Water Logging Prediction

Nitin Pandey
nitin.pandey@ves.ac.in
D17B , 52
CMPN , VESIT

Neeraj Premani
neeraj.premani@ves.ac.in
D17B , 56
CMPN , VESIT

Mihir Wagle
mihir.wagle@ves.ac.in
D17B , 73
CMPN , VESIT

1. Problem Statement

During waterlogging each sector officer is assigned multiple areas to inspect. However, doing that in-person is inefficient and time consuming. There's no provision for real-time identification of waterlogging areas. The existing system depends heavily on phone calls and SMS to identify waterlogged areas and lacks the ability to provide water level data(in water logging areas) to officers through a web portal. Our project aims to identify the waterlogged areas and do a detailed analysis of it.The system will provide predictions and suggest appropriate measure to cope with the situation.Water logging causes considerable damage to human lives and property every year. However, in the current situation, BMC relies on phone calls and SMS to identify the water logging areas. This inefficient method causes delay in water removal, leading to huge loss to human life and property. Our motivation behind this project is to build a system that would reduce this delay by enabling BMC officials to identify water logging areas as soon as they occur and guide the working team to respective locations.

2. Introduction

Urban-Waterlogging refers to the rainwater that gets accumulated and remains stagnant in a particular area and has no place to drain-off, which causes various problems. Waterlogged streets are a fairly well-known setbacks in cities and villages throughout India. Basically it is caused when inflow of water is much more than runoff. Monsoon is a dominant and extended season in India. The south-west summer monsoon continues for a four-month-long period from June through September, followed by the northeast or retreating monsoon from October to December. Drainage is often outdated, mainly in the older parts of the metro cities and most of the non-metro cities. This leads to severe perpetuated water logging situation. The current

system to get rid of water-logging is based either on its passage through manholes on its own, or starting of some pumping stations which are switched on by operators on getting notice from officials. This system is highly dependent and inefficient also there's no provision for real-time identification of waterlogging areas. Our system aims at identifying such waterlogged areas in real time and do a detailed analysis of it, notify the authorities about it and suggest ways to cope up with the situation. We also intend to predict the waterlogged areas.

3. Dataset explanation

Our dataset is composed of readings from Rain and Tide APIs alongside a given classification of low, medium or high. For predicting the same, we have used supervised learning. Specifically decision trees for the benefits they offer.

4. Algorithm

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.

In decision analysis, a decision tree and the closely related influence diagram are used as a visual and analytical decision support tool, where the expected values (or expected utility) of competing alternatives are calculated.

A decision tree consists of three types of nodes:

Decision nodes – typically represented by squares
Chance nodes – typically represented by circles
End nodes – typically represented by triangles
Decision trees are commonly used in operations research and operations

management. If, in practice, decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision trees is as a descriptive means for calculating conditional probabilities. Decision trees, influence diagrams, utility functions, and other decision analysis tools and methods are taught to undergraduate students in schools of business, health economics, and public health, and are examples of operations research or management science methods.

5. Analysis

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        14319              99.4306 %
Incorrectly Classified Instances         82               0.5694 %
Kappa statistic                          0.4583
Mean absolute error                      0.0057
Root mean squared error                  0.0557
Relative absolute error                 64.6463 %
Root relative squared error             84.6977 %
Total Number of Instances            14401

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
              0.999    0.574    0.996      0.999   0.997      0.539  0.768     0.997     1
              0.343    0.002    0.490      0.343   0.403      0.407  0.757     0.274     2
              0.250    0.000    0.667      0.250   0.364      0.408  0.666     0.309     3
Weighted Avg. 0.994    0.571    0.993      0.994   0.994      0.539  0.768     0.992

