

Oct 31, 13 20:06

573:Proj2:Naveen Kumar Lekkalapudi

Page 1/1

```

+Title: Proj 2:CS573
+=====
+
+I would like to test the Test Essential Assumption Knowledge(TEAK)
+approach on a naive bayes tree (nb-tree) and calculate its performance.
+
+
+Prologue
+-----
+
+To increase the accuracy, inconsistent values of a data set must be handled mor
e efficiently.
+An nb-tree has naive bayes classifiers as the leaf nodes and is built by using
uniivariate splits at each node.
+A tree of data can be built based on that approach, this tree of data has naive
bayes classifiers at the leaves.
+The split on trees are done based on variance for discrete values and entropy
for continuous values.
+An estimate can be performed on such data by using the TEAK's design.
+The design states to pick a prediction system and prune outliers based on that
"essential" assumption to result an efficient classifier.
+
+
+
+Support
+-----
+
+Design steps in TEAK:
+Select a prediction system: nb-tree classifier can be used as an initial predic
tion system.
+Identify essential assumption: Though it is not mentioned, locality plays a key
role in nb-tree.
+
+           Data that is similar to each other ends up in sa
me leaf of naive bayes and predicts the same outcome.
+Identify assumption violation: The data that is similar but predicting a differ
ent result is considered inconsistent and is marked to remove.
+
+           Different theories can be used to identify such
violations such as comparing variance or entropy.
+Execute TEAK: Run above steps on training data to build an nb-tree and test the
built tree on testing data.
+
+
+
+Approximate Steps
+-----
+
+1 For each incoming record in a test dataset:
+2     Add the record to the root node.
+3     Run the nb classifer on the node.
+4     Calculate variance or entropy and decide on the split.
+5     If split comes true, divide the node into two child nodes.
+6     Run step 4 on each node till there are no more splits.
+7 Using TEAK at node:
+8     Test for violations of TEAK assumption and prune the data that deems to
have more variance or entropy(other theories can be used)
+9 Test the resulting nb-tree with testing data.
+
+
+
+Conclusion
+-----
+
+I assume that TEAK improves the efficiency of nb-tree by not focusing on the in
consistent values and improving efficiency.
+
+

```