

River Flooding Detection and GINI Tree Generation

D. Gayathri, P. Laxman, N.Anudeep

Electronics and communication engineering, Sreenidhi institute of science and technology, Hyderabad, Telangana.

ABSTRACT

Floods don't have the same effect throughout the length of the river. Depending on the rainfall, and water volumes coming from the tributaries, floods will have more devastating effect in some areas and less effect in other areas. This project aims at classifying the flood water levels data using Gini classification. The flood data comprises of water levels measured at a couple of points across the width of a river. Gini index calculated based on the water levels of data measured at 100 different points across the length of the river. The water levels are classified into three categories: '0' stands for low risk, '1' stands for medium risk and '2' stands for high risk.

Keywords

Date of Submission: 15-06-2021

Date of Acceptance: 30-06-2021

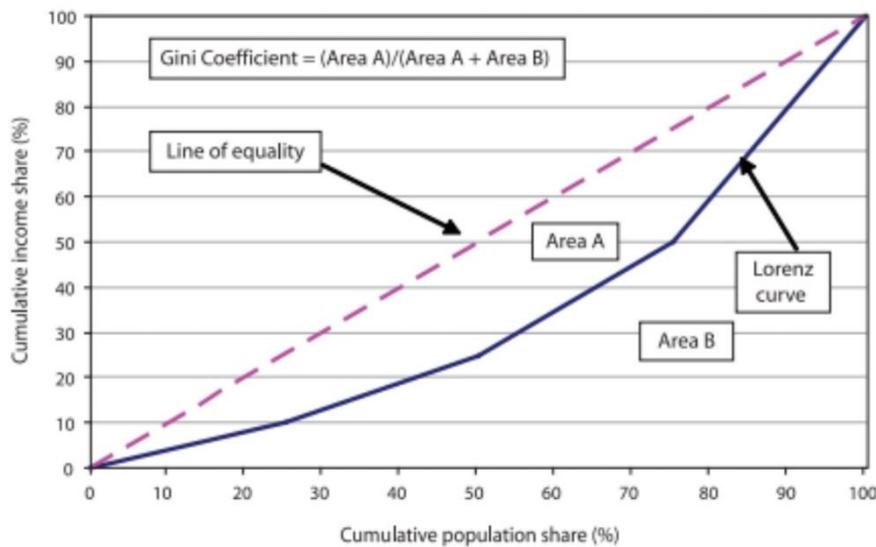
I. INTRODUCTION

Floods don't have the same effect throughout the length of the river. Depending on the rainfall, and water volumes coming from the tributaries, floods will have more devastating effect in some areas and less effect in other areas. This project aims at classifying the flood water levels data using Gini classification. The flood data comprises of water levels measured at a couple of points across the width of a river. Gini index calculated based on the water levels of data measured at 100 different points across the length of the river. The water levels are classified into three categories: 0 stands for low risk, 1 stand's for medium risk and '2' stands for high risk. The project comprises of four modules. The first module deals with adding the river details and point's details to the DB. The second module deals with calculating the Gini indexes based on the attributes of river points. The third module deals with generating the decision tree as a .dot file, the final module deals with generating the decision tree as a .pdf file. The Gini index is a statistical measure of distribution developed by the Italian statistician Corrado Gini in 1912. It is also called as a Gini coefficient, or, Gini Ratio. The Gini coefficient measures the inequality among values of a frequency distribution. A Gini coefficient of zero expresses perfect equality, where all values are the same (for example, same water level at the points). A Gini coefficient of 1 expresses maximal inequality among values. Gini index is used in CART (Classification and Regression Trees) methodology to generate the decision trees. A decision tree is a decision support tool that uses a

tree-like model of decisions and their possible consequences. CART builds decision trees from a set of training data, using the concept of Gini Index. Gini index-based discretization is a supervised, top-down splitting technique. It explores class distribution information in its calculation and determination of split-points (data values for partitioning an attribute range). To discretize a numerical attribute, A, the method selects the value of A that has the maximum Gini-index as a split-point, and recursively partitions the resulting intervals to arrive at a hierarchical discretization. Such discretization forms a concept hierarchy for that numerical attribute A.

II. LITERATURE SURVEY

Equitable access to water, or to the benefits derived from using water, is critical to eradicating poverty and promoting growth. This is particularly important in India, which is still facing significant inequalities in access to and use of water. The present study explores the application of the Gini Coefficient, which has hitherto only been used for income and land distribution, to quantify the distribution of water resources. The Gini Coefficient is one of the most commonly used indicators for measuring distribution. It is traditionally applied to the measurement of income inequality, but has also been applied to measure land inequality. As yet, it has not been applied to measure water use inequality. The Gini Coefficient is calculated from un-ordered size data as the "relative mean difference", i.e., the mean difference between every possible pair of individuals, divided by the mean size and is defined as follows.



