International Journal for Electronic Crime Investigation

ISSN: 2522-3429 (Print)
ISSN: 2616-6003 (Online)

DOI: https://doi.org/10.54692/ijeci.2024.0804211

Research Article

Vol. 8 issue 4 Oct-Dec 2024

A Novel Methodology for Classifying Wikipedia Articles: Insights into Digital Forensics and Content Integrity

Imran Khan[1], Nadia Tabassum[1*], Muhammad Asim[2], Muhammad Hassan Ghulam Muhammad[3], Mushtaq Niazi[4], Muhammad Farrukh Khan[5], Umer farooq[6]

[1]Department of Computer Science, Virtual University of Pakistan, Lahore, Pakistan,
[2]National College of Business Administration & Economics, Multan, Pakistan,
[3]Department of Computer Science, IMS Pak Aims Lahore, Pakistan,
[4]National College of Business Administration & Economics, Lahore, Pakistan,
[5]Department of Computing, NASTP Institute of Information Technology Lahore Pakistan,
[6]Department of computer science, Hamdard university, Karachi, Pakistan

Corresponding Author: Nadiatabassum@vu.edu.pk

Received: October 15, 2024; **Accepted:** October 28, 2024; **Published:** December 17, 2024

ABSTRACT

Well-written articles shape readers interaction with information, as top-ranked articles are more likely to be seen than those further down the ranks. We present a new approach to classifying Wikipedia articles across various quality dimensions, harnessing knowledge gained from expert assessments. The study also includes an attempt to develop a solid framework that meets the evaluation of the quality of an article for the sole purpose to ensure the integrity of the content in the multipoint structure of the Internet, and to have input for the applications of Digital Forensics. The suggested method: the article details is gathered using the Wikipedia API and a set of metrics is well-defined to store and analyze this information. The methodology then explores the relationship between independent variables (metrics of the articles) and the dependent variable (quality level as rated by the experts). Three machine learning algorithms (RF, J48, and NB) are then used to classify the articles. The classification is dragged along with the expert reviews to determine whether quality level of Wikipedia articles. The empirical evidence illustrates the

effectiveness of the proposed approach, with average accuracies greater than 70% for the J48 algorithm. The precision, recall and F-measure values corresponding to the classification models' accuracy exceed 0.7, representing a strong performance model. Overall, these findings indicate that the method uses reliable criteria, which classifies Wikipedia articles in accordance with experts' opinions, making it a reliable tool for quality assessment. In addition, the study underscores the significance of the combined focus on precision and recall for assessing the quality of a model, thereby demonstrating how useful this method is in ensuring that content can be trusted and as part of digital forensics.

Keywords: Wikipedia, Article Length (in word), Article Age (in days), Number of Edits, Article Viewer, Feature Articles, Good Articles, B-Class Articles, and C-Class Articles

1. INTRODUCTION

There are many types of different encyclopedias available on the internet. However, Wikipedia is the most popular and matchless one, and it is browsed by numerous readers. Wikipedia is a multilingual encyclopedia the Wikipedia articles are available in more than 250 languages. These Wikipedia articles are ranked by the number of articles containing a language. English Wikipedia ranks first because it has a massive number of articles compared to Wikipedia's other languages. Numerous readers use it because it contains massive data on multi-dimensional topics or almost every aspect of life. Moreover, the readers can edit the article making it more meaningful and useful for all and sundry if they want. It is the greatest natural source of information. Because our research is about the English Wikipedia article, if we talk about only the English Wikipedia articles, there are more than 6 million articles, more than 937 million edits are available only in the English version of Wikipedia, and more than 38 million registered users. All published articles in Wikipedia are ranked in different categories like Feature Article (FA), A-Class, B-Class, C-Class, Stub Articles, or Good Articles (GA). The FA has been the highest-ranking due to the eight different things (1) inclusive article, (2) good writing, (3) unbiased, (4) demonstrable and good research, (5) concentration on the article topic and proper article length, (6) unchanging, (7) use the proper style (Wikipedia style standard), (8) suitable pictures [1]. If these things match with an article that article categories with high-ranking mean Feature Article. The expert team categorizes the articles into different

categories. Each Wikipedia language has different grading but these four categories are the same in each Wikipedia language, Feature Article, Good Articles, B, and C class article. Wikipedia has its manual article publish process first the writer submits their article, the initial editor check that article and give them to the junior editors then give the feedback and finally publish the article rank wise. So, there is no automation method to forecast the quality of articles corresponding to the expert opinion. The domain reader read the high-ranked article and ignore the low-ranked article. The novice reader also ignores the other quality dimensions. These quality dimensions help to examine and explain the Wikipedia article quality. The main aim of this research to make a methodology to evaluate the quality of Wikipedia articles to the quality dimensions. In this regard, our research imagined three-fold. Firstly, we design a crawler to extract the article's information. We introduce a set of metrics to record the article's information. Secondly, examine the relationship between the dependent variable and the independent variable. Third, we apply classification algorithms (RF (Random Forrest), J48, and Naïve Bayes) to classify the article according to the quality level suggested by the reviewer expert. The classifier's performance was to be examined with the usage of well-known performance procedures Precision and Recall, Accuracy, F-measure. Classification of Articles: The Wikipedia expert team reviews the article and publishes it in concerned classes. All published articles in Wikipedia are ranked in different categories or classes like Feature Article (FA), A-Class, B-Class, C-Class, Stub Articles, or Good Articles (GA), etc. all of the Wikipedia article language has some common classes/ranking like FA, Good article, etc. In our research, we use only four well-known classes for classification. Wikipedia and its Importance: Wikipedia is a multilingual encyclopedia the Wikipedia articles are available in more than 250 languages. These Wikipedia articles are ranked by the total number of articles that contain a language. It is the greatest natural source of information. Because our research is about the English Wikipedia article if we talk about only the English Wikipedia articles there are more than 6 million articles, more than 937 million total number edits are available only in the English version of Wikipedia, and also more than 38 million of registered users. Because most of the users prefer the Wikipedia article as comparing the other encyclopedia even that the new user also prefers the Wikipedia article. There are some supplementary quality aspects like the length of an article, edits, age, and viewer/watcher of an article. We find that quality dimensions with the help of some literature review [2]. These

quality dimensions help to examine and realize the quality of an article Pearson correlation one method of estimation of the association between two variables there score in an interval or ratio level. Here we want to examine the correlation between the independent variables (AL, NV, NoE, and AA). Linear regression is a model-based technique that is an extension of the Pearson correlation. The regression allows us to do one or more independent variables and see how would predict the score of one dependent variable. We can also call the independent as predicted variables and dependent variables an outcome variable. Why the reader mostly prefers Wikipedia articles? There is some reason the users prefer Wikipedia articles. In Wikipedia there is a massive amount of data available if we only talk about the English Wikipedia article there are more than 6 million articles available and also have a million edits that are the reason the popularity of Wikipedia articles is too higher than the other type of encyclopedias. The article reader can easily edit the article in English Wikipedia articles there are more than 937 million total edits are available. Wikipedia is the most natural resource of information. Wikipedia has its article review process; the review team is competent about its subject. The Wikipedia expert team gets feedback from the other editors. The article that is received from the editor appears on the peer review list. After the review process, the peer review expert nominates an article for a feature article or good article. It is manually processing and articles ranked in different categories. We are introducing a new method that classifies Wikipedia articles concerning the quality dimensions [3]. Often whenever new reader comes to read or targeting the Wikipedia articles, they only observe the Wikipedia article ranking/classes either the article is good, stub, B, C or the feature article. The feature article is a high-ranked article than a good article and so on. The novice reader read the high-ranked article and ignore the low-ranked article. The novice reader also ignores the supplementary quality aspects like the length, edits, age, and viewer/watcher of an article. These quality dimensions help to examine and realize the quality of an article. The learner also faces the difficulty to understand or observe the Wikipedia article quality before pointing the article for reading [4]. Some parameters influence the Wikipedia article quality [5]. We first want to identify these parameters and then use the parameters to define the Wikipedia article classification for the novice reader, so the novice reader understands and observes the Wikipedia article quality and easily selects the article with the help of these quality parameters.

