**Q. Classify Nepali characters ( Ka to na - 5 character). The following points need to include in your report.**

To identify 5 Nepali characters from (क to ङ)

**1. Develop a few training dataset.**

First we need to develop a training set. There must be many training sets (more than 100 sets for each letter). The training sets should include character and each character’s feature. For example, the training set can be:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Label | Feature 1 | Feature 2 | Feature 3 | ….. | ….. | Feature n |
| क | x1 | x2 | x3 |  |  | xn |
| ... | …. | …. | …. | …. | … | … |
| क | y1 | y2 | y3 |  |  | yn |
| ख | …. | …. | …. | …. | … | … |
| ख | …. | …. | …. | …. | … | … |
| …. | …. | …. | …. | …. | … | … |
| …. | …. | …. | …. | …. | … | … |
| ङ | …. | …. | …. | …. | … | xxn |

**2. Specify the features for machine learning.**

Features are too essential component for machine learning. In machine learning and pattern recognition, a feature is an individual measurable property of a phenomenon being observed. Choosing informative, discriminating and independent features is a crucial step for effective algorithms in pattern recognition, classification and regression. In our Nepali character classification, the features may include histograms counting the number of black pixels along horizontal and vertical directions, stroke detection, edge detection, Histogram Of Gradient (HOG) of the image, contours detection, etc.

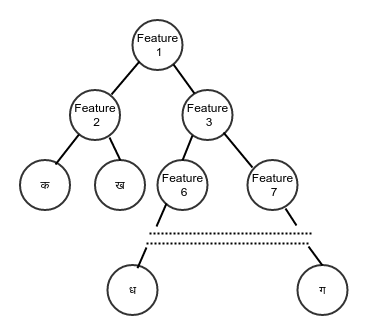
More the feature selection, better will be the classification or the classification and prediction accuracy will increase. In our case, the feature value is generally continuous value.

**3. Apply machine learning ( decision tree ) for making model.**

To implement a machine learning we have many algorithms like KNN, Support Vector machine, Decision Tree, Boosted Decision tree, Gradient descent, etc. In addition to that, we can also employ Neural Network and Deep Learning techniques for better result.

In our context, we’ll use Decision Tree Classifier.

To implement machine learning, we first build a Decision Tree learning model from our training datasets. Out decision tree includes a root node, which is a feature name. Branching of tree is done in the basis of the decision value of the features. The leaf node contains the labels.

  
*Figure: Decition Tree Model from training data sets.*

For example, if the feature is about whether there is a vertical line in the character or not. Then the decision can be made as: if there is a vertical line in the character, then it is not the 5th letter, i.e. the (ङ). Another feature can be, whether the character has a circle or not. If if has a circle, then the character can be either क or ख but, not other characters.

In this way, we build up the decision tree in order to build the training model.

**4. How can model be used for classifying test dataset.**

Model build in the previous part can be now used to predict or classify the new and unknown data.

Lets say, we have a new character that machine need to identify. For that, we first extract features from the new character. The feature should be same and exactly equal to the features that are in the training sample.

For example:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Label | Feature 1 | Feature 2 | Feature 3 | ….. | ….. | Feature n |
| ? | p1 | p2 | p3 |  |  | pn |

Now, we fit the test samples feature in the trained Decision Tree model. Then, according to the feature’s value, we traverse the decision tree model, until leaf node is reached which finally determines the label of the test character.