FUNCTIONAL REQUIREMENTS DOCUMENT

Date	15 October 2022
Team ID	PNT2022TMID39965
Project Name	Project - Emerging Methods for Early Detection of Forest Fires
Maximum Marks	4 Marks

FUNCTIONAL REQUIREMENTS DOCUMENT

(FOREST FIRE DETECTION UNSING AI)

Overview
The functional requirements document (FRD) is a formal statement of an application's functional requirements. It serves the same purpose as a contract. The developers agree to provide the capabilities specified. The client agrees to find the product satisfactory if it provides the capabilities specified in the FRD.

Quality is meeting requirements. For that reason, the FRD is the central document in system development. It is used for the following:

- Designing and developing tile application system.
- Evaluating the product in all subsequent phases of the life cycle.
- Determining the success of the project.

The FRD has the following characteristics:

- It demonstrates that the application provides value to the State in terms of the safety of forest and saving lives
- It contains a complete set of requirements for the application. It leaves no room for anyone to assume anything not stated in the FRD.
- It is solution independent. The ERD is a statement of what the application is to do—not of how it works. The FRD does not commit the developers to a design. For that reason, any reference to the use of a specific technology is entirely inappropriate in an FRD.
- It run with artificial intelligence and provide automated result

A requirement is a condition that the application must meet for the customer to find the application satisfactory. A requirement has the following characteristics:

- It provides a benefit to the government and private sectors.
- It describes the capabilities the application must provide in people safety in terms.
- It does describe how the application provides that capability.

FUNCTIONAL REQUIREMENTS DOCUMENT

- It does describe such design considerations as computer hardware, operating system, and database design.
- It is stated in unambiguous words. Its meaning is clear and understandable.
- It is verifiable and automated using AI

1 GENERAL

1.1 Project Description

Forest fires are a major environmental issue, creating economic and ecological damage while endangering human lives. There are typically about 100,000 wildfires in the United States every year. Over 9 million acres of land have been destroyed due to treacherous wildfires. It is difficult to predict and detect Forest Fire in a sparsely populated forest area and it is more difficult if the prediction is done using ground-based methods like Camera or Video-Based approach. Satellites can be an important source of data prior to and also during the Fire due to its reliability and efficiency. The various real-time forest fire detection and prediction approaches, with the goal of informing the local fire authorities.

1.1.1 Background

- . Evolution emerges in the processing, computation, and algorithms, this strives many researchers to pay attention in many domains where they work in the processing of surveillance video streams so that abnormal or unusual actions could be detected.
- The usage of UAVs is recommended in the detection of forest fire due to the high mobility and ensue the coverage areas at various altitudes and locations at a low cost.
- This can be done even in the 3D model for capturing the scenes.
- The 3D modeling techniques can also be extended to various natural disaster prediction models.
- Focus to meet practical detection and meet the necessity of early detection including the generation of the mixed reality model of the forest fire area that gives more information, and prevention analysis will be made easy.

The some appears as the same as fog, and the model can be classify the fog as smoke so it can be developed with some extra advanced feature in the future.

FUNCTIONAL REQUIREMENTS DOCUMENT

1.1.2 Purpose

2 THERE IS NO NEED FOR THE EXPOSURE OF HUMANS TO PERILOUS ACTIVITIES WHEN REMOTE SENSING IS DEPLOYED.

3 USAGES OF SATELLITE IMAGES TO OBSERVE, DETECT AND REPORT FIRE EVENTS.

4 IMPLEMENTATIONS OF THE WIRELESS SENSOR NETWORK TO OBSERVE THE FIRE EVENTS EXIST IN ALL AREAS.

5 THE AIM OF THIS PROJECT IS TO DEVELOP A MODEL TO DETECT THE FIRE AND ITS COVERAGE AREA, AND IN ADDITION, IT ALSO OBSERVES THE FIRE IN THE LOW REGION.

6 IT CAN BE SIGNIFICANTLY SHORTEN THE REACTION TIME AND ALSO REDUCE THE POTENTIAL DAMAGE AS WELL AS THE COST OF FIRE FIGHTING.