There are main three objectives of

this research.

- The leverage metrics contain some parameters such that Article Viewer/Watcher, Length, Age, and Edits of an article, and that metrics also contain the four ranking/classes C, B, Good, and Feature.

- Analysis and validation examine the relationship between the dependent variable and the independent variable.

- The usage of suggested leverage metrics to classify Wikipedia articles to the quality parameters/dimensions.

There are three research questions given below to fill-full the investigation.

- Research Question 1 (RQ1): Is there an association between the proposed leverage metric (article age, number of edits, number of the viewer, and article length with the article quality?

- Research Question 2 (RQ2): Can we use the quality dimension (Article Length, Article Page Viewer, Number of Edits, and Article Age) to forecast Wikipedia article quality?

- Research Question 3 (RQ3): Does the proposed methodology contain the parameters for the forecasting of the article's quality according to the professional quality analysts?

The fundamental of that research is to investigate and experiment with the Wikipedia articles quality as well as classify the Wikipedia articles with the help of a machine learning algorithm, to facilitate the novice reader to read the famous article to the quality dimensions [6]. The Encyclopedia is the source of information on all the branches and reference work in a comprehensive manner. Encyclopedias are alienated into category-wise, articles or set alphabetically by article name. Encyclopedias have significant worth in the sense of articles, covering articles in almost every field [7]. The role of digital forensics in modern-day tech world revolves around the investigation, recovery, and analysis of digital data stored on electronic devices. It helps solve cybercrimes, protect data, and make sure those involved in legal proceedings get justice. Content integrity (part of digital forensics) Content integrity emphasizes authenticity and the unaltered condition of data. To maintain integrity, it employs cryptographic hash functions, tamper-proof logging, and rigorous chain-of-custody procedures to safeguard evidence for presentation in court. The landscape of digital forensics is evolving with the ever-changing nature of cyber threats,

advancements in technologies such as artificial intelligence and encryption, which pose both opportunities and challenges for digital forensic experts. Cybercriminals use complex techniques to access content and systems, making content integrity essential to fostering trust, security and accountability in digital ecosystems [8]. The correlation between digital forensics and content integrity as tools for protecting the digital ecosystem from rising cyber threats. A systematic identification, preservation, and analysis of electronic data is known as digital forensics, which is used for criminal investigations, incident response, and compliance monitoring. Content integrity, on the other hand, saims to certify data authenticity, proving it is unchanged and trustworthy, which is vital for confidence in digital transactions and communications. The integrity of businesses is a question asked from both sides, for emerging technologies like blockchain, which uses immutable records and transparent data trails, gaining traction. But, there are some challenges put forth such as encrypted data, anti-forensic techniques and the scale of digital environments, which are always in need of some innovative solutions. Hence, digital forensics and content integrity are essential to cyber resilience, enabling the foundations of justice, security, and accountability in the digital age [9] [10].

2. LITERATURE REVIEW

In this paper, the authors [11] examined more than 40 million Wikipedia articles from 55 Wikipedia languages to find out data about more than 200 million reliable and popular sources and references. The author designs theirs owns algorithm in python to extract data from the semantic database for classification which discovers the most consistent in the precise areas. They introduced the 10 models for valuation consistency and admiration of metadata about the author of articles, page views, and references used in Wikipedia articles. One of those models is based on frequency existence, which has already been used. The other nine models were used in several mixtures of events that connected with the admiration of Wikipedia articles and their quality. By using Wikipedia and DBpedia, they automatically recognized the position of the beginning in an explicit field. However, they examine the changes of admiration and consistency in time recognized growing leader in every month. The output result can be used to enhance the quality of Wikipedia articles for several languages [12].

The authors [13] extant their earlier research in which they examined the Wikipedia article editing performance concerning Wikipedia article quality categories and articles

editing and now they increase their examination to edit conflict in Wikipedia articles network, high and low-quality content, and controversial problems. They additionally validate the forecasting potential calculated article metrics, high quality vs. low quality, and conflicted vs. none conflicted. They initiated some research questions and they want to find the answer [14].

They find out how to describe editing performance/behavior in Wikipedia articles.

They also want to find out the relationship between the Wikipedia hyperlink, Wikipedia article quality, and editing behavior on Wikipedia articles and the strength of relation if occur [15].

Are Wikipedia hyperlink and editing behavior for introducing articles edit war, Wikipedia article quality, or disagreement?

For the first question, they classify the 4941 Wikipedia articles via ML algorithms. For the second question, make first-order Markov for every Wikipedia article by calculating comparative frequencies of the edit act. At the last perform some statistical tests to describe the editing behavior, calculate and associate system metrics, Wikipedia hyperlink for article sample, and perform classification.

In this research, the author [9] conduct a wide survey of earlier studies and sum up a broad feature outline with editing history, network, text statistics, readability, writing style, and article structure. They choose the modern deep learning models to measure the Wikipedia article quality. These models are a deep neural network (DNN), CNN-LSTMs, stacked LSTMs, convolutional neural network (CNN), bidirectional LSTMs, and long short-term memory (LSTMs) network. According to the author, the main object of this research is to fill some breaks. First, some articles accept complete features. Second, the thing is there is no such automated system they depend on the human resource to measure the Wikipedia article quality. Third, the lack of proper classification, we cannot identify the performance comparison in Wikipedia article quality. Forth, because of lacking deep learning, anyone not measures the Wikipedia article quality in true characters.

For automatic quality valuation in different languages, the authors [10] present the classification models. According to the authors, that system is based on self-tested and state-of-the-art examinations. In the given research also describe the methods for excellence calculation of infoboxes. Infoboxes are the physical part of Wikipedia articles that hold the information of an article. [16] and their team design the system to find the quality of the ASEAN Wikipedia article. They use the statically feature that can help the user to calculate the quality of an article. The author suggests the feature set for the ASEAN Wikipedia article. They

examine the statically feature so that # of files, # of headings, # of link, length of an article, # of the infobox, and language of an article by the usage of Decision Tree and Naïve Bayes algorithms. They use that same statically feature to evaluate the different types of top five ASEAN languages such as Indonesian, Vietnams, Thai, Philippines, and Malaysian and originate the significant role in quality clustering classification. The [17] and all of the suggested ontology-based classification frameworks for Wikipedia article quality. By the use of ontology, they created the key idea of articles in three areas as an information representation. They use the OAM tool for creating the information, do facts plotting, and by using the set of rules classify the Wikipedia article quality [18] and all of the other authors suggest a new automatic assessment technique of Wikipedia article quality via examining their content in terms of readability score and format features. They use the Wikipedia article quality classes/ranking orderly from low to high, stub article, start, C, B class, Good article, and Feature articles. In this paper introduce the new nine feature and also introduce the new classification model. The research hypothesis is the writing style matter for measuring the article quality. They also use some variables such as Article length in the byte, references, links, the total number of citation templates, etc. for classification. The dataset contains 20489 Wikipedia articles and uses KNN, CART, SVM, RF algorithms for classification. AUC and NDCG are used for performance measurement [19]. Digital forensics have made significant strides in the recent past, presenting new ways of ensuring the body of digital evidence is sound and valid. One such area where there has been some development includes use of blockchain technology within the context of e-discovery market. Researchers suggest a block-based forensic framework sandbox paper "A Framework for Digital Forensics Using Blockchain to Secure Digital Data" that utilizes an immutable ledger system of data stored on a blockchain to address duplication as well as support the security of digital information. Doing so solves the problem of integrity and evidence provenance across jurisdictions, serving as a tamper-proof means of recording forensic evidence [20].

Furthermore, as they employ anti-forensics methods, they have also contributed to the study of the effect of anti-forensics techniques on digital forensics investigations. Overview of Digital Forensics and Anti-Forensics Techniques This study explores the techniques to hide, alter, or destroy digital evidence so that forensic analysis loses its credibility. This information helps forensic professionals to deepen their knowledge about anti-forensics strategies, and develop the proper

methods to counteract and remove them to recover the authenticity of the digital evidence [21, 22].

