Gift This Course

A-Z Machine Learning using Azure Machine Learning (AzureML)

Hands on AzureML: From Azure Machine Learning Introduction to Advance Machine Learning Algorithms. No Coding Required.

BEST SELLER ★★★★★ 4.3 (215 ratings) 1,597 students enrolled

Created by Jitesh Khurkhuriya Last updated 3/2018 English English

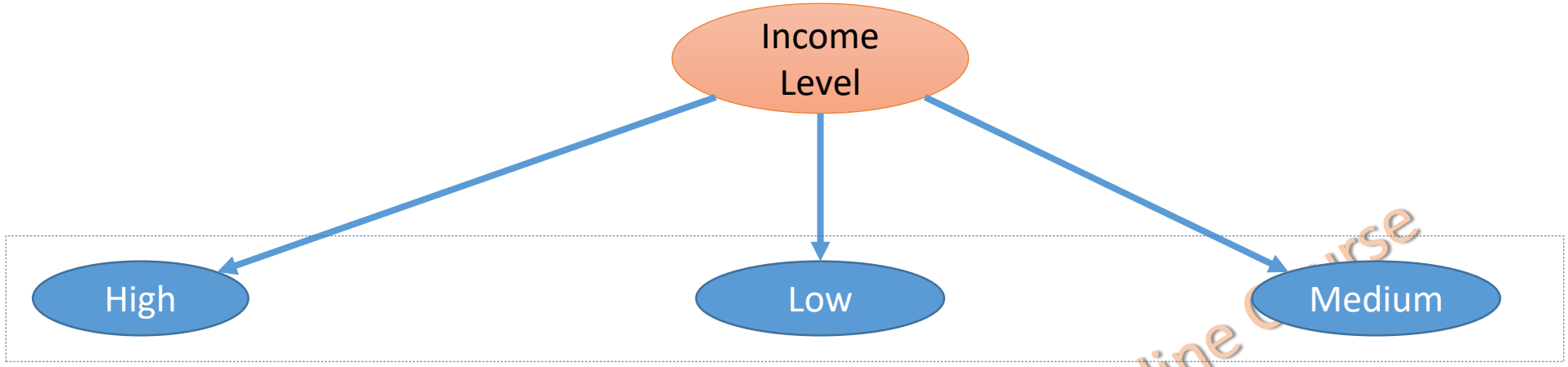
Classification

What is a Decision Tree?

What is Decision Tree?

- Supervised learning method
- Decision support tool that uses a tree-like graph or model of decisions and their possible consequences
- Various variations such as Boosted Decision Tree, Decision Forest, Decision Jungle
- Can be used for categorical as well as continuous variables

Loan ID	Income Level	Credit Score	Employment	Approved?
L1	Medium	Low	Self-Employed	No
L2	High	Low	Self-Employed	Yes
L3	High	High	Salaried	Yes
L4	Medium	Low	Salaried	Yes
L5	Low	High	Salaried	No
L6	Low	Low	Self-Employed	No
L7	High	Low	Salaried	Yes
L8	Medium	Low	Self-Employed	No
L9	High	High	Self-Employed	Yes
L10	Medium	High	Self-Employed	Yes
L11	High	Low	Salaried	Yes
L12	Medium	High	Salaried	Yes
L13	Medium	High	Self-Employed	Yes
L14	Low	Low	Self-Employed	No
L15	Low	High	Self-Employed	No



LID	IL	CS	ET	Status
L2	High	Low	SE	Yes
L3	High	High	Salaried	Yes
L7	High	Low	Salaried	Yes
L9	High	High	SE	Yes
L11	High	Low	Salaried	Yes

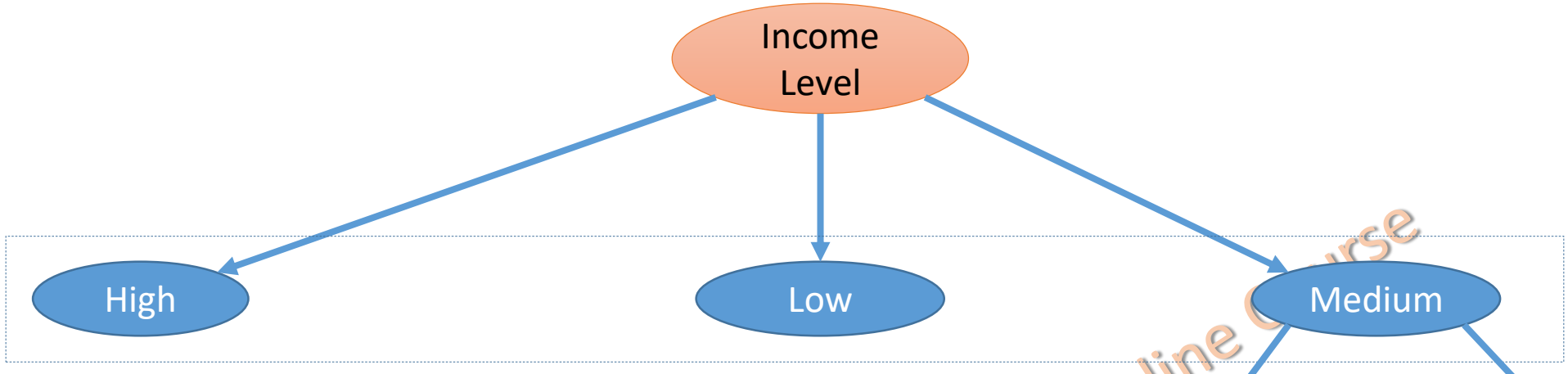
Pure Subset

LID	IL	CS	ET	Status
L5	Low	High	Salaried	No
L6	Low	Low	SE	No
L14	Low	Low	SE	No
L15	Low	High	SE	No

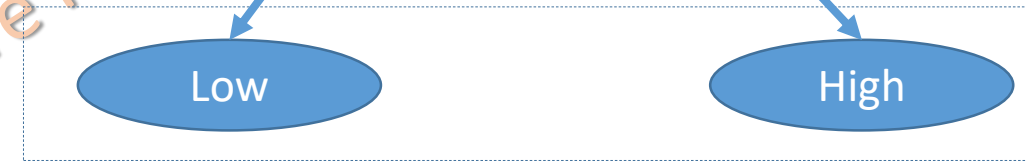
Pure Subset

LID	IL	CS	ET	Status
L1	Medium	Low	SE	No
L4	Medium	Low	Salaried	Yes
L8	Medium	Low	SE	No
L10	Medium	High	SE	Yes
L12	Medium	High	Salaried	Yes
L13	Medium	High	SE	Yes