7 BY DETECTING THE FOREST FIRES WE CAN ABLE TO SAVE MORE THAN 1000,000 ACRES , WILD LIVES NATURE.

It saves our future environment for future generation by detecting the forest fires in the earlier manner

7.1.1 Assumptions and Constraints

Forest fires have become one of the major disasters occurring in recent years. The effects of forest fires have a lasting impact on the environment as it led to deforestation and global warming, which is also one of its major cause of occurrence. Forest fires can be dealt by collecting the satellite images of forest and if there is any emergency caused by the fires then the authorities are notified to neutralize its effects. By the time the authorities get to know about condition, the fires would have already caused a lot of damage to the specific sector. By adopting Data mining and machine learning techniques it can provide an efficient prevention approach where data associated with forests can be used for predicting the places with high possibility of forest fires. Numerous algorithms like Logistic regression, Support Vector Machine, Random forest, K-Nearest neighbors in addition to Bagging and Boosting predictors are used, both with and without Principal Component Analysis (PCA). Among the models in which PCA was applied, Logistic Regression gave the highest F-1 score of 68.26 and among the models where PCA was absent, Gradient boosting gave the highest score of 68.36. Geostationary satellite remote sensing systems are a useful tool for forest fire detection and monitoring because of their high temporal resolution over large areas. These computerized system is capable of capturing, storing, analyzing, and displaying geographically referenced information that is, data identified according to location

1) Several techniques for making an intelligent decision were introduced to predict forest fire automatically. Fuzzy reasoning system is one of that is capable of making real time decision and Fuzzy triangular number can better express fuzzy linguistic terms, it has been integrated with multi-attribute decision-making which have been applied in fields such as risk evaluation and

performance evaluation. A fuzzy logic algorithm uses five membership functions as temperature, smoke, light, humidity and distance was introduce to forecast the probability of fire. A decision-making tool was developed using fuzzy logic to designate a fuel model for forest-fires, which is subject to surface fire spread techniques to develop a real time fire prediction system. 2) Before applying the machine learning techniques from it is necessary to investigate the case including the provided raw data-set and the existing method provided by forest fire records. Based on conditions, we introduce four simple traditional machine learning methods with their algorithmic descriptions. In addition, necessary data pre-processing is provided. The fourth algorithm is the k-mean clustering algorithm. It is an unsupervised learning mechanism, so the number of clusters and initial centroid point(s) need to be pre-defined. Algorithm 4 shows the mechanism in detail. 3) Data mining is one such efficient approach in which the forest fires can be predicted based on their past occurrences. Data mining requires an authentic and a clean set of data for prediction. If the dataset is not clean or if there are many unknown values then those values must be taken care of before we use them for modeling. The dataset present in the UCI Machine learning packages about the forest fires is used for prediction. proposed a related work to predict the area burned by 2 the forest fires using the dataset. Initially, the feature 'area' was transformed using ln (1+ x) function. Data mining models were applied and fitted. Post-processing was done on the outputs with the inverse of transform. The experiment ln(x+1) was conducted using 10-fold (cross-validation) x 30 runs.

7.1.2 Interfaces to External Systems

1) The current system consists of Data Mining and sensor which are capable of sensing the smoke and fire. In effect, meteorological conditions (e.g. temperature, wind) are mostly the cause of forest fires and several fire indexes, such as the forest Fire Weather Index (FWI), use such data. In this work, we explore a Data Mining (DM) approach to predict the burned area of forest fires.
2) Five different DM techniques, e.g. Support Vector Machines (SVM) and Random Forests, and four distinct feature selection setups (using spatial, temporal, FWI components and weather attributes), were tested on real-world data collected from the northeast region of Portugal. The best configuration uses a SVM and four meteorological inputs (i.e. temperature, relative humidity, rain and wind) and it is capable of predicting the burned area caused by small fires, which are more frequent. Such knowledge is particularly useful for improving firefighting resource management (e.g. prioritizing targets for air tankers and ground crews)
. 3) Our system consists of high temporal and spatial image to prevent these destructions. Using of Geostationary satellite remote sensing systems which are a useful tool for forest fire detection and monitoring because of their high temporal resolution over large areas. In this, we propose a combined 3-step forest fire detection algorithm (i.e., thresholding, machine learning-based modeling, and post processing) with the help of geostationary satellite.