3. MATERIALS AND METHODS

Wikipedia has its article review process; the review team is competent about its subject. The Wikipedia expert team gets feedback from the other editors. The article that is received from the editor appears on the peer review list. After the review process, the peer review expert nominates an article for a feature article or good article. It is manually processing and articles ranked in different categories. We are introducing a new method that classifies Wikipedia articles to the quality dimensions. Our research methodology is different from the previous researches. We use the Decision Tree and Naïve Bayes algorithm to classify the Wikipedia articles in previous research the most authors use those algorithms. But the methodology is different from ours.

It is shown in Figure I our proposed research model. There are three main phases, Data Analysis, Classification, and Results of the proposed solution discussed in the section of result and discussion.

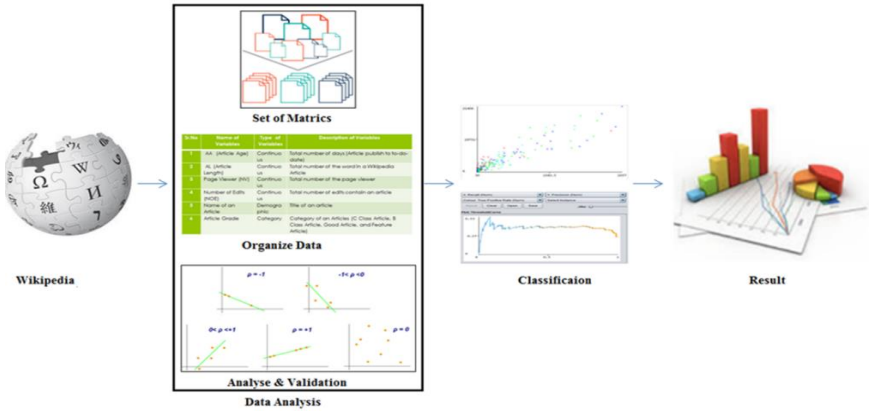

Figure 1: Proposed Methodology

3.1. Phase 1: Data Analysis

In this phase collect the data from the English Wikipedia with the help of Wikimedia API. That phase also consists of three segments. After the collection of Wikipedia article data, design a set of three different types of matrices. Each matric has an equal number of dependent and independent variables but the number of articles different. Then organize the given set of matrices/corpuses if we have required to change or rearrange the data. Finally, we analyze and validate these datasets

[23].

3.2. Set of Matrices

Wikipedia provides the article with statistical information on any article in any language. Wikipedia provides different statistic information gives openly Display title, Page length (in bytes), Page ID, Page content-language, Number of page watchers, Page creator, Date of page creation, Latest editor, Date of latest edit, Total number of edits, Characters, Words of an article, Sections numbers, References, Unique references, Links to this page, Page contributors, Editors, Top editors, etc. The English Wikipedia also provides their article information. We collect statistical information about an article from the Wikipedia statistic page. The English Wikipedia article categorizes the article topic wise the main category is divided into subcategories. These topic wise categories are Government, Health, History, Nature, Business, Crime, Environment, Education, Sports, Biology, Life, Law, etc. We design the three different types of datasets and each dataset has a different topic (the type of article are different from each dataset), these three datasets consist of a total number of 600 Wikipedia English articles with the help of Wikimedia API: Wikipedia API provides the data access of Wikipedia article in different formats (PHP, JSON, none, XML, phpfm, jsonfm).[en.wikipedia.org/w/api.php ?]. Each dataset has a different type of variable, dependent variable, independent variable the article title, and article ranking. The independent variable consists of variables that are Article Edits (total number of edits contain an article), Article Age, Article Page Viewer, and Article Length. When an article publishes by the Wikipedia article expert team the article is ranked (Grading) with different tags in English Wikipedia article it's ranked with Feature Article (FA), Good Article (GA), A-Class, B Class, C Class, Stub, and Start Articles. In this research, we select only high grading first four articles the Feature Article, Good Articles, B Class, and C Class articles in the same size (which are common for all languages of Wikipedia). The dependent variables consist of article grading/tagging that are FA, GA, B, C class articles [24]. Create a corpus that consists of three different topics and 600 total number of articles, 100 articles for education topic, 100 articles for biology topic, and 400 miscellaneous topics articles (Sports, Science, Politics, Technology, Government, Health, History, Nature, Business, etc.). Then design the set of metrics with the help of the literature and perform some tests to classify the Wikipedia article quality. In the given matrices we have three different types of datasets. We describe these datasets one by one. Dataset_1. The dataset-1 contain the total number of 100 Wikipedia article of Biology related topic. The dataset has four dependent variables that are

grading by Feature Article, Good Article, B-Class, and C-Class articles and four quality dimensions variables Article Age (AA), Number of an Edits (NOE), Article Viewer (NV), and the Article Length (AL), also called the independent variable. The dataset has 25 articles of each four grading articles. Dataset_2.

The dataset-2 also contain the total number of 100 Wikipedia article of Education related topic. The dataset has also four dependent variables that are grading by Feature Article, Good Article, B-Class, and C-Class articles and four quality dimensions variables Article Age (AA), Number of an Edits (NOE), Article Viewer (NV), and the Article Length (AL) also called the independent variable. The dataset has 25 articles of each four grading articles. Dataset_3. The dataset-3 consists of 400 miscellaneous topics article including the politic, sports, sciences, technology, and education, and so on. Dataset-3 has also four dependent variables that are grading by Feature Article, Good Article, B-Class, and C-Class articles and four quality dimensions variables Article Age (AA), Number of an Edits (NOE), Article Viewer (NV), and the Article Length (AL) also called the independent variable. The dataset has 100 articles of each four grading articles.

3.3. Organize the Data

After collecting the dataset, we organize it. In the given below Table shown the Wikipedia article variable names, variables type, and description Length of an Article (AL): The length of an article means the total size in the word of an article and its type is continuous.

For example, if an article has 3526 words in the whole article it means the article length is 3526. That value increases or decreases with time because when an editor edits an article it adds or deletes some article text.

Table 1: Organize Data

Sr. No	Variables	Type	Description
1	Length of an Articles (AL)	Continuous	Total numbers of Word the Wikipedia article containing
2	Age of an Articles (AA)	Continuous	Article publish days (publish date)
3	Edits Number (NOE)	Continuous	Total numbers of Edits the Wikipedia article containing
4	Article Viewer (NV)	Continuous	Total numbers of page watcher the Wikipedia article containing
5	Article Title	Demographic	Article name/title
6	Ranking	Category	Article ranking class (FA, GA, C, and B class articles)

Age of an Article (AA): Every Wikipedia publish article has a published data we compute the total number of days. The publish date subtracts from the current date. Its type is continuous.

Total Number of Edits (NoE): In Wikipedia, there are thousands of article editors, one article written or edit from different editors and any person can easily edit an article the Wikipedia computes that edits according to the date, time, editor name, etc. Wikipedia articles have their total number of edits and their type is continuous.

Article Page Viewer (AV): It's also a continuous type. Article page viewer means the total number of users who watch the article, the registered user, or IP users.

Article Title: Its demographic type. Every Wikipedia article has its title. Without the title, the Wikipedia team cannot publish any article. It is the compulsory part of an article.

Ranking/Grading: When an article publishes from the Wikipedia article expert team the article ranked (Grading) with different tags. In English, Wikipedia article is ranked with classes, Star, Stub, A_Class, B_Class, C_Class, Good Articles, and Feature Articles. In this research, we selected/use only high grading the first four articles the Feature Article, C_Class, B_Class, and Good Articles in equal quantity. The dependent variables consist of article grading that are FA, GA, B, C class articles. We assign some value to the selected rank article. We know that the feature article is highly ranked, therefore, assign the value four,

then the second-highest ranked is Good Articles we assign the value three, to B-Class articles give the value two, and last the C-Class article give the value one.

