***Generare arbori pentru evaluarea de binarizări optime***

**Concepts**

1. Binarization

* Binarization is the method of converting any gray – scale image (multi tone image) into black – white image (two tone image).
* The main goal of image binarization is the segmentation of document into foreground text and background.
* The simplest approach to binarization is thresholding[2].
* Generally, binarization is categorized as Global and Local.

|  |  |
| --- | --- |
| Global Binarization | Local Binarization |
| * The technique in which a single threshold value is applied to binarize the entire image * Fast, but fails for images with complex background. | * Instead of single threshold for the whole image, a different value of threshold is chosen for every pixel. * The threshold is chosen depending upon the neighbourhood pixels. * Do not give good results for the documents suffering from background noise. |

1. Thresholding

* Thresholding is a type of image segmentation, where we change the pixels of an image to make the image easier to analyze.
* The main parameter of this conversion is the threshold t, with the value of which the brightness of all is then compared. After the comparison, the pixel is assigned one of two possible values: 0 - "object boundary" or 1 - "remaining area".
* Methods: Otsu’s Method, Niblack’s Method, Sauvola Method, Bernsen, etc.

1. Image Histogram

* The histogram is used for graphical representation of a digital image.
* The horizontal axis of the graph represents the tonal variations, while the vertical axis represents the total number of pixels in that particular tone.
* The left side of the horizontal axis represents the dark areas, the middle represents mid-tone values and the right hand side represents light areas. The vertical axis represents the size of the area (total number of pixels) that is captured in each one of these zones.
* Thus, the histogram for a very dark image will have most of its data points on the left side and center of the graph.

1. Ground Truth

* The objective verification of the particular properties of a digital image, used to test the accuracy of automated image analysis processes.

**Monte Carlo Algorithm**

Monte Carlo approximation (which is one of the MC methods) is a technique to approximate the expectation of random variables, using samples.

It uses Random Sampling, which is used to generate multiple possible outcomes and calculate de average result.

Monte Carlo techinques involve three steps:

1. Set up the Predictive Model:identify the dependent variable to be predicted and the independent variables(input risk of predicted variables that will drive the predictions).
2. Probability Distribution of the independent variables: define a range of values and assign probabilites for each
3. Run simulations repeatedly to generate random values of the independent variables.

**Random Tree Generator Using Prüfer Sequence**

Explanation: <https://www.geeksforgeeks.org/random-tree-generator-using-prufer-sequence-with-examples/>

**Histogram Generation**

<https://datacarpentry.org/image-processing/05-creating-histograms/>

Using OpenCV Library: <https://www.geeksforgeeks.org/opencv-python-program-analyze-image-using-histogram/>