7.2 Points of Contact

the names, titles, and roles of the major participants in the project. At a minimum, list in following:

- Project Manager
- Development project leader

- User contacts
- Agency employee whose signature constitutes acceptance of the FRD

7.3 Document References

[1] https://ieeexplore.ieee.org/document/6527492 [2] https://www.researchgate.net/publication/ 309758635 _A_symbolic_distributed_event_detection_scheme_ for_Wireless_Sensor_Networks [3] http://agritech.tnau.ac.in/agriculture/agri_majorareas _disastermgt_forestfire.html [4] Distributed Event Detection in Wireless Sensor Networks for Forest Fires Yashwant Singh and Suman Saha Urvashi Chugh and Chhavi Gupta Department of Computer Science and Engineering Jaypee University of Information and Technology Waknaghat, Solan-173245,INDIA {yashwant.singh & suman.saha}@juit.ac.in [5] Detection of Forest Fires using Machine LearningTechnique: A Perspective Aditi Kansal1, Yashwant Singh2, Nagesh Kumar3, Vandana Mohindru4 Department of Computer Science & Engineering Jaypee University of Infonnation Technology Waknaghat, Solan- 173234, (H.P), India 1 aditi.kansaI4@gmail.com, 2yashu _ want@yahoo.com

8 FUNCTIONAL REQUIREMENTS

A.Hardware Specifications
1) MQ2 Gas Sensor
(. MQ2 Gas Sensor Operating Voltage: 5V Preheat Duration: 20 seconds Sensitivity of Digital pin can be varied using potentiometer. Detection Gas: Hydrogen, CO, methane, Alcohol etc. The enveloped MQ2 have 6 pins, 4 of them are used to fetch signals, and other 2 used for providing heating current.)

2) DHT22 Humidity and Temperature Sensor
(Each node is comprised of microcontroller which has the following sensors and peripherals interfaced: • Wi-Fi Module • Smoke/Gas Sensor • Humidity and Temperature Sensor The idea is to create several such nodes. These nodes are distributed over the entire forest area. In our demonstration, we are presenting the functioning of a single node only.)
3) NodeMCU ESP8266 Microcontroller
(• Gini Index—The Gini Index is calculated by subtracting the sum of squared probabilities of each class from one. It satisfies larger partitions and easy to implement whereas information gain satisfies smaller partitions with distinct values [12]. Gini Index= 1—(P(x=k))2 (7) To choose a split a feature with low Gini Index is selected. For the construction of decision tree a classic CART algorithm uses Gini Index.

8.1 Data Requirements

We have created total 5 cluster named as cluster 0, cluster 1, cluster 2, cluster 3, cluster 4 each cluster data are trained using regression model and the features selected are used to find which attributes we need to use in our prediction model. We find the attributes Temperature and RH are the are the best features for doing a prediction model. By including the attributes like temperature,RH,DMC and DC in our prediction model and by splitting into training and testing set help us to get better prediction accuracy By doing this the overfitting of the model is reduced. The training time of the model is reduced because we are eliminating the attributes that are less contributing to the output variable(Area burnt).Here we are using random forest classifiers for

the feature ranking of the attributes. Before building a machine learning algorithm we need to split the data into Train and Test.This split is used to validate the model. The training part is used to create a model and testing part is used to verify the model created.Here the data split is done by separating 70 % to the test data and 30 % to the train data. Then a standard scalar function is introduced on the training set. Standardization of the dataset is the common usage for many machine learning algorithms. The standard transform is applied to both test and training set. Now the these data is loaded into a dataframe.Now the prediction models are implemented using this dataFrame.A confusion matrix is build from which we can calculate True positive.

8.2 Functional Process Requirements

Google Colaboratory — Colaboratory is a free Jupyter notebook environment provided by Google where one can use free GPUs and TPUs which requires no setup and runs entirely in the cloud. The Jupyter Notebook is an open-source web application which allows to create and share documents that contain live code, equations, visualizations and narrative text[11]. A notebook is a list of cells. Cells contain either explanatory text or executable code and its output. With Colaboratory one can write and execute code, save and share their analyses, and access powerful computing resources

- Context
- Detailed view of the processes
- Data (attributes) input to and output from processes
- Logic used inside the processes to manipulate data
- Accesses to stored data
- Processes decomposed into finer levels of detail

9 OPERATIONAL REQUIREMENTS

Operational requirements describe the non-business characteristics of an application.

