CHAPTER 6
SYSTEM METHODOLOGY

This chapter explains the system methodology, system workflow, algorithm expressions and diagrammatic design for the proposed e-voting system. The system methodology works to predict the best accuracy from the voter's dataset. It also works to identify the possibility of e-voting system based on the growth of literacy and internet by state wise. The main attention of the system is to visualize the statistics report to ECI with minimal classification error.

6.1 PROPOSED ARCHITECTURE

The proposed methodology encourages the e-voting for the public using efficient decision tree algorithm. It increases the voter's turnout, save the voter's time and travelling cost. The method designs a systematic framework for prediction by using the voter's related data. The method is focusing to find out all possible way to improve the voter's turnout with the implementation of a proper e-voting system. Figure 6.1 explains the system workflow of efficient decision tree algorithm with implementation steps and various modules.
6.2 IMPLEMENTATION PRE-PROCESSING STEPS

The method is designed to read ECI- MP (Member of Parliament) data sets for processing the further level. It extracts all related attributes from the ECI data set in particular it concentrates on ECI MP Election 2014 dataset to predict the accuracy and reduce the classification error. The following modules are expressed in designing a framework to implement the e-voting system.
6.2.1 Voters Information

A graphical user interface is designed to record voters information with genetic and communication details. Here, basic and genetic information is recorded with state wise voter’s information. Also, the system asks the voters to enter his/her basic information's with unique identifications numbers to proceed for e-voting.

6.2.2 Feature Selection

Feature Selection is a process to choose some appropriate features which are relevant to the specific criterion. It also removes irrelevant data and thus reduces the dimensionality of space. It is sometimes considered to be NP-hard problems. It helps to increase the performances of e-voting turn out. There are three types of feature selection techniques:

- Filters
- Wrappers
- Embedded Methods

This feature selection process is done with a proper mechanism; it contains 4 steps:
- **Subset Generation** - create the features of subsets of original data set for performance evaluation.

- **Subset Evaluation** - the created subsets evaluation process is done based on some specific relevant criteria.

- **Stopping Criteria** - The feature selection procedure needs to stop with some criteria. Otherwise, it could run exhaustively. Ex. Pre-defined no. of iterations, predefined number feature selection.

- **Validation** - The validation is done by testing both selected and original features to see if there is any improvement.

### 6.2.3 Clustering

The advantages of clustering approaches are as follows:

- Clustering is a partition of data into groups of similar objects.

- Expressing the data by fewer clusters necessarily loses certain fine details, but achieves simplification.

- The similar documents are grouped in a cluster if their cosine similarity measure is less than a specified threshold.
6.2.4 Classification

Classification is the process to predict the voter's turnout from ECI large database objects. It is used to get the classifier parameters where classifier training determines the classifier parameters for each candidate or each type of class and predicts the classification accuracy. The main important step in the classifier system is perhaps feature extraction stage. This stage produces a set of salient features which represents the information needed for the later training stage. In the training stage, those salient features belong to the same classes/objects/patterns are grouped and will be used as the reference for further classification. Then in the classification stage, the same kind of features are extracted and compared with the references obtained in the training stage. Therefore the performance of the classification system depends on the extracted features.

The goal of this research work is how to efficiently extract the salient features for the classification of ECI MP election data set and then to give the classification results to evaluate the corresponding features and hence to visualize the data prediction with predicted accuracy rate and classification error details.

6.2.5 Result Visualization

Finally, this module displays the classes’ entity and graphical representation of voter’s turn out with minimal classification error which
encourages the e-voting systems. It visualizes the result based on hop and node representations. Here, the predicted attributes can be viewed in graphical representations. Also, the classified data can be viewed with their related attribute group wise.

6.3 EFFICIENT DECISION TREE ALGORITHM

The main objective of proposed e-voting algorithm is to reduce the classification error and minimize the retrieval process in comparison with an available data set. This system helps to minimize misclassification error pruning. For each leaf in a tree it is possible to make some instances which are misclassified on a training set by propagating errors. It could be compared with error rate if the leaf was changed by the most common class resulting form. To evaluate the error reduction, sub tree leaf can be considered for pruning. The measurement of error reduction is done for all leaf in a tree and one which have higher reduction error value. This process will continue with new cutting tree until they are unable to find error reduction rate at any leaf. The error is compared with a total instance and training set instances.

