IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

RIVER WATER AVAILABILITY FOR THE CONSTRUCTION OF DAM USING GINI INDEX

[1]S.G. Subramania Siva, [2].Y. Phanindra, [3] S. Sai Manish [4] Abhishek Vijay Kumar Waghmare

[1]Assistant Professor, [2]Student, [3]Student, , [4]Student
[1]Department of Information Technology,
[1]JBIET, Hyderabad, India

Abstract: In order to construct a dam, across a river, there should be enough water available in the river, where the dam is proposed to be constructed. This project is supposed to verify the place between two given points whether it is suitable for the construction of dam or not using Gini Index. This project measures the water availability level, between two given points of a river, using Gini index calculated based on the availability levels of data of last 100 seasons. The availability levels are classified into three categories: 0 stands for low availability, 1 stands for medium availability and 2 stands for high availability. A dam can be constructed, if the availability level is 2. 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). The project comprises of four modules. The first module deals with adding the river details and points details to the DB. The second module deals with calculating the Gini indexes based on the attributes of river point. 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 Front end module is used to create the needed GUI screens for the project.

IndexTerms **- Decision tree Classifier, Gini Index, Top-Down Splitting Technique.**

I. INTRODUCTION

Dam is a barrier that stops or restricts the flow of water or underground streams. Reservoirs created by dams not only suppress floods but also provide water for activities such as irrigation, human consumption, industrial use, aquaculture, and navigability. Hydropower is often used in conjunction with dams to generate electricity. A dam can also be used to collect water or for storage of water which can be evenly distributed between locations. Dams generally serve the primary purpose of retaining water, while other structures such as floodgates or levees (also known as dikes) are used to manage or prevent water flow into specific land regions. It is necessary to construct the dam across the river between two points where the availability of water is high, Because if more water is available more probability to construct the dam or if less water available there is less chances of construction of dam.

We use Gini index to check that the water availability is high or low. Gini index shows the availability of water in three ways, they are level 0: The availability of water is very low, level 1: The availability of water is medium, level 3: The availability of water is high. And the dam can be constructed if the availability is 2. Gini index finally generates the decision tree which tells about the construction of dam.

Data collection is very important in this aspect we rely on the data very much to get the accurate results. The data that given to Gini index consists of high and low level of water availability at the two pints we considered in that season, and also the distance between the two points from the sea shore. The difference between the high and low availability of water considered as the levels of availability of water at that given two points.

This Project uses PyQt tool to create the needed Graphical User Interfaces, PyUIC module to automatically generate the code. PyUic tool is used automatically generate the code for the Front end user interfaces created by PyQt. All the front end python code is automatically generated by this tool, by converting the user interface (.ui) files into .py files.

A decision tree is a tree whose internal nodes can be taken as tests (on input data patterns) and whose leaf nodes can be taken as categories (of these patterns). These tests are filtered down through the tree to get the right output to the input pattern. Decision Tree algorithms can be applied and used in various different fields. It can be used as a replacement for statistical procedures to find data, to extract text, to find missing data in a class, to improve search engines and it also finds various applications for prediction. Many Decision

tree algorithms have been formulated. They have different accuracy and cost effectiveness. It is also very important for us to know which algorithm is best to use. The decision support tool analyses the variations in the levels of water in the season. The decision tree is constructed based on the Gini index value which is high among in the given input. The tool builds decision trees from a set of training data, using the concept of Gini index.

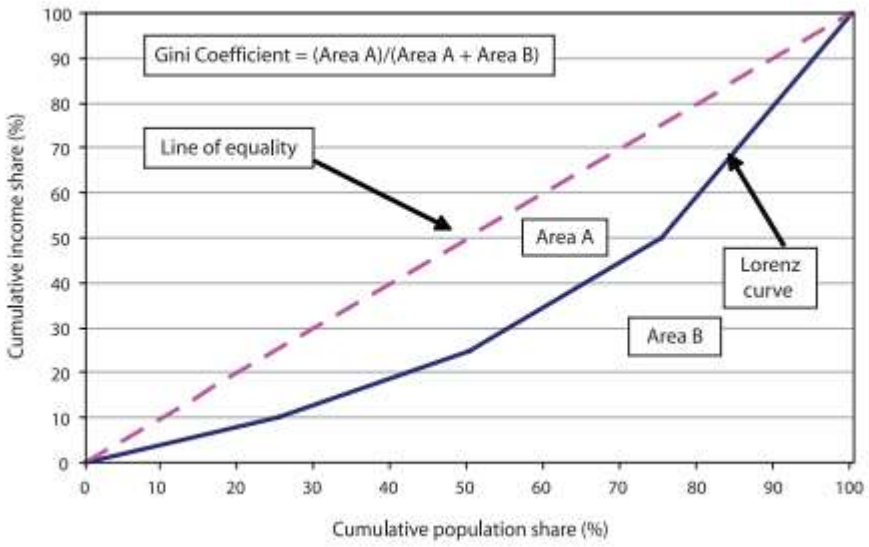
LITERATURE SURVEY

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 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.

$$Gini = \frac{1}{2n^2\bar{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j|$$

The Gini Coefficient can be displayed graphically as a plot of the distribution of the size fractions of ordered individuals. This is termed the Lorenz curve and is shown in below figure.

Graphical example of the Lorenz curve and Gini Coefficient.



