dataset, and the meta-model is trained on the out-of-fold predictions. It then learns which models to trust and to what degree and trains itself on the out-of-fold predictions.

5.3 BENEFITS OF USING ENSEMBLE LEARNING

- Performance: An ensemble can achieve better performance and make more accurate predictions than any single model that it contributes.
- Robustness: The spread or dispersion of the predictions, as well as the performance of the model. [29]

5.4 NATURAL LANGUAGE PROCESSING

In the branch of Artificial Intelligence and computer science known as Natural Language Processing (NLP), the text is analyzed by a machine trained to derive meaning from unstructured or extremely variable human-written content [30].

5.5 TERM FREQUENCY, INVERSE DOCUMENT FREQUENCY (TF-IDF)

Information retrieval and text mining frequently use TF-IDF, the term frequency-inverse document frequency. This weight is a mathematical metric to assess a word's significance to a group of documents. Every time a word appears in a document, its significance increases proportionately [31].
The TF-IDF weight method is a popular tool used by search engines to score and rank the content of a page in relation to a search request. Stop-words filter with TF-IDF can also be applied effectively in a variety of domains, including text summarization and categorization.

5.6 GRID SEARCH

The term grid search refers to an approach used in traditional hyperparameter optimization. This strategy includes a thorough search for the training set over a subset of the hyperparameter space. It is a trial-and-error, brute-force algorithm. The hyper-parameters that can be combined in any way that wants to employ this approach must be specified. The number of search iterations might increase fast. Therefore, the user must be careful in choosing [32].
This technique begins by partitioning the hyperparameter domain into a discrete grid. Next, it tests each possible value merged using a current collection of the hyperparameter values within the grid while simultaneously calculating performance metrics to evaluate the model with each combination through cross-validation. Finally, after evaluating each combination, the model with the set of parameters that gives us the most accurate results overall is deemed the best to use [33].

6. EVALUATION CRITERIA

The machine learning model must be regularly evaluated and adjusted to achieve optimal performance. Several real-world contexts use these metrics [34].
1- Accuracy:
An algorithm could be evaluated using test data, with test predictions divided into four sets. The detection of True Positives (TP) was positive and also expected that be positive. In contrast, the True Negatives (TN) detection was negative and is expected to be negative. False Positives (FP) were observed as negative but projected to be positive. False Negatives (FN) were observed as positive but projected to be negative. [35]

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

2- Precision
It is the ratio of the correct class predictions divided by the total number of class predictions [36].

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

3- Recall
It measures how many accurately predicted positive observations were compared to all of the observations made in the actual class. [37]

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

4- F1- Score
The F1- score is a combinational harmonic of the Precision Sensitivity metrics that describe the model's ability to identify class faults. [38]

$$\text{F1}_{\text{Score}} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

5- Confusion Matrix
A table with four different values, including actual and predicted, is presented here. The column in the table represents the real class, and the row represents the predicate class. [39]
The values of the confusion matrix table are:

• True Positive (TP): The class is positive, as is the model prediction.

• True Negative (TN): The class is negative, as is the model prediction.

• False Positive (FP): The model was incorrectly classified as negative.

• False Negative (FN): The model was incorrectly classified as positive.

Where real positive class (P) = TP + FN, real negative class (N) = FP + TN

7. METHODOLOGY OF THE PROPOSED FRAMEWORK

This section provides the methodology of the proposed framework that consists of three stages with different steps to implement the process of webpage classification.

8. PROPOSED FRAMEWORK

The adopted framework in this thesis uses two different strategies. The first strategy used machine learning algorithms to classify webpages into different classes. In contrast, the second strategy involved stacking ensemble models to classify the webpages.
This section presents the proposed framework's general architecture, which consists of the following stages: preparing the dataset to be trained, then the training stage to build the models for classification. These evaluation metrics assess the model's performance and the accuracy of the different adopted calculated, as depicted in Figure 3

Figure 3. The General Stages of the Proposed Framework

8.1. Data Preparation Stage

This stage consists of three main steps, as depicted in Figure 4.
The inputs for this stage were the raw data; the output will be the pre-processed data ready to be trained.

