# Quality In-Use of Mobile Geographic Information Systems for Data Collection

Badr El Fhel[1]*, Ali Idri[2]
Mohammed V University in Rabat, Rabat, Morocco[1, 2]
Mohammed VI Polytechnic University[2]

***Abstract*—Mobile Geographic Information Systems (GIS) plays a vital role in data collection, offering diverse functionalities for spatial data handling. Despite advancements, accurately determining the usage environment during development remains challenging. This study uses machine learning and natural language processing to automatically classify user reviews based on the ISO 25010 quality-in-use model. Motivated by the challenge of gauging user experience during development, stakeholders analyze user reviews for insights. An experimental study compares Support Vector Machine (SVM), Random Forest, Logistic Regression, and Naive Bayes classifiers, revealing superior performance by SVM and Random Forest, particularly in efficiency evaluation. Findings underscore the efficacy of SVM in classifying user reviews, emphasizing its effectiveness in evaluating efficiency within mobile GIS applications. Moreover, it provides valuable insights for stakeholders, contributing to the enhancement of software quality of mobile GIS apps.**

***Keywords*—Mobile GIS for data collection; machine learning; software product quality; ISO/IEC 25010; natural language processing; user experience**

## I. Introduction

Mobile GIS has known a significant rise in recent years as a method for data acquisition across diverse disciplines including, but not limited to, environmental monitoring [1], urban planning [2], and emergency management [3]. These GISs allow users to efficiently capture, analyze, and store spatial data related to space, resulting in an increase in productivity compared to traditional methods [4]. The implementation of Mobile GISs can provide significant benefits in terms of cost-effectiveness and real-time data acquisition [5]. In fact, mobile GIS is widely considered for data collection purpose, primarily due to the set of sensors supported by mobile devices that enable capturing positions especially Global Positioning System (GPS) and the Global Navigation Satellite System GNSS. In addition, mobile GIS enable orientation measure through the compass sensor [6]. Moreover, from a data quality point of view, mobile GISs functionalities allow controlling data quality during collection activities [7]; thereby aspect of data quality can be ensured. For instance; the accuracy of data is verified by implementing data validation rules that prevent users from inputting data when the positioning system provides values out of tolerance. Another aspect of data quality is the completeness of data which can be achieved by ensuring that all required items are collected. Finally, the verification of data consistency is achieved through the application of spatial constraints. These constraints serve to alert the user when collected data conflicts with information from other data sources. For instance, an area may be collected as a building, whereas in another data source, it is classified as a farm. These functionalities and features have the potential to influence the attractiveness of the application by partially or fully meeting user's needs. In fact, multiple mobile GIS apps, specifically designed for data collection, are currently available for public use in app repositories [8]. These repositories allow users to provide their feedbacks in the form of ratings and reviews, which are crucial for app developers and designers to improve their services and tailor the applications to meet user needs. However, due to the large number of feed backs and the diversity of wording used, reading and analyzing all reviews and ratings is time consuming manually, thus the need for the automation of this process. Moreover, the quality-in-use evaluation of these apps from the user point of view with respect to (International Standardization Organization) ISO 25010 standard [9] can be a tedious and a difficult task.

Besides, recent technological advancements have resulted in the proliferation of frameworks and libraries for natural language processing (NLP) [10], a specific area within the field of computer science and artificial intelligence that focuses on the comprehension, interpretation, and generation of human language by computers. One widely employed technique in NLP is the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization, which represents text as numerical vectors [11]. When combined with machine learning (ML) classification methods, this technique enables the automated categorization of natural language into predefined classes.

For software quality, the ISO 25010 model provides two distinct models: The first is a software product quality model, which outlines eight characteristics pertaining to the static and dynamic properties of a given specific system or software product. The second is a quality-in-use model which defines the quality in use as the extent to which a product or system can be used by specific users to meet their needs and achieve specific goals with effectiveness, efficiency, freedom from risk, and satisfaction in specific contexts of use. In addition, the quality in use model defines five quality characteristics: (1) effectiveness, which refers to the accuracy and completeness with which users achieve their specified goals; (2) efficiency, which refers to the resources expended in relation to the accuracy and completeness with which users achieve their goals; (3) satisfaction, which refers to the degree to which user needs are satisfied when a product or system is used in a specified context of use; (4) freedom from risk, which refers to the degree to which a product or system mitigates potential

risks to economic status, human life, health, or the environment; and (5) context coverage, which refers to the degree to which a product or system can be used with effectiveness, efficiency, freedom from risk, and satisfaction in both specified contexts of use and in contexts beyond those initially identified.

This study assesses the quality-in-use of mobile GIS for data collection by employing manual labeling, NLP techniques, and term frequency-inverse TF-IDF as pre-processing steps on collected reviews and ratings. Subsequently, ML classification techniques are applied to the pre-processed reviews through an experimental process to identify the most suitable classifier for the specific domain of mobile GIS data collection. The classification of reviews aligns with the quality-in-use model of the ISO 25010 standard.

The study's novel contributions in the field of mobile GIS for data collection can be summarized as follows:

*1)* Proposing a novel application of natural language processing techniques, specifically IF-IDF, for analyzing user reviews in the context of mobile GIS. This approach enables the extraction of valuable insights from a large volume of user-generated data.

*2)* Evaluating the performance of four machine learning techniques - Logistic Regression, Support Vector Machine, Random Forest, and naïve bayes - in classifying user reviews based on the ISO 25010 quality characteristics, with a particular focus on the "efficiency" class (characteristic).

*3)* Comparing the performance metrics of SVM and Random Forest in identifying reviews belonging to the "efficiency" class, showcasing the superior performance of SVM.

*4)* Underlining the significance of SVM as a suitable classifier for classifying mobile GIS user reviews according to ISO 25010, offering better performance in accurately categorizing reviews related to "efficiency."

The paper is organized as follow: Section II provides an overview of the related works. Section III presents the method. Section IV outlines the experimental process, and Section V presents the results of the study. Section VI discusses the findings, and Section VII addresses potential threats to validity. Finally, Section VIII encompasses Conclusion and potential future works.

## II. Related Work

In order to identify the used approaches for analyzing and classifying user reviews and ratings in mobile GISs for data collection, an analysis of previous relevant studies was conducted, with a focus on the type of study (i.e., review or empirical study, etc.), the scope (i.e., the mobile applications of GIS for data collection, or mobile applications in general, etc.), the quality aspects (i.e., quality attributes from ISO 25010 or others), NLP techniques, and ML techniques.