4. ANALYSIS AND VALIDATION

In the analysis and validation phase, we examine some tests on the dependent variable and independent variables. We do statistical methods such as correlating and linear regression to respond to the RQI. When we performing a correlation typically performing an analysis of the hypothesis that one variable is associated with another variable. It could be a positive correlation and the negative correlation and the Pearson correlation one method of estimation of the association between two variables there score in an interval or ratio level. Here we want to examine the correlation between the independent variables (AL, NV, NoE, and AA). The correlation output tables 5.1, 5.3, 5.5 have shown the three figures the total number of items, significance, and the Pearson correlation value. The N value shows the total number of items. Then we have a significant value does this relationship is significant. The value of Pearson correlation differs between 1 and -1. 1.0 is a perfect positive correlation and -1 is a perfect negative correlation [25].

Linear regression is a model-based technique that is an extension of the Pearson correlation. The regression allows us to do one or more independent variables and see how would predict the score of one dependent variable. We can also call the independent a predicted variable and dependent variables an outcome variable. We have a dependent variable here in a category and four independent variables AA, AL, NV, and NoE. We have a total 600 number of records and it's divided into three categories 400 Wikipedia miscellaneous articles, 100 Biology, and 100 education-related Wikipedia articles.

We assign the four different values to the dependent variable because the Feature Article is high ranking, we assign the value four, and Good Article to assign value three, B-Class assign two, and C-Class articles assign the value one. That phase is very important for our research because we depend upon those two variables the dependent variable and the independent variable. To find the relationship between the dependent variable and the independent variable with the help of four hypotheses.

We find the answer given four hypotheses in all the given categories.

Hypothesis I (H-I): Weighty effect of article length (AL) on Wikipedia article rating?

Hypothesis II (H-II): Weighty effect of the number of edits (NoE) on Wikipedia article rating?

Hypothesis III (H-III): Weighty effect of the number of viewers (NV) on Wikipedia article rating?

Hypothesis IV (H-IV): Weighty effect

of age of an article (AA) on Wikipedia article rating?

In multiple regression, there is one dependent variable called category, and the four independent variables AA, AL, NV, and NoE.

4.1. Classification

The classification is the second phase of our research methodology. After the analysis and validation phase, the classification phase makes an important role in that research. First, we find analysis and validation results if the result satisfies or makes a positive relationship between the dependent variable and the independent variable then we classify the quality dimension with article grading/ranking. This phase aims to use the set of proposed metrics to predict the quality levels of Wikipedia articles. There is a lot of classification algorithm Bayes, tree, Meta, lazy, etc., but we use only Bayes and tree algorithm's because the response of the algorithms is well known and well performance. Apply the classification algorithm on datasets. Then we classification the proposed matric to investigate the effectiveness of metrics to measure the Wikipedia article's quality according to quality dimensions. In the classification phase, we find the response of RQ-2 and RQ-3[26] [27] . To respond to the RQ-2 we apply classification algorithms (RF (Random Forrest), J48, and Naïve Bayes) to classify the article according to quality dimensions. The classifier's performance examines (respond to the RQ-3) with the usage of well-known presentation methods Precision, Recall, and Accuracy [28][29].

4.2. Result

The result consisting of hypotheses and research questions. The RQ1 result examines the relationship between the quality dimensions and the Wikipedia article quality. The RQ2 classification of the quality dimensions and finally the RQ3 consist of performance measures. The result is the final phase of our given methodology. In the result phase, we provide all the given results. First, we provide analysis and validation results. In analysis and validation, we perform two tested first to find the correlation and then find the relation between the dependent and independent variables. That result is shown in the graph, selector plots, or table form. Second, we provide the classification results that result also be shown in the graph, selector plots, or table form. We also provide the result of algorithm comparison in the table form.

4.3. Experiment Procedure

In this chapter, we introduce the different pseudocode codes for the experimental procedure. What pseudocode helps us to experiment procedure of our research proposal. The purpose of that pseudocode is to define the process that how the proposed methodology can be working. We follow that pseudocode to experiment with our model.

Relationship between quality dimension variables.

The significant impact of quality

dimension (independent variable) variable on article categories (dependent variable).

Select an accurate classifier according to the expert opinion.

Pseudocode_1 Relationship between quality dimension variables: To find the relationship between the AL, NoE, NV, and AA we use pseudocode_1. That pseudocode gives the help to find out the RQ1.

Input:

- Dataset: Three different types of datasets. Dataset-1, Dataset-2, Dataset-3
- Categories: Feature articles, Good articles, B class articles, C class articles.
- AL: Article length, the total number of words that are contained by an article.
- NV: Total number of page/article viewers.
- NoE: Total number of edits that are contained by an article.
- AA: Article age, the difference between the articles published, and the to-do date. (We can say that the number of days)
- N: The number of articles in the collection. The total number of 600 articles which are divided into three different types of dataset

Procedure Begin:

i. Select dataset-1.
ii. Apply the correlation between AL with NV.
iii. Analysis of the result and find the positive, negative correlation
iv. Repeat that process (1 to 3) for all variables
v. Also, repeat that process for all datasets (dataset-2 and dataset-3).

Procedure End

Output:

The correlation is the best practice to find the relationship between the variables.

Pseudocode_2 Significant impact of quality dimension variable on article categories: The desire pseudocode_2 helps us to find out the relationship between the independent variable with the dependent variable. The independent variable consists of AL, NoE, NV, and AA. The dependent variable consists of article category FA, GA, B, and C class article. That also provides the answers to the desire's hypotheses.

Input:

- Dataset: Three different types of datasets. Dataset-1, Dataset-2, Dataset-3
- Categories: Feature articles, Good articles, B class articles, C class articles.
- AL: Article length, the total number of words that are contained by an article.
- NV: Total number of page/article viewers.
- NoE: Total number of edits that are contained by an article.
- AA: Article age, the difference between the articles published, and the to-do date. (we can say that the number of days)

- N: The number of articles in the collection. The total number of 600 articles which are divided into three different types of dataset

Procedure Begin:

i. With the help of statistical tools find the Linear Regression between the Wikipedia article categories with quality dimension.
ii. Apply the linear regression on categories (Feature articles, Good articles, B class articles, C class articles) with quality dimensions (AL, NV, NoE, AA)
iii. Check out the model summary to find out how much categories value increases or decreases to the quality dimension.
iv. Check out the ANOVA table either the model is significant or not
v. Check out the coefficient table to find out the significant impact of quality dimension on categories.
vi. Apply these processes to each typed of dataset (dataset-2 and dataset-3).

Procedure End

Output:

Linear regression is the best practice to find a significant relationship between the dependent variable and the independent variable.

Pseudocode_3 Select an accurate classifier according to the expert opinion: The classification makes an important role in our research. We find the RQ2 and RQ3 answers in that section. We select three different algorithms to classify the Wikipedia article. In our model, we use the quality dimension parameter to classify the Wikipedia article quality according to the expert opinion.

Input:

- Dataset: Three different types of datasets. Dataset-1, Dataset-2, Dataset-3
- Categories: Feature articles, Good articles, B class articles, C class articles.
- AL: Article length, the total number of words that are contained by an article.
- NV: Total number of page/article viewers.
- NoE: Total number of edits that are contained by an article.
- AA: Article age, the difference between the articles published, and the to-do date. (We can say that the number of days)
- N: The number of articles in the collection. The total number of 600 articles which are divided into three different types of dataset.

Procedure Begin:

i. Select dataset-1
ii. Apply the j48 tree algorithm to the N article in the collection.
iii. Apply the assessment criteria. (F-measure, Accuracy, precision, and recall)
iv. Repeat these processes on each typed of the dataset (dataset-1, dataset-2)
v. Apply and repeat Random Forest, and Naïve Bayes algorithm on each dataset.
vi. Compare the classifier output
vii. Select the best classifier

according to the expert opinion.

Procedure End

Output:

With the help of the Weka statistical tool find the Wikipedia classification. We can easily classify the dependent and independent variables by the tree algorithm.

5. RESULT AND DISCUSSION

Based on the analysis of information, a framework is proposed. The objectives of this framework are to overcome the issues found and to make our system more Secure, Vigilant, and Resilient in [15]terms of security.