Feasibility analysis

As the name implies, a feasibility study is used to determine the viability of an idea, such as ensuring a project is legally and technically feasible as well as economically justifiable. It tells us whether a project is worth the investment—in some cases, a project may not be doable. There can be many reasons for this, including requiring 9 too many resources, which not only prevents those resources from performing other tasks but also may cost more than an organization would earn back by taking on a project that isn't profitable.

Economical feasibility

This assessment typically involves a cost/benefits analysis of the project, helping organizations determine the viability, cost, and benefits associated with a project before financial resources are allocated. It also serves as an independent project assessment and enhances project credibility—helping decision makers determine the positive economic benefits to the organization that the proposed project will provide. Our project is economically feasible because in this we have used “UBUNTU”, “PYTHON”, “PYQT” designer tool and “PYUIC” which are all available as an open source.

Technical Feasibility

This assessment focuses on the technical resources available to the organization. It helps organizations determine whether the technical resources meet capacity. Technical feasibility also involves evaluation of the hardware, software, and other technology requirements of the proposed system. A prototype of the tool was developed to verify the technical feasibility. The prototype is

working successfully and hence the project is feasible.

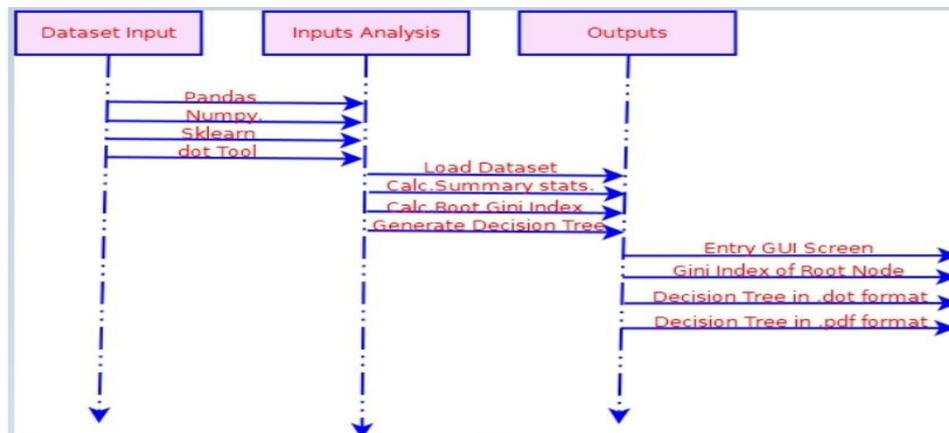
Python

Python is a widely used high level programming language for general purpose programming, created by Guido Van Rossum and first released in 1991. An interpreted language, Python has a design philosophy that emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly brackets or keywords), and a syntax that allows programmers to express concepts in fewer lines of code than might be used in languages such as C++ or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale. Python features a dynamic type system and automatic memory management and supports multiple programming paradigms, including object-oriented, imperative, functional programming, and procedural styles. It has a large and comprehensive standard library. Python interpreters are available for many operating systems, allowing Python code to run on a wide variety of systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations

Sample Code

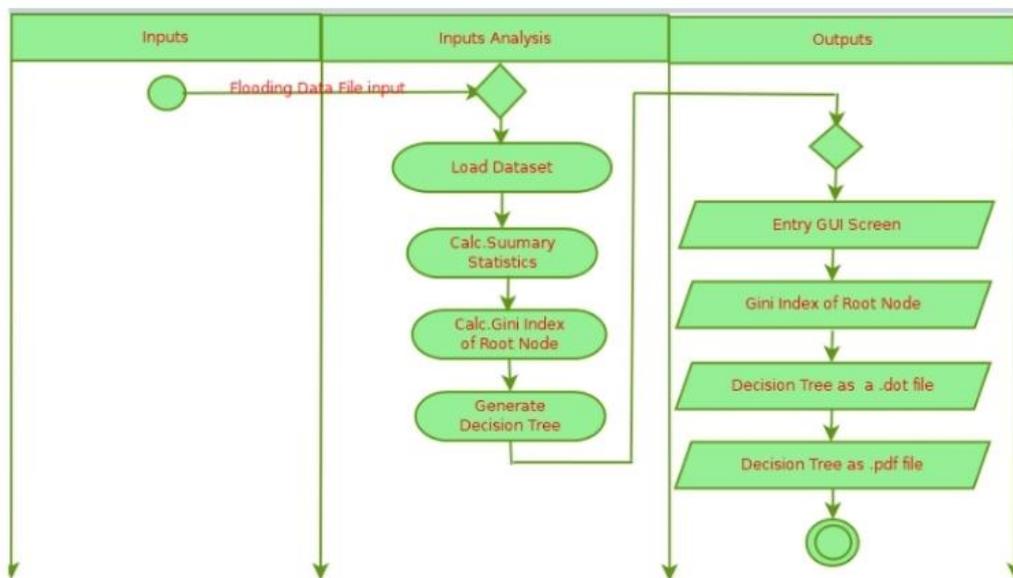
Python Code (To be filled in after coding)
 rivgin1.py
 river1.py
 points1.py
 samp2.py
 datafile1.py
 samp5.py