The aforementioned relevant studies are presented in Table I, which indicates that there have been diverse approaches employed to tackle the issue of software quality for both mobile apps in general and mobile GIS specifically for data collection purposes. For instance, Lew et al. [12] employed a modeling framework, 2Q2U (Internal/External Quality, Quality in Use, Actual Usability, and User Experience), to evaluate the quality of a desktop GIS application. This framework adopts a flexible approach to integrate and establish connections between the usability and user experience in order to evaluate software applications. Rahman et al. [13] conducted a study to validate the reliability and validity of an instrument aimed at assessing the influence of GIS quality and user satisfaction on individual work performance. The researchers drew upon an extensive analysis of existing literature and sought input from experts to develop a comprehensive questionnaire consisting of 68 items specifically related to GIS quality, user satisfaction, and individual work performance. In addition, Moumane et al. [14] conducted an empirical study with the objective of assessing the usability of mobile applications on different mobile operating systems. The study aimed to evaluate a framework specifically designed for mobile environments, based on the usability characteristic outlined in the ISO 9126 Software Quality Standard. Meng et al. [15] conducted an assessment of the usability of a Web-based Public Participatory GIS (Web-PPGIS) in a practical application setting. The researchers administered a questionnaire to participants and discovered notable disparities in system usability. These variations were observed based on the users' levels of experience and education. Other related studies have focused on the quality of data in mobile GIS as part of the system. Wang et al. [7] outlined the open architecture of field-based Mobile GIS and emphasized the importance of spatial data quality considerations. The study further elucidated how spatial data quality issues were tackled within the Mobile GIS context, in accordance with internationally recognized geoinformatics standards like ISO and Open Geospatial Consortium (OGC) standards. Furthermore, in another study by Song et al. a linear evaluation model utilizing Geographical Weighted Regression (GWR) and a nonlinear evaluation model based on random forest (RF) were developed [16]. These models were employed to quantitatively assess the relationship between geographical factors and the positioning bias of mobile phone locations.

With respect to the application of ML classification and NLP, Oyebode et al. [17] used ML classification, NLP, and TF-IDF techniques to evaluate and classify 88,125 user reviews in 104 mental health apps based on predefined classes. Five techniques were involved in this study and they are RF, Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), Logistic Regression (LR), and Stochastic Gradient Descent (SGD). Dos et al. [18] conducted a user feedback classifier based on ML of Decision Tree (DT), Naïve Bayes (NB), LR, RF, and SVM for the classification of reviews on mobile apps across various domains. The classification was performed in accordance with software quality characteristics defined by the ISO 25010 standard. In addition, Dias et al. [19] applied ML techniques and NLP in the context of software requirements classification. The study employed four algorithms: LR, SVM, MNB, and kNN. The results indicated that the use of TF-IDF in conjunction with LR produced the best classification results in differentiating requirements.

TABLE I. RELATED STUDIES

| Study ID | Type of study | Scope | Quality aspects | NLP techniques | ML techniques |
|---|---|---|---|---|---|
| Lew et al. [12] | Modelling framework | Desktop GIS | Learnability | - | - |
| Rahman et al. [13] | survey research | GIS | Validity<br>Reliability | - | - |
| Moumane et al. [14] | Empirical study | Mobile apps | Usability | - | - |
| Meng et al. [15] | Empirical study | Web GIS | Usability | - | - |
| Song et al. [16] | Qualitative study | Mobile apps | Spatial accuracy | - | GWR, RF |
| Oyebode et al. [17] | Comparative study | Mobile health apps | Thematic and sentiment analysis | TF-IDF | SVM, MNB, SGD, LR, RF |
| Dos et al. [18] | Algorithm development and evaluation study | mobile apps | External quality characteristics of ISO 25010 | TF-IDF | NB, LR, DT, RF, and SVM |
| Dias et al. [19] | Algorithm development and evaluation study | Software Requirements | (Functional of non-functional) | Bag of Words and TF-IDF | LR,SVM,MNB,KNN |
| Elfhel et al. [20] | Requirements engineering | Mobile GIS for data collection | External quality characteristics of ISO 25010 | - | - |
| Elfhel et al. [21] | Requirements engineering | Mobile GIS for data collection | usability<br>internationalization (i18n)<br>performance efficiency<br>reliability<br>sustainability | - | - |

The authors in [20] has presented a measure of the external quality of mobile GIS for data collection by assessing the degree of impact of requirements related to mobile GIS for data collection on each external quality characteristic, aligned with ISO/IEC 25010. In a separate study, the authors in [21] presented a catalog of requirements for mobile GIS data collection, and demonstrated how it can be used to evaluate such applications.

This study diverges from the aforementioned related work by integrating various dimensions. Notably, while prior studies have explored diverse aspects such as Mobile GIS for data collection, algorithm development, and evaluation, the current study uniquely incorporates and merges these facets. Specifically, the investigation delves into the intersection of Mobile GIS for data collection and the application of both machine learning and natural language processing techniques. In contrast to certain previous studies that addressed the scope of Mobile GIS for data collection but refrained from employing machine learning techniques, this study bridges the gap by incorporating advanced methodologies to automatically classify user reviews based on the ISO 25010 quality-in-use model. This integration enables a more comprehensive understanding of the user experience, contributing a novel perspective to the existing body of literature in this domain. Through the integration of the Mobile GIS scope for data collection with the refined application of machine learning techniques, this study presents a distinctive and valuable contribution to the field, laying the foundation for more refined insights and progress in the evaluation of software quality for mobile GIS applications.

To the best of our knowledge, there have been no prior assessments conducted on the quality in use of mobile GIS for data collection using the ISO 25010 standard, natural language processing (NLP), and machine learning (ML) techniques.

## III. METHOD

The methodology employed in this study comprises five stages, as illustrated in Fig. 1: data collection, data preprocessing, data labeling, data vectorization, automated classification, and evaluation. The subsequent subsections offer a detailed overview of each step in the methodology:

### *A. Data Collection*

During the data collection step, a two-fold approach is used to gather users' reviews on mobile GIS applications for data collection.

- First, a pre-existing list of apps obtained from [8] was utilized, and specific inclusion criteria were applied to determine their selection. Each app needed to satisfy the following inclusion criteria: (1) relevance to mobile GIS for data collection, (2) an update date of 2020 or later, and (3) a minimum of five user reviews.
- Second, a combination of the Google Play API [22] and a Java program, developed by the research team, was utilized to gather user reviews from the selected applications.