In this chapter, we find the answer to the given three research questions and also find the answer to the hypothesis there are three research questions to the fill-full investigation. In the case of studies 1 to 3, we find out the answer of RQ1 with four hypotheses and find out the answer of RQ2 and RQ3 in the classification section. We design three different types of data set with the help of Wiki API. Wikipedia API provides the data access of Wikipedia articles in different formats (PHP, JSON, none, XML, phpfm, jsonfm). [en.wikipedia.org/w/api.php?] and we design a PHP app to fetch the data from Wikipedia.

5.1. Case Study 1:

In this case, the study considers dataset_1, the N value of dataset_1 is 400. It means there are four hundred Wikipedia articles is to be examined.

We want to find the answer to RQ1 from dataset_1. In our model we want two tests to find out the answer of RQ1, the Pearson correlation, and the linear regression. First, we find the relationship between the quality dimensions variables. We are expecting a positive correlation between these variables. If one variable increases the other variable also be increased. The AL with NV the correlation value is 0.628 means its positive correlation is quite strong. If the AL value increases the NV value also be increased. The correlation between AL with NoE also positive and quite strong because the value of correlation is 0.703. The relationship between the AL and AA also be a positive but not good or less positive relationship. The correlation between NV and the NoE is almost perfect because the value between the NV and NoE is 0.934. The value of NV with AA is 0.456 it means the correlation is positive but it's less positive and the value of NoE between AA is 0.494 it’s also a positive but less positive correlation.

We have four hypotheses. For that, we want to find the linear regression between the leverage metric with article quality.

In this case, we can see AA, AL, NV, and NoE variables use to predict the Wikipedia article category. The adjusted R square gives us the percentage of the variant in the dependent variable or outcome variable explained by the independent variable. In this case, 2.2% of all of the variability in the category can be

explained by the independent variable.

ANOVA table says that our model is the significant model if the model is significant nothing else mater. What it means the model is significant we just looking at the sig. value, the sig. value is .013 this is less than .05, therefore, the model is significant

The result shows the Unstandardized, standardized, and p-value first we look at the p-value. The H-I answer is 'yes' because the AL p-value is .001 the given value less than the .05 it means the variable AL is a statistically significant impact on the outcome variable. The H-II, H-III, and H-IV answer is 'no' because the p-value of NV, NoE, and AA are .811, .265, and .980 it means the given all three variables are statically not a significant impact on the outcome variable.

In the unstandardized, if the AL increases the value 1, one unit of change provides the 7.5 change in the dependent variable. In the unstandardized, the NV increases by a value of 1, one unit of change provide the 0.000109 change in the dependent variable. NoE however, works differently from the AL NV. In NoE, we have a negative value of unstandardized coefficient -5.665 so this tells us the NoE increases by one unit the category value decreases by -5.665. The AA works the same as NoE it decreases the value of the dependent variable in the unstandardized and standard coefficient. The AA we have a negative value of unstandardized coefficient -9.292 so this tells us the AA increases by one unit the category value decreases by -9.292. In standard deviation, the AL increases by one standard deviation we have also increase the dependent variable by .246 standard deviation.

In standard deviation, the NV increases by one standard deviation we have also increase the dependent variable by .033 standard deviation. In every standardized deviation of the moment, we see the NoE one standardized deviation of the moment we see variable the dependent variable decreases by -.172 standard deviation. In every standardized deviation of the moment, we see the AA one standardized deviation of the moment we see variable the dependent variable decreases by -.001 standard deviation.

5.2. Case Study 2

This case study considers dataset_2, in this dataset the N value is 100 means there are one hundred education-related Wikipedia articles are to be examined.

We find the relationship between the quality dimensions variables and linear regression to find out the answer to RQ1. The AL with NoE correlation is good positive relation because the value between the variable is 0.748 means it positive correlation and quite a strong relation. When one variable increases the other variable also be increased. The correlation between AL with NV also positive and quite less relationship because the value of correlation is 0.563. The correlation between AL with AA is also positive but quite less relationship because the value between

the AL and AA is 0.482. The relationship between NoE with NV is a strong positive correlation because its correlation value is 0.786 it's an almost linear correlation and the relationship between the NoE with AA is also positive its value is 0.593. The value of NV with AA is 0.510 it means the correlation is positive.

In dataset-2 we have also four hypotheses. For that, we want to find the linear regression between the leverage metric with article quality. The adjusted R square gives us the percentage of the variant in the dependent variable or outcome variable explained by the independent variable. In this case, 24.9% of all of the variability in the category can be explained by the independent variable.

ANOVA resul says that our model is a significant model or not. What it means the model is significant we just looking at the sig value, the sig value is 0.000003 this is less than .05 therefore the model is significant. The unstandardized, standardized, and p-value. The H-I answer is 'yes' because the AL p-value is 0.023. The given value less than 0.05 means the variable AL is a statistically significant impact on the outcome variable.

The H-II answer is 'yes' because the NoE p-value is 0.007. The given value less than 0.05 means the variable NoE is a statistically significant impact on the outcome variable. The H-III answer is 'yes' because the NV p-value is 0.000073. The given value less than 0.05 means the variable NV is a statistically significant impact on the outcome variable. The p-value of AA is 0. 533 that value is greater than the .05 that means the variable is statically not a significant impact on the outcome variable the answer of H-IV is 'no'.

In the unstandardized, if the value of AL increases by 1 value or one unit of change provide the 0.000114 change in the category variable. In the unstandardized variable, NoE increases by a value of 1 or one unit of change provide the 0.000299 means it increases the value of a dependent variable. If the NV increase the one unit it the dependent variable is decreased by 0.006 and if the AA increase by one unit the dependent variable increase by 4.488E-5. In AL, NoE, and the AA if the value increases one unit the outcome value also increases but, in the NV, if the value increases the outcome variable value decrease.

In standard deviation, the AL increases by one standard deviation we have also increase the dependent variable by .305 standard deviation. In every standardized deviation of the moment, we see the NoE one standardized deviation of the moment we see variable the dependent variable increases by 0.503 standard deviations, and in every standardized deviation of the moment we see the NV one standardized deviation of the moment, we see variable the dependent variable decreased by -0.587 standard deviation. The AA one standardized deviation of the moment we see variable the dependent variable also increases

respectively 0.068 standard deviations. In this regression, the AL, NoE, and AA variable works the same means to increase the dependent variable but NV different outcomes from that.

5.3. Case Study 3

In this case study considers dataset_3, in this section, we examine the 100 biology-related Wikipedia articles. That section is also the same variable as above mention.

We also find out the relationship between the quality dimensions variable for RQ-1. We know that significant value tells us either the variable is significant with each other or not. Here our main focus on the correlation because we want to acknowledge that the relationship is positive or negative between the variables.

The correlation value is 0.602 means it positive correlation and is quite a strong relation. If the AL value increases the NoE value also be increased. The correlation between AL with NV also positive and quite less relationship because the value of correlation is 0.471. The correlation between AL with AA is also positive but quite less relationship because the value between the AL and AA is 0.426. The relationship between NoE with NV is a strong positive correlation because its correlation value is 0.883 it's an almost linear correlation and the relationship between the NoE with AA is also positive its value is 0.547. The value of NV with AA is 0.509 it means the correlation is positive. We can also see the scatter plot in figure 5.3 in which we can easily examine the correlation of biology-related Wikipedia article variables. We can see the AA with all other variables, not a good correlation.

We find the linear regression to find out the answer of four hypotheses from dataset-3. The adjusted R square value, in this case, 24.7% of all of the variability in the category can be explained by the independent variable. The model is significant we just looking at the sig. value, the sig. value is 0.000003 this is less than .05 therefore the model is significant.

The unstandardized, standardized, and p-valve. The answer to H-I is 'yes' because the AL p-value is 0.000003. The given value less than 0.05 means the variable AL is a statistically significant impact on the outcome variable. But the H-II, H-III, and H-IV answer is 'no' because the p-value of NV, NoE, and AA is .163, .260, and .448. These values are greater than the .05 that means the given all three variables are statically not a significant impact on the outcome variable.

