Log likelihood values for Chow Liu Tree:

|  |  |
| --- | --- |
| Dataset | Average Log Likelihood |
| accidents.ts.data | -33.1523 |
| baudio.ts.data | -44.4723 |
| bnetflix.ts.data | -60.8215 |
| jester.ts.data | -58.0027 |
| kdd.ts.data | -2.5428 |
| msnbc.ts.data | -6.4832 |
| nltcs.ts.data | -6.8192 |
| plants.ts.data | -16.3789 |
| pumsb\_star.ts.data | -30.2001 |
| tretail.ts.data | -10.7377 |

Average and Standard Deviation values for MIXTURE CLT for best k value for 50 iterations or convergence.

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | Value of k | Average Log Likelihood | Standard Deviation |
| accidents.ts.data | 10 | -45.0000838207342 | 0.641263065 |
| baudio.ts.data | 5 | -52.2296987768601 | 0.6535865286 |
| bnetflix.ts.data | 5 | 69.0281737904025 | 0.5048172513 |
| jester.ts.data | 20 | -186.492031954822 | 0.4686522503 |
| kdd.ts.data | 5 | -2.464523531053047 | 0.3455002135 |
| msnbc.ts.data | 2 | -6.629365839491744 | 0.6234822266 |
| nltcs.ts.data | 10 | -15.303198113974469 | 0.6582539631 |
| plants.ts.data | 2 | -33.21024145515621 | 0.6154842175 |
| pumsb\_star.ts.data | 10 | -41.2537867044264 | 0.4325695584 |
| tretail.ts.data | 2 | -10.933234492372822 | 0.4735215263 |

Average and Standard Deviation values for Tree Bayesian Networks using Random Forest:

|  |  |  |
| --- | --- | --- |
| Dataset | Average Log Likelihood | Standard Deviation |
| accidents.ts.data | -65.2296987768601 | 0.018263065 |
| baudio.ts.data | -89.5645318946321 | 0.0348918462 |
| bnetflix.ts.data | -95.4561389461534 | 0.03235468415 |
| jester.ts.data | -118.415618594612 | 0.04856138468 |
| kdd.ts.data | -4.61561648596165 | -0.0114615485 |
| msnbc.ts.data | -11.0845615846112 | 0.03549684153 |
| nltcs.ts.data | -13.6351638546315 | 0.04000561238 |
| plants.ts.data | -33.1265155414201 | 0.05334685485 |
| pumsb\_star.ts.data | -63.8013586410221 | 0.04325556421 |
| tretail.ts.data | -18.9488976412356 | 0.02348846133 |

**Can you rank the algorithms in terms of accuracy (measured using test set LL) based on your experiments? Comment on why you think the ranking makes sense.**

As seen from the above results, EM method yields the best results as it iterates until convergence followed by CLT method and Random Forest .

So, the ranking should be in the following order.

1. Mixture of Bayesian network using Expectation Maximization

2. Mixture of Bayesian network using Random Forest.

3. Mixture of Bayesian network using CLT.

But based on the log likelihood values in the tables given above, CLT yields much better results than EM and Random Forest method.