As a result, a set of 19 apps were selected in the data collection step with a total of 8,793 reviews collected from these apps (see Table II) for comprehensive list of the selected applications and detailed of collected reviews).

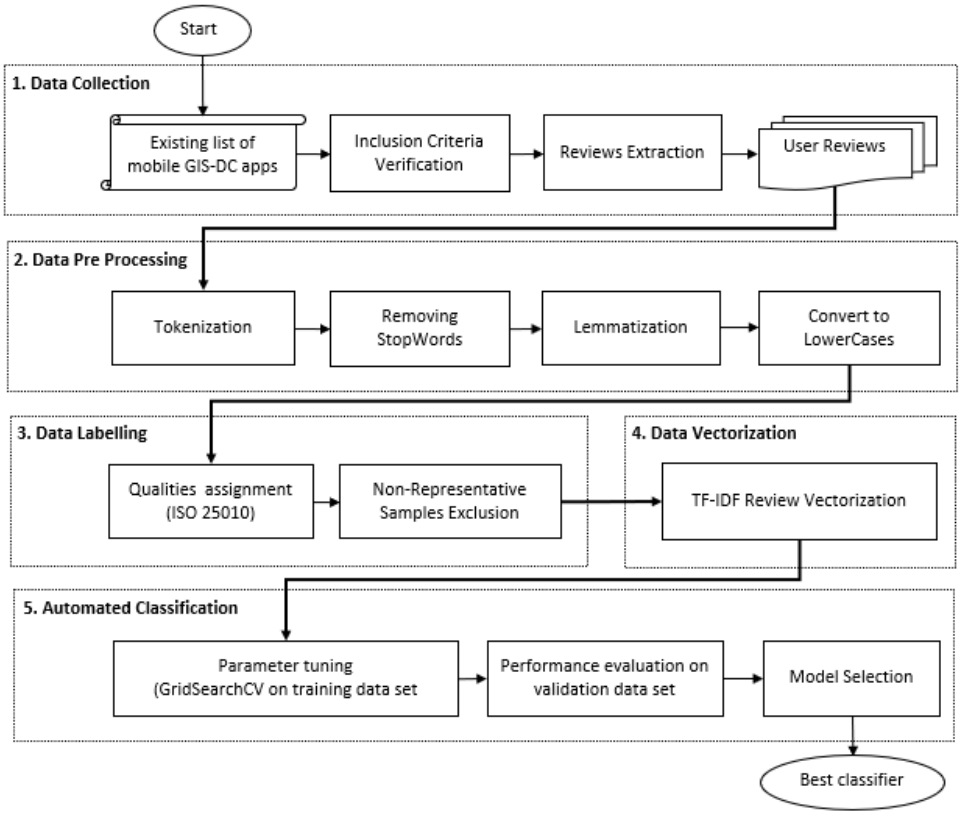


Fig. 1. User reviews classification pipeline.

TABLE II. MOBILE GIS APPS SELECTED

| Application Name & Link | Number of Reviews |
|---|---|
| Mappt: GIS Data Collection | 57 |
| QField for QGIS | 467 |
| Mobile Topographer GIS | 148 |
| My GPS Coordinates | 1570 |
| Mobile Data Collection | 64 |
| GIS Mapper - Surveying App for | 5 |
| GIS Surveyor - Land Survey and | 64 |
| Land Map - GPS Land Survey & M | 34 |
| GPS Coordinates | 3780 |
| Measure map | 109 |
| Mapit Spatial - GIS Data Colle | 15 |
| Geo Survey | 5 |
| NextGIS Mobile | 8 |
| Mapit GIS - Map Data Collector | 456 |
| MapPad GPS Land Surveys | 286 |
| Save Location GPS | 1056 |
| Locus GIS offline land survey | 179 |
| SW Maps - GIS & Data Collector | 388 |
| Epicollect5 Data Collection | 102 |

### *B. Data Preprocessing*

Data preparation is a crucial step in natural language processing (NLP), involving the cleaning and preprocessing of raw text data to eliminate irrelevant information. In order to achieve this, the following well-known steps were followed [23]:

- Tokenization: In NPL, tokenization involves segmenting words into units called tokens based on certain rules such as removing punctuation or capitalization. The resulting tokens are intended to convey a semantic meaning. The tokenization of the collected reviews was achieved by removing punctuation marks, digits, and foreign characters (non-Latin) from the text data.
- Removing stop words: Stop words are commonly occurring words within text data that have little semantic value, such as "the" or "is", and are removed during preprocessing for NLP. The Natural Language Tool Kit (NLTK) package contains a pre-built list of stop words that can be downloaded and used [24]. However, to ensure the inclusion of domain-specific terms in the data analysis, the authors of this study have compiled a list of words related to mobile GIS to prevent them from being removed during preprocessing. This list included for instance "GPS" - a widely-known sensor used for positioning that facilitates data collection via mobile GIS. Other term of “Accuracy" was included in the list as it relates to the precision of positioning, and consequently, the quality of data collected through mobile GIS. Additionally, “Map” was involved in the list as it’s an important component in GIS that allow data presentation.

- Lemmatization in order to reduce words to their base form. For instance, words of "running," "ran," and "run" will be reduced to their base form "run".
- Convert words to lowercase.

The aforementioned steps of data preprocessing were achieved using a python program developed by the authors of this study. For each review in the data set, the program executes successively the operations of Tokenization, removing stop words, Lemmatization and converting to lowercase. The output of these steps is then stored into new column of 'pre-processed-review'.

### C. Data Labelling

The data labelling step consists on the classification of the user reviews (resulted from step 2) through a manual process, which was carried out by the primary author, with respect to the quality characteristics specified in the ISO 25010 model for quality-in-use. For each review, the corresponding predefined quality characteristics are affected by the primary author and then validated by the others authors for relevance and consistency. In cases of disagreement, a consensus was achieved through collective discussion among all authors. The manual process was conducted through a web application that was specifically developed by the research team for this purpose. Fig. 2 depicts the interface of this application, which enables users to navigate through reviews and manually assign quality characteristics to each review by clicking the button related to the corresponding quality. At the end of the data labelling, a comma-separated values (CSV) file that contains the pre-processed-review with the corresponding label is generated using the button CSV. It is noteworthy that during the data labeling process, certain reviews were deemed ambiguous due to their unclear meanings or the presence of non-Latin characters that remained from the data preparation stage. As a result, these reviews were excluded from the data set, resulting in a reduction in the total number of reviews from 7322 to 6904. Table III shows the detailed results in term of reviews and quality characteristics.

