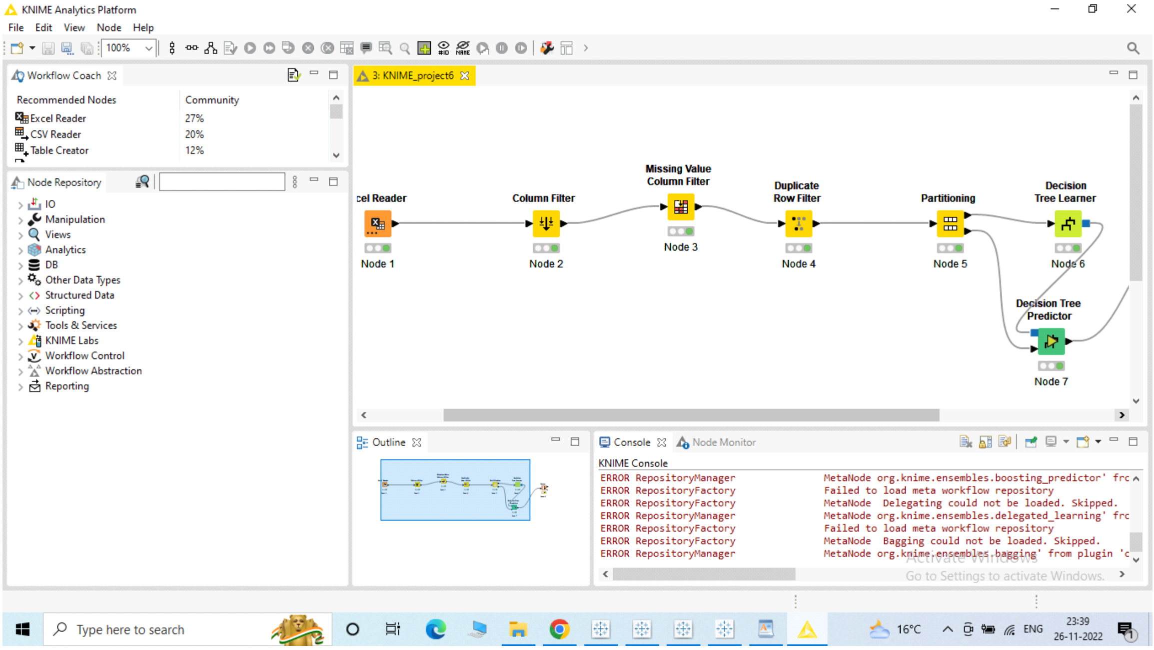
BIDV KNIME PROJECT REPORT

DEVESH RANA (20BSM007)