The proposed algorithm evaluates normalized information gain that results from selecting an attribute for splitting the data. The splitting process interrupts if all instances in a subset belong to the similar class. Here, the leaf nodes represent the class. In this scenario, the proposed algorithm creates a decision node higher in the tree by using excepted class value. The value is
regenerated for every iteration with a minimal tree. The ranges are measured based on the value and the minimal attributes. This system reduces the retrieval time and data classification times.

The classification is performed on the instances of the training set and a tree is a formation of instances. The main work of pruning is to minimize the classification error. It creates a binary tree. After tree formation, proposed classifier is applied on every tuple in a data set and predicts the classification result for every tuple. The proposed algorithm does not consider the missing value.

### 6.3.1 Features of Proposed e-Voting System

The features of the proposed e-voting system are as follows:

- It accounts missing values and decision tree pruning.
- Proposed algorithm handles both numeric and textual datasets
- It reduces the error rate of data.
- It very effective to predict quality of data
- It designed the rule to identify the particular data generation.
- This system is highly dedicated to predict better accuracy.
This algorithm also handles the missing values in the training data.

This system performs pruning after a design of tree forms. It removes the unwanted branches in a tree which are assisting in reaching the leaf.

**Pseudo code of proposed system**

**Input:** MP Election ARFF datasets

**Output:** Predict (P) the Precision, Recall(R), F-Measure (F), Mean Absolute Error MAE, Root Mean Square Error (RMSE)

**Procedure:**

Step1: Pre-process MP Datasets D

Step2: Extract the MP Election Dataset D;

Step3: Read the nominal, numeric and missing attributes from Dataset D;

Step4: Verify the instance and attribute;

Step5: Extract the feature of attribute from Dataset D;
Step 6: Select the attributes Total Education, male & female education, education growth, computer user, internet user, internet user growth;

Step 7: Check the base class;

Step 8: Find the features of selected attributes a;

Step 9: Check the class label of data set;

if classes have multiple categories

Step 10: Perform tree formation;

Step 11: Find normalized data for every attribute from splitting attribute;

Step 12: Consider highest normalized data;

Step 13: Perform the logistic regression with linear function;

Step 14: Consider every leaf; left tree should contain low value and right tree leaf high value;

Step 15: Allow design the branch;

Step 16: Achieve the leaf node;

Step 17: Reduce the un-wanted sub-tree to improve the instance classification power;
Step18: Combine the formatted tree with linear function to predict accuracy;

Step19: Apply the pruning techniques to reduce the classification complexity;

Step20: Predict the accuracy;

Step21: Predict classification error;

Step22: Display the P, R, F and MAE & RMSE;

Else

Step23: Display exception unable to predict the class;

6.4 DIAGRAMMATIC REPRESENTATION OF PROPOSED APPROACH

A system diagrammatical representation contains an identity of classes, relationship between classes and also their attributes. In objector, the classes are serialized into substance classes and control classes. Computer aided software engineering apparatuses are accessible to outline a class diagram with an assistance of Meta demonstrating. There are some current components available such as a fusion strategy for object modeling technique, classes and exercises.
Here, Objector used the benefit of specialist for the coordinated effort of equipment and programming framework. In Fusion technique, there is no such kind of design, where a customer will supply their requirement in order to complete a bounded period. There are different works where analyst and also creator need to cooperate.

The system proceeds with Rational Rose, an instrument which is feasible to system analyst that makes client case diagram and fashioner for the class outline. The architect engineer can start his work once he/she utilizes case outline from the expert. Once the outline is over, at that point, they continue to select programming for building up the application.

6.4.1 UML Diagrams for e-Voting System

UML stands for Unified Modeling Language for detail, perception and programming design documentation. This system was offered by object management group in 1997. This device contains the following components such as structural, behavioral and architecture structure.