Split Further



Credit Score

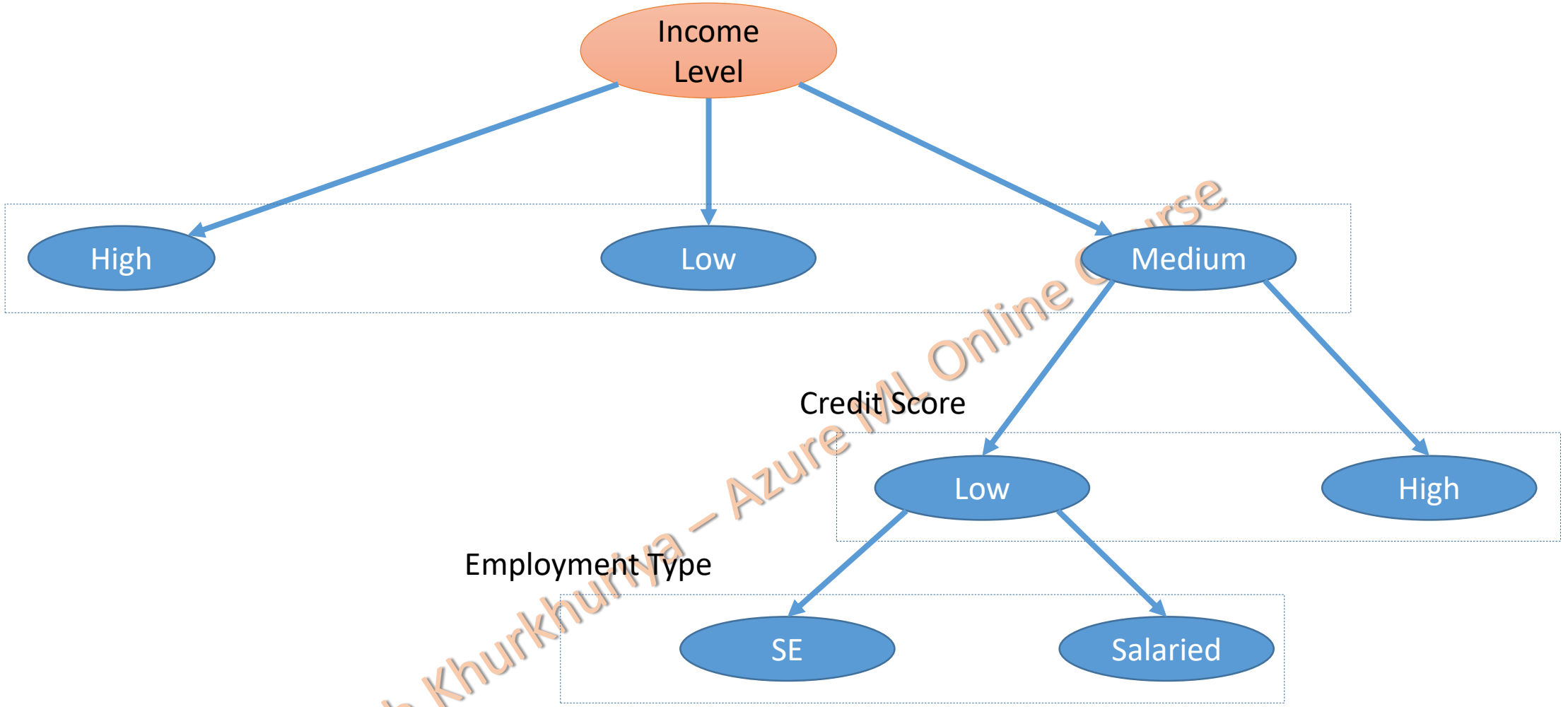


LID	IL	CS	ET	Status
L1	Medium	Low	SE	No
L4	Medium	Low	Salaried	Yes
L8	Medium	Low	SE	No

Split Further

LID	IL	CS	ET	Status
L10	Medium	High	SE	Yes
L12	Medium	High	Salaried	Yes
L13	Medium	High	SE	Yes