=== Confusion Matrix ===

     a     b    c   <-- classified as
 14289    18    0 |    a = 1
    43    24    3 |    b = 2
    11     7    6 |    c = 3
```
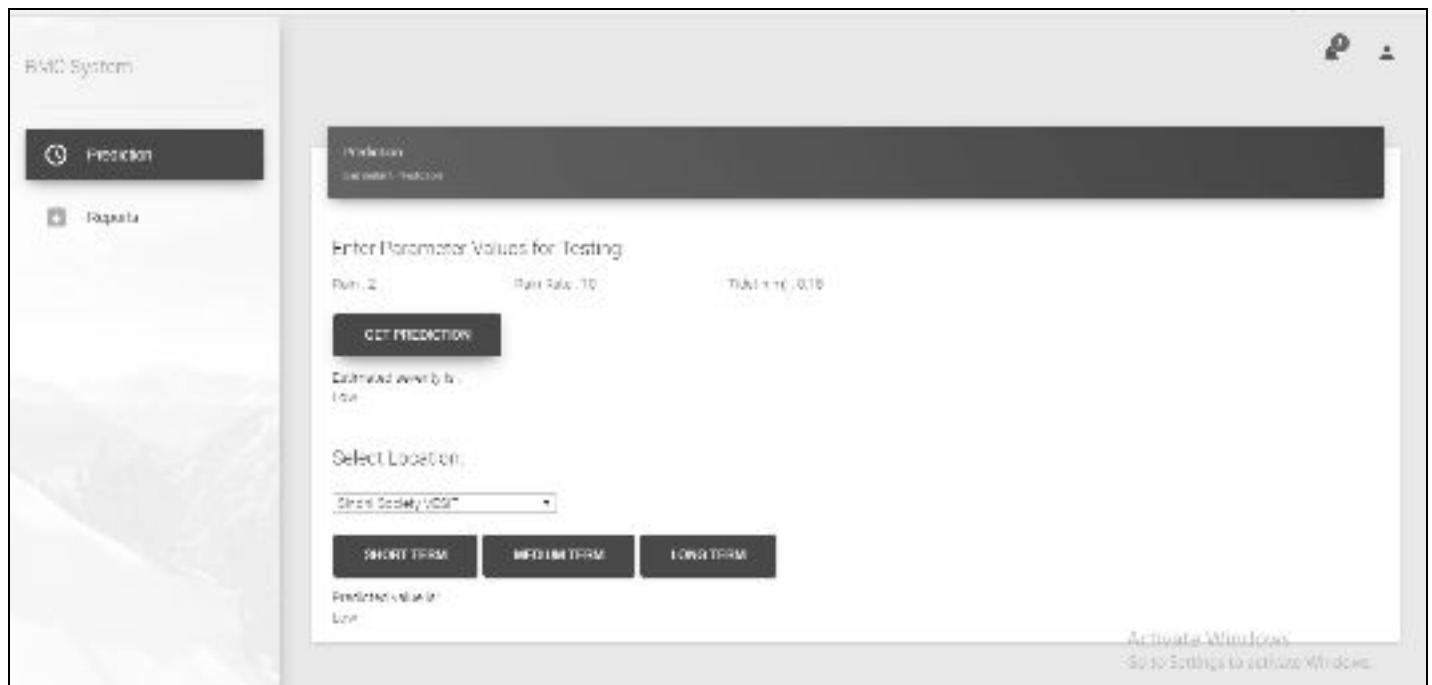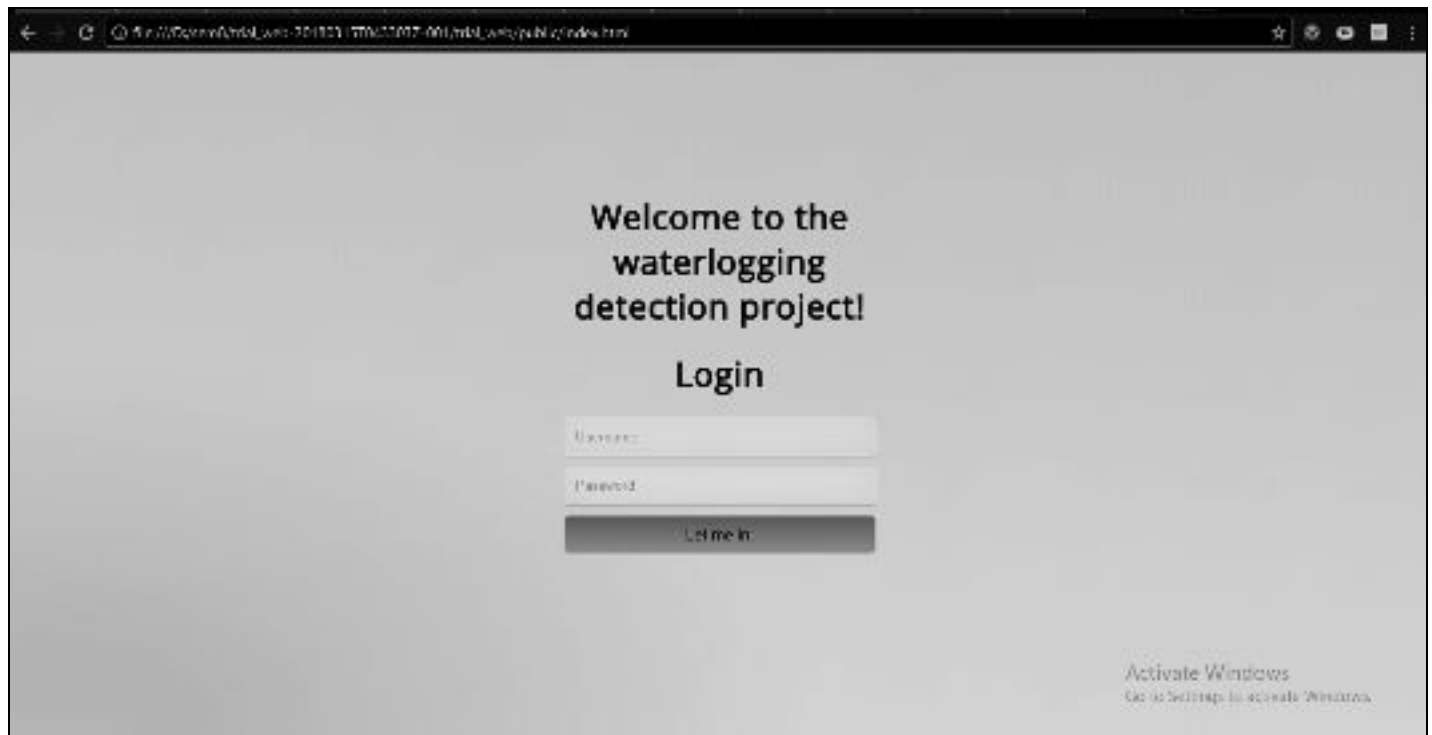
6. GUI screenshots

BMC Waterlogging Report

Region wise max water logging reported in 2017
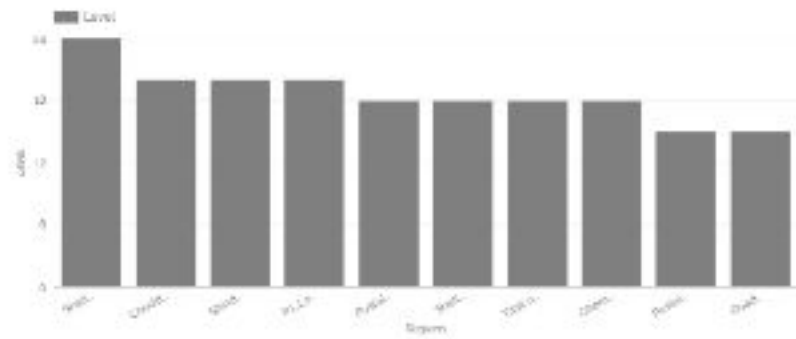
7. Output results
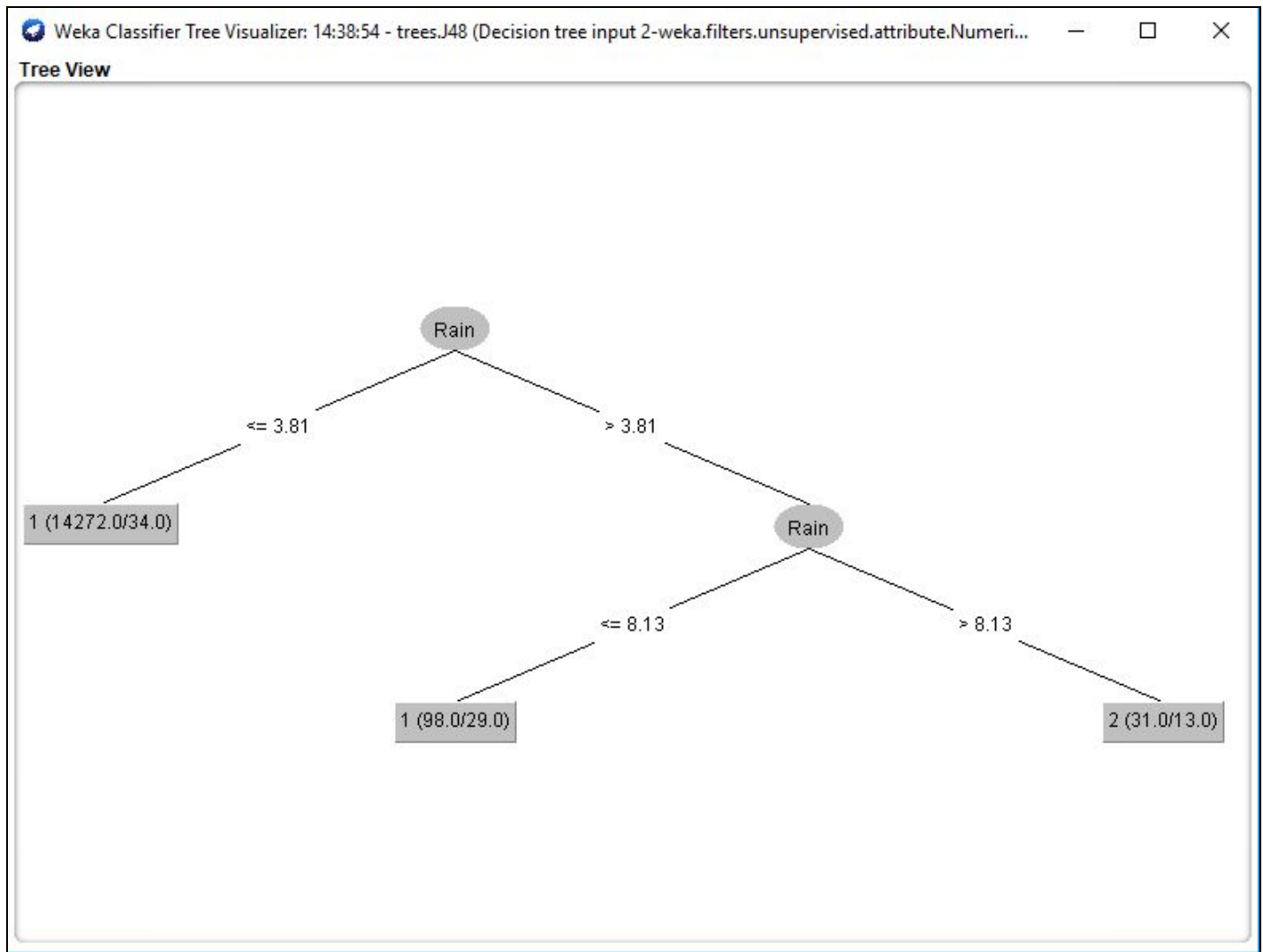
8. Conclusion

- Waterlogging is one the most common natural disasters and have widespread effect. Waterlogging forecasting is hence an important research area, and several possible solutions have been proposed in literature.
