Term Project Report (Data Mining)

GROUP MEMBER

M ZAIN MEHMOOD (19122051)

HAFIZ M ATIF SHAMS (19122118)

BSCS (Semester 6 “A”)

CRICKET MATCH WINNING PREDICTION MODEL

COURSE INSTRUCTOR: Miss UZMA

INTRODUCTION:

Cricket is one of the most popular sports in the world, with billions of fans following the game. Predicting the outcome of cricket matches is a challenging task. Traditional methods of match prediction rely on expert opinions and statistical analysis, but these methods are often subjective and can be inaccurate.

Data mining techniques offer a promising approach for predicting cricket match outcomes. Data mining involves extracting knowledge from large datasets to identify patterns and relationships. This knowledge can then be used to build predictive models that can forecast future outcomes.

Major Columns of Dataset:

- **Ground Name**
- **Opposition**
- **Inning**
- **Score**
- **RPO**
- **Over**
- **Average Score in Last 10 matches**
- **Probability of Winning**

OBJECTIVE:

- Collect and pre-process a comprehensive dataset of historical cricket match data.
- Identify relevant features from the dataset that influence match outcomes.
- Develop decision tree and random forest models to predict match outcomes.
- Evaluate the performance of the models using various metrics.

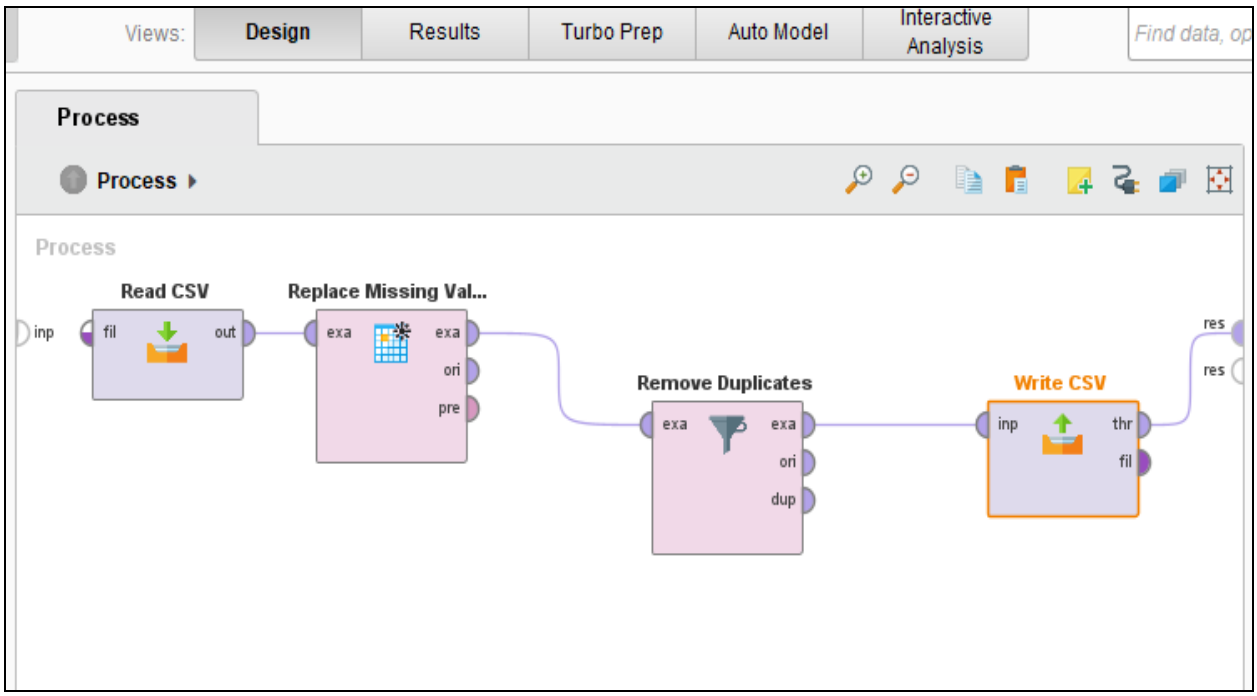
DATA COLLECTION:

Data Set has collected from the official website of cricinfo

https://www.espncricinfo.com/

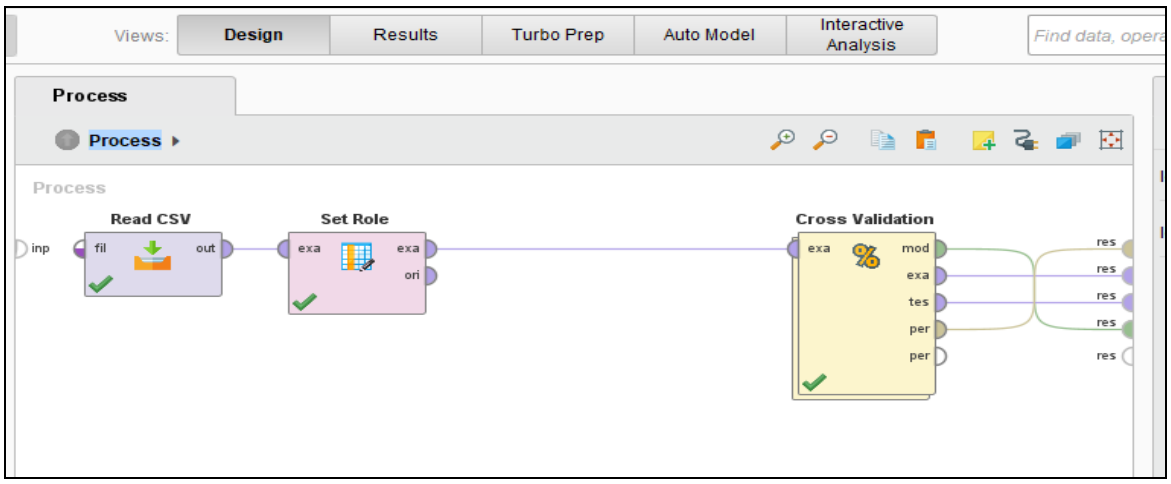
Data Preprocessing:

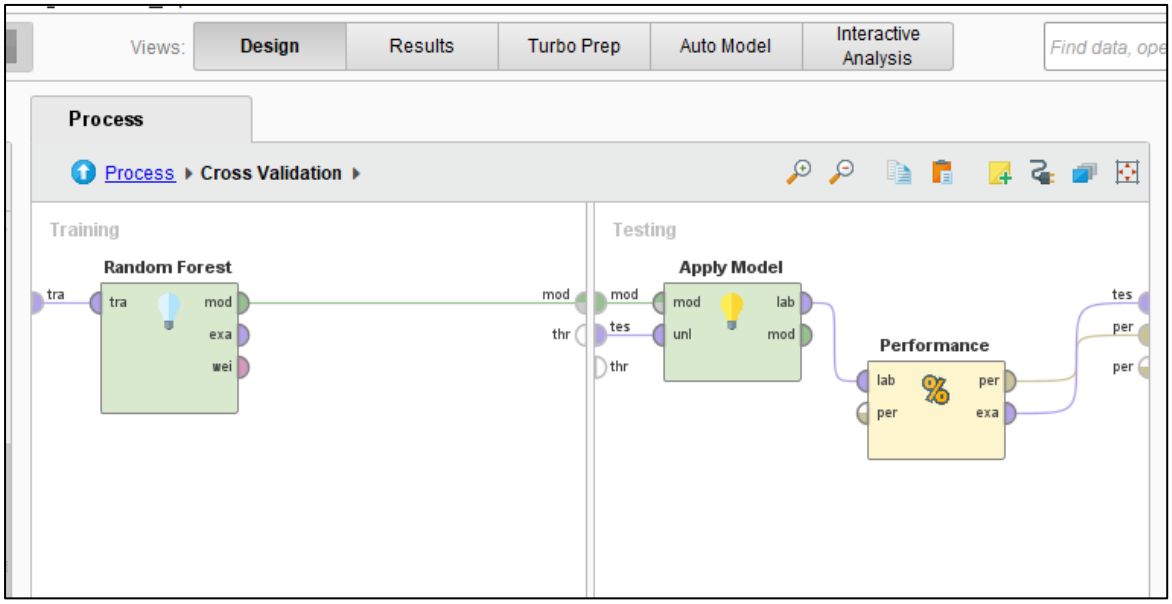
In this Step We check for the missing Values if any missing values found in the data set so it has replaced by using Replace Missing Value operator, also remove the duplicate data from data set through Remove Duplicate Operator.



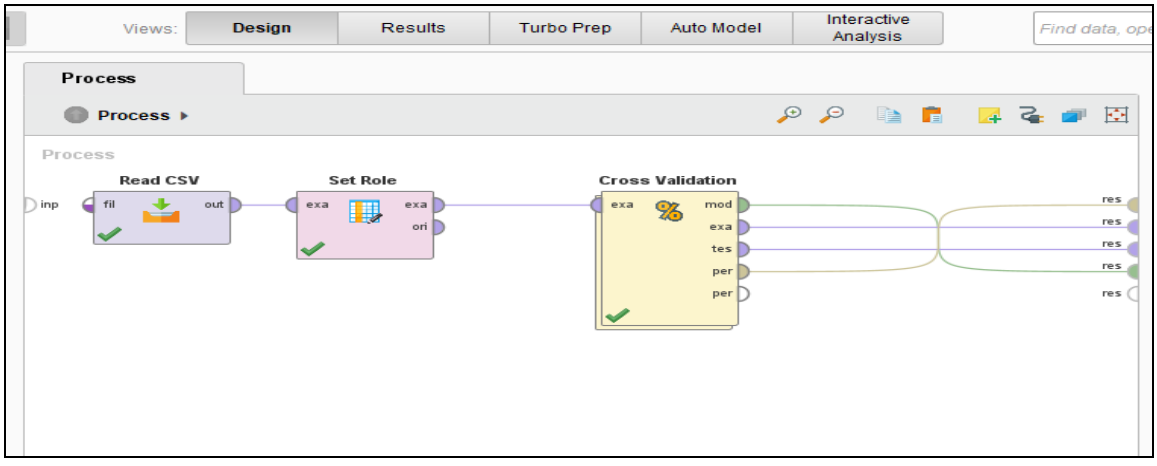
Modeling and Evaluation:

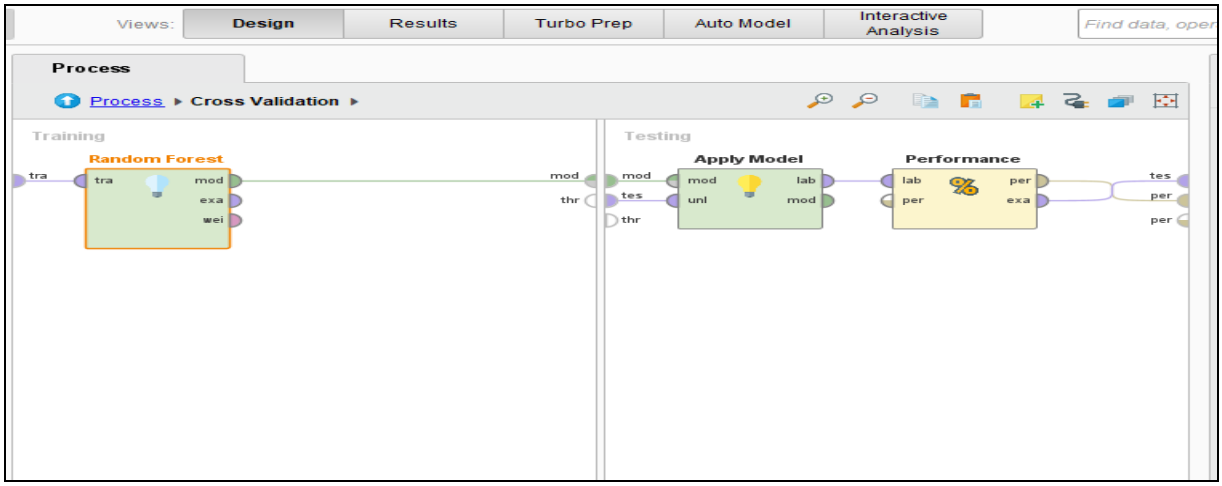
Decision Tree:





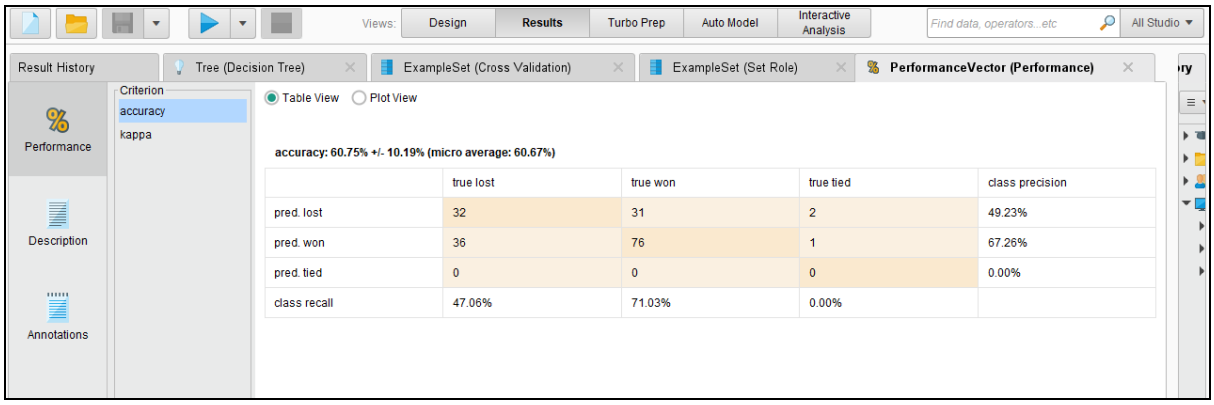
Random Forest:

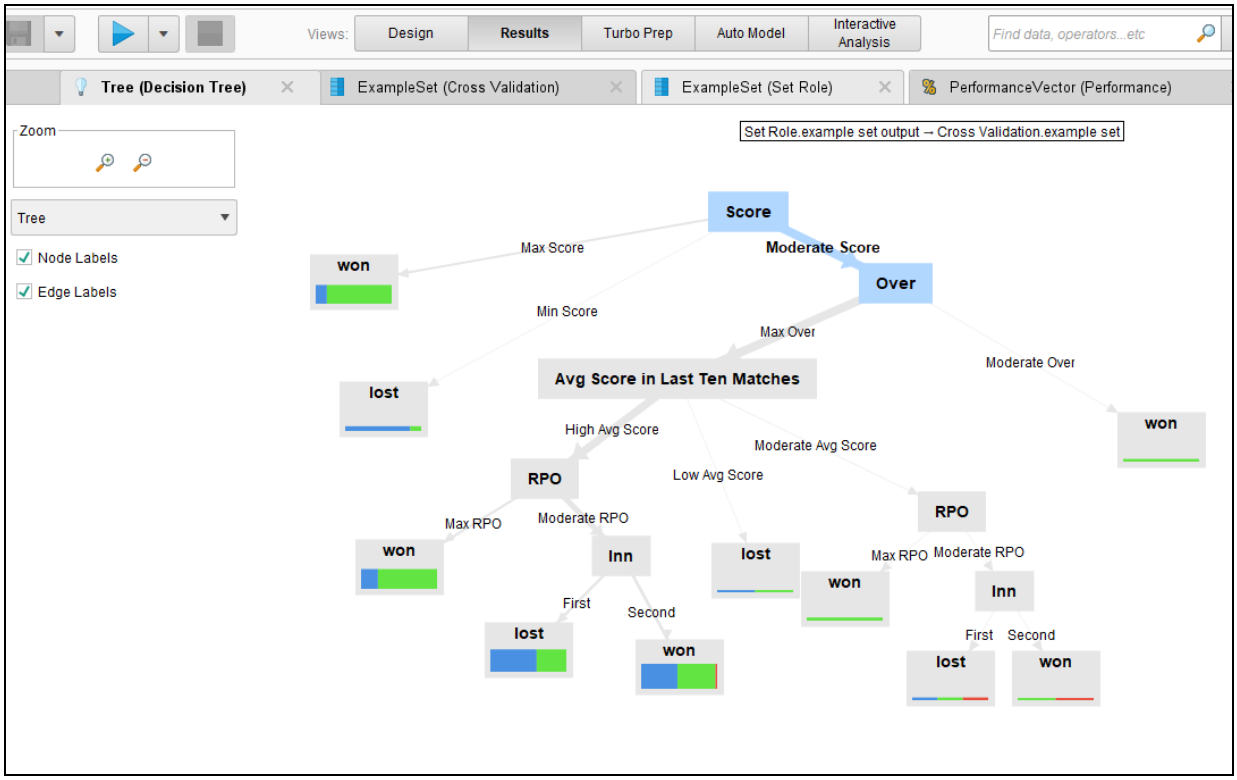




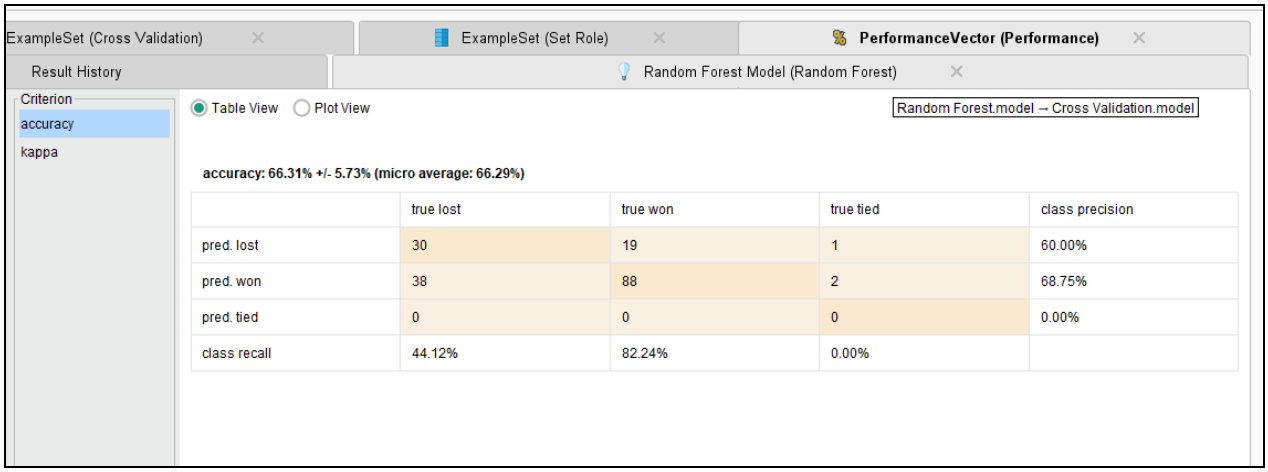
RESULTS:

T20 Decision Tree:

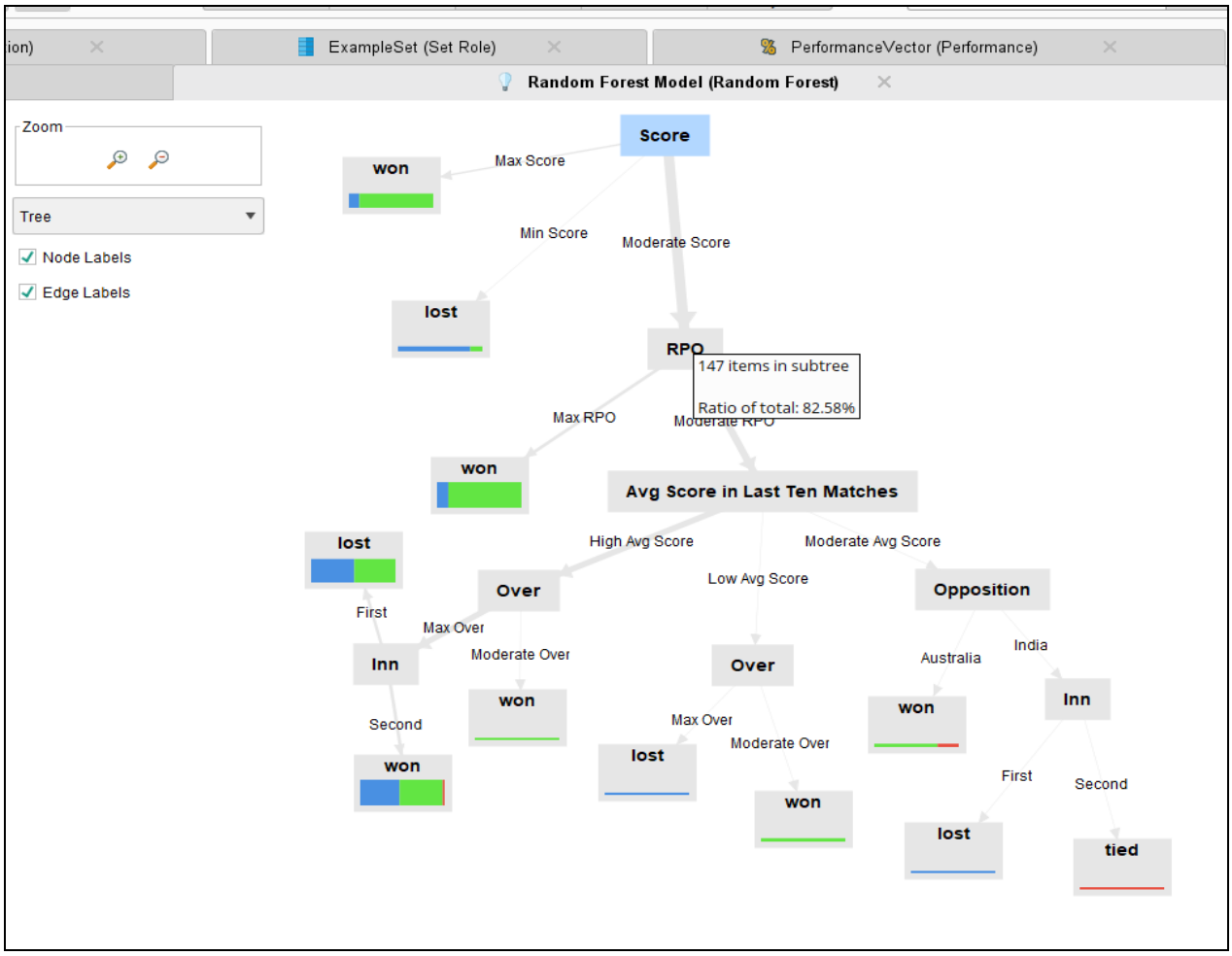




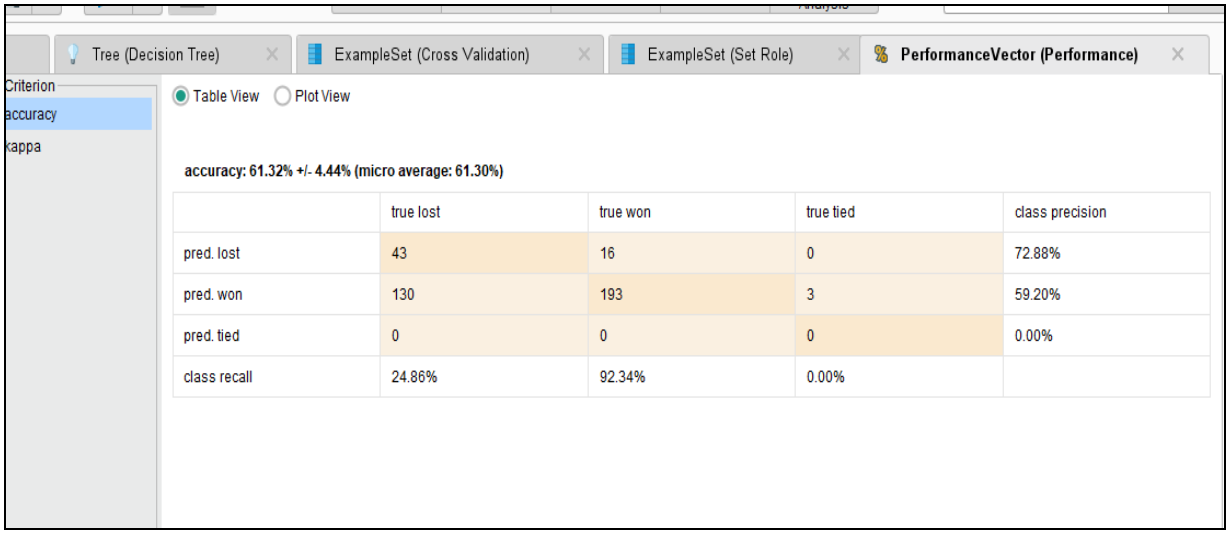
T20 Random Forest:



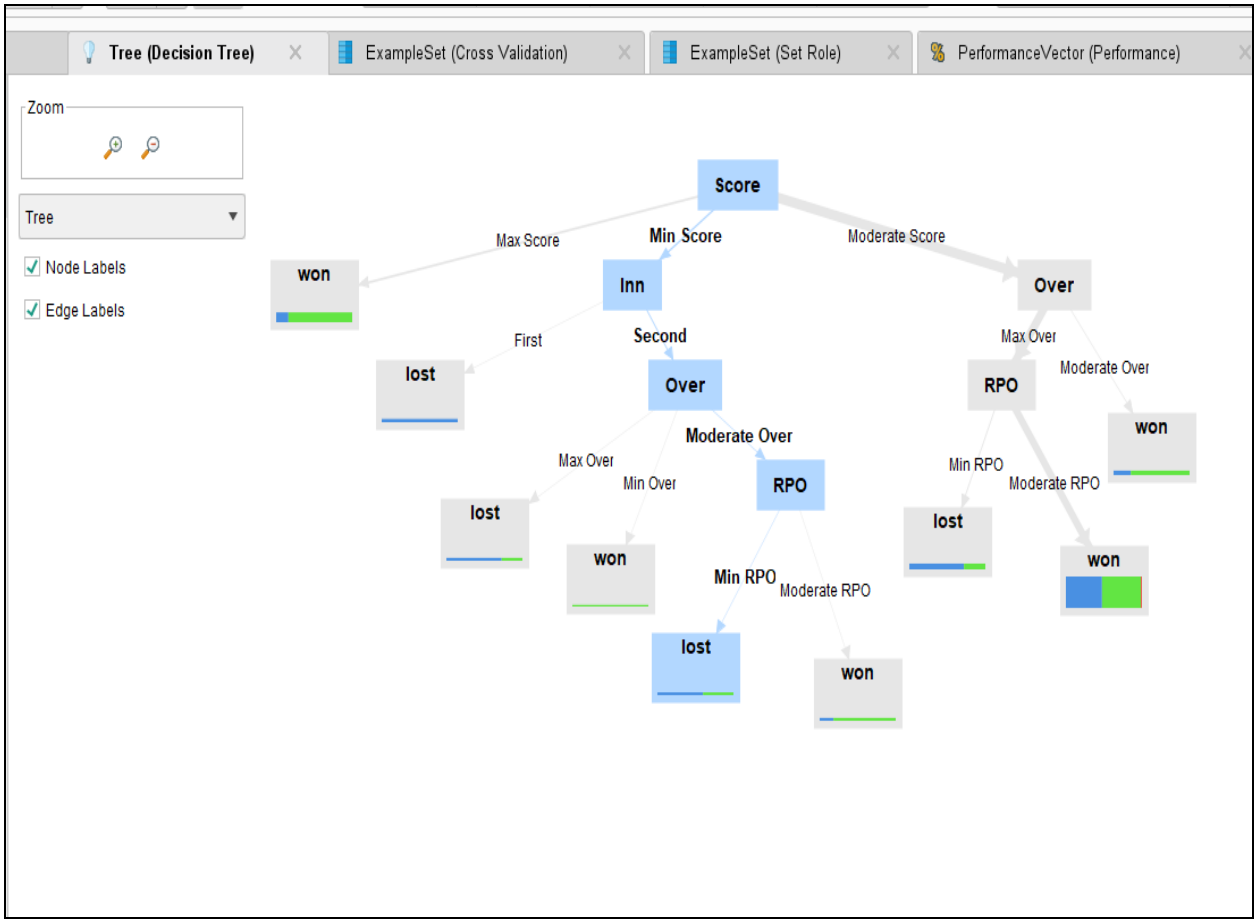
	true lost	true won	true tied	class precision
pred. lost	30	19	1	60.00%
pred. won	38	88	2	68.75%
pred. tied	0	0	0	0.00%
class recall	44.12%	82.24%	0.00%	