Pure Subset



LID	IL	CS	ET	Status
L1	Medium	Low	SE	No
L8	Medium	Low	SE	No

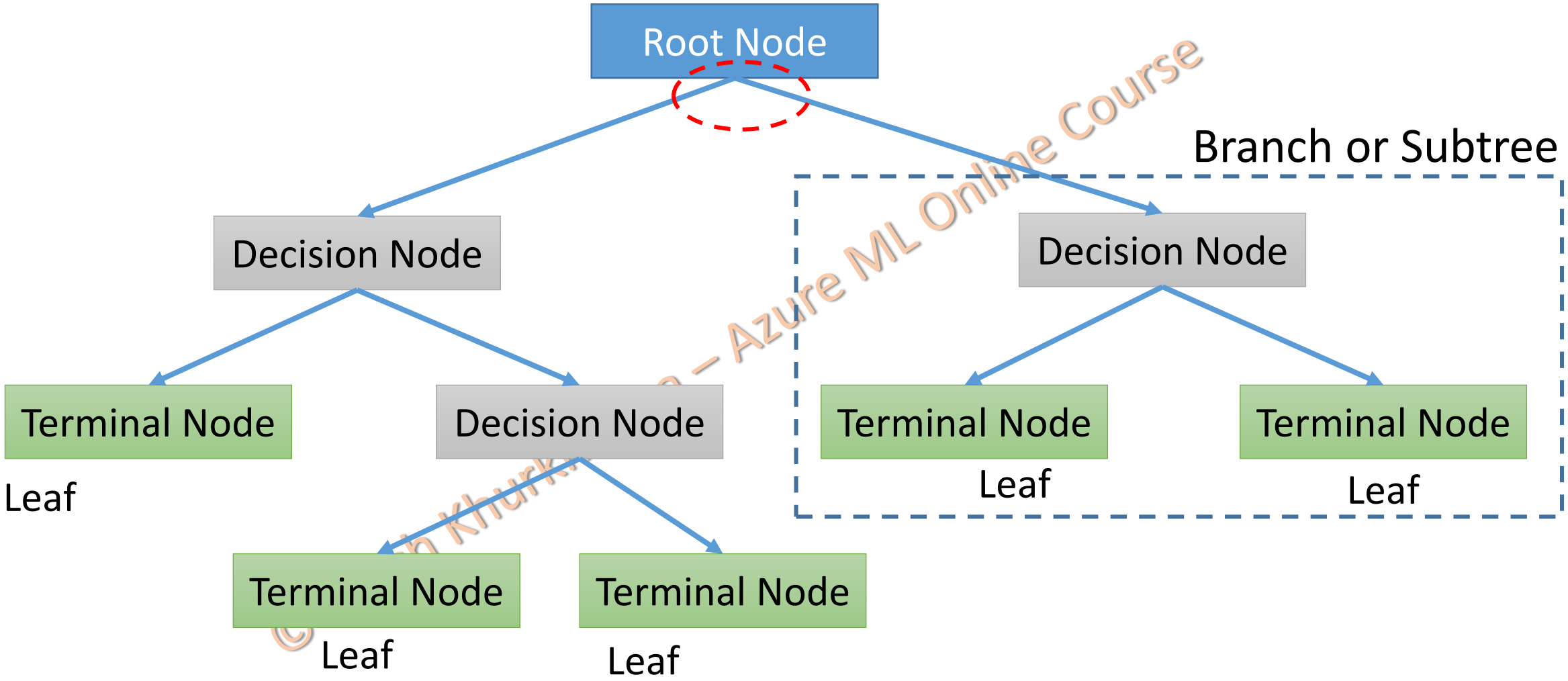
Pure Subset

LID	IL	CS	ET	Status
L4	Medium	Low	Salaried	Yes

Pure Subset

© Jitesh Khurkhuriya – Azure ML Online Course

Decision Tree Terms



Definitions

- **Root Node:** It represents entire population or sample and this further gets divided into two or more homogeneous sets.
- **Splitting:** It is a process of dividing a node into two or more sub-nodes.
- **Decision Node:** When a sub-node splits into further sub-nodes, then it is called decision node.
- **Leaf/ Terminal Node:** Nodes do not split is called Leaf or Terminal node.
- **Branch / Sub-Tree:** A sub section of entire tree is called branch or sub-tree.

© Jitesh Khurkhuriya – Azure ML Online Course

Ensemble Learning

Everyday Ensemble Learning

Decision?

Is this price fair?

Appreciation of price?

Construction Quality?

Neighbourhood?

Location appropriate?

Decision?

Broker or real estate portal to check fair price, price appreciation

Friend or colleague who stays nearby or stayed in the neighbourhood

Inspection by an architect for quality checks and structural defects.

Decision?

Is this price fair?

Appreciation of price?

Construction Quality?

Majority

Weighted Average

Neighbourhood?

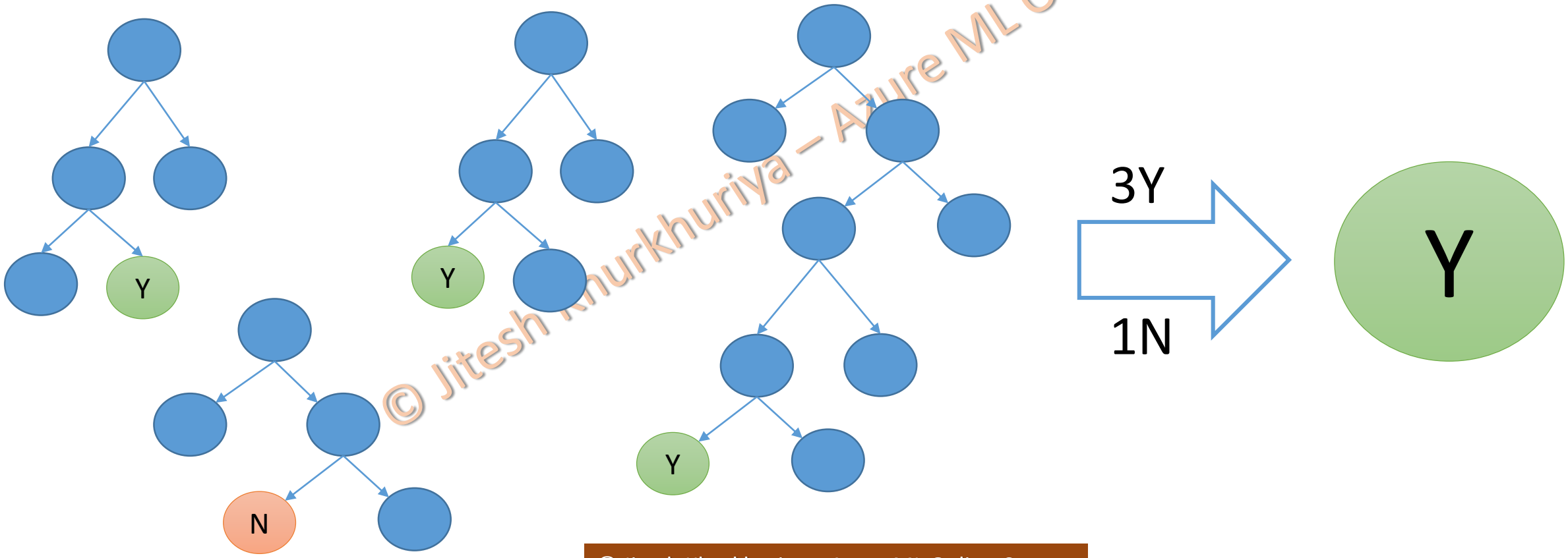
Location appropriate?