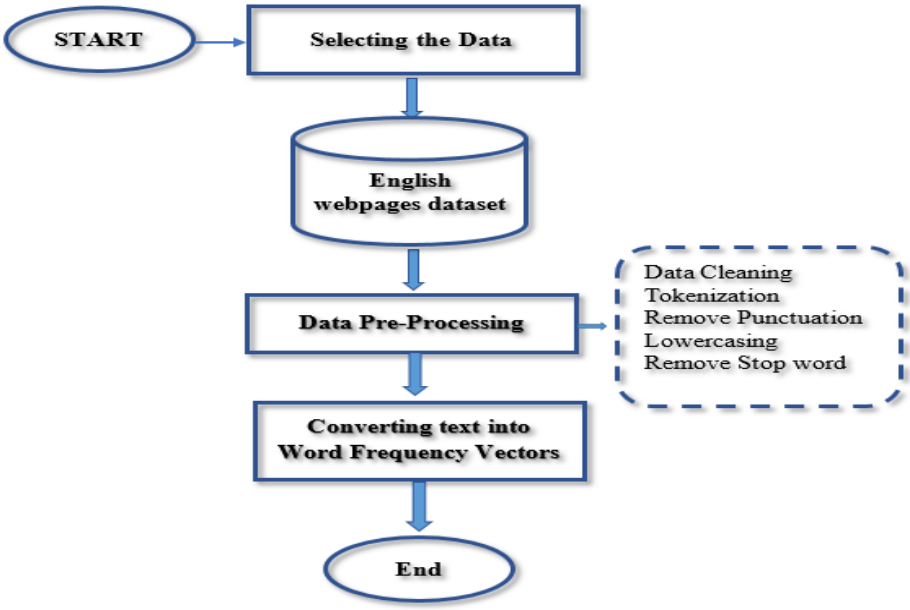

Figure 4. Illustrates the Data Preparation Stage

- **Describe The Dataset**

This step involves understanding and exploring the preliminary-level data. The English webpages dataset was used, which was obtained from the Kaggle website. The dataset was created by scraping different web pages and then classifying them based on the extracted text. It is a file that contains 1408 different rows classified into 16 categories. The categories' names and the total number of words that are included in each class is demonstrated in Table 1, Table 2 shows a sample of the selected dataset.

Table 1. The classes names and number of words in each class in the English dataset

No.	Category Types	Number of words
1	Educational Websites	774075
2	Business/Corporate Website	387336
3	Travel Websites	615047
4	Streaming Services Websites	385775
5	Sports Websites	854152
6	E-Commerce Website	480333
7	Games Websites	299445
8	News Websites	1020091
9	Health and Fitness Websites	534152
10	Computers and Technology Website	409889
11	Photography Websites	401760
12	Food Websites	437253
13	Law and Government Websites	443235
14	Social Networking and Messaging Website	177552
15	Adult Websites	156038
16	Forums Websites	50604

Table 2. shows a sample of the selected dataset.

No	Webpage_url	Webpage_text	Category
0	https://www.booking.com/index.htmlaid=1743217	“official site good hotel accommodation big …”	Travel
1	https://travelsites.com/expedia/	“expedia hotel book sites like use vacation wor...”	Travel
2	https://travelsites.com/tripadvisor/	“tripadvisor hotel book sites like previously d...”	Travel
3	https://www.momondo.in/?ispredir=true	“cheap flights search compare flights momondo f...”	Travel
4	https://www.ebookers.com/?affcid=ebookers-uk.n...	“bot create free account create free account si...”	Travel

8.2 Training Stage

The model was trained using a set of 985 and 1126 rows in cases of 70% and 80% of the dataset. This was performed to determine how increasing the set size would affect the results. see Figure 5 that clarified the stage:

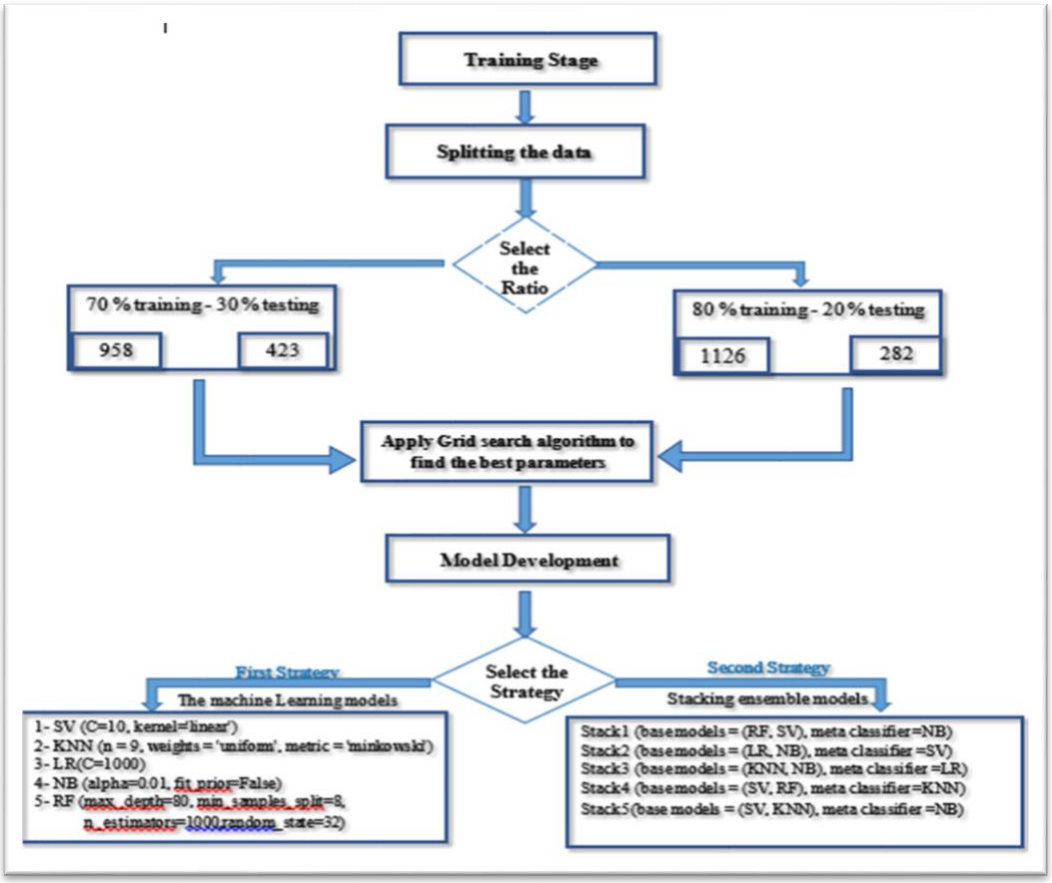

Figure 5. Illustrates Training Stage

- **Models Development**

This step involves training the models to classify the dataset, it has two different proposed strategies:

1- Machine Learning Algorithms:

This strategy has implemented a set of machine learning algorithms to assess their efficiency in the classification problem. Before implementing any algorithm, the crucial step is to fine-tune its hyperparameters to ensure optimal model performance. Figure 6 illustrates the required sequence to build the machine learning models.

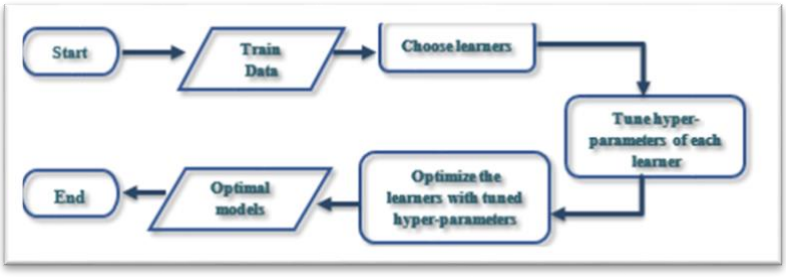

Figure 6. illustrates the sequence of building the machine learning models.

2- Ensemble Stacking Learning

The development of a framework for classifying web pages is the focus of this section. It involves extracting the features from a webpage and then categorizing it based on those features. Stacking is a technique for classifier combination that merges the results of base learners using a meta-level classifier**.**

The strategy has used five different webpage classification models have been created to classify the selected dataset as explained below in detail:

a. Stack1

Random Forest and Support Vector Machines have been used in the first stack as base learners and Naïve Bayes as meta-classifiers. Figure 7 depicts the machine-learning algorithms that have been utilized in this stack.

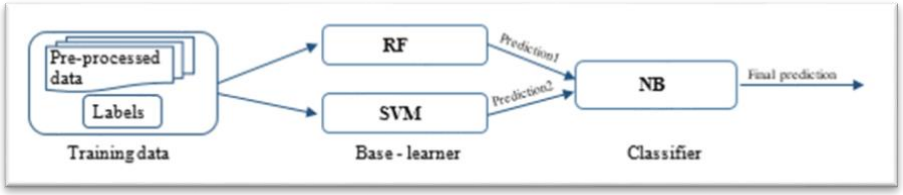

Figure 7: Stack1

b. Stack2

Logistic Regression and Naïve Bayes algorithms have been used in the second stack as base -learners and Support Vector Machines a meta-classifier. Figure 8 depicts the machine-learning algorithms utilized in this stack.