A. Stage 1 (Data Collection) To realize a supervised learning algorithm the first requirement is a dataset. Hence the first step towards implementation is data collection. An environment resembling the forest situations at the time of initiation of fire. The data of each node is recorded. After certain weeks, set of CSV file is generated. Detection of forest fire requires the features such as temperature, humidity, smoke etc. For our project, we tried two methods to collect the data. We created dummy dataset by taking some samples at room temperature then by burning the leaves and finally the samples were taken when the fire was stopped

B. . B. Stage 2 (Model Generation) This stage is dedicated to training the agent to generate an accurate and flexible model. The same dataset is divided into two parts. One is used for training purpose whereas the other is used for validation of the model. It uses an Decision tree based approach for the classification purpose.

C. C. Stage 3 (Nodes Deployment) All the sensor nodes are deployed over the entire area under surv eillance and are launched. Once commenced, they periodically update the data from sensors on the cloud

D. . D. Stage 4 (Prediction) At the base station, the updated values are used to predict whether a fire like situation is created. An alarm is generated to alert the fire fighter in

case of emergency.

9.1 Security

The security Section describes the need to control access to the data. This includes controlling who may view and alter application data.

the consequences of the following breaches of security in the subject application:

- Erasure of contamination of application data
- Disclosure of Government secrets
- Disclosure of privileged information about individuals

the type(s) of security required. Include the need for the following as appropriate:

there is a need to control access to the facility housing the application.

- the need to control access by class of users. For example, "No user may access any part of this application who does not have at least a (specified) clearance."
- State the need to control access by data attribute. State, for example, if one group of users may view an attribute but may not update it while another type of user may update or view it.
- the need to control access based on system function. State for example, if there is a need to grant one type of user access to certain system functions but not to others. For example, "This function is available only to the system administrator."
- there is a need for accreditation of the security measures adopted for this application. For example, C2 protection must be certified by an independent authorized organization.

9.2 Audit Trail

The correlation matrix is used to find which attribute has a significant correlation on the output target variable. The correlation ranges from Negative and positive. If the correlation is zero then there is no relationship between two attributes. We can see that the temperature has more positive correlation with the area burnt. The wind has more negative correlation with the output variable. The correlation of each cluster are given below

9.3 Data Currency

A. Method 1: Linear Regression The problem here is modelled into a Regression task since over motive is to predict the area of the land burnt.The variable that we need to predict is in numerical value. The regression allows us to model mathematically the relationship between two or more variable. This linear regression model is used to find whether there is a positive or negative relationship between the variables. Normally a regression equation is Y(Dependent variable) = a (intercept) + b (slope of the line) * X (Independent or explanatory variable) B. Method 2: Gradient boosting Gradient boosting is a technique for producing regression models consisting of collections of regressor. It is an ensemble algorithm where the repressor predictions are combined usually by some sort of weighted average or vote in order to provide an over all

prediction. Boosting is a method in which learners are learned sequentially with early learners fitting simple models to the data and analyzing the data for errors . C. Method 3: Bagging Bagging is a classic ensemble method known as bootstrap aggregation. Bagging algorithm consist of many classifiers each uses only some portions of data in each iterations and then combining them through a model averaging techniques. The idea behind this is reduce the over fitting in the class of models. The bootstrap method in bagging creates a random subset of data from a given dataset by sampling D. Method 4: Random forest Random forest is a powerful algorithm which can be used for both regression and classification. The algorithm first creates bootstrap samples from the original data. A regression tree is developed from the each bootstrap samples. Then it randomly sample the number of predictors and the best split is chosen from the variables. Now the aggregation method is used for predicting the new data E. Method 5: SVM regression SVM looks at the extremes of the dataset and draws a decision boundary known as hyper plane near the extreme points in the dataset. It is a method which uses epsilon loss function and performs linear regression in high dimensional space. The SVM always follows a kernel trick where we can use different kernels like RBF, linear, polynomial, Sigmoid