Sequence diagram



A sequence diagram is an interaction diagram that shows how objects operate with one another and in what order. It is a construct of a message sequence chart. A sequence diagram shows object interactions arranged in time sequence. From above mentioned sequence diagram we have to go

in sequence: Enter the needed details as shown in the above figure, Provide the river’s details, measurement points details, calculate the gini indices, and then generate the decision trees in .dot format, and finally in .pdf format



Testing

Software Testing

Software testing is the process of evaluation a software item to detect differences between given input and expected output. Testing assesses the quality of the product. Software testing is a process that should be done during the development process. In other words, software testing is a verification and validation process.

Verification

Verification is the process to make sure the product satisfies the conditions imposed at the start of the

development phase. In other words, to make sure the product behaves the way we want it to.

Validation

Validation is the process to make sure the product satisfies the specified requirements at the end of the development phase. In other words, to make sure the product is built as per farmer requirements.

Basics of software testing

Black box Testing

Black box testing is a testing technique that ignores the internal mechanism of the system and focuses on

the output generated against any input and execution of the system. It is also called functional testing.

White box Testing

White box testing is a testing technique that takes into account the internal mechanism of a system. It is also called structural testing and glass box testing. Black box testing is often used for validation and white box testing is often used for verification

Types of testing

There are many types of testing like

Unit Testing

Integration Testing

Functional Testing

System Testing

Regression Testing

Unit Testing

Unit testing is the testing of an individual unit or group of related units. It falls under the class of white box testing. It is often done by the programmer to test that the unit he/she has implemented is producing expected output against given input.

Integration Testing

Integration testing is testing in which a group of components are combined to produce output. Also, the interaction between software and hardware is tested in integration testing if software and hardware components have any relation. It may fall under both white box testing and black box testing.

Functional Testing

Functional testing is the testing to ensure that the specified functionality required in the system requirements works. It falls under the class of black box testing.

System Testing

System testing is the testing to ensure that by putting the software in different environments (e.g., Operating Systems) it still works. System testing is done with full system implementation and environment. It falls under the class of black box testing.

Regression Testing

Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing.

System Testing Results

The project is thoroughly tested by testing the each and every text box and push buttons of the GUI Screens, and verifying the corresponding results in the Data Base. Following is the rivers Details screen along with data.

III. CONCLUSION

This project entitled “River Flood data classification and Gini tree generation”, is useful to the river boards, in accurately determining the peak and low flood points to construct the dams. The project is also useful to the general public, in particular to the farmers, as enough water can be provided to the needed fields, by constructing dams at high water availability points. This project finally leads to the improvement of flood precautionary measures by studying the previous floods data.

Future scope

As of now, the system is successfully tested with the data set comprising of the data of Betwa River.. The applicability of the projects to the data sets of other rivers, need to be explored.

REFERENCES

- [1]. Vinay Kumar M, Lalu Prasad M, Kolhapuram Shreyaa et al. : AN EFFICIENT FLOOD IDENTIFICATION AND CAUTIONING SYSTEM, International Journal of Advanced Research in Computer Science.
- [2]. An Assessment of Environmental Flow Requirements of Indian River Basins, by V. Smakhtin and M. Anputhas.
- [3]. <https://www.python.org/>
[4]<https://github.com/baoboa/pyqt5/blob/master/pyuic/uic/pyuic.py>
- [4]. <https://www.numpy.org/>
- [5]. <https://riverbankcomputing.com/software/pyqt/intro>
- [6]. SCOTT MIAU AND WEI-HSI HUNG: River Flooding Forecasting and Anomaly Detection Based on Deep Learning, IEEEAccess