In the unstandardized, if the value of AL 1 or one unit of change provides the 0.000211 change in the category variable. In the unstandardized variable, NoE increases by a value of 1 or one unit of change that provides the -.000141 its means it decreases the value of a dependent variable. In AL if the value increases one unit the outcome value also increases but in the NoE if the value increases the outcome variable value decrease. The value of

NV and AA also increase the dependent variable value because these variables have a positive value if the NV increase the one unit it the dependent variable also be increased by 0.000682 and if the AA increase by one unit the dependent variable increase by 4.729E-5.

In standard deviation, the AL increases by one standard deviation we have also increase the dependent variable by. 559 standard deviations.

In every standardized deviation of the moment, we see the NoE one standardized deviation of the moment we see variable the dependent variable decreases by -.295 standard deviation. In this regression, the NoE variable works differently from the other variable. The NV and the AA one standardized deviation of the moment we see variable the dependent variable also increase respectively by 0.201 and 0.081standard deviation.

5.4. Classification

In our model, we have selected only three algorithms. The two algorithms belong to decision tree J48 and Random Forrest and one Bayes Naïve algorithm. Each algorithm applies to each dataset/corpus. As we know that we design the three different types of data set the first data set consist of 400 miscellaneous Wikipedia articles states the second dataset consists of 100 Education related Wikipedia articles and the third dataset consist of also 100 Biology related Wikipedia articles. In the above section, we find the RQ1 by the use of correlation of independent variable and linear regression of the dependent variable between the independent variable. The result of these experiments is mention above. Know in this phase we want to classify these variables and find out the answer to RQ2 and RQ3. Here we have some rules to find out the answer of RQ2 and RQ3, for RQ2 if any algorithm classification average result is more than the 70% it means the RQ2 answer is ‘yes’ and by using that algorithm, we can be used the quality dimensions to forecast Wikipedia article quality. For RQ3 if the average performance of precision, recall, and accuracy are more than 0.70 it means the proposed methodology contains the parameters for the forecasting of the article’s quality according to the professional quality analysts.

Dataset_1: J48 Algorithm: Here first we apply the j48 algorithms on dataset one which has 400 miscellaneous Wikipedia articles. The correctly classified instance is 291 its means there are 72.75% of Wikipedia articles that are true classified according to expert reviewers. The incorrectly classified instances are 109 its means there is 27.75% of the classification is incorrect. The precision is 0.740 and the recall is 0.728 and the f-measure is 0.729. When we see the confusion matrix of the feature article out of 100 Wikipedia articles 67 articles are well or same classify the same class. In table 5.7 the good articles' true positive is 85, B class articles are 69 and the C class article is 70 true positive value. In the feature article, 4 articles are predicted

as a good article, 26 articles that are predicated as B class articles, and 3 articles which are predicated as C class articles. In the good articles, 7 articles are predicted as feature articles, 4 articles that are predicated as B class articles, and 4 articles which are predicated as C class articles. In the B class article, 8 articles are predicted as a feature article, 16 articles that are predicated as good articles, and 7 articles which are predicated as C class articles. In the C class article, there is only one article that is predicted as a feature article, 14 articles that are predicated as good articles, and 15 articles which are predicated as B class articles.

Random Forest: Know apply the Random Forest algorithms on the datasets_1 one which has 400 miscellaneous Wikipedia articles.

The correctly classified instance is 198 its means there is 49.5% of Wikipedia articles are truly classified according to expert reviewers. The incorrectly classified instances are 202 its means there are 50.5% classification is incorrect. The precision is 0.491 and the recall is 0.495 and the f-measure value is 0.492. When we see the confusion matrix of the feature article out of 100 Wikipedia articles 61 articles are well or same classify the same class. The good articles' true positive is 52, B class articles are 34 and the C class article is 51 true positive value. In the feature article, 13 articles are predicted as a good article, 19 articles that are predicated as B class articles, and 7 articles which are predicated as C class articles. In the good articles, 17 articles are predicted as feature articles, 17 articles that are predicated as B class articles, and 14 articles which are predicated as C class articles. In the B class article, 25 articles are predicted as a feature article, 20 articles that are predicated as good articles, and 21 articles which are predicated as C class articles. In the C class article, 12 articles are predicted as a feature article, 19 articles that are predicated as good articles, and 17 articles which are predicated as B class articles.

Naïve Bayes Algorithm: It is not a tree algorithm it is a Bayes algorithm. Know we apply that algorithm on the datasets one which has 400 miscellaneous Wikipedia articles. The correctly classified instance is 134 its means there are 33.5% of Wikipedia articles are truly classified according to the expert reviewers. The incorrectly classified instances are 266 its means there is 66.5% of the classification is incorrect. The precision is 0.357 and the recall is 0.335 and the f-measure is 0.278. When we see the confusion matrix of the feature article out of 100 Wikipedia articles only 15 articles are well or same classify the same class. The good articles' true positive is 89, B class articles are 17 and the C class article is 13 true positive value. In the feature article, 15 articles are predicted as good articles, 14 articles that are predicated as B class articles, and 14 articles which are predicated as C class articles. In the good article, 4 articles

are predicted as a feature article, 1 article which is predicated as B class articles, and 6 articles which are predicated as C class articles. In B class article there are 15 articles which are predicted as a feature article, 61 articles which are predicated as good articles and 7 articles which are predicated as C class article. In the C class article there 8 articles which are predicted as a feature article, 71 articles which are predicated as good articles, and 8 articles which are predicated as B class article.

We apply all three algorithms on the Dataset_2 and also on Dataset_3 the result is shown in Table 2 below:

Table 2: Classification Result Table

Datasets	**J48**	**RF**	**NB**
Dataset1	72.75	49.5	33.5
Dataset2	69	36	41
Dataset3	76	51	43

For RQ2 if any algorithm classification average result is more than 70% it means the RQ2 answer is “yes” and by the use of that algorithm, we can be used the quality dimensions to forecast Wikipedia article quality. We selected three different algorithms to apply in our model. The performance of these algorithms is different from each other. When we see the average accuracy of the J48 decision tree algorithm is 72.5% it means our model has a significant classification. In this research in our model, rule-1 is for RQ2 and the result of the J48 algorithm is more than 70% the answer of RQ2 is “yes” we can by using that algorithm we can be used the quality dimensions to forecast Wikipedia article quality. By using this algorithm our model does not identify the 27.5% value correctly. The performance of the Random Forest algorithm is different from the other algorithm. When we see the average accuracy of the Random Forest decision tree algorithm is 45.5% it means our model has not significant classification according to the Wikipedia expert classifier. The answer to RQ2 is ‘no’. By using this algorithm our model does not identify the 54.5% value correctly. The Naïve Bayes performance is too low as compare to the j48 and random forest tree algorithm. When we see the average accuracy of the Naïve Bayes algorithm is 39% it means our model has not significant classification according to the Wikipedia expert classifier. By using this algorithm our model does not identify the 61% value correctly. The answer to RQ2 is ‘no’ we cannot use the desired algorithm to

classify the quality dimensions to forecast Wikipedia article quality, because the average accuracy result of all datasets of the Naïve Bayes algorithm is less than 70%.

For RQ3, if the average performance of precision, recall, and accuracy are more than 0.70 it means the proposed methodology contains the parameters for the forecasting of the article's quality according to the professional quality analysts. When we analyze the p-r curve of the J48 algorithm it is not a bad one it is good. The 1, 1 is the desired graph to produce any good classifier. The classifier is a more upper right corner that classifier is the best classifier that is a perfect classifier. This p-r curve is more to the upper right corner this is not a perfect one but this is not also a bad one so this p-r curve is a really good one. The precision is a true positive divided by a true positive plus a false positive.

$$tp/tp + fp$$

Equation -I

The recall is the opposite measure of precision true positive divided by the true positive plus false negative.

$$tp/tp + fn$$

Equation -II

So, f-measure is a combination of precision and recall. The average result of precision, recall, and f-measure in all three datasets by using the J48 algorithm are respectively 0.751, 0.726, and 0.714 shown Figure.2. These average results affirm rule-2 because it competes for the 0.7, for that reason the RQ3 answer is 'yes' it means the proposed methodology contains the parameters for the forecasting of the article's quality according to the professional quality analysts.