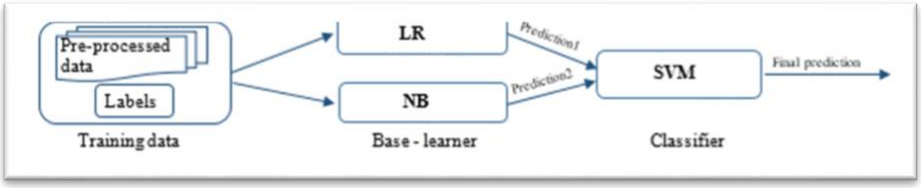

Figure 8. stack2

c. Stack3

In the third stack, K- nearest neighbours and Naïve Bayes algorithms have been used as base learners and Logistic Regression as meta-classifiers. Figure 9 depicts the machine-learning algorithms utilized in this stack.

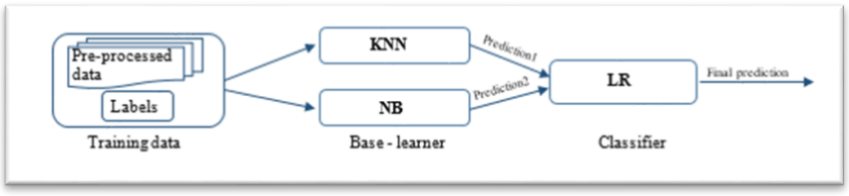

Figure 9. Stack3

d. Stack4

Support Vector Machines and Random Forest algorithms have been used in this stack as base learners and K- Nearest Neighbours as meta-classifiers. Figure 10 depicts the machine-learning algorithms utilized in this stack.

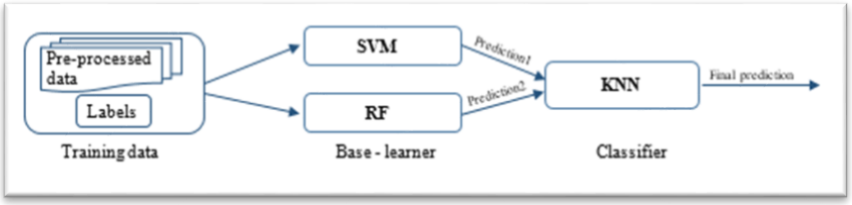

Figure 10 stack 4

e. Stack5

In the final stack, Support Vector Machines and K- Nearest Neighbours algorithms have been used as base learners and Naïve Bayes as meta-classifiers. Figure 11 depicts the machine-learning algorithms utilized in this stack.

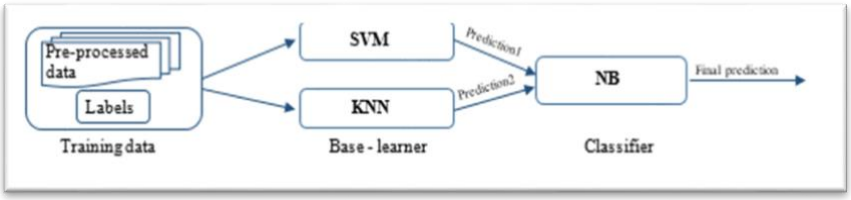

Figure 11 stack 5

8.3 MODELS EVALUATION

It involves evaluating the model using the relevant metrics. The model will be generalized to the newly acquired data and assessed at this stage. This involves determining how well the model can classify webpages into sixteen categories based on unseen data that has not been presented yet.

9. RESULTS AND DISCUSSION

This section will discuss all results that have been obtained from each model. First, the machine learning algorithms will be explained, then ensemble stacking learning algorithms will be explained.

9.1 Experiments and The Results:

This section will explain the machine learning algorithms' results and discuss their results separately based on how well the model performs with training and testing splitting percentages.

1- Experiment 1 with 70:30 percent with five machine learning algorithms (SVM, KNN, NB, RF, LR) with tuned hyperparameters.

2- Experiment 2 with 80:20 percentage with five machine learning algorithms (SVM, KNN, NB, RF, LR) with tuned hyperparameters

3- Experiment 3 with 70:30 percentage with stacking algorithm with tuned hyperparameters.

4- Experiment 4 with 80:20 percentage with stacking algorithm with tuned hyperparameters, see Table 3. This clarified the experiment.

