**Name : Juhi Barai**

**Net id : jkb140130**

**I. Clustering without PCA:**

**(1.) KMEANS, K=2**

**SSE : 12562.193775711039**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 <-- assigned to cluster

0 1 | ALL-T-cell-2

1 0 | ALL-T-cell-3

1 0 | ALL-T-cell-6

1 0 | ALL-T-cell-9

1 0 | ALL-T-cell-10

1 0 | ALL-T-cell-11

1 0 | ALL-T-cell-14

1 0 | ALL-T-cell-23

1 0 | ALL-B-cell-1

1 0 | ALL-B-cell-4

1 0 | ALL-B-cell-5

1 0 | ALL-B-cell-7

1 0 | ALL-B-cell-8

1 0 | ALL-B-cell-12

1 0 | ALL-B-cell-13

1 0 | ALL-B-cell-15

1 0 | ALL-B-cell-16

1 0 | ALL-B-cell-17

1 0 | ALL-B-cell-18

1 0 | ALL-B-cell-19

1 0 | ALL-B-cell-20

1 0 | ALL-B-cell-21

1 0 | ALL-B-cell-22

1 0 | ALL-B-cell-24

1 0 | ALL-B-cell-25

1 0 | ALL-B-cell-26

1 0 | ALL-B-cell-27

1 0 | AML-28

1 0 | AML-29

0 1 | AML-30

0 1 | AML-31

1 0 | AML-32

1 0 | AML-33

0 1 | AML-34

0 1 | AML-35

1 0 | AML-36

1 0 | AML-37

1 0 | AML-38

Cluster 0 <-- ALL-T-cell-3

Cluster 1 <-- ALL-T-cell-2

**(2.) KMEANS, K=3**

**SSE : 11151.928018360504**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 2 <-- assigned to cluster

0 0 1 | ALL-T-cell-2

1 0 0 | ALL-T-cell-3

1 0 0 | ALL-T-cell-6

1 0 0 | ALL-T-cell-9

0 0 1 | ALL-T-cell-10

0 0 1 | ALL-T-cell-11

1 0 0 | ALL-T-cell-14

1 0 0 | ALL-T-cell-23

1 0 0 | ALL-B-cell-1

1 0 0 | ALL-B-cell-4

0 0 1 | ALL-B-cell-5

1 0 0 | ALL-B-cell-7

1 0 0 | ALL-B-cell-8

0 0 1 | ALL-B-cell-12

0 0 1 | ALL-B-cell-13

0 0 1 | ALL-B-cell-15

0 0 1 | ALL-B-cell-16

1 0 0 | ALL-B-cell-17

0 0 1 | ALL-B-cell-18

0 0 1 | ALL-B-cell-19

1 0 0 | ALL-B-cell-20

0 0 1 | ALL-B-cell-21

0 0 1 | ALL-B-cell-22

1 0 0 | ALL-B-cell-24

0 0 1 | ALL-B-cell-25

0 0 1 | ALL-B-cell-26

1 0 0 | ALL-B-cell-27

1 0 0 | AML-28

1 0 0 | AML-29

0 1 0 | AML-30

0 1 0 | AML-31

1 0 0 | AML-32

1 0 0 | AML-33

0 0 1 | AML-34

0 1 0 | AML-35

1 0 0 | AML-36

1 0 0 | AML-37

1 0 0 | AML-38

Cluster 0 <-- ALL-T-cell-3

Cluster 1 <-- AML-30

Cluster 2 <-- ALL-T-cell-2

**(3.) HIERARCHICAL CLUSTERING, K=2**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 <-- assigned to cluster

1 0 | ALL-T-cell-2

1 0 | ALL-T-cell-3

1 0 | ALL-T-cell-6

1 0 | ALL-T-cell-9

1 0 | ALL-T-cell-10

1 0 | ALL-T-cell-11

1 0 | ALL-T-cell-14

1 0 | ALL-T-cell-23

1 0 | ALL-B-cell-1

1 0 | ALL-B-cell-4