### D. Data Vectorization

This step consists of transforming text reviews into numerical values which can then be utilized as input for machine learning classification algorithms. TF-IDF [25] an extensively utilized technique in natural language processing, facilitates the transformation of text data into numerical vectors with a focus on classifying user reviews. This method computes multiplication of the term frequency (TF) with the inverse document frequency (IDF) for each term present in the review, yielding a numerical representation of the significance and rarity of the terms. This numerical representation enables the detection of patterns and trends within user reviews and the subsequent categorization of these reviews according to specific quality characteristics.

TABLE III. DISTRIBUTION OF REVIEWS ACROSS QUALITY CHARACTERISTICS

| Quality Characteristic | Number of reviews |
|---|---|
| Context Completeness | 78 |
| Flexibility | 6 |
| Effectiveness | 2236 |
| Efficiency | 952 |
| Economic Risk Mitigation | 25 |
| Environmental Risk Mitigation | 2 |
| Health and Safety Risk Mitigation | 6 |
| Comfort | 1128 |
| Pleasure | 1653 |
| Trust | 190 |
| Usefulness | 628 |

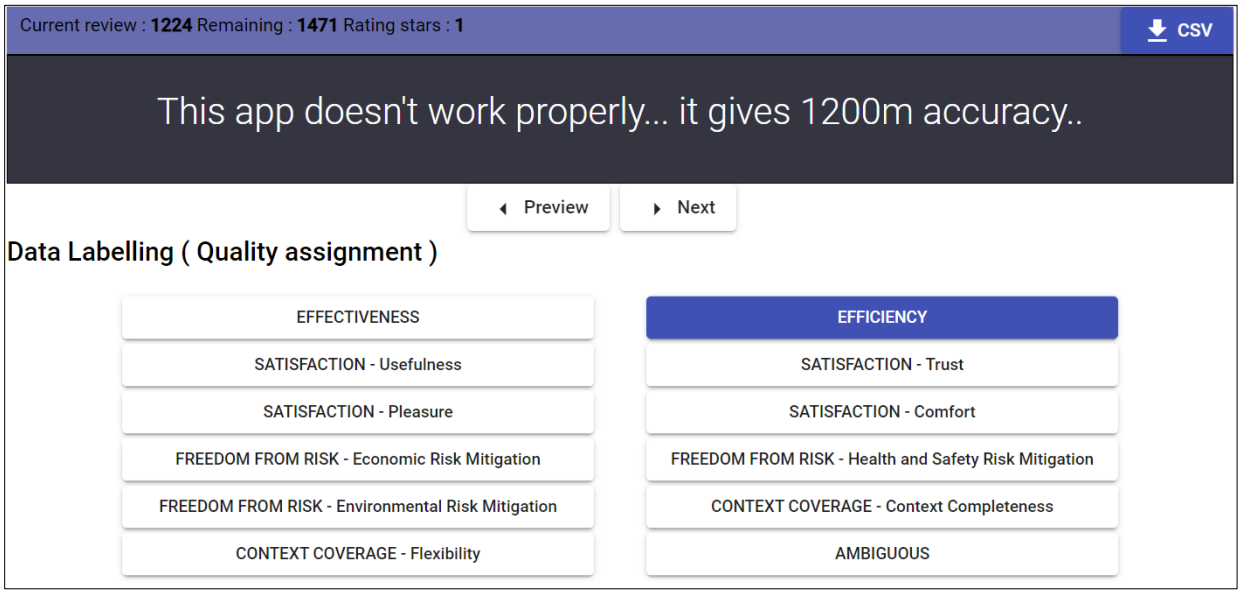


Fig. 2. Screenshot of the data labelling web interface.

In order to apply the TF-IDF vectorization technique on the user reviews, the authors developed a Python script that makes uses of the Scikit-learn [26]. This script reads the CSV file generated during the preceding data labeling phase and computes the frequency of each term in the reviews along with their respective importance scores. The resulting TF-IDF matrix comprises the user reviews in the rows and the overall terms in the columns. Moreover, the script stores the quality characteristic of each review, obtained from the data labeling step, in an additional column labeled "labels". Finally, the output of the script is produced in a new CSV file named "TF-IDF.csv".

### *E. Automated Classification and Evaluation*

The objective of this step is to identify the most suitable machine learning algorithm for classifying user reviews related to mobile GIS for data collection based on quality-in-use characteristics of ISO. To achieve this, the datasets generated through steps 1 to 3 were used as input for the classification methods. Given the impracticality of testing all potential combinations of classification techniques, an experimental study was conducted to automate the testing and evaluation process for each machine learning algorithm's performance.

To summarize, in this study, a dataset of user reviews related to a set of mobile GIS for data collection was obtained. These reviews were subjected to preprocessing utilizing natural language processing methodologies, followed by vectorization utilizing the TF-IDF vectorization technique. A manual labelling process was carried out to classify reviews based on the quality-in-use model of ISO. A dataset with 6904 reviews was obtained and will be used in the experimental study performed in the next section.

## IV. Experimental Study

In this section, an experimental study is conducted to explore the application of machine learning (ML) classification techniques on the pre-processed reviews (obtained from steps 1 to 3 in the previous section). The objective is to identify the best classifier for mobile GIS data collection.

### *A. Dataset Preprocessing*

As shown in Table III, A few quality characteristics within the quality-in-use model have limited or insignificant representation due to the small number of available samples. These qualities are: Context Coverage – Flexibility with only six reviews, Freedom from Risk quality with 2, 6, and 25 reviews respectively to Environmental, Health and Safety, and Economic Risk Mitigation. To maintain the validity and reliability of the model, reviews associated with these particular qualities were subsequently excluded from further analysis. Thus, the dataset has undergone a reduction in the total number of samples from 6904 to 6815.

Furthermore, Fig. 3 presents a statistical analysis of the data related to this study, revealing a notable discrepancy in the sample distribution across different quality characteristics. This discrepancy gives rise to an imbalanced data challenge. To mitigate the issue of imbalanced data, the Synthetic Minority Over-sampling Technique (SMOTE) [27] was utilized to generate synthetic samples.