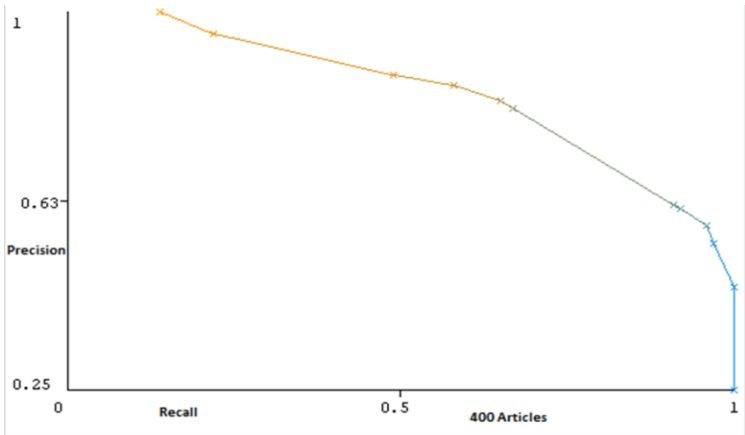

Figure 2: Precision-Recall Graph of J48 Algorithm

Below the Figure II shows a detailed analysis of the Random Forest algorithm. When we analyze the p-r curve it is not perfect because the precision, recall, and f-measure average value are 0.456, 0.455, and 0.451. The 1, 1 is the desired graph to produce any perfect classifier. The classifier is a more upper right corner that classifier is the best classifier that is a perfect classifier. This p-r curve is not perfect because it is shown in the zig-zag graph and less than the center of the graph. These average results did not affirm the rule-2 because the average performance value of the RF algorithm is less than 0.7, for that reason, the RQ3 answer is 'no' its means the proposed methodology does not contain the parameters for the forecasting of the article's quality according to the professional quality analysts with the usage of the desired algorithm.

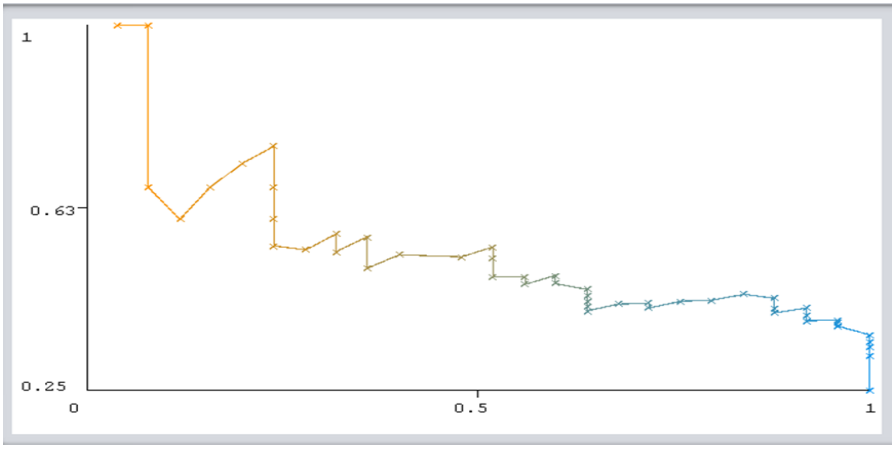

Figure 3: Precision-Recall Graph of the Random Forest Algorithm.

When we analyze the p-r curve of NB it is a bad one it is not good shown in Figure 3. The 1, 1 is the desired graph to produce any good classifier. This p-r curve is more to the upper left corner and in the middle of the graph, this is not good. The average result of all performance measure factor precision, recall, and f-measure explained by figure 4, 5 in all three datasets is less than 0.7, for that reason, the RQ3 answer is 'no' it means the proposed methodology does not contain the parameters for the forecasting of the article's quality with the desired algorithm according to the professional quality analysts

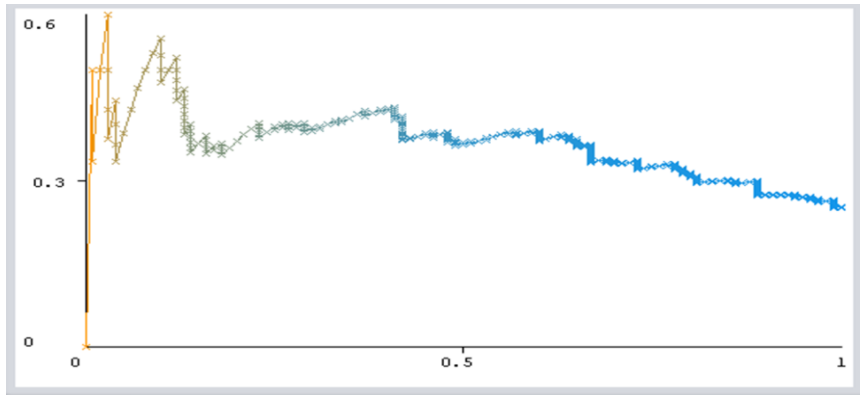

Figure 4: Precision-Recall Graph of Naive Bayes Algorithm.

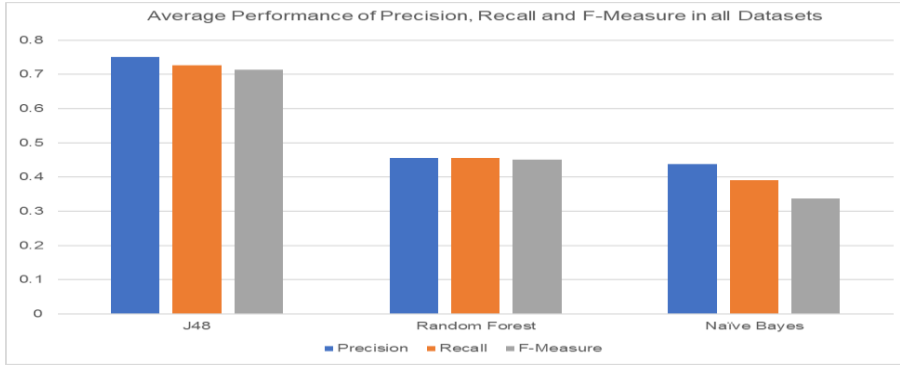

Figure 5: Average Performance of Precision, Recall and F-Measure in all Databases

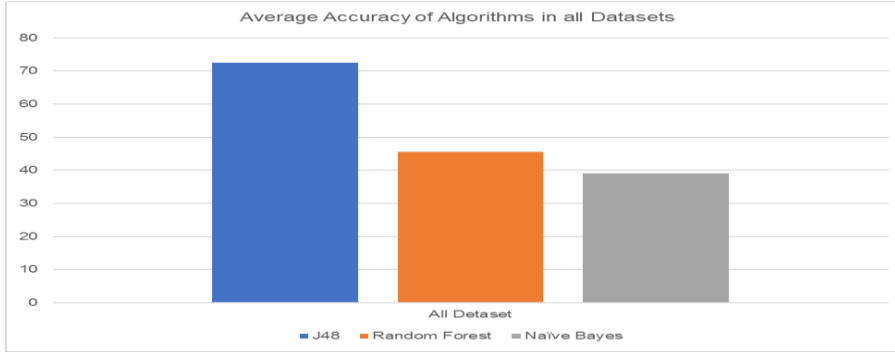

Figure 6: Average Accuracy of Algorithms in All Databases

We can see the accuracy of the J48 algorithm in our model is too much better than the Naïve Bayes and the RF and the precision. The average accuracy is more than 70%. The precision, recall, and f-measure values are also more than 0.7 it means the performance of that model with the J48 algorithm is suitable for Wikipedia article quality according to the expert opinion explained in figure 6.