Table 3. clarified the strategies and results

Number of Experiment	The Splitting Ratio	Algorithms	Results in Accuracy
Experiment 1	70:30	(SVM, KNN, NB, RF, LR) with tuned hyperparameters.	SVM=91.72 % KNN=88.65 % NB=90.42 % RF=88.41% **LR=93.38%**
Experiment 2	80:20	(SVM, KNN, NB, RF, LR) with tuned hyperparameters.	SVM=93.26 % KNN=89.007 % NB=90.78 % RF=90.42 % **LR=94.68 %**
Experiment 3	70:30	stacking algorithm with tuned hyperparameters.	Stack1=93.38% Stack2=92.90% Stack3=92.90% Stack4=92.67% **Stack5=93.61%**
Experiment 4	80:20	stacking algorithm with tuned hyperparameters.	Stack1=94.68% Stack2=93.26% Stack3=92.43% Stack4=93.97% **Stack5=95.035%**

9.1.1 Results of experiments of Machine Learning Algorithms

The results of Experiments 1 and 2 in accuracy are clarified in Table 4, and Figure 12 clarified the results of Experiments 1 and 2 with the Evaluation Matrix.

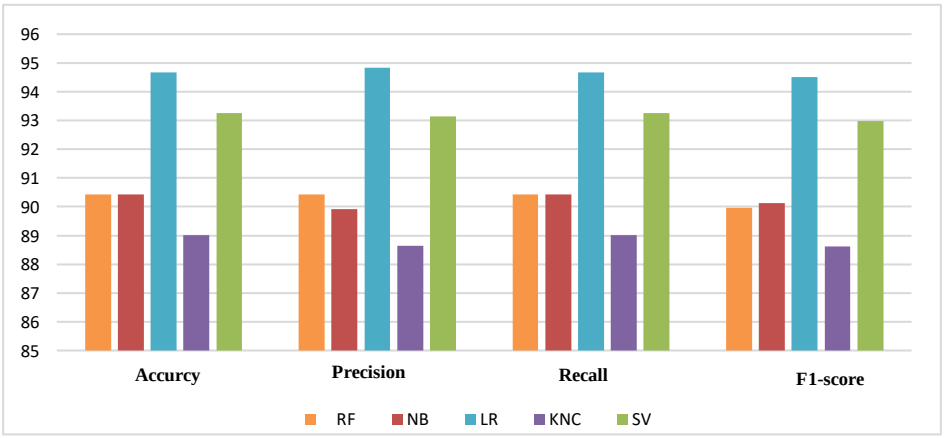

Figure 12. Comparison between the five algorithms with percentage (80:20) with Evaluation Matrix

Table 4. Comparison between the results of the five algorithms with percentage (70:30) and (80:20)

Algorithm	Accuracy 70:30	Accuracy 80:20
SVM	91.72 %	93.26 %
LR	**93.38%**	**94.68 %**
KNN	88.65 %	89.007 %
RF	88.41%	90.42 %
NB	90.42 %	90.78 %

9.2.3 DISCUSSION of THE EXPERIMENT MACHINE LEARNING ALGORITHMS

- Firstly, the machine learning models, Support Vector Machines, K Nearest Neighbors, Naïve Bayes, Logistic Regression, and Random Forest, were trained on 70:30 percent of the dataset to see which models were most effective. The findings that have been obtained by evaluating each model using the test data to determine its strength and efficiency in classifying categories, the Logistic Regression – model that has been discovered is superior to the others in terms of all the values of its performance metrics. In addition to his ability to classify the largest possible number of classes correctly, this was noticed by plotting the confusion matrix in Figure 12
- To get better outcomes than those previously obtained by training machine learning algorithms with a split of 70:30 from the data set. The same algorithms were trained with a percentage of 80:20, where the training data set was increased. This effort improved the results of all algorithms, and it increased the number of correctly classified classes. However, the Logistic Regression model is still the most efficient and effective.

According to the results, the best model for automatic webpage classification was the Logistic Regression model, regardless of the size of the two training data samples. This does not mean that the other models are inefficient because they may be the best and most appropriate for other tasks; however, the LR was the most effective model with the data that have been used. According to what is understood about machine learning, there is no best algorithm unless it has been tested and trained on the data of the problem to be created to solve it. This is because the data is the foundation for any model.