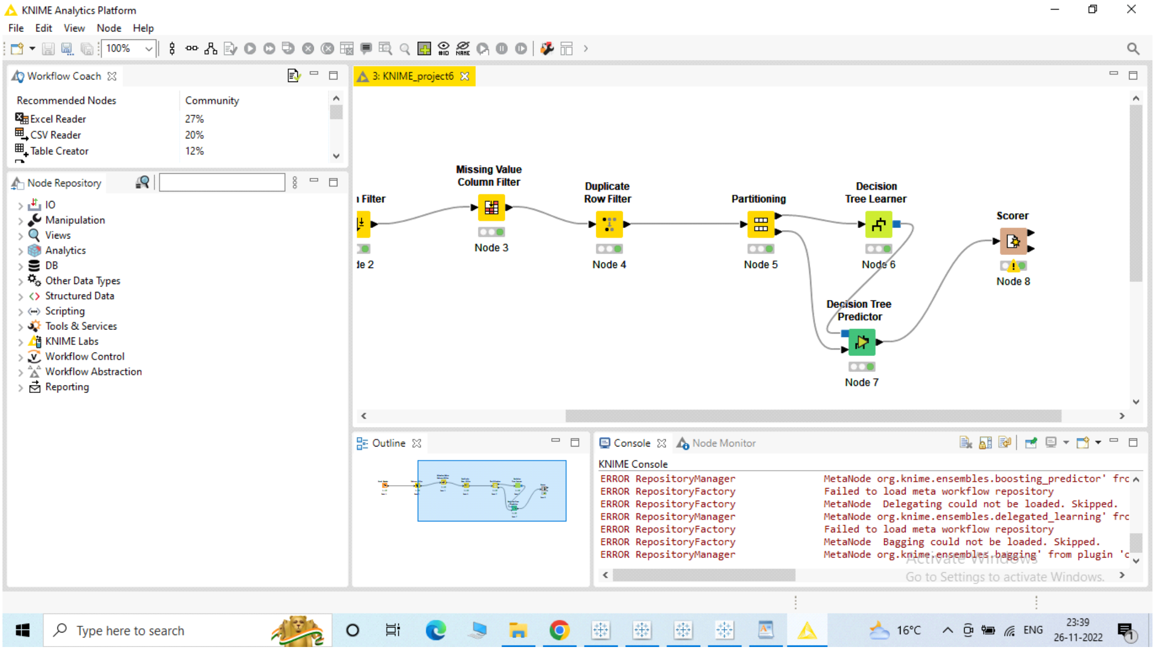
KNIME PROJECT -

Insert all the required nodes and filter the data(pre processing) one by one by configuring the nodes and then execute each node resp. to get the final results. During partioning data will be trained and tested and will be divided into 80&20 ratio. after all the nodes are successfully executed we can find the accuracy of the data and the confusion matrix.

Insertion of nodes-: Firstly, I created the nodes that i required and I used the data of HR ANALYTICS, then read the data with the help of excel reader.

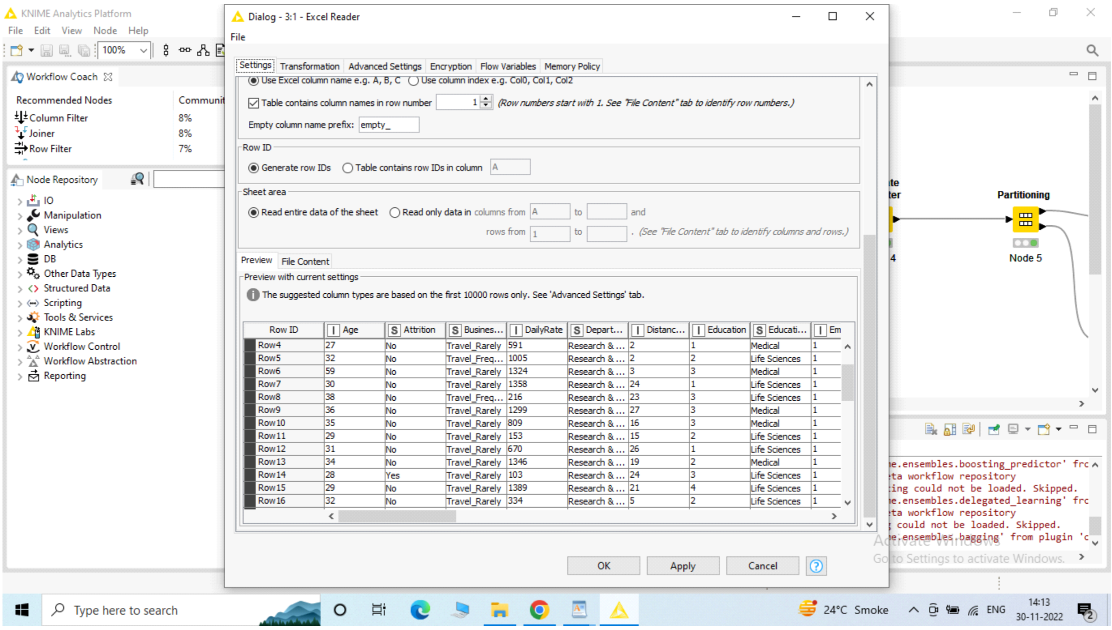


```
ERROR RepositoryManager         MetaNode org.knime.ensembles.boosting_predictor' frc
ERROR RepositoryFactory         Failed to load meta workflow repository
ERROR RepositoryFactory         MetaNode  Delegating could not be loaded. Skipped.
ERROR RepositoryManager         MetaNode org.knime.ensembles.delegated_learning' frc
ERROR RepositoryFactory         Failed to load meta workflow repository
ERROR RepositoryFactory         MetaNode  Bagging could not be loaded. Skipped.
ERROR RepositoryManager         MetaNode org.knime.ensembles.bagging' from plugin 'c
```



```
ERROR RepositoryManager          MetaNode org.knime.ensembles.boosting_predictor' frc
ERROR RepositoryFactory          Failed to load meta workflow repository
ERROR RepositoryFactory          MetaNode  Delegating could not be loaded. Skipped.
ERROR RepositoryManager          MetaNode org.knime.ensembles.delegated_learning' frc
ERROR RepositoryFactory          Failed to load meta workflow repository
ERROR RepositoryFactory          MetaNode  Bagging could not be loaded. Skipped.
ERROR RepositoryManager          MetaNode org.knime.ensembles.bagging' from plugin 'c
```