### *B. Experimental Process*

The experimental process steps used is summarized as the following:

- Four ML techniques are used, namely: (1) Support Vector Machine was introduced by (Vapnik and coworkers) as “a training algorithm that maximizes the margin between the training patterns and the decision boundary” [28]. The SVM classifiers can be improved by modifying the kernel functions (Linear, Polynomial…) and its parameters (C: regulation, gamma: kernel coefficient …) [29]. (2) Logistic Regression is a statistical method applied for classification tasks by analyzing the relationship between a binary variable and one or more independent variables using a logistic function. [30]. (3) Naive Bayes is defined as a simple probabilistic model for classification that assumes that the features are conditionally independent given the class label [31]. The method models the probability of each class given the observed features using Bayes' theorem, and selects the class with the highest probability as the predicted class for a given input. (4) Finally, Random Forest is defined as an ensemble learning method that perform classification by aggregating the predictions of multiple decision trees [32].
- A Grid Search [33] tuning parameter method with five-fold cross-validation was employed to identify the optimal set of hyper-parameters for each technique (see Table IV for the values for GS parameters).

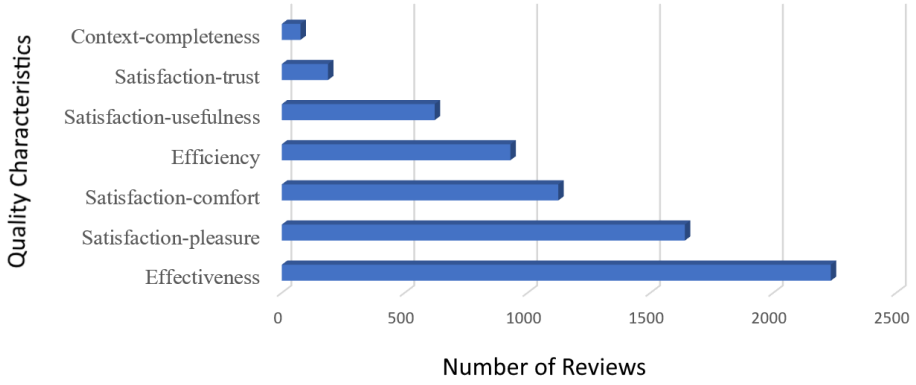


Fig. 3. Distribution of the dataset into quality classes.

TABLE IV. VALUES OF GRID SEARCH PARAMETERS

| Model | Parameter | Values |
|---|---|---|
| Logistic Regression | Regularization Parameter C | 0.1, 1, 10 |
| | Optimization Algorithm | liblinear, lbfgs, saga |
| Support Vector Machine | Regularization Parameter C | 0.1, 1, 10 |
| | Kernel Function | linear, rbf, sigmoid |
| | Gamma | scale, auto |
| | Degree | 2, 3 |
| Random Forest | Number Of Decision Trees | 50, 100, 200 |
| | Criterion | gini, entropy |
| | Max Depth | None, 10, 20 |
| | Min Samples Split | 2, 5 |
| | Min Samples Leaf | 1, 2 |
| | Max Features | sqrt, log2 |
| | Bootstrap | True, False |
| naïve bayes | Alpha | 0.1, 0.5, 1.0 |
| | Algorithm | Multinomial Naïve Bayes |

- A Python script was developed using the Scikit library to achieve optimal classifier performance. The script implements the algorithm depicted in Algorithm 1 and is available upon email request to the author.

**Algorithm 1:** Grid Search for ML Algorithms

Initialize a model-params dictionary of the four ML algorithms and their parameters

Create an empty report array

Compute

For each model in model-params

  Create gvv instance of GridSearchCV with model params and five-fols cross-validation

  Fit gcv with the training set to find the best hyper parameters

  Test the fitted model on the test dataset

  Compute the confusion matrix

  Compute the evaluation metrics

  Add the confusion matrix and the evaluation metrics to the report

  End

Display report

- The performance of the four-classifier experimented in this study was evaluated using four commonly used accuracy criteria [34]: (1) Precision, which quantifies the proportion of true positive predictions among all positive predictions made by the classifier. (2) Recall, which quantifies the proportion of true positive predictions among all actual positive instances. (3) Accuracy, which quantifies the proportion of correct predictions made by the classifier among all instances. (4) F-score, which combines precision and recall into a single score.

## V. RESULTS ANALYSIS

Table V displays the performance of each classifier with respect to all the utilized performance metrics, along with the corresponding optimal values for the hyperparameters.

The results indicated that:

- The Random Forest classifier achieved a precision of 0.81, indicating that, out of all instances that were predicted as positive, 81% were actually positive. The classifier also achieved a recall of 0.79, indicating that, out of all true positive instances, 79% were correctly identified by the classifier. The overall accuracy of the classifier was found to be 0.79, indicating that 79% of the predictions made by the classifier were correct. The F1-score, which is a harmonic mean of precision and recall, was found to be 0.80, indicating that the precision and recall of the classifier was balanced.
- The SVM classifier obtained scores that were slightly different from those of the Random Forest classifier, with a precision score of 0.79, an accuracy score of 0.80, a recall score of 0.80, and an F1-score of 0.79.
- The Logistic Regression classifier performed slightly worse in terms of accuracy and recall, but obtained 0.81 in precision and 0.79 in F1-score.
- The Naive Bayes classifier had the lowest scores across all accuracy criteria, indicating that it performed less well than the other three classifiers.

Moreover, the confusion matrices scores related to SVM and Random Forest were calculated and presented respectively in Table VI and Table VII. As depicted in the confusion matrices, both models demonstrate strong performance. This is evidenced by the majority of entries being located along the diagonal of the matrices.

TABLE V. GLOBAL CLASSIFICATION SCORES AND HYPER PARAMETERS

| Model | Performance scores | | | | Hyper Parameters | |
|---|---|---|---|---|---|---|
| | *Precision* | *Accuracy* | *Recall* | *F1-score* | *Parameter* | *Value* |
| Random Forest | 0.81 | 0.79 | 0.79 | 0.8 | Number Of Decision Trees | 200 |
| | | | | | Criterion | entropy |
| | | | | | Max Depth | None |
| | | | | | Min Samples Split | 2 |
| | | | | | Min Samples Leaf | 1 |
| | | | | | Max Features | log2 |
| | | | | | Bootstrap | False |
| Support Vector Machine | 0.79 | 0.8 | 0.8 | 0.79 | Regularization Parameter C | 10 |
| | | | | | Kernel Function | rbf |
| | | | | | Gamma | scale |
| | | | | | Degree | 2 |
| Logistic Regression | 0.81 | 0.77 | 0.77 | 0.79 | Regularization Parameter C | 10 |
| | | | | | Optimization Algorithm | saga |
| naïve bayes | 0.77 | 0.73 | 0.73 | 0.75 | Alpha | 0.1 |
| | | | | | Algorithm | Multinomial Naïve Bayes |

