**Aim** : To find the best subset of features using Ant Colony Optimization (ACO) algorithm.

We’ve used ACO for finding the best feature subset from the given dataset. In the implementation of ACO, we first initiate 759 ants, one from each feature(considering the destination as 760th feature).

Currently we have implemented the solution in java and will convert it into MapReduce code.

The java program AcoTest.java reads the input dataset file “train\_data\_test.csv”, which contains 9 features(ants) and uses ACO algorithm to find the feature subset for each ant.

To **execute the program**, we need to compile the program using

**command :** ***javac AcoTest.java***

Upon successful compilation, the ***.class*** files are created. The program can be executed using

**command : *java AcoTest***

To find the accuracies of the feature subset, we have implemented the Machine Learning algorithms Naïve Bayes, Decision Trees and are working on Logistic Regression, Support Vector Machines (SVM).

The commands to execute the pyspark programs are

***spark-submit Naive.py***

***spark-submit Decisiontree.py***

These programs learn the entire train dataset and calculate the accuracies for the test dataset.

The MapReduce job will determine the features subset, which is the optimal path traversed by the Ants from source to destination.

**Mapper** – Reads the whole dataset, and generates a correlation matrix, which holds the weights to traverse from a node to another.

**Mapper** – Probabilistically determine the next node to be traversed.

The path traversed by ants is determined by predicting the probability of a node(feature) to be selected, among all the untraversed features present in the subset.

**Mapper** – After each successful iteration, we update the Pheromone. Pheromone evaporation rate p=0.01.

**Mapper** – Get the features subset of each ant. For each ant, the set of features traversed from source to destination is obtained.

**Reducer** – Each Machine Learning algorithm will find the accuracies for subset of each ant.

**Reducer** – Determine the best accuracy out of the accuracies obtained for all the ants.