Ensemble Learning

- All algorithms have errors
- Collective wisdom is higher than the individual intelligence
- Generate a group of base learners and combined result gives higher accuracy
- Different base learners can use different,
  - Parameters
  - Sequence
  - Training sets etc
- Two major Ensemble Learning Methods
  - Bagging
  - Boosting

Bagging

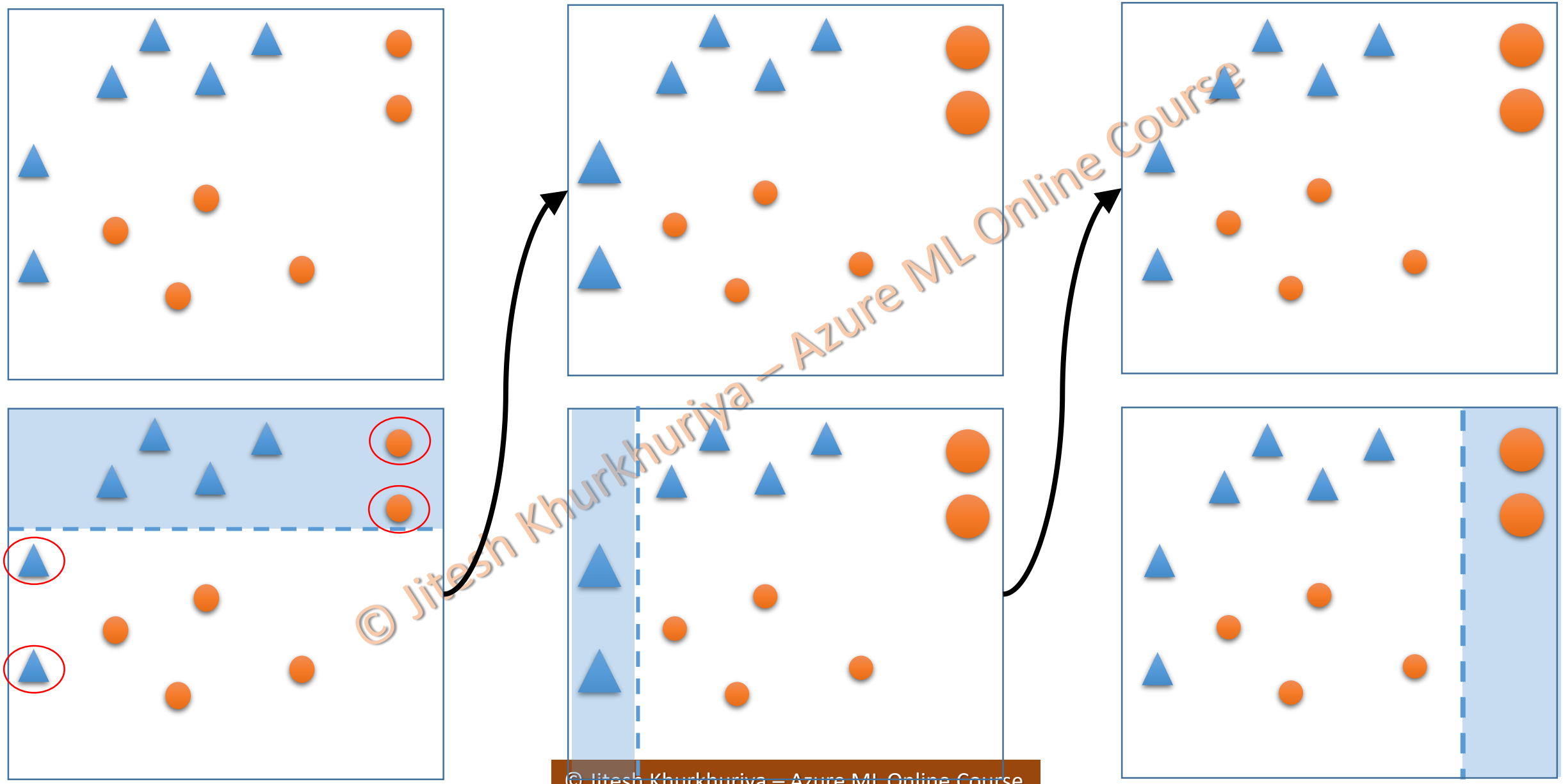
- Various models are built in parallel
- All models vote to give the final prediction



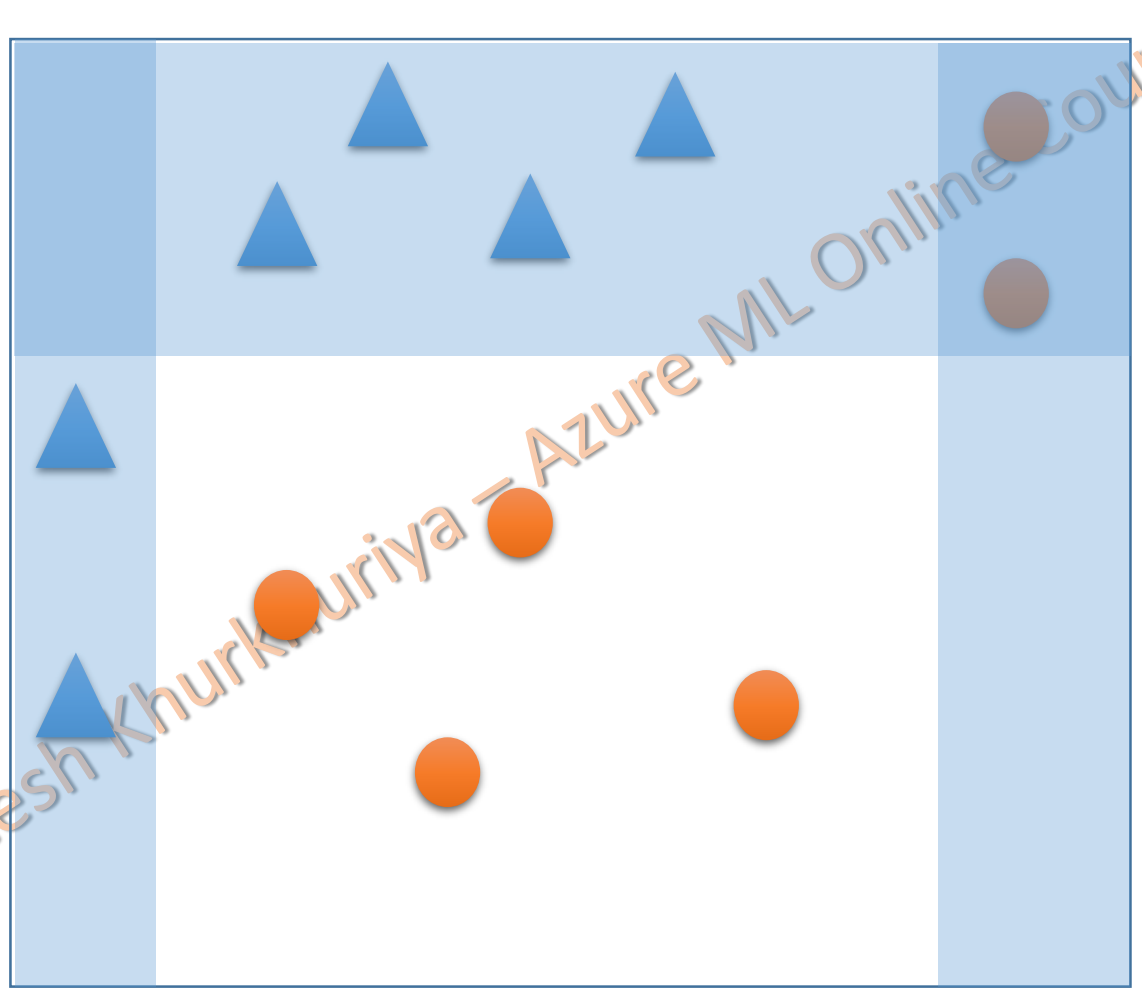
Boosting

- Train the Decision Tree in a sequence
- Learn from the previous tree by focussing on incorrect observations
- Build new model with higher weight for incorrect observations from previous sequence