TABLE VI. SVM CONFUSION MATRIX

| | | Actual Values | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **Context Completeness** | **Effectiveness** | **Efficiency** | **Comfort** | **Pleasure** | **Trust** | **Usefulness** |
| Predicted Values | Context Completeness | 2 | 3 | 3 | 3 | 0 | 0 | 3 |
| | Effectiveness | 0 | 402 | 9 | 14 | 7 | 3 | 13 |
| | Efficiency | 0 | 14 | 164 | 16 | 6 | 2 | 4 |
| | Comfort | 0 | 12 | 19 | 178 | 12 | 2 | 10 |
| | Pleasure | 1 | 3 | 5 | 19 | 275 | 1 | 4 |
| | Trust | 0 | 4 | 10 | 9 | 7 | 7 | 2 |
| | Usefulness | 0 | 10 | 11 | 19 | 10 | 1 | 64 |

TABLE VII. RANDOM FOREST CONFUSION MATRIX

| | | Actual Values | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **Context Completeness** | **Effectiveness** | **Efficiency** | **Comfort** | **Pleasure** | **Trust** | **Usefulness** |
| Predicted Values | Context Completeness | 2 | 3 | 1 | 4 | 1 | 3 | 0 |
| | Effectiveness | 0 | 399 | 10 | 10 | 5 | 13 | 11 |
| | Efficiency | 0 | 11 | 160 | 10 | 3 | 14 | 8 |
| | Comfort | 0 | 10 | 18 | 169 | 6 | 19 | 11 |
| | Pleasure | 2 | 0 | 11 | 13 | 262 | 12 | 8 |
| | Trust | 1 | 1 | 6 | 4 | 6 | 18 | 3 |
| | Usefulness | 1 | 4 | 12 | 12 | 9 | 11 | 66 |

TABLE VIII. SVM AND RANDOM FOREST QUALITY CLASS SCORES

| Classifier | Quality Class | precision | recall | F1_score |
|---|---|---|---|---|
| SVM | Context Completeness | 0.67 | 0.14 | 0.24 |
| | Effectiveness | 0.9 | 0.9 | 0.9 |
| | Efficiency | 0.74 | 0.8 | 0.77 |
| | Comfort | 0.69 | 0.76 | 0.73 |
| | Pleasure | 0.87 | 0.89 | 0.88 |
| | Trust | 0.44 | 0.18 | 0.25 |
| | Usefulness | 0.64 | 0.56 | 0.6 |
| Random Forest | Context Completeness | 0.33 | 0.14 | 0.2 |
| | Effectiveness | 0.93 | 0.89 | 0.91 |
| | Efficiency | 0.73 | 0.78 | 0.75 |
| | Comfort | 0.76 | 0.73 | 0.74 |
| | Pleasure | 0.9 | 0.85 | 0.87 |
| | Trust | 0.2 | 0.46 | 0.28 |
| | Usefulness | 0.62 | 0.57 | 0.59 |

Furthermore, the performance scores related to SVM and Random Forest were calculated for each quality class and the results are presented in Table VIII. As demonstrated, the precision, recall, and F1-score demonstrate heterogeneity across various categories, providing valuable insights into the classification performance of each algorithm. Subsequently, in the following section, these outcomes will be discussed in the context of the criteria for mobile GIS for data collection to select the best classifier from SVM and RF.

## VI. Discussion

Table V illustrates that the accuracy metric for SVM and RF classifiers achieved high values of 0.80 and 0.79, respectively. These results suggest that both classifiers were successful in correctly classifying a high proportion of instances, indicating that the vectorization process utilizing TF-IDF was successful in identifying relevant terms within the corpus of user reviews. Note that TF-IDF was previously identified in research as a strong vectorization method among user reviews [17, 18].

The effectiveness of TF-IDF in mobile GIS for data collection reviews can be explained by its adeptness at capturing term significance through frequency calculations. Within this domain, where reviews frequently incorporate specialized terminology and jargon pertaining to geographic information, mobile devices, and associated technologies, TF-IDF stands out by recognizing and assigning importance to these specific terms based on their frequency. This emphasis on the frequency of domain-specific terms contributes to a more precise representation of the data, aligning with the high accuracy metrics observed in the classifiers' performance as highlighted in Table V.

The SVM classifier and Random Forest classifier were evaluated using precision, accuracy, recall, and F1-score metrics. The results revealed that the Random Forest classifier obtained scores of 0.81, 0.79, 0.79, and 0.80, respectively, while the SVM classifier obtained scores of 0.79, 0.80, 0.80, and 0.79, respectively. These results indicate that the Random Forest classifier performed slightly better in terms of precision and F1-score, while the SVM classifier performed better in terms of accuracy and recall.

The SMOTE technique has been employed to mitigate the issue of class imbalance. however, a detailed analysis of class scores is still necessary to reveal any performance variations of classifiers on specific classes and provide a more comprehensive understanding of their capabilities. Although no significant differences were observed in the four performance scores of the two classifiers, Random Forest and SVM, an exhaustive evaluation of their performance was conducted, taking into account the specific domain of mobile GIS for data collection. In fact, the requirements of mobile GIS for data collection regarding the positioning accuracy is crucial, as it affects directly the quality of collected data [21], which subsequently impacts the overall data collection process. Moreover, a real challenge is associated with GPS positioning accuracy in smartphones [35] and extensive investigations were conducted to identify factors that influence the accuracy of mobile GIS positioning [36-38]. In this light, various solutions have been adopted to enhance the positioning accuracy in mobile mode [39, 40]. Therefore, comparing the performance of Random Forest and SVM classifiers on the class of efficiency can aid in selecting the best classifier for mobile GIS data collection purposes.

Based on the evaluation of the classifiers scores presented in Table VIII, the SVM classifier appears to be a more suitable option for identifying a maximum number of user reviews belonging to the "Efficiency" class in mobile GIS data collection. The SVM classifier exhibits a higher F1-score (0.77), recall score (0.80), and precision (0.74) as compared to the Random Forest classifier (F1-score: 0.75, recall: 0.78, precision: 0.73) for this class. These findings suggest that the SVM classifier has a greater ability to detect positive samples of the "Efficiency" class while maintaining a good balance between precision and recall. Furthermore, the SVM classifier has a higher precision score (0.74) than the Random Forest classifier (0.73) for this class, indicating that the SVM classifier generates fewer false positive predictions. Thus, the SVM classifier may be the optimal choice for this classification task in the mobile GIS data collection domain.