1 0 | ALL-B-cell-5

1 0 | ALL-B-cell-7

1 0 | ALL-B-cell-8

1 0 | ALL-B-cell-12

1 0 | ALL-B-cell-13

1 0 | ALL-B-cell-15

1 0 | ALL-B-cell-16

0 1 | ALL-B-cell-17

1 0 | ALL-B-cell-18

1 0 | ALL-B-cell-19

1 0 | ALL-B-cell-20

1 0 | ALL-B-cell-21

1 0 | ALL-B-cell-22

1 0 | ALL-B-cell-24

1 0 | ALL-B-cell-25

1 0 | ALL-B-cell-26

1 0 | ALL-B-cell-27

1 0 | AML-28

1 0 | AML-29

1 0 | AML-30

1 0 | AML-31

1 0 | AML-32

1 0 | AML-33

1 0 | AML-34

1 0 | AML-35

1 0 | AML-36

1 0 | AML-37

1 0 | AML-38

Cluster 0 <-- ALL-T-cell-2

Cluster 1 <-- ALL-B-cell-17

**(4.)HIERARCHICAL CLUSTERING, K=3**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 2 <-- assigned to cluster

1 0 0 | ALL-T-cell-2

1 0 0 | ALL-T-cell-3

1 0 0 | ALL-T-cell-6

1 0 0 | ALL-T-cell-9

1 0 0 | ALL-T-cell-10

1 0 0 | ALL-T-cell-11

1 0 0 | ALL-T-cell-14

1 0 0 | ALL-T-cell-23

1 0 0 | ALL-B-cell-1

1 0 0 | ALL-B-cell-4

1 0 0 | ALL-B-cell-5

1 0 0 | ALL-B-cell-7

1 0 0 | ALL-B-cell-8

1 0 0 | ALL-B-cell-12

1 0 0 | ALL-B-cell-13

1 0 0 | ALL-B-cell-15

1 0 0 | ALL-B-cell-16

0 1 0 | ALL-B-cell-17

1 0 0 | ALL-B-cell-18

1 0 0 | ALL-B-cell-19

0 0 1 | ALL-B-cell-20

1 0 0 | ALL-B-cell-21

1 0 0 | ALL-B-cell-22

1 0 0 | ALL-B-cell-24

1 0 0 | ALL-B-cell-25

1 0 0 | ALL-B-cell-26

1 0 0 | ALL-B-cell-27

1 0 0 | AML-28

1 0 0 | AML-29

1 0 0 | AML-30

1 0 0 | AML-31

1 0 0 | AML-32

1 0 0 | AML-33

1 0 0 | AML-34

1 0 0 | AML-35

1 0 0 | AML-36

1 0 0 | AML-37

1 0 0 | AML-38

Cluster 0 <-- ALL-T-cell-2

Cluster 1 <-- ALL-B-cell-17

Cluster 2 <-- ALL-B-cell-20

**II. Clustering after PCA:**

**List of Selected 50 attributes**

Gene Accession Number

AFFX-BioB-M\_at

AFFX-BioC-5\_at

AFFX-CreX-5\_at

AFFX-BioB-5\_st

AFFX-DapX-3\_at

AFFX-LysX-M\_at

AFFX-PheX-M\_at

AFFX-TrpnX-3\_at

AFFX-HUMGAPDH/M33197\_3\_at

AFFX-HUMTFRR/M11507\_3\_at

AFFX-M27830\_M\_at

AB000460\_at

AB001325\_at

AB003177\_at

AC002045\_xpt1\_at

AD000092\_cds2\_at

AF000562\_at

AF005775\_at

AF007111\_at

AF009368\_at

AJ001047\_at

D00762\_at

D11094\_at

D13305\_at

D13636\_at

D13900\_at

D14659\_at

D14811\_at

D16294\_at

D17391\_at

D17716\_at

D21205\_at

D23660\_at

D25328\_at

D26135\_at

D26443\_at

D26579\_at

D28114\_at

D28589\_at

D30756\_at

D31886\_at

D38553\_at

D45248\_at

D49817\_at