DATASET-: HR Analytics data used



● Use Excel column name e.g. A, B, C ○ Use column index e.g. Col0, Col1, Col2

☑ Table contains column names in row number 1 *(Row numbers start with 1. See "File Content" tab to identify row numbers.)*

Empty column name prefix: empty_

Row ID

● Generate row IDs ○ Table contains row IDs in column A

Sheet area

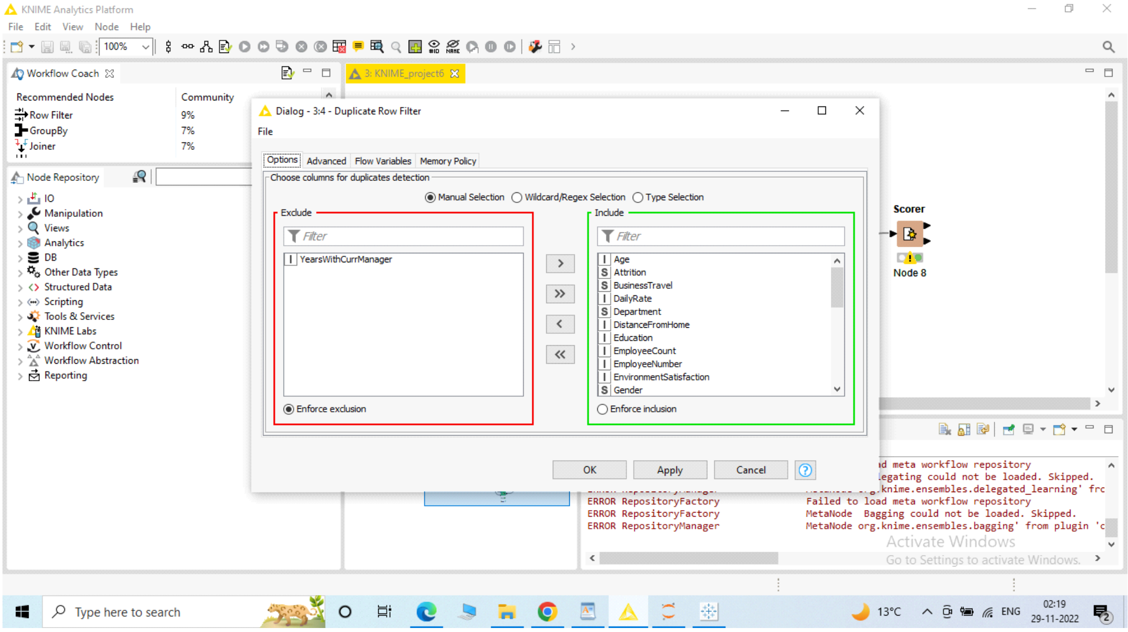
● Read entire data of the sheet ○ Read only data in columns from A to and rows from 1 to . *(See "File Content" tab to identify columns and rows.)*