In addition, the complexities inherent in user reviews within the mobile GIS for data collection domain introduce a level of intricacy marked by complex and nonlinear relationships between linguistic expressions and corresponding sentiments. These reviews serve as reflections of nuanced discussions prevailing in this specialized technical domain. Leveraging their unique capacity to define optimal hyperplanes within high-dimensional spaces, SVM exhibit notable proficiency in capturing the nuanced patterns embedded in these reviews. The algorithm's adeptness in recognizing subtle differences and correlations within the technical language of user reviews establishes SVMs as a resilient and effective choice for classifying user-generated content within the intricate realm of mobile GIS for data collection. This underscores their efficacy in addressing the inherent complexities specific to mobile GIS for data collection.

## VII. Threats to Validity

Although objectivity was applied during the research process, there may still be limitations to this study:

- In the natural language processing phase, certain terms may have been erroneously categorized as stop words and consequently eliminated from the dataset. This could impact the construct validity of the study. To address this issue, a specialized GIS term dictionary was constructed to ensure that relevant terms are not automatically removed during the data preprocessing stage, thus improving construct validity.
- The automated classification in this study concerned mobile GIS user reviews, which could pose potential challenges to external validity. To address this concern, the set of studied reviews was carefully chosen to ensure a representative sample. This limitation may have slightly affected the performance metrics, but optimism exists that the results may be utilized in forthcoming studies related to mobile GIS.

- User reviews were assigned manual classifications based on the quality-in-use model. However, there is a possibility that a review may belong to more than one class which impact the internal validity. To address this issue, only the clearest classification was considered.

## VIII. Conclusion and Future Work

This study involved an experiment aimed at identifying the best classifier for analyzing user reviews of mobile GIS applications in the context of data collection. The process involved five steps: data collection, data preprocessing, data labeling, data vectorization, automated classification, and evaluation.

The evaluation of classifiers unveiled notable performance metrics. The Random Forest classifier showcased balanced performance, exhibiting a precision of 0.81, a recall of 0.79, an accuracy of 0.79, and an F1-score of 0.80. The SVM classifier, with slightly differing yet competitive scores, achieved a precision of 0.79, accuracy of 0.80, recall of 0.80, and an F1-score of 0.79. Likewise, the Logistic Regression classifier demonstrated a precision of 0.81, accuracy of 0.79, recall of 0.79, and an F1-score of 0.79, while the Naive Bayes classifier showed lower scores across accuracy criteria. Notably, when honing in on the "efficiency" class, the SVM classifier outperformed the Random Forest classifier, displaying superior precision (0.74), recall (0.80), and F1-score (0.77) compared to the Random Forest classifier (precision: 0.73, recall: 0.78, F1-score: 0.75). These results underscore the effectiveness of the TF-IDF vectorizer and SVM classifier combination within the specific domain of mobile GIS for data collection, emphasizing the significance of efficiency requirements in this context. The implications of this study extend to developers and designers of mobile GIS applications, providing insights for automatic quality evaluation using the ISO 25010 quality-in-use model.

In future investigations, the aim is to expand the scope of the study by increasing the number of experiments conducted. This expansion will enable a more extensive gathering of relevant and accurate results. Additionally, we intend to investigate the correlation between external quality and the quality-in-use of mobile GIS applications specifically designed for data collection purposes, with the ultimate goal of developing a predictive model for quality-in-use. This may have practical implications for enhancing the user experience and satisfaction of mobile GIS applications for data collection by ensuring that external quality meets the requirements of quality-in-use.

## Declaration of Competing Interest

The authors declare that the publication of this article does not involve any conflicts of interest.

## References

[1] M. M. Nowak, K. Dziób, Ł. Ludwisiak, and J. Chmiel, "Mobile GIS applications for environmental field surveys: A state of the art," Global Ecology and Conservation, vol. 23, p. e01089, 2020/09/01/ 2020, doi: https://doi.org/10.1016/j.gecco.2020.e01089.

[2] B. Yang, "Developing a Mobile Mapping System for 3D GIS and Smart CityPlanning,"Sustainability,vol.11,no.13,2019,doi:10.3390/su11133713

[3] I. H. El-Gamily, G. Selim, and E. A. Hermas, "Wireless mobile field-based GIS science and technology for crisis management process: A case study of a fire event, Cairo, Egypt," The Egyptian Journal of Remote Sensing and Space Science, vol. 13, no. 1, pp. 21-29, 2010/06/01/ 2010, doi: https://doi.org/10.1016/j.ejrs.2010.07.003.

[4] A. Jayasinghe, N. Sanjaya, and Y. Chemin, "Application of Mobile GIS for Mobility Mapping," 06/01 2014.

[5] F. Döner, "Examination and comparison of mobile GIS technology for real time Geo-data acquisition in the field," Survey Review, vol. 40, no. 309, pp. 221-234, 2008/07/01 2008, doi: 10.1179/003962608X291013.

[6] Z. Ma, Y. Qiao, B. Lee, and E. Fallon, Experimental evaluation of mobile phone sensors. Signals and Systems Conference (ISSC 2013), 24th IET Irish (in en), 2013.

[7] F. Wang and W. Reinhardt, "Spatial data quality concerns for field data collection in mobile GIS," in Proc.SPIE, 2006, vol. 6420, p. 64201C, doi: 10.1117/12.712733.[Online].Available: https://doi.org/10.1117/12.712733.

[8] B. E. Fhel, A. Idri, and L. Sardi, "Free Mobile Geographic Information Apps Functionalities: A Systematic Review," (in en), RACSC, vol. 16, no.3,p.e200722206911,2023,doi:10.2174/2666255816666220720113157.

[9] Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models, I. S. ISO/IEC/IEEE-25010, 2011.

[10] V. N. Gudivada and K. Arbabifard, "Chapter 3 - Open-Source Libraries, Application Frameworks, and Workflow Systems for NLP," in Handbook of Statistics, vol. 38, V. N. Gudivada and C. R. Rao Eds.: Elsevier, 2018, pp. 31-50.

[11] "Data Mining," in Mining of Massive Datasets, A. Rajaraman and J. D. Ullman Eds. Cambridge: Cambridge University Press, 2011, pp. 1-17.

[12] P. Lew, L. Zhang, and L. Olsina, "Usability and user experience as key drivers for evaluating GIS application quality," in 2010 18th International Conference on Geoinformatics, 18-20 June 2010 2010, pp. 1-6, doi: 10.1109/GEOINFORMATICS.2010.5567803.