- Inside prediction module based upon the current and past rainfall data , previous year prediction data value for water level will be predicted along with the comparisons with actual value afterwards to calculate error rate .

9. References

[1] A CASE STUDY ON WATER LOGGING PROBLEMS IN AN URBAN AREA (2014,December) . Retrieved from https://www.researchgate.net/publication/265848376_A_CASE_STUDY_ON_WATER_LOGGING_PR

OBLEMS_IN_AN_URBAN_AREA_OF_BANGLADESH_AND_PROBABLE_ANALYTICAL_SOLUTIONS

[2] Central Water Commision (2017,October) , Retrieved from http://www.cwc.gov.in/

[3]Strom Water Management Model (SWMM)(2017,October) , Retrieved from https://www.epa.gov/water-research/storm-water-management-model-swmm

[4] ESP8266 (2017, October) , Retrieved from https://en.wikipedia.org/wiki/ESP8266

[5] MQTT (2017,October) , Retrieved from http://mqtt.org/

[6] Twilio (2017,September) , Retrieved from https://www.twilio.com/

[7] Amazon Elastic Compute Cloud (Amazon EC2) (September,2017) , Retrieved from https://aws.amazon.com/ec2/

[8] https://www.kaggle.com/datasets

[9] www.camo.com/multivariate_analysis.html

[10] http://scikit-learn.org/stable/modules/svm.html

[11] http://dni-institute.in/blogs/cart-decision-tree-gini-index-explained/

[12] https://en.wikipedia.org/wiki/Decision_tree_learning

[13] http://www.learnbymarketing.com/481/decision-tree-flavors-gini-info-gain/

[14] https://www.researchgate.net/publication/289009275_Prediction_of_diabetes_using_decision_trees