Preview with current settings

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	Age	Attrition	Busines...	DailyRate	Depart...	Distanc...	Education	Educati...	Em
Row4	27	No	Travel_Rarely	591	Research & ...	2	1	Medical	1
Row5	32	No	Travel_Freq...	1005	Research & ...	2	2	Life Sciences	1
Row6	59	No	Travel_Rarely	1324	Research & ...	3	3	Medical	1
Row7	30	No	Travel_Rarely	1358	Research & ...	24	1	Life Sciences	1
Row8	38	No	Travel_Freq...	216	Research & ...	23	3	Life Sciences	1
Row9	36	No	Travel_Rarely	1299	Research & ...	27	3	Medical	1
Row10	35	No	Travel_Rarely	809	Research & ...	16	3	Medical	1
Row11	29	No	Travel_Rarely	153	Research & ...	15	2	Life Sciences	1
Row12	31	No	Travel_Rarely	670	Research & ...	26	1	Life Sciences	1
Row13	34	No	Travel_Rarely	1346	Research & ...	19	2	Medical	1
Row14	28	Yes	Travel_Rarely	103	Research & ...	24	3	Life Sciences	1
Row15	29	No	Travel_Rarely	1389	Research & ...	21	4	Life Sciences	1
Row16	32	No	Travel_Rarely	334	Research & ...	5	2	Life Sciences	1

Filtering the data in nodes -: After reading the data using excel reader I filtered the data according to my requirement using different column and row filters.



Dialog - 3:4 - Duplicate Row Filter

Options | Advanced | Flow Variables | Memory Policy

Choose columns for duplicates detection

● Manual Selection ○ Wildcard/Regex Selection ○ Type Selection

Exclude:
- I YearsWithCurrManager

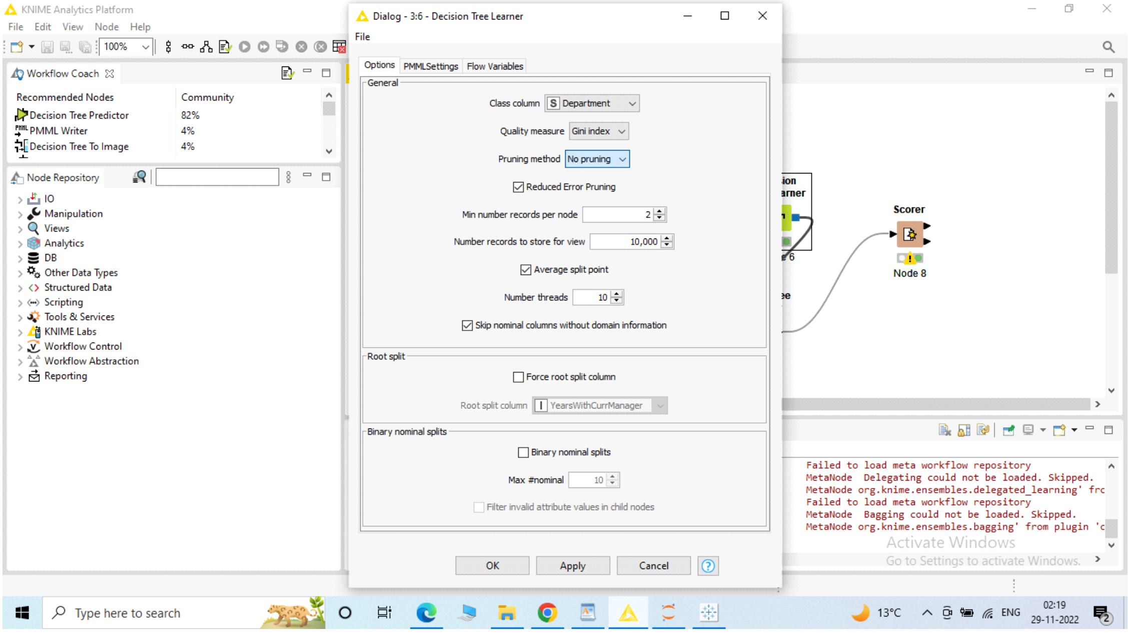
● Enforce exclusion

Include:
- I Age
- S Attrition
- S BusinessTravel
- I DailyRate
- S Department
- I DistanceFromHome
- I Education
- I EmployeeCount
- I EmployeeNumber
- I EnvironmentSatisfaction
- S Gender