[13] M. S. Rahman, S. M. Shuhidan, M. N. Masrek, and M. F. Baharuddin, "Validity and Reliability Testing of Geographical Information System (GIS) Quality and User Satisfaction towards Individual Work Performance,"Proceedings,vol.82,no.1,2022,doi:10.3390/proceedings2022082068.

[14] K. Moumane, A. Idri, and A. Abran, "Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards," SpringerPlus, vol. 5, p. 548,2016, doi: 10.1186/s40064-016-2171-z.

[15] Y. Meng and J. Malczewski, "Usability evaluation for a web-based public participatory GIS: A case study in Canmore, Alberta," cybergeo, 2009/12/17/ 2009, doi: 10.4000/cybergeo.22849.

[16] X. Song, Y. Long, L. Zhang, D. G. Rossiter, F. Liu, and W. Jiang, "Spatial Accuracy Evaluation for Mobile Phone Location Data With Consideration of Geographical Context," IEEE Access, vol. 8, pp. 221176-221190, 2020, doi: 10.1109/ACCESS.2020.3043317.

[17] O. Oyebode, F. Alqahtani, and R. Orji, "Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews," IEEE Access, vol. 8, pp. 111141-111158, 2020, doi: 10.1109/ACCESS.2020.3002176.

[18] R. dos, "A Practical User Feedback Classifier for Software Quality Characteristics," in The 33rd International Conference on Software Engineering and Knowledge Engineering, 2021/07/06/ 2021, pp. 340-345,doi:10.18293/SEKE2021-055.[Online].Available: http://ksiresearch.org/seke/seke21paper/paper055.pdf.

[19] E. Dias Canedo and B. Cordeiro Mendes, "Software Requirements Classification Using Machine Learning Algorithms," (in en), Entropy, vol. 22, no. 9, p. 1057, 2020/09// 2020, doi: 10.3390/e22091057.

[20] B. E. Fhel, L. Sardi, A. Idri, and A. Idri, "Quality Evaluation of Mobile GIS for Data Collection," in 17th International Conference on Evaluation of Novel Approaches to Software Engineering, 2023/01/25/ 2023, pp. 309-316.[Online].Available: https://www.scitepress.org/Link.aspx?doi=10.5220/0011033900003176.

[21] B. El Fhel, L. Sardi, and A. Idri, "A Requirements Catalog of Mobile Geographic Information System for Data Collection," Á. Rocha, H. Adeli, G. Dzemyda, F. Moreira, and A. M. Ramalho Correia, Eds., 2021 2021, Cham: Springer International Publishing, in Advances in

Intelligent Systems and Computing, pp. 324-336, doi: 10.1007/978-3-030-72651-5_32.

[22] Google."API Google Play Developer." https://developers.google.com/android-publisher?hl=fr (accessed March 28, 2023, 2023).

[23] A. Occhipinti, L. Rogers, and C. Angione, "A pipeline and comparative study of 12 machine learning models for text classification," Expert Systems with Applications, vol. 201, p. 117193, 2022/09/01/ 2022, doi: https://doi.org/10.1016/j.eswa.2022.117193.

[24] "NLTK::Natural Language Toolkit.". Available: https://www.nltk.org/. (accessed March 28, 2023).

[25] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: a literature survey," International Journal on Digital Libraries, vol. 17, no. 4, pp. 305-338, 2016/11/01, doi: 10.1007/s00799-015-0156-0.

[26] "scikit-learn: machine learning in Python — scikit-learn 1.2.1 documentation." [Online]. Available: https://scikit-learn.org/stable/. accessed March 28, 2023).

[27] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," jair, vol. 16, pp. 321-357, 2002/06/01/ 2002, doi: 10.1613/jair.953.

[28] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," 1992/07/01/ 1992, New York, NY, USA: Association for Computing Machinery, in COLT '92, pp. 144-152, doi: 10.1145/130385.130401.

[29] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," Neural Networks, vol. 12, no. 6, pp. 783-789,1999/07/01, doi: https://doi.org/10.1016/S0893-6080(99)00032-5.

[30] A. Agresti, Foundations of linear and generalized linear models (Wiley series in probability and statistics). Hoboken, New Jersey: John Wiley & Sons Inc, 2015, p. 1.

[31] M. Lintean and V. Rus, "Large scale experiments with naive bayes and decision trees for function tagging," Int. J. Artif. Intell. Tools, vol. 17, no. 03, pp. 483-499, 2008/06// 2008, doi: 10.1142/S0218213008004011.

[32] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001/10/01 2001, doi: 10.1023/A:1010933404324.

[33] A. Zakrani, A. Najm, and A. Marzak, "Support Vector Regression Based on Grid-Search Method for Agile Software Effort Prediction," in 2018 IEEE 5th International Congress on Information Science and Technology (CiSt), 21-27 Oct. 2018, pp. 1-6, doi: 10.1109/CIST.2018.8596370.

[34] A. Tharwat, "Classification assessment methods," Applied computing and informatics, vol. 17, no. 1, pp. 168-192, 2021.

[35] F. Zangenehnejad and Y. Gao, "GNSS smartphones positioning: advances, challenges, opportunities, and future perspectives," Satellite Navigation, vol. 2, no. 1, p. 24, 2021/11/16/ 2021, doi: 10.1186/s43020-021-00054-y.

[36] C. Bauer, "On the (In-)Accuracy of GPS Measures of Smartphones: A Study of Running Tracking Applications," in International Conference, 2013 2013, Vienna, Austria: ACM Press, pp. 335-341, doi: 10.1145/2536853.2536893.

[37] K. Merry and P. Bettinger, "Smartphone GPS accuracy study in an urban environment," (in en), PLoS ONE, vol. 14, no. 7, p. e0219890, 2019/07/18/ 2019, doi: 10.1371/journal.pone.0219890.

[38] P. A. Zandbergen, "Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning," (in en), Transactions in GIS, vol. 13, pp. 5-25, 2009/06// 2009, doi: 10.1111/j.1467-9671.2009.01152.x.

[39] Z. Peng, Y. Gao, C. Gao, R. Shang, and L. Gan, "Improving Smartphone GNSS Positioning Accuracy Using Inequality Constraints," (in en), Remote Sensing, vol. 15, no. 8, p. 2062, 2023/04/13/ 2023, doi: 10.3390/rs15082062.

[40] J. Hwang, H. Yun, Y. Suh, J. Cho, and D. Lee, "Development of an RTK-GPS Positioning Application with an Improved Position Error Model for Smartphones," (in en), Sensors, vol. 12, no. 10, pp. 12988-13001, 2012/09/25/ 2012, doi: 10.3390/s121012988.