9.4 Reliability

Four machine learning techniques were applied to develop the prediction model i.e. Decision Tree, Classification, Logistic Regression and Random Forest. Model evaluation was performed by calculation some of the performance parameters i.e. Accuracy, Sensitivity, Recall and RMSE. Analysis revealed that Classification is best prediction model based on higher accuracy and lower Root Mean Square value. Soil moisture is a deciding factor whether ignition due to Lightening, Debris and Arson will occur or not So far no major research work has been done in order to assess the dryness of soil in the forest land. Even measurement for dryness of soil is possible but acquiring the data to measure the soil dryness over the vast land which covers the densely populated forest is very difficult. Moreover mobilization of resources and budget allocation needs to be taken care

9.5 Recoverability

Recoverability is the ability to restore function and data in the event of a failure. Answer the following questions in this section:

- In the event the application is unavailable to users (down) because of a system failure, how soon after the failure is detected must function be restored?
- In the event the database is corrupted, to what level of currency must it be restored? For example “The database must be capable of being restored to its condition on no more than one hour before the corruption occurred.”
- If the process site (hardware, data, and onsite backup) is destroyed how soon must the application be able to be restored?

9.6 System Availability

System will be active all the time software and working process is in cloud storage and process 24 hours
If any forest fire accure the detection happen and process through online and sends notification to user

9.7 Fault Tolerance

Fault tolerance is the ability to remain partially operational during a failure. Describe the following in this section:

- Which functions need not be available at all times?
  Internet
- If a component fails what (if any) functions must the application continue to provide? What level of performance degradation is acceptable?
  Level 3 internal erorr and shutdown

For most applications, there are no fault tolerance requirements. When a portion of the application is unavailable there is no need to be able to use the remainder for the application.

9.8 Performance

In order to assess the performance of our model we considered Accuracy, Sensitivity, Precision and RMSE for evaluating our model. The output revealed that Classification has the highest Accuracy compared to other four techniques. Root mean Square parameter is very important when assessing the performance of the prediction model. It basically reveals the error encountered by prediction model based on the predictor variable. Both Decision Tree and Classification has the lowest RMSE value i.e. (0.1) followed Random Forest (1.3).Logistic Regression yields the highest RMSE value (2.9). Lower values of RMSE indicate better fit. RMSE is a good measure of how accurately the model predicts the response, and is the most important criterion for fit if the main purpose of the model is prediction.

9.9 Capacity

List the required capacities and expected volumes of data in business terms. For example, state the number of cases about which the application will have to store data. For example, "The project volume is 600 applications for naturalization per month." State capacities in terms of the business. Do not state capacities in terms of system memory requirements or disk space.

9.10 Data Retention

Dataset has been acquired from kaggle.com which contains a repository for dataset. The dataset is mainly used for research and academic purposes. The dataset for our research contains data related to the occurrence of Wild fires in US. The dataset was extracted as SQLite file from kaggle.com which is labelled as FPA_FOD_20170508.sqlite. The file was opened with SQLite Studio. SQLite Studio is a basically a Graphical User Interface for writing and debugging SQL queries using SQLite database.FPA_FOD_20170508.sqlite file contain several database tables.

10 REQUIREMENTS TRACE ABILITY MATRIX

The requirements trace ability matrix (RTM) provides a method for tracking the functional

requirements and their implementation through the development process. Each requirement is included in the matrix along with its associated section number. As the project progresses, the RIM is updated to reflect each requirement's status. When the product is ready for system testing, the matrix lists each requirement, what product component addresses it, and what test verify that it is correctly implemented.

Include columns for each of the following in the RTM:

- Requirement description
- Requirement reference in FRD
- Verification Method
- Requirement reference in Test Plan

11 GLOSSARY

Forest fires are occurring throughout the year with an increasing intensity in the summer and autumn periods. Regardless of the reasons for the ignition of the forest fires,they usually cause devastating damage to both nature and humans. To fight forest fires, different solutions were employed throughout the years.They were primarily aimed at the early detection of the fires.The simplest of the solutions is the establishment of a network of observation posts both cheap and easy to accomplish, but also time-consuming for the involved people.ICT-based networks of cameras and sensors and even satellite-based solutions were developed and used in the last decades.These solutions have greatly decreased the forest fires in the forest, but have also proven to be expensive and hard to maintain.En-number of algorithms and techniques were used here to detect the fire.This project will involve the development and in the use of the both systems and will analyze the advantages and the benefits, were everyone will gain the best simple solutions.