○ Enforce inclusion

```
ERROR RepositoryFactory          Failed to load meta workflow repository
ERROR RepositoryFactory          MetaNode  Bagging could not be loaded. Skipped.
ERROR RepositoryManager          MetaNode org.knime.ensembles.bagging' from plugin 'c
```

Configurations-: I configured the required nodes and made some changes



Dialog - 3:6 - Decision Tree Learner

Options | PMMLSettings | Flow Variables

General

Class column: S Department

Quality measure: Gini index

Pruning method: No pruning

☑ Reduced Error Pruning

Min number records per node: 2

Number records to store for view: 10,000

☑ Average split point

Number threads: 10

☑ Skip nominal columns without domain information

Root split

☐ Force root split column

Root split column: I YearsWithCurrManager

Binary nominal splits

☐ Binary nominal splits

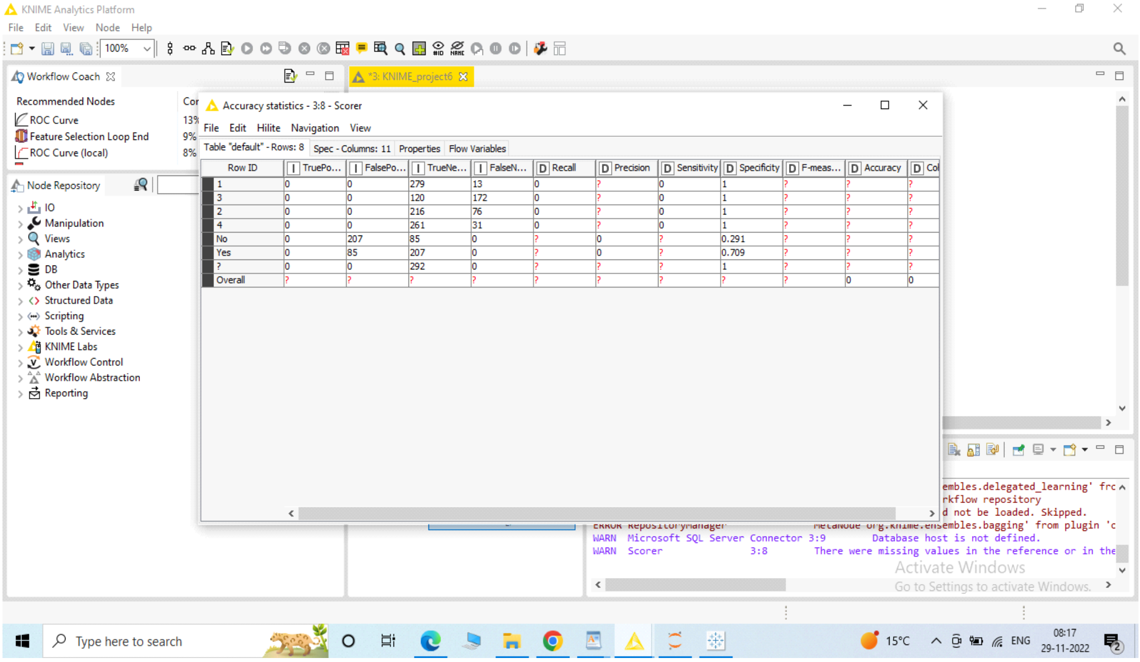
Max #nominal: 10

☐ Filter invalid attribute values in child nodes

```
Failed to load meta workflow repository
MetaNode  Delegating could not be loaded. Skipped.
MetaNode org.knime.ensembles.delegated_learning' frc
Failed to load meta workflow repository
MetaNode  Bagging could not be loaded. Skipped.
MetaNode org.knime.ensembles.bagging' from plugin 'c
```

- After all the nodes are successfully configured, filtered and executed we can get the final results in the SCORER NODE where we can find the Accuracy and Confusion Matrix of the data.

Accuracy -:



Accuracy statistics - 3:8 - Scorer

File Edit Hilite Navigation View

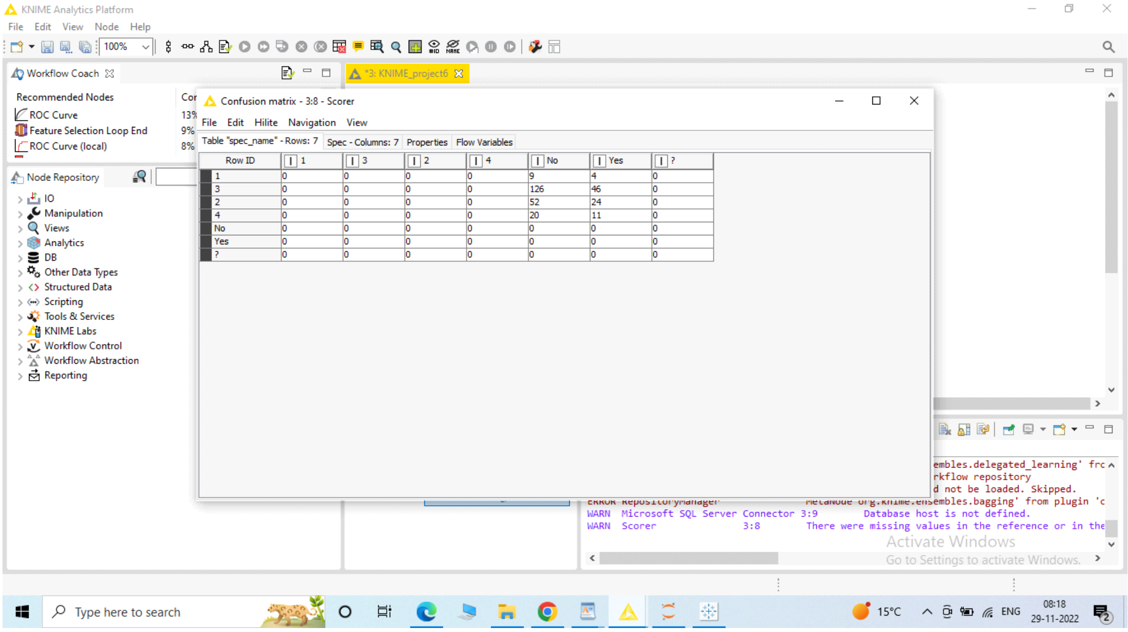
Table "default" - Rows: 8 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specificity	F-meas...	Accuracy	Co
1	0	0	279	13	0	?	0	1	?	?	?
3	0	0	120	172	0	?	0	1	?	?	?
2	0	0	216	76	0	?	0	1	?	?	?
4	0	0	261	31	0	?	0	1	?	?	?
No	0	207	85	0	?	0	?	0.291	?	?	?
Yes	0	85	207	0	?	0	?	0.709	?	?	?
?	0	0	292	0	?	?	?	1	?	?	?
Overall	?	?	?	?	?	?	?	?	?	0	0

```
embles.delegated_learning' frc
rkflow repository
d not be loaded. Skipped.
ERROR RepositoryManager             MetaNode org.knime.ensembles.bagging' from plugin 'c
WARN  Microsoft SQL Server Connector 3:9       Database host is not defined.
WARN  Scorer                   3:8        There were missing values in the reference or in the
```

Activate Windows
Go to Settings to activate Windows.

Confusion matrix-:



Confusion matrix - 3:8 - Scorer

File Edit Hilite Navigation View

Table "spec_name" - Rows: 7 | Spec - Columns: 7 | Properties | Flow Variables

Row ID	1	3	2	4	No	Yes	?
1	0	0	0	0	9	4	0
3	0	0	0	0	126	46	0
2	0	0	0	0	52	24	0
4	0	0	0	0	20	11	0
No	0	0	0	0	0	0	0
Yes	0	0	0	0	0	0	0
?	0	0	0	0	0	0	0

```
embles.delegated_learning' frc
rkflow repository
d not be loaded. Skipped.
ERROR RepositoryManager             MetaNode org.knime.ensembles.bagging' from plugin 'c
WARN  Microsoft SQL Server Connector 3:9       Database host is not defined.
WARN  Scorer                   3:8        There were missing values in the reference or in the
```

Activate Windows
Go to Settings to activate Windows.