© Jitesh Khurkhuriya – Azure ML Online Course



Boosted Model



Two Class Boosted Decision Tree

Bank Telemarketing

- Goal is to predict if the client will subscribe to a product or not

- Number of instances – 45, 211

1. Age
2. Job Type
3. Marital Status
4. Education Level
5. Credit Default?
6. Housing Loan?
7. Personal Loan
8. Contacted Type
9. Contacted Month
10. Last Contacted day
11. Contact Duration
12. Campaign Type
13. P-Days
14. Previous
15. P-Outcome
16. Emp-Var-Rate
17. Consumer Price Index
18. Consumer Confidence Index
19. Euribor 3 Month Rate
20. Number of employees
21. Subscribed?

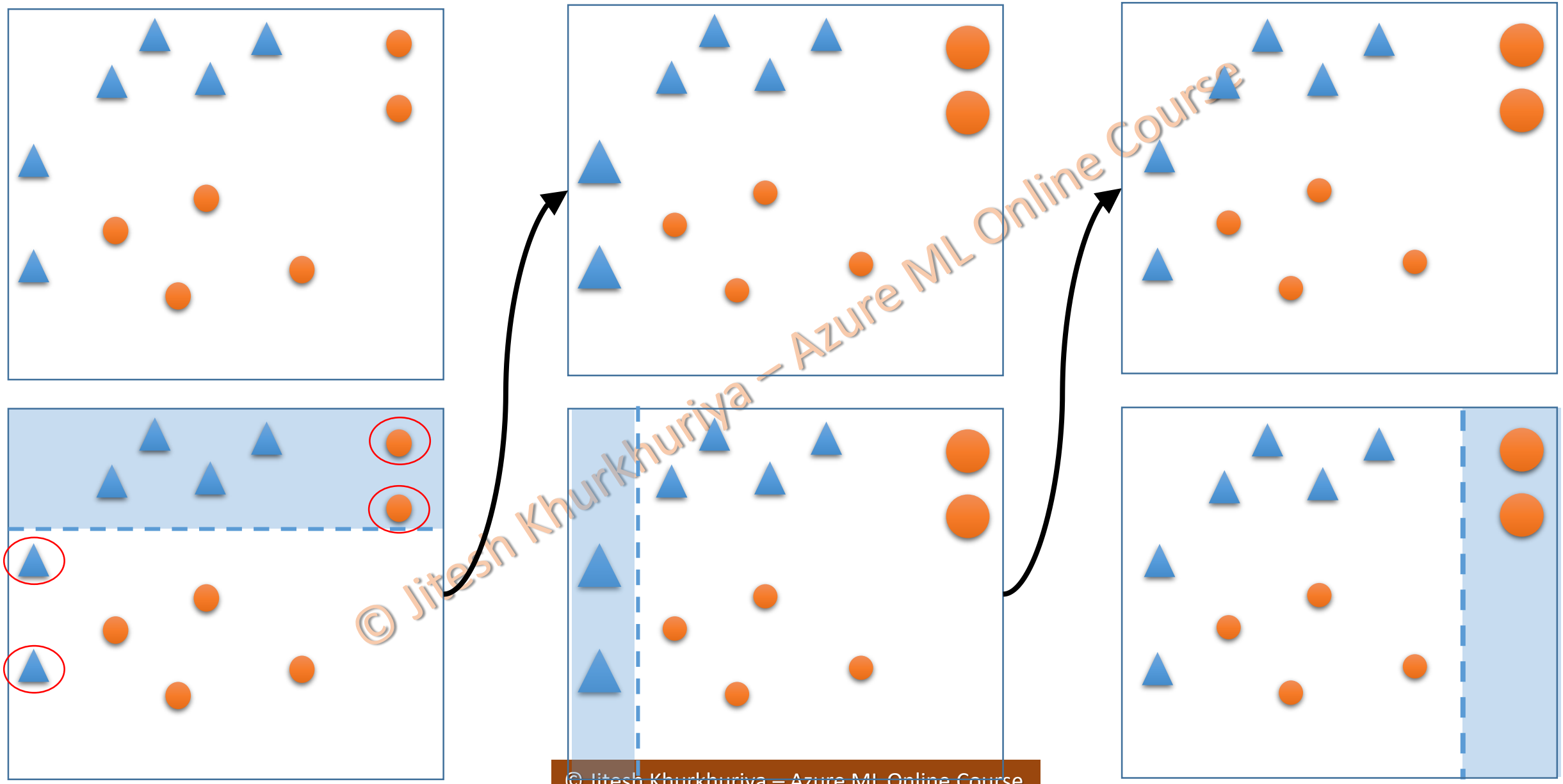
https://archive.ics.uci.edu/ml/datasets/bank+marketing

Source: [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

Boosting

- Train the Decision Tree in a sequence
- Learn from the previous tree by focussing on incorrect observations
- Build new model with higher weight for incorrect observations from previous sequence