6. CONCLUSION

We have three case studies each case study has a different dataset. In each case study, we experiment with the Pearson correlation, and linear regression to find out the answer to RQ1. The experimental result of the Pearson correlation on each dataset (dataset-1, dataset-2, and dataset-3) indicates that the correlation between the quality dimensions is significant (positive). If one quality dimension is increased the other variable also be increased. To find out the hypothesis of our research we perform some statistical tests like linear regression. The experimental result of the linear regression on case study 1, the only AL is a significant impact on the category. The H-I answer is 'yes' because the AL p-value is .001 the given value less than the .05 it means the variable AL is a statistically significant impact on the outcome variable. The H-II to H-IV answer is 'no' because the p-value of NV, NoE, and AA are .811, .265, and .980 it means the given all three variables are statically not a significant impact on the outcome variable. The adjusted R square gives us the percentage of the variant in the dependent variable or outcome variable explained by the independent variable. In dataset-1 the 2.2% of all of the variability in the category can be explained by the independent variable. Our model is significant because the ANOVA table value is .013 which is less than .05, therefore, the model is significant. The experimental result of the linear regression on case study 2 there is only AA is not a significant impact on the category it means the H-IV answer is 'no' but the other variables AL, NoE, and AA make a significant impact on the category the H-I, H-II, and H-III answer is 'yes'. The adjusted R square value is different from the dataset-1. The dataset-2, 24.9% of all of the variability in the category can be explained by the independent variable. That model is also significant because the output value is 0.000003. Dataset-3 also a significant model and the R square value is and the only AL that makes a significant impact on the quality of the H-I answer is 'yes'. The other variable that did not make any significant impact on the quality of the H-II to H-IV answer is 'no'. The dataset-3, 24.7% of all of the variability in the category can be explained by the independent variable.

The last section of our research is the classification of the Wikipedia article according to the expert opinion and check the performance of the desired output. For that purpose, we select the

three algorithms J48, Random Forest, and Naïve Bayes. In our model from all three datasets, the average output of the j48 algorithm is or correctly classified instances are 72.5. It means our model is 72.5% correctly classified instances. The average accuracy of the J48 decision tree algorithm is 0.73. It means our model is 73% accurate if we use the J48 tree algorithm. The classification of J48 is better than the other two algorithms. The average output of J48 is above 70% accurate according to the expert opinion. The average output of all three datasets of the RF algorithm is 45.5% and the Naïve Bayes is 39% accurate according to the expert opinion on each dataset. The performance of the J48 algorithm with our model is better because the precision, recall, and f-measure value are more than 0.7 in all three datasets but the other algorithm value is less than 0.5.

REFERENCES

[1] F. F. Acosta, "whoVIS: Visualizing Editor Interactions and Dynamics in Collaborative Writing Over Time," 2015.

[2] A. Kittur and B. Suh, "He Says, She Says: Conflict and Coordination in Wikipedia," CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, AMC, 2007.

[3] B. de La Robertie, Y. Panzarasa, "Measuring Article Quality in Wikipedia using the Collaboration Network," IEEE/ACM International Conference, 2015.

[4] H.-H. Chen, "How to use readability formulas to access and select English reading materials," *Journal of Educational Media and Library*, vol. 50, no. 2, 2012.

[5] G. Wu, M. Hall, "Classifying Wikipedia Articles Using Network Motif Counts and Ratios," ACM, 2012.

[6] H. Liu, E.-P. Lim, "Predicting Trusts Among Users of Online Communities - An Epinions Case Study," Institutional Knowledge at Singapore Management University, Singapore, 2008.

[7] Q. Ignat and V. -L. Lefèvre, "Quality Assessment of Wikipedia Articles Without Feature Engineering," IEEE/ACM Joint Conference on Digital Libraries (JCDL), 2016.

[8] I. Khan and S. Hussain, "An Empirical Study to Predict the Quality of Wikipedia Articles," World CIST, Springer, pp. 8, 2019.

[9] J. Liu, S. Rosen, "Using Big Data and Network Analysis to Understand Wikipedia," 2017.

[10] K. Saengthongpattana and T. Sano, "Ontology-Based Classifiers for Wikipedia Article Quality Classification," Springer Nature Switzerland AG, 2019.

[11] W. Lewoniewski, "Measures for Quality Assessment of Articles and Infoboxes in Multilingual Wikipedia," International Conference on Business Information Systems, Springer, 2019.

[12] X. Li, "Assessing the Quality of Information on Wikipedia: A Deep-Learning Approach," *Journal of the Association for Information Science*

and Technology, vol. 13, 2019.
[13] S. Lim, "How and Why Do College Students Use Wikipedia?," *Journal of the American Society for Information Science and Technology*, 2009.
[14] L. Flekova and O. Ferschke, "What Makes a Good Biography?: Multidimensional Quality Analysis Based on Wikipedia Article Feedback Data," Proceedings of the 23rd International Conference on World Wide Web, AMC, pp. 855–866, 2014.
[15] M. Hu and E.-P. Lim, "Measuring Article Quality in Wikipedia: Models and Evaluation," CIKM '07: Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, AMC, pp. 10, 2007.
[16] Y. Nakamura and S. Shibasaki, "Assessing the Quality of Wikipedia Editors Through Crowdsourcing," *Nara Institute of Science and Technology*, 2016.
[17] N. Japkowicz and M. Shah, "Evaluating Learning Algorithms: A Classification Perspective," Cambridge University Press, 2011.
[18] Q. Su and P. Liu, "A Psycho-Lexical Approach to the Assessment of Information Quality on Wikipedia," IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE, 2015.
[19] R. Conti and E. Marconi, "Maturity Assessment of Wikipedia Medical Articles," IEEE 27th International Symposium on Computer-Based Medical Systems, pp. 6, 2014.
[20] S. Albota and A. Panchenko, "Requirements for the Linguistic Quality Control of Wikipedia Articles," 2019.
[21] S. A. M. Naqvi, T. Alyas, N. Tabassum, A. Namoun, and H. H. Naqvi, "Post Pandemic World and Challenges for E-Governance Framework," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 3, pp. 2630–2636, 2021.
[22] W. Khalid, M. W. Iqbal, T. Alyas, N. Tabassum, N. Anwar, and M. A. Saleem, "Performance Optimization of Network Using Load Balancer Techniques," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 3, pp. 2645–2650, 2021.
[23] T. Alyas, I. Javed, A. Namoun, A. Tufail, S. Alshmrany, and N. Tabassum, "Live Migration of Virtual Machines Using a Mamdani Fuzzy Inference System," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 3019–3033, 2022.
[24] M. A. Saleem, M. Aamir, R. Ibrahim, N. Senan, and T. Alyas, "An Optimized Convolution Neural Network Architecture for Paddy Disease Classification," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 6053–6067, 2022.
[25] J. Nazir., "Load Balancing Framework for Cross-Region Tasks in Cloud Computing," *Computers, Materials and Continua*, vol. 70, no. 1, pp. 1479–1490, 2022.

[26] N. Tabassum, T. Alyas, M. Hamid, M. Saleem, S. Malik, and S. B. Zahra, "QoS Based Cloud Security Evaluation Using Neuro Fuzzy Model," *Computers, Materials and Continua*, vol. 70, no. 1, pp. 1127–1140, 2022.

[27] M. I. Sarwar, K. Nisar, and A. Khan, "Blockchain – From Cryptocurrency to Vertical Industries - A Deep Shift," in IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pp. 537–540, Sept. 20–23, 2019, Dalian, China.

[28] M. Hamza, "Optimizing early detection of diabetes through retinal imaging: A comparative analysis of deep learning and machine learning algorithms," *Journal of Computational Informatics and Business*, vol. 1, no. 1, pp. 1-12, 2024.

[29] M.-H. Zia, A. Hussain, and M. Hamza, "Comparative Analysis of Random Forest and Support Vector Machine Classifiers for unjustified malware detection of Android Devices Data Consuming SMOTE and ROC-AUC Metrics," 2024 Horizons of Information Technology and Engineering (HITE), Lahore, Pakistan, pp. 1-4, 2024