9.3 RESULTS of EXPERIMENTS of ENSEMBLE STACKING ALGORITHMS

The results of experiments 3 and 4 in accuracy are clarified in Table 5, and Figure 13 clarifies the results of experiments 3 and 4 with the Evaluation Matrix

Table 5 Comparison between the five stacks with percentage (70:30) and (80:20)

Stack number	Accuracy 70:30 tuned parameters	Accuracy 80:20 tuned parameters
1	93.38%	94.68%
2	92.90%	93.26%
3	92.90%	92.43%
4	92.67%	93.97%
5	**93.61%**	**95.035%**

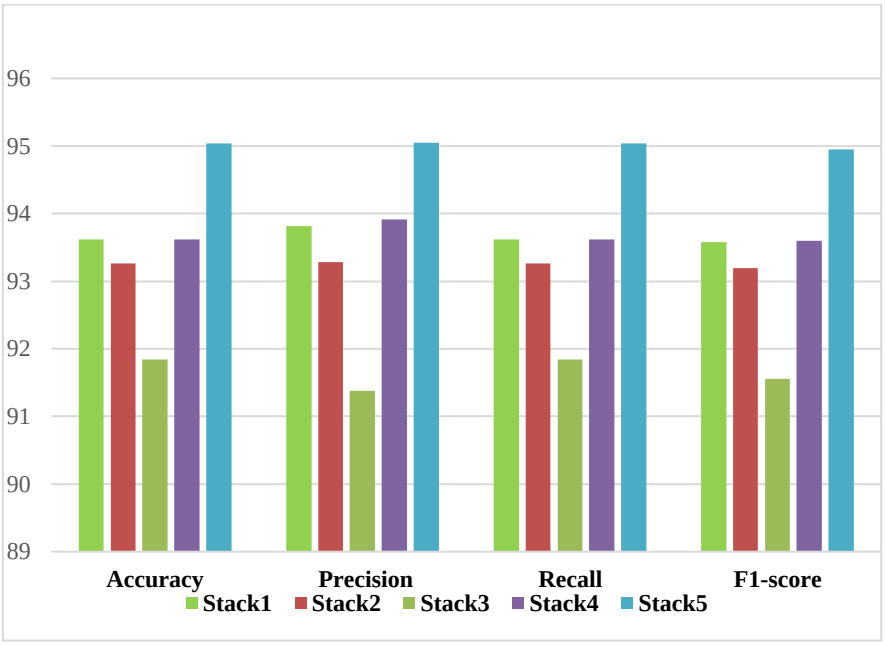

Figure 13. Comparison between the five stacks with percentage (80:20) with Evaluation Matrix

9.3.1 DISCUSSION of THE ENSEMBLE STACKING ALGORITHMS

In order to improve the results acquired in the first research strategy and to achieve a model capable of accurately classifying webpages, the second research method (Ensemble Stacking learning) was used. Its results will be reviewed in this section.

- When 70% of the total data size was used for training and the default hyperparameters were used for all the algorithms on which the stacks were built, the results show that Stack5, built using algorithms (SVM, RF, LR), is the best among all the stacks, as evidenced by the results of the metrics obtained by testing the model that has been built using test data. It obtained 0.93 accuracies, 0.92 Precision, 0.93 Recal1, 0.92, and F1-Score, allowing it to classify as many diverse webpages as possible.

The stack5 also performed the best out of all the stacks when the training data size was increased to 80% of its original size and the default values for the hyperparameters were used. It had an accuracy of 0.94, a precision of 0.95, a recall of 0.94, and an F1-score of 0.94.

- In this experiment, stack5 (SVM, KNN, NB) achieved higher accuracy compared to other stacks by training 70% of the total data size. The accuracy rate of the stack5 model after computation from the confusion matrix was 94% on the test set 0.94 Precision, 0.94 Recal1, 0.93, and F1-Score. When the training data size was expanded to 80% of its original size, stack 5 performed the best of all stacks. It had an accuracy of 0.95, a precision of 0.95, a recall of 0.95, and an F1-score of 0.95.

Based on the results shown above, we conclude that the stack model with tuned - hyperparameters is the best, with a training sample size estimated to be 80% of the total data size. Compared to the prior model, which was trained using the default hyperparameter values, its performance in categorizing webpages is superior and advanced. This is why it is regarded as the most effective.

10. CONCLUSION

A classifier is learned using labeled data examples with specified classes, then used to estimate the classes of fresh instances. This is a supervised learning problem.

One of the critical difficulties in web mining is categorizing web pages. Considering the enormous amount of information presented, Web applications demand the development of effective classifiers with high prediction performance.

The classification of Webpage means much more than merely categorizing and putting webpages into preset categories using identified data. Focusing on crawling is essential since it facilitates web search and advertising, making it a significant and popular subject. Managing the higher dimensional space and obtaining high predictive performance are two important difficulties that should be adequately handled for effective and reliable web mining applications.