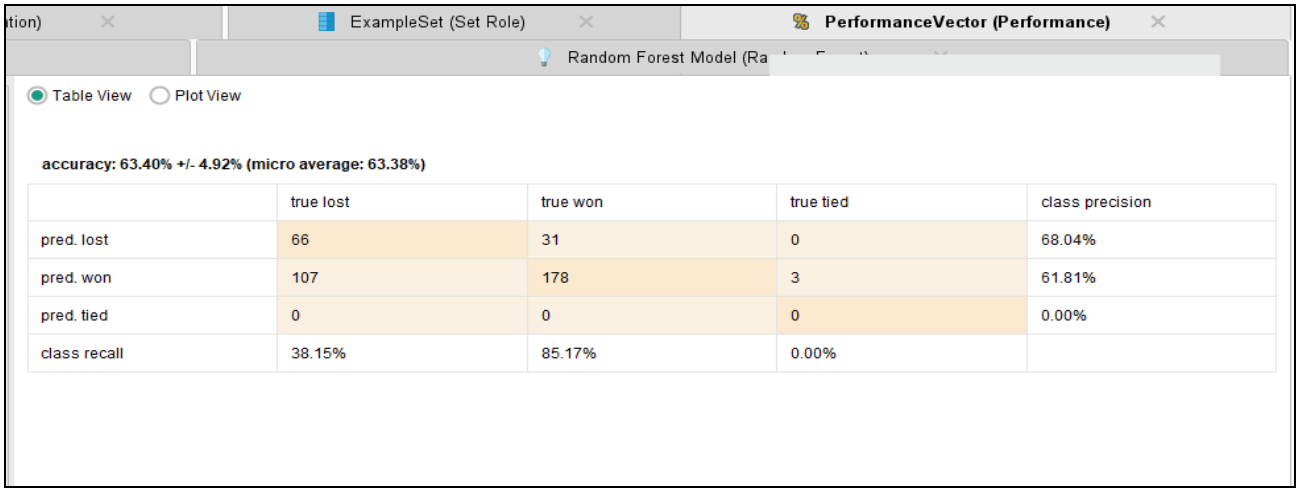
ODI Decision Tree:



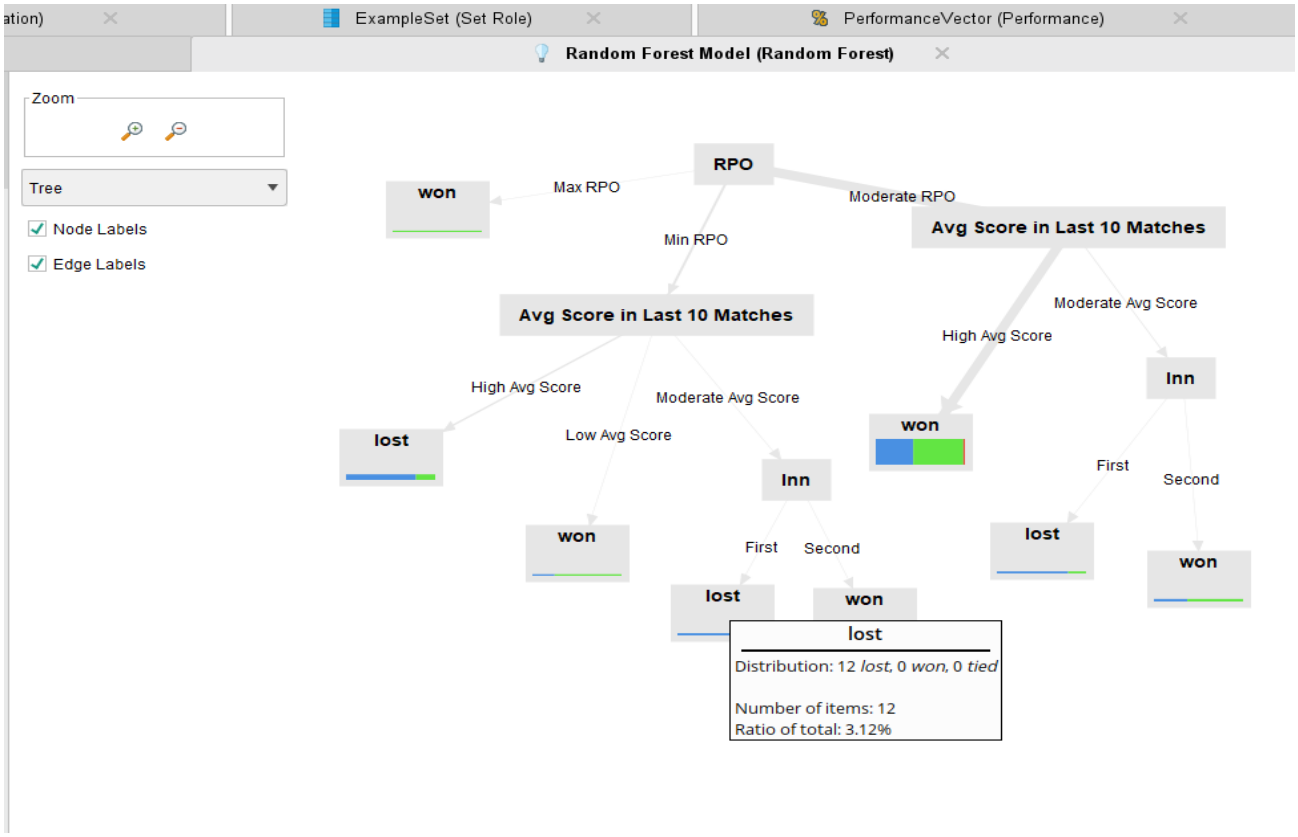
	true lost	true won	true tied	class precision
pred. lost	43	16	0	72.88%
pred. won	130	193	3	59.20%
pred. tied	0	0	0	0.00%
class recall	24.86%	92.34%	0.00%	