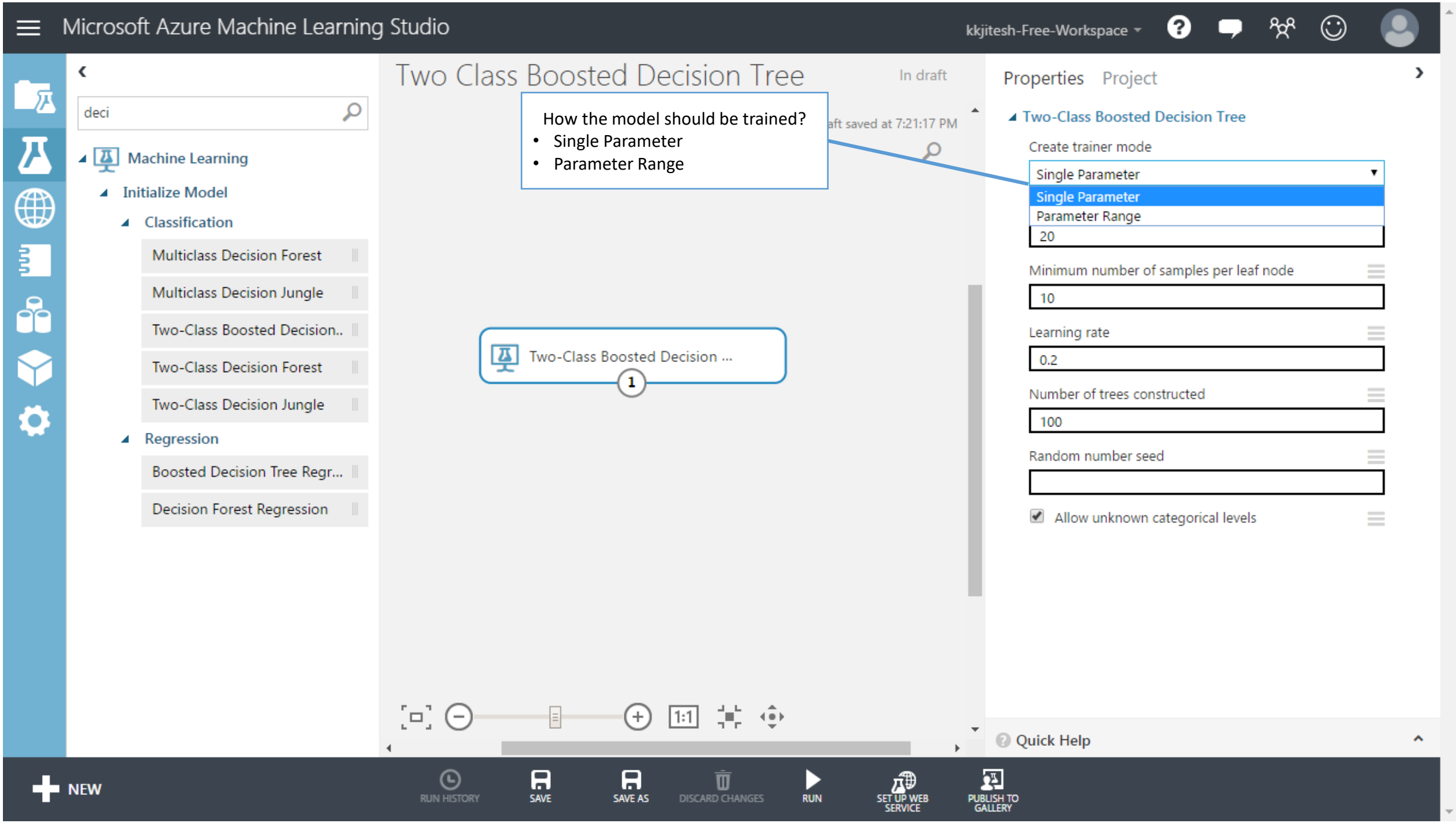
© Jitesh Khurkhuriya – Azure ML Online Course