D50863\_at

D55640\_at

D63483\_at

D64158\_at

D78586\_at

D86965\_at

**(1.) KMEANS K=2**

**SSE : 83.97363052711**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 <-- assigned to cluster

1 0 | ALL-T-cell-2

0 1 | ALL-T-cell-3

0 1 | ALL-T-cell-6

1 0 | ALL-T-cell-9

1 0 | ALL-T-cell-10

1 0 | ALL-T-cell-11

1 0 | ALL-T-cell-14

1 0 | ALL-T-cell-23

1 0 | ALL-B-cell-1

1 0 | ALL-B-cell-4

1 0 | ALL-B-cell-5

1 0 | ALL-B-cell-7

0 1 | ALL-B-cell-8

1 0 | ALL-B-cell-12

1 0 | ALL-B-cell-13

1 0 | ALL-B-cell-15

1 0 | ALL-B-cell-16

1 0 | ALL-B-cell-17

1 0 | ALL-B-cell-18

1 0 | ALL-B-cell-19

0 1 | ALL-B-cell-20

1 0 | ALL-B-cell-21

1 0 | ALL-B-cell-22

0 1 | ALL-B-cell-24

1 0 | ALL-B-cell-25

1 0 | ALL-B-cell-26

1 0 | ALL-B-cell-27

1 0 | AML-28

0 1 | AML-29

1 0 | AML-30

0 1 | AML-31

1 0 | AML-32

0 1 | AML-33

0 1 | AML-34

0 1 | AML-35

1 0 | AML-36

0 1 | AML-37

0 1 | AML-38

Cluster 0 <-- ALL-T-cell-2

Cluster 1 <-- ALL-T-cell-3

**(2.) KMEANS K=3**

**SSE : 72.91972564024564**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 2 <-- assigned to cluster

1 0 0 | ALL-T-cell-2

0 1 0 | ALL-T-cell-3

0 1 0 | ALL-T-cell-6

1 0 0 | ALL-T-cell-9

1 0 0 | ALL-T-cell-10

1 0 0 | ALL-T-cell-11

1 0 0 | ALL-T-cell-14

1 0 0 | ALL-T-cell-23

1 0 0 | ALL-B-cell-1

1 0 0 | ALL-B-cell-4

1 0 0 | ALL-B-cell-5

1 0 0 | ALL-B-cell-7

0 1 0 | ALL-B-cell-8

1 0 0 | ALL-B-cell-12

1 0 0 | ALL-B-cell-13

1 0 0 | ALL-B-cell-15

1 0 0 | ALL-B-cell-16

1 0 0 | ALL-B-cell-17

1 0 0 | ALL-B-cell-18

1 0 0 | ALL-B-cell-19

0 1 0 | ALL-B-cell-20

1 0 0 | ALL-B-cell-21

1 0 0 | ALL-B-cell-22

0 1 0 | ALL-B-cell-24

1 0 0 | ALL-B-cell-25

1 0 0 | ALL-B-cell-26

1 0 0 | ALL-B-cell-27

1 0 0 | AML-28

0 1 0 | AML-29

1 0 0 | AML-30

0 1 0 | AML-31

1 0 0 | AML-32

0 1 0 | AML-33

0 0 1 | AML-34

0 1 0 | AML-35

1 0 0 | AML-36

0 1 0 | AML-37

0 0 1 | AML-38

Cluster 0 <-- ALL-T-cell-2

Cluster 1 <-- ALL-T-cell-3

Cluster 2 <-- AML-34

**(3.)HIERARCHICAL K=2**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 <-- assigned to cluster

1 0 | ALL-T-cell-2

1 0 | ALL-T-cell-3

1 0 | ALL-T-cell-6

1 0 | ALL-T-cell-9

1 0 | ALL-T-cell-10

1 0 | ALL-T-cell-11

1 0 | ALL-T-cell-14

1 0 | ALL-T-cell-23

1 0 | ALL-B-cell-1

1 0 | ALL-B-cell-4

1 0 | ALL-B-cell-5

1 0 | ALL-B-cell-7