In the experiments that have been done, On the English webpages dataset, ensemble stacking exceeded the individual classifiers for the two mentioned techniques, providing the best (highest) average prediction performance.

Precision, Recall, F1-score, and accuracy provided with the thises on the dataset are 95.044%, 95.035%, 94.952%, and 95.035%, respectively, achieved with the use of tuned hyperparameters and Ensemble Stacking learning.

In addition, among all the compared results, Ensemble Stacking learning with (Support vector machine, KNearest Neighbor, and Naïve Bayes) algorithms yielded better performances than base learning algorithms.

For future suggestions that may be useful to develop a more robust model which leads to automatic webpages classification, the following can be expanded to investigate:

1- This research method used text from the Webpage's body to classify it according to Subject classification. It offers a quick outcome while maintaining accuracy. Future functional and sentiment classification evaluations of this method will need to be done using different data sets.
2- Utilize one of the feature choices, such as the Chi-Squared test or Mutual Information, and continue to analyze the data to discover other characteristics that may enhance classification performance.
3- To get a better outcome, investigations on cutting-edge machine learning algorithms like deep learning will also be considered.

REFERENCES

[1] D. Yogeshwaran and D. Yuvaraj, "Text Classification using Recurrent Neural Network in Quora," *International Research Journal of Engineering and Technology (IRJET)*, vol. 06, no. 2395-0056, 2019.

[2] O. Djuraskovic, "How Many Websites Are There? - Web Stats 2022", *FirstSiteGuide*, 2022. [Online]. Available: https://firstsiteguide.com/how-many-websites/.

[3] M. Mughal, "Data Mining: Web Data Mining Techniques, Tools and Algorithms: An Overview," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, 2018. DOI: 10.14569/ijacsa.2018.090630.

[4] T. Mahmoud, T. Abd-El-Hafeez and D. El-Deen, "A Design of an Automatic Web Page Classification System", *British Journal of Applied Science & Technology*, vol. 18, no. 6, pp. 1-14, 2016. DOI: 10.9734/bjast/2016/30376.

[5] P. Vinod and P. Prajapati, "Comparative Study of Web Page Classification Approaches", *International Journal of Computer Applications*, vol. 179, no. 45, pp. 6-9, 2018. DOI: 10.5120/ijca2018916994.

[6] N. Gali, R. Istodor and P. Fränti, "Functional Classification of Websites", *Proceedings of the Eighth International Symposium on Information and Communication Technology*, 2017. DOI: 10.1145/3155133.3155178.

[7] S. Liu, T. Forss and Kaj-Mikael Bjork. "Web Content Classification with Topic and Sentiment Analysis". *Terminology and Knowledge Engineering*, Jun 2014, Berlin, Germany.

[8]https://www.analyticssteps.com/blogs/binary-and-multiclass-classification-machine-learning.

[9] S. Liu, T. Forss and Kaj-Mikael Bjork. "Web Content Classification with Topic and Sentiment Analysis". *Terminology and Knowledge Engineering*, Jun 2014, Berlin, Germany.

[10] C. Arya and S. Dwivedi, "News web page classification using url content and structure attributes", *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, 2016. DOI: 10.1109/ngct.2016.7877434.

[11] K. Nagaraj, B. Bhattacharjee, A. Sridhar, and S. GS, "Detection of phishing websites using a novel twofold ensemble model," *Journal of Systems and Information Technology*, vol. 20, no. 3, pp. 321–357, 2018.

[12] S. Matic, C. Iordanou, G. Smaragdakis and N. Laoutaris, "Identifying Sensitive URLs at Web-Scale", *Proceedings of the ACM Internet Measurement Conference*, 2020.

[13] D. Deeksha, R. Bhatia, S. Bhardwaj, M. Kumar, K. Bhatia and S. Gill, "Stacking Ensemble-based Automatic Web Page Classification", *2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, 2021. DOI: 10.1109/ccict53244.2021.00042.

[14] K. Kersting, "Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines", *Frontiers in Big Data*, vol. 1, 2018. Available: 10.3389/fdata.2018.00006.

[15] G. S. Ohannesian and E. J. Harfash, "Epileptic Seizures Detection from EEG Recordings Based on a Hybrid System of Gaussian Mixture Model and Random Forest Classifier," *Inform.*, vol. 46, no. 6, pp. 105–116, 2022, doi: 10.31449/inf.v46i6.4203.