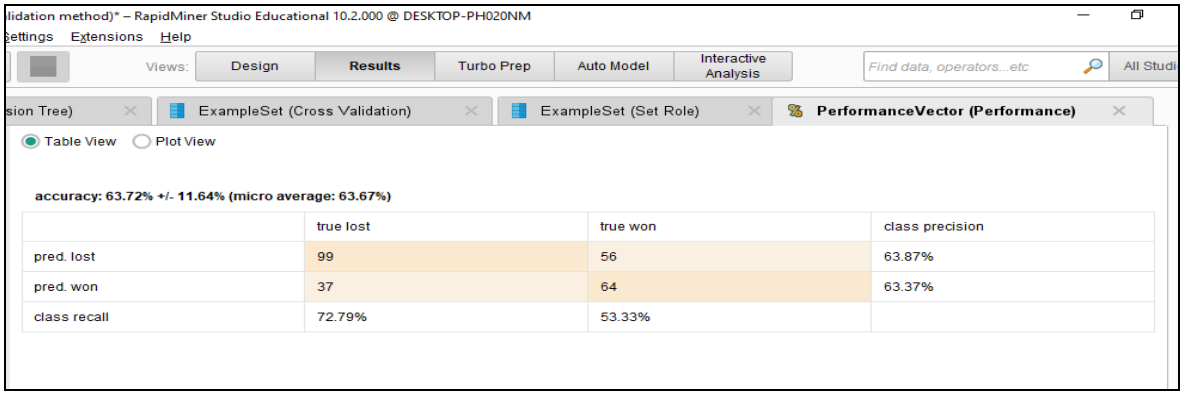
ODI Random Forest:



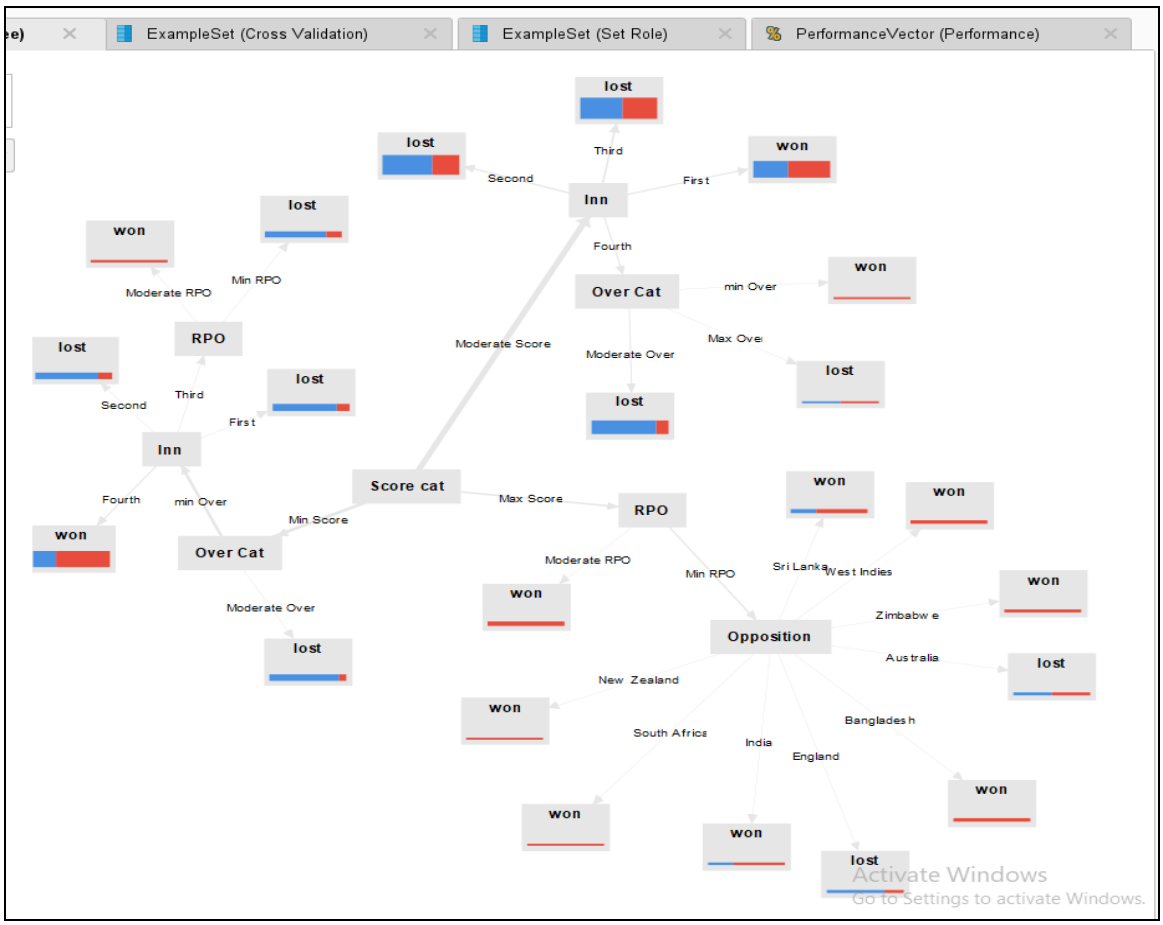
	true lost	true won	true tied	class precision
pred. lost	66	31	0	68.04%
pred. won	107	178	3	61.81%
pred. tied	0	0	0	0.00%
class recall	38.15%	85.17%	0.00%	