1 0 | ALL-B-cell-8

1 0 | ALL-B-cell-12

1 0 | ALL-B-cell-13

1 0 | ALL-B-cell-15

1 0 | ALL-B-cell-16

1 0 | ALL-B-cell-17

1 0 | ALL-B-cell-18

1 0 | ALL-B-cell-19

1 0 | ALL-B-cell-20

1 0 | ALL-B-cell-21

1 0 | ALL-B-cell-22

1 0 | ALL-B-cell-24

1 0 | ALL-B-cell-25

1 0 | ALL-B-cell-26

1 0 | ALL-B-cell-27

1 0 | AML-28

1 0 | AML-29

1 0 | AML-30

1 0 | AML-31

1 0 | AML-32

1 0 | AML-33

0 1 | AML-34

1 0 | AML-35

1 0 | AML-36

1 0 | AML-37

0 1 | AML-38

Cluster 0 <-- ALL-T-cell-2

Cluster 1 <-- AML-34

**(4.) HIERARCHICAL K=3**

Class attribute: Gene Accession Number

Classes to Clusters:

0 1 2 <-- assigned to cluster

1 0 0 | ALL-T-cell-2

1 0 0 | ALL-T-cell-3

1 0 0 | ALL-T-cell-6

1 0 0 | ALL-T-cell-9

1 0 0 | ALL-T-cell-10

1 0 0 | ALL-T-cell-11

1 0 0 | ALL-T-cell-14

1 0 0 | ALL-T-cell-23

1 0 0 | ALL-B-cell-1

1 0 0 | ALL-B-cell-4

1 0 0 | ALL-B-cell-5

1 0 0 | ALL-B-cell-7

0 1 0 | ALL-B-cell-8

1 0 0 | ALL-B-cell-12

1 0 0 | ALL-B-cell-13

1 0 0 | ALL-B-cell-15

1 0 0 | ALL-B-cell-16

1 0 0 | ALL-B-cell-17

1 0 0 | ALL-B-cell-18

1 0 0 | ALL-B-cell-19

1 0 0 | ALL-B-cell-20

1 0 0 | ALL-B-cell-21

1 0 0 | ALL-B-cell-22

1 0 0 | ALL-B-cell-24

1 0 0 | ALL-B-cell-25

1 0 0 | ALL-B-cell-26

1 0 0 | ALL-B-cell-27

1 0 0 | AML-28

1 0 0 | AML-29

1 0 0 | AML-30

1 0 0 | AML-31

1 0 0 | AML-32

1 0 0 | AML-33

0 0 1 | AML-34

1 0 0 | AML-35

1 0 0 | AML-36

1 0 0 | AML-37

0 0 1 | AML-38

Cluster 0 <-- ALL-T-cell-2

Cluster 1 <-- ALL-B-cell-8

Cluster 2 <-- AML-34

**III. Classification**

**Training Data**

|  |  |  |
| --- | --- | --- |
| Algorithms | Parameters | %Training Accuracy |
| Random Forest | Maxdepth=5 Numtrees=220 Seed=15 | 100 |
| Boosting | NumIterations=100 Seed=10 Weightthreshold=70 | 100 |
| Bagging | NumIterations=100 Seed=100 maxdepth=5  Numtrees=100 | 100 |
| J48 | confidenceFactor=0.25  minNumObj=1  minFolds=1  seed=50 | 100 |

**Test Data**

|  |  |  |
| --- | --- | --- |
| Algorithms | Parameters | %Training Accuracy |
| Random Forest | Maxdepth=1 Numtrees=100 Seed=10  minFeature=3 | 7.8947 |
| Boosting | NumIterations=100 Seed=10 Weightthreshold=70 | 2.6316 |
| Bagging | NumIterations=100 Seed=100 maxdepth=5  Numtrees=100 | 3.6316 |
| J48 | confidenceFactor=0.25  minNumObj=1  minFolds=1  seed=50 | 4.7368 |