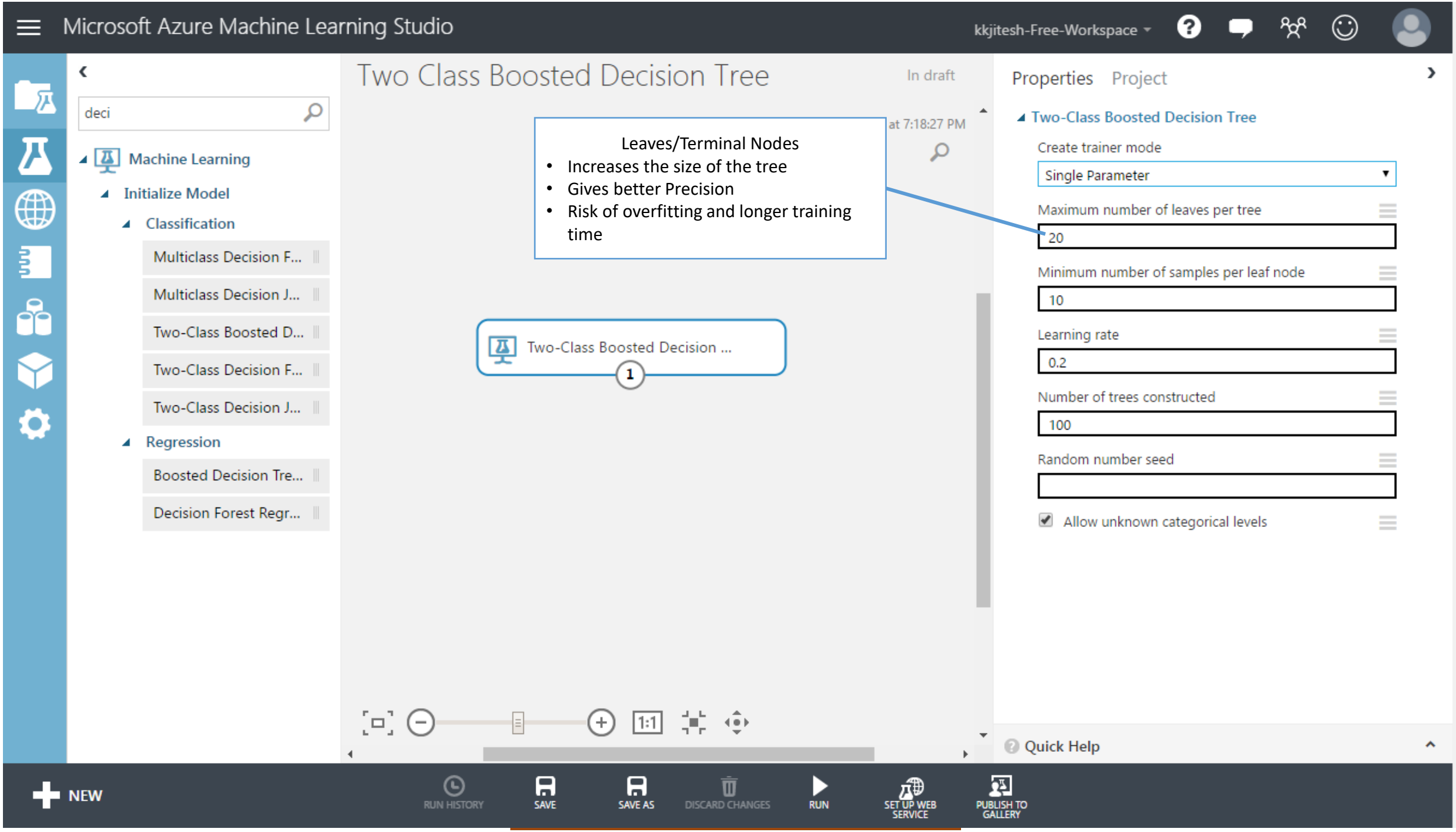


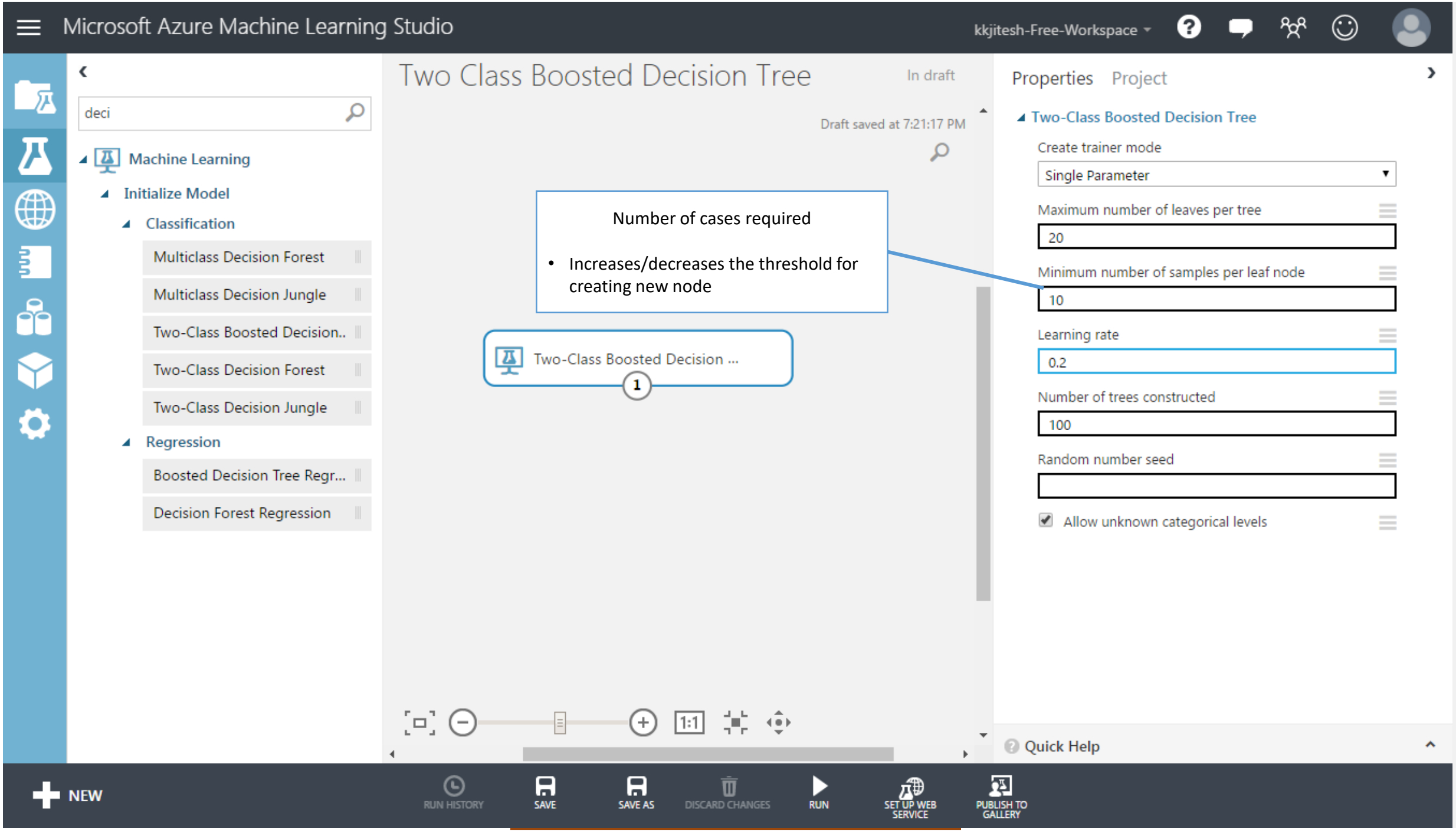
Two-Class Boosted Decision Tree?