The goal of the system is to make cooperation amongst outside and inward element with dynamic conduct. These inside or outside process are known as performing the activity. It contains performing the activity, their relationship, use case and their process. These diagrams contain the assortment of diagram use case, class, sequence, activity, collaboration, state
transition and deployment diagram. These are very effective for non-IT ability clients who want specialized capacity and process stream of venture or applications. It is not just enlightening for non-IT clients but it's additionally useful for business necessity expert, programming analyzer and despite the fact that for software engineer likewise to comprehend business or customer prerequisite effectively. At last, it gives the ideal opportunity for engineer or developer to find out the stream, condition and operation of the activities.

- **Use Case Diagram for e-Voting System**

Use case diagram is used to present the utilization of the framework amid customer necessities elicitation and examination handle. The most of the part in use case diagram focus on the conduct of the framework from the outer or untouchable perspective. In this outline, an actor is outside the limit, while the utilization cases are inside the limit of the framework. At long last, this diagram helps to the client to perceive the performing activity with their exercises all together shrewd. The details of the interaction between actor and activity are explained in Figure 6.2.
Figure 6.2 Use case diagram for e-Voting System

- **Sequence Diagram for e-Voting System**

  Sequence diagrams make relationship the between formulation and visualization for communicating all objects. The diagram is utilized to identify the extra object which contributes in the use cases. The objective of this diagram is to establish interaction among objects. Figure 6.3 expresses the sequential activity along with objects.
Communication Diagram for e-Voting System

A Communication diagram is one type of diagram that gives connections between objects or concerning sequenced messages. Communication diagram shows the coordination of data taken from class, sequence and use case diagrams to clarify the static structure and dynamic conduct of a framework. It keeps up the grouping of messages, which are recognized in a sequential request and set close-by connect content where subtle elements are given in Figure 6.4.
The objective of class diagram is to clarify the structure of the framework with their properties and operations. It contains certain data about every segment or on-screen characters which help database overseer to make a table for performing activity or segment with particular attributes. Classes are a reflection that determines the normal structure and conduct of a set Class diagram explains the framework regarding objects, classes, characteristics, operations and their affiliations. Figure 6.5 shows the interaction among the object, attributes and operations in details.
Activity Diagram for e-Voting System

An activity diagram explains conditional stream of a framework as far as on-screen characters with their exercises. Activities are states that present to the execution of an arrangement of operations. Activity diagrams are practically similar to flowchart outline and information stream of activities. Here, the action outlines are ordered independently as: The activity diagram details are elaborated in Figure 6.6.
The state diagram is one type of the UML diagram which is used to express the conduct of frameworks. It requires that the framework clarified is formed of a limited number of states. In some situation, this is to be sure, amid other sensible reflection. Many types of state diagram are accessible which contrast somewhat and contain different semantics whose points of interest are explained in Figure 6.7.
Figure 6.7 State diagram for e-Voting System

- Package Diagram for e-Voting System

A package diagram demonstrates the connection between the namespace and bundle. The bringing in namespace determines the names of the bundle procedures of the bundles. A bundle blend has demonstrated the connections between two bundles which show the substances of the two bundles that are to be combined and subtle elements appear in Figure 6.8.
A deployment diagram illustrates the physical organization of artifacts on hubs. The hubs show up as boxes and the artifacts distributed to every hub show up as rectangles inside the cases. Hubs may have sub hubs, which show up as settled boxes. A solitary hub in an organisation outline may reasonably present to different physical hubs, for example, a bunch of database servers. Figure 6.9 explains the execution procedure of proposed strategy.
Figure 6.9 Deployment diagram for e-Voting System

6.4.2 Data Flow Diagram for e-Voting System

The data flow diagram is a graphical expression for communicating framework necessities in an outline. The objective of data flow diagram is to identify the framework necessities and to distinguish the real changes. The data flow diagram comprises a set of bubbles joined by lines. The bubbles are the identification of data transformations and the presentation of the line to data streams in the data flow. A data flow diagram expounds what that information stream in as opposed to how they are handled. Hence data flow diagram can be said that it covers from starting purpose of configuration work
with their usefulness to the most reduced level plan subtle elements. This diagram helps to a client to clarify the customer necessity in a most straightforward way. The data flow diagram of proposed framework is explained in Figure 6.10

![Data flow diagram for e-Voting System](image)

**Figure 6.10 Data flow diagram for e-Voting System**