[16] A. K. Ali and A. M. Abdullah, "Fake accounts detection on social media using stack ensemble system," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 3013–3022, 2022, doi: 10.11591/ijece.v12i3.pp3013-3022.

[17] M. H. Altamimi, M. A. Aljabery, and I. S. Alshawi, "Big Data in E-government : Classification and Prediction using Machine Learning Algorithms," vol. 1, no. December, pp. 41–55, 2022, doi: 10.52940/ijici.v1i2.11.

[18] Abdulhamit Subasi, Practical Machine Learning for Data Analysis Using Python. Academic Press, 2020.

[19] X. Song, X. Liu, F. Liu and C. Wang, "Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis", International *Journal of Medical Informatics*, vol. 151, p. 104484, 2021. Available: 10.1016/j.ijmedinf.2021.104484 .

[20] Simeone, O. A very brief introduction to machine learning with applications to communication systems. *IEEE Transactions on Cognitive Communications and Networking*, volume 4, no. 4, 2018: pp. 648–664.

[21] B. D. Ripley, " Pattern recognition and neural networks.," *Cambridge University press*, p. (cit. on p. 11). 2007.

[22] G. Matošević, J. Dobša and D. Mladenić, "Using Machine Learning for Web Page Classification in Search Engine Optimization", *Future Internet*, vol. 13, no. 1, p. 9, 2021. Available: 10.3390/fi13010009.
[23] J. P. Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques Third Edition. 2020, Elsevier.
[24] E. Alpaydin, "Neural Networks and Deep Learning," *Mach. Learn.*, 2021, doi: 10.7551/mitpress/13811.003.0007.
[25] I. Hull, *Machine Learning for Economics and Finance in TensorFlow 2: Deep Learning Models for Research and Industry*. Apress, Berkeley ,2021.
[26] Q. Lin, Y. Zhu, S. Zhang, P. Shi, Q. Guo, and Z. Niu, "Lexical based automated teaching evaluation via students' short reviews," Computer Applications in Engineering Education, vol. 27, no. 1, pp. 194–205, 2019, doi: 10.1002/cae.22068.
[27] H. Teguim Kamdjou, "Prediction of Student PerformanceUsing Machine Learning Algorithms", Master Thesis, Heinrich-Heine-University, 2017.
[28] J. Brownlee, *Ensemble Learning Algorithms With Python Make Better Predictions with Bagging, Boosting, and Stacking*, v.1.1. Machine Learning Mastery, 2020.
[29] H. Teguim Kamdjou, "Prediction of Student PerformanceUsing Machine Learning Algorithms", Master Thesis, Heinrich-Heine-University, 2017.
[30] Y. Gupta, G. Raghuwanshi and A. Tripathi, "A New Methodology for Language Identification in Social Media Code-Mixed Text", *Advances in Intelligent Systems and Computing*, pp. 243-254, 2020. Available: 10.1007/978-981-15-3383-9_22.
[31] J. Heaton, *Artificial Intelligence for humans, volume 3: Deep Learning and Neural Networks*, 1st ed. s.l.: Createspace Independent Publishing Platform, 2015.
[32] J. H. Learning, *Python Machine Learning A Crash Course for Beginners to Understand Machine learning, Artificial Intelligence, Neural Networks, and Deep Learning with Scikit-Learn, TensorFlow, and Keras*, 2019.
[33] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents", *International Journal of Computer Applications*, vol. 181, no. 1, pp. 25-29, 2018. Available: 10.5120/ijca2018917395.
[34] L. Cuadros-Rodríguez, Pérez-Castaño, E. & Ruiz-Samblás, C. J. T. T. I. A. C, "Quality performance metrics in multivariate classification methods for qualitative analysis," vol. 80, pp. 612-624, 2016.
[35] F. Hanoon and A. Hassin Alasadi, "A modified residual network for detection and classification of Alzheimer's disease", *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, p. 4400, 2022. Available: 10.11591/ijece.v12i4.pp4400-4407 .
[36] P. RUMMAN, "Detecting fake news with linguistic models and classification algorithms," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, pp. 431-448, 2017.
[37] S. Gilda, "Evaluating machine learning algorithms for fake news detection," *IEEE 21st* International *Conference*, pp. 1-8, 2017.
[38] L. Cuadros-Rodríguez, Pérez-Castaño, E. & Ruiz-Samblás, C. J. T. T. I. A. C, "Quality performance metrics in multivariate classification methods for qualitative analysis," vol. 80, pp. 612-624, 2016.