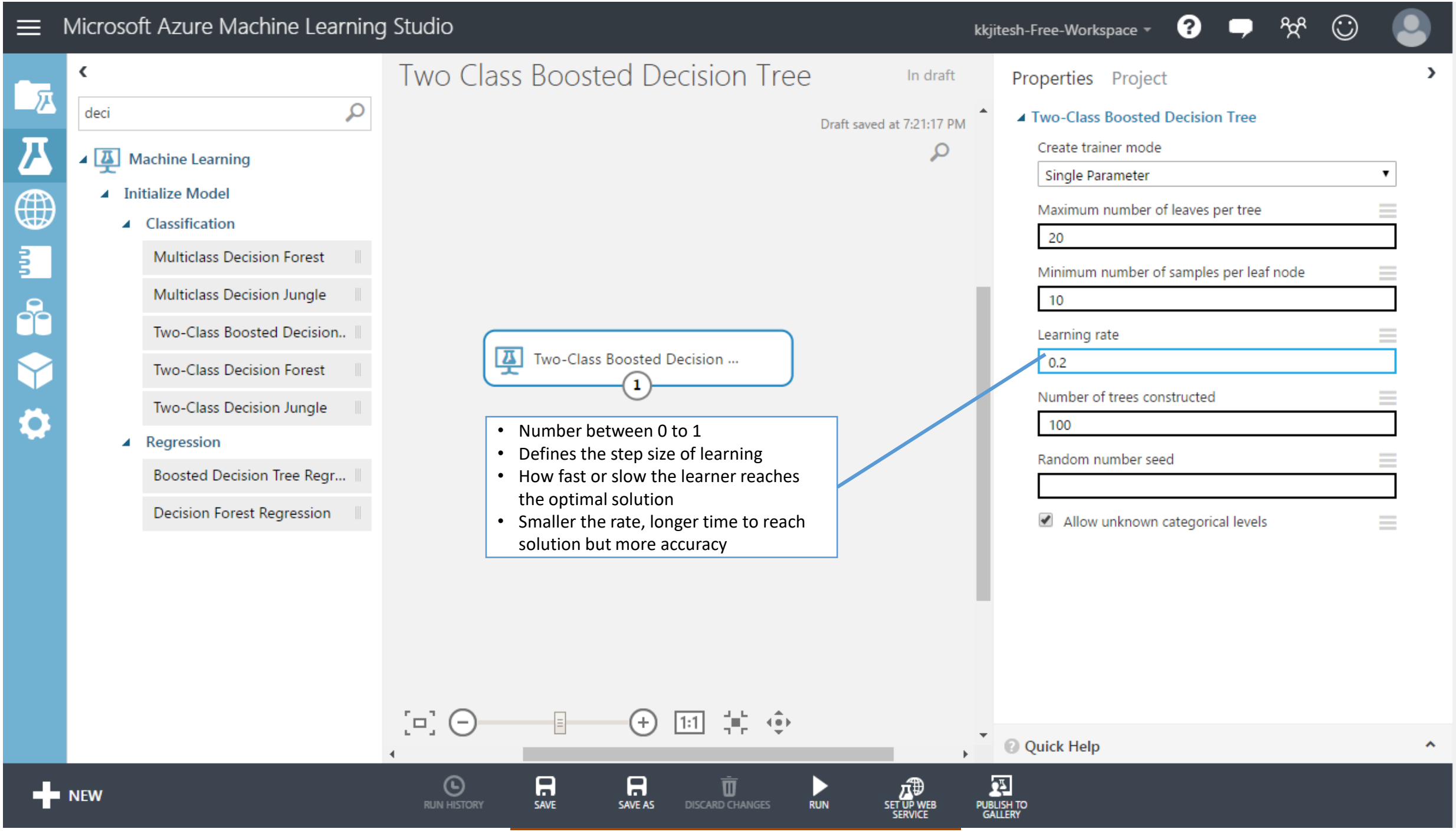
- Machine learning model based on the boosted decision trees algorithm
- Based on ensemble learning method
- Among the easiest methods to get top performance
- One of the more memory-intensive learners.

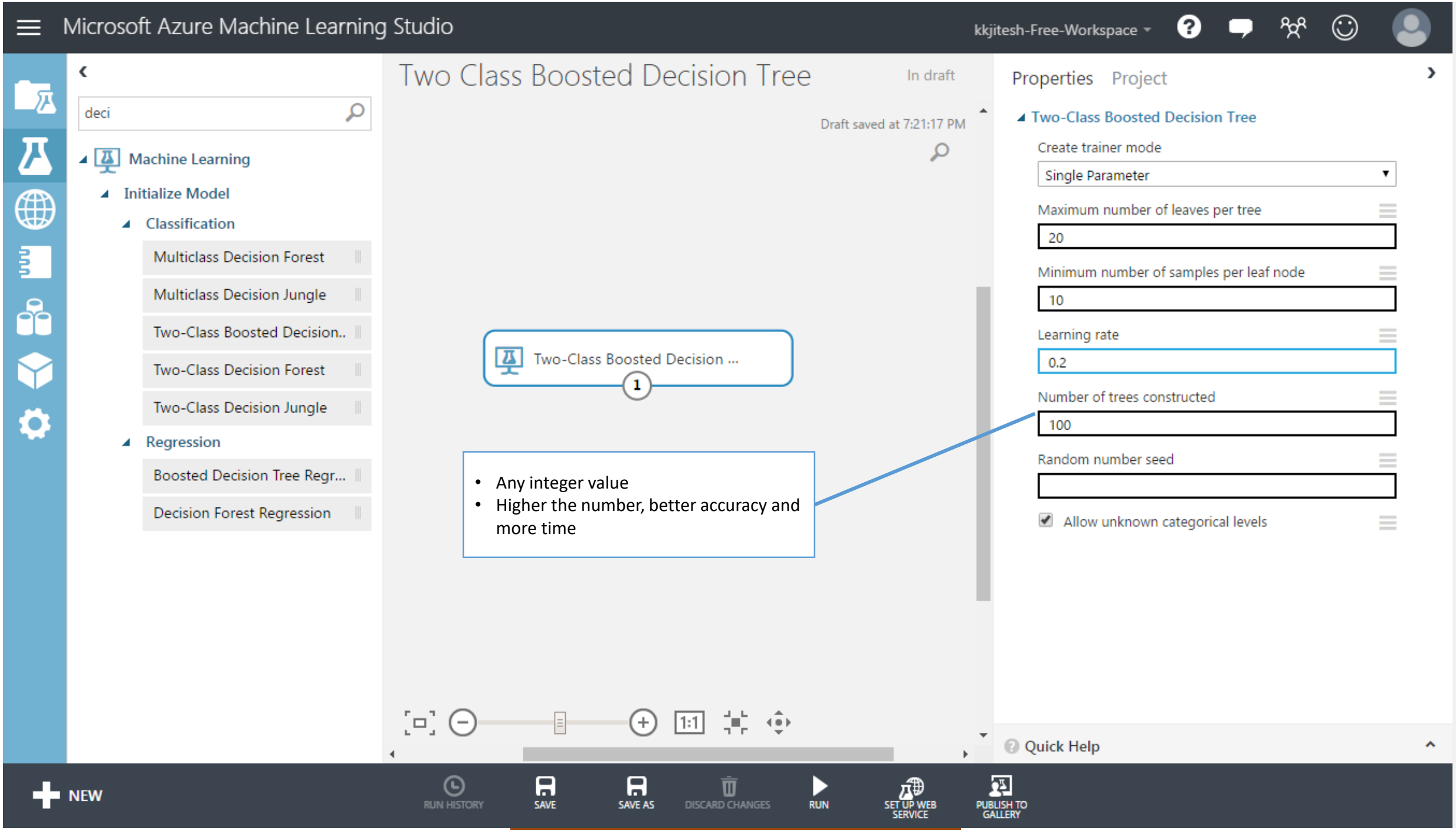
© Jitesh Khurkhuriya – Azure ML Online Course