In a perfectly equal society the Lorenz curve would plot as a straight line. This is termed the line of equality. In most cases, however, the Lorenz curve plots below this line of equality, showing the inequality in the distribution of income, land or, now, water between members of a community. In the example shown in above figure, the poorest 50 percent of the population account for only 25 percent of the total income of all individuals in the area, while the richest 20 percent account for 50 percent of the total income. The Gini Coefficient is calculated as the ratio of the area between this Lorenz curve and the line of equality (Area A) and the total triangular area under the line of equality (Are A + B). The closer to 1, the more unequal is the distribution of income, and the closer to 0, the more equal is the distribution of income.

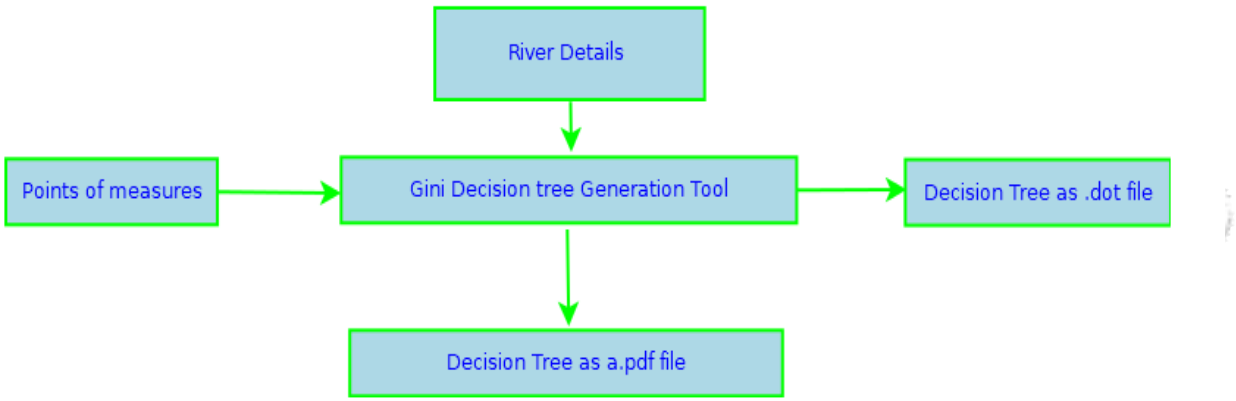
PROPOSED SYSTEM

In this proposed system, we focus on the water availability level, between two given points of a river, using Gini index calculated based on the availability levels of water at that points of last 100 seasons. The availability levels are classified into three categories: 0 stands for low availability, 1 stands for medium availability and '2' stands for high availability. A dam can be constructed, if the availability level is 2.The project comprises of four modules. The first module deals with adding the river details and points details to the DB. The second module deals with calculating the Gini indexes based on the attributes of river point. 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 dataset we used is consists of levels of the water at that points about past 100 years. By using the qt designer we created the screens to take the input to the Gini index. The attributes of the input are highest level of water at the point1, and point2 and also the lowest level of the water available at the point1, and point2 in the season. And the difference between the highest level and the lowest level of the water levels also is given as input. Then the Gini index calculates the Gini index value for all the inputs and create the .dot file and finally the decision tree is created. 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 point. The proposed system gives the levels of water available.

SYSTEM ARCHITECTURE

The first step in this is giving the details of the river to the database for which the name of the river he is given initially. The data consist of the levels of the water at the selected points across the selected river and the distance of the points from the seashore. In the second step, Gini index is used for the calculation and the decision tree generation in the form of the .dot file and finally the .pdf file of decision tree is generated.



ALGORITHM :

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.

1. Importing Data manipulation library files.
//Pandas

2. Initialize attributes. #col_names=
['low', 'high', 'low1', 'high1',' availability']

3. Load dataset and read using read_csv.

4. Use Gini index for the calculation of levels of water availability.
Formula:
Giniroot = 1.0 - (float(trcnt1)/float(trcnt))2 - (float(trcnt2)/float(trcnt))**2 – (float(trcnt3)/float(trcnt))**2**

5. Gini index creates the .dot file.

6. finally .pdf file is generated

MODULES

I. *Data Collection*

Firstly, Dataset can be collected from various sources of any organization. The right dataset helps for the calculation of Gini index and it can be manipulated as per our requirement. Our data mainly consists of the levels of the water at the selected point about past 100 years. The data can be collected from the organization based on the water levels, distance between two points, distance of the points from the shore, availability. By collecting these it makes accurate in calculation

II. *Data Processing*

At the beginning, when the data was collected, all the values of the attributes selected were continuous numeric values. Data transformation was applied by generalizing data to a higher-level concept so as all the values became discrete. The criterion that was made to transform the numeric values of each attribute to discrete values depended on the Gini index of the points.

III. *Training the Data*

After the data has been prepared and transformed, the next step was to build the classification model using the decision tree technique. The decision tree technique was selected because the construction of decision tree classifiers does not require any domain knowledge, we can done by using the Decision Tree Classifier() which is in the sklearn library. By using the attribute we have considered in the dataset we train the model by using the algorithm. The training sets are used to tune and fit the models.

IV. *Deploying the Model*

The trained data can be used for the Testing the data. It helps to give the output or accurate water levels at the points we selected used for the dam construction.

CONCLUSION

This project entitled **"River water availability using Gini index for construction of dam**" , is useful to the river boards, in accurately determining the river 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 water availability to the people, living in the vicinity of a river.

REFERENCES

[1] J. Srivastava, P. Desikan, and V. Kumar. Web mining—concepts, applications, and research directions. pp. 278-282..

[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

[5] https://www.numpy.org/

[6] https://riverbankcomputing.com/software/pyqt/intro

[7] https://www.ubuntu.com/