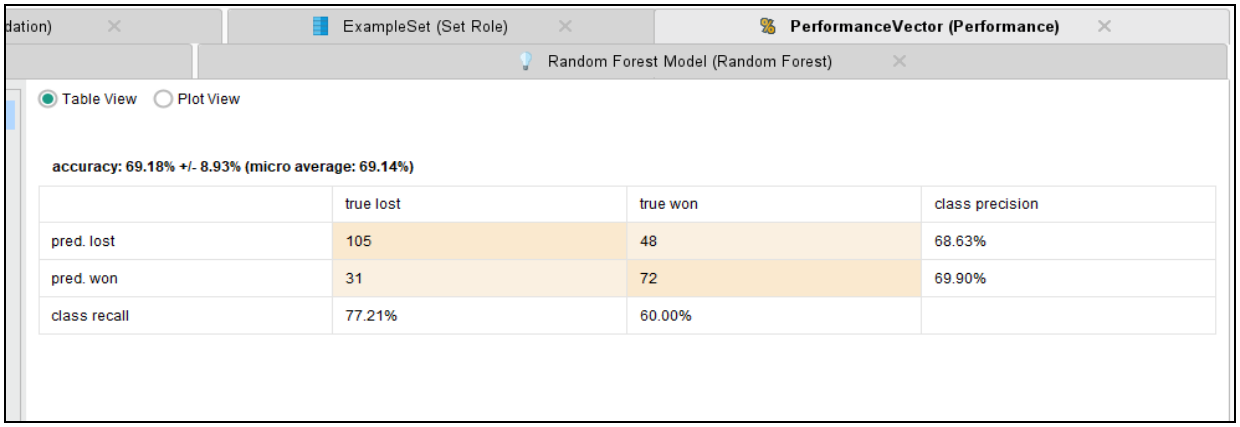
Test Decision Tree:



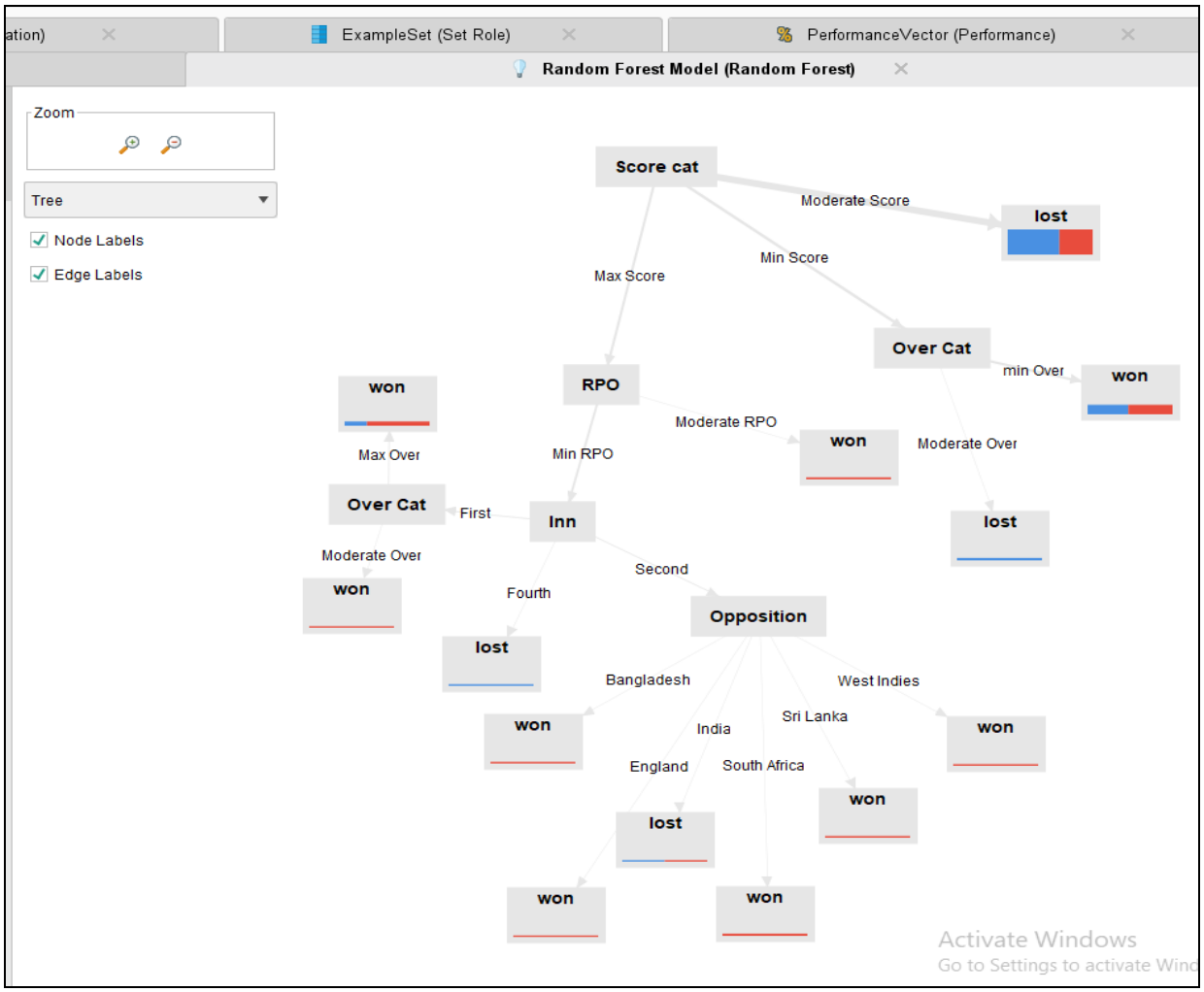
	true lost	true won	class precision
pred. lost	99	56	63.87%
pred. won	37	64	63.37%
class recall	72.79%	53.33%	



Test Random Forest:



	true lost	true won	class precision
pred. lost	105	48	68.63%
pred. won	31	72	69.90%
class recall	77.21%	60.00%	



Conclusion:

This Data Mining project utilizes Decision Tree and Random Forest algorithms to predict the winning team in Cricket matches. It has a high potential to provide stakeholders with valuable information for insightful decision-making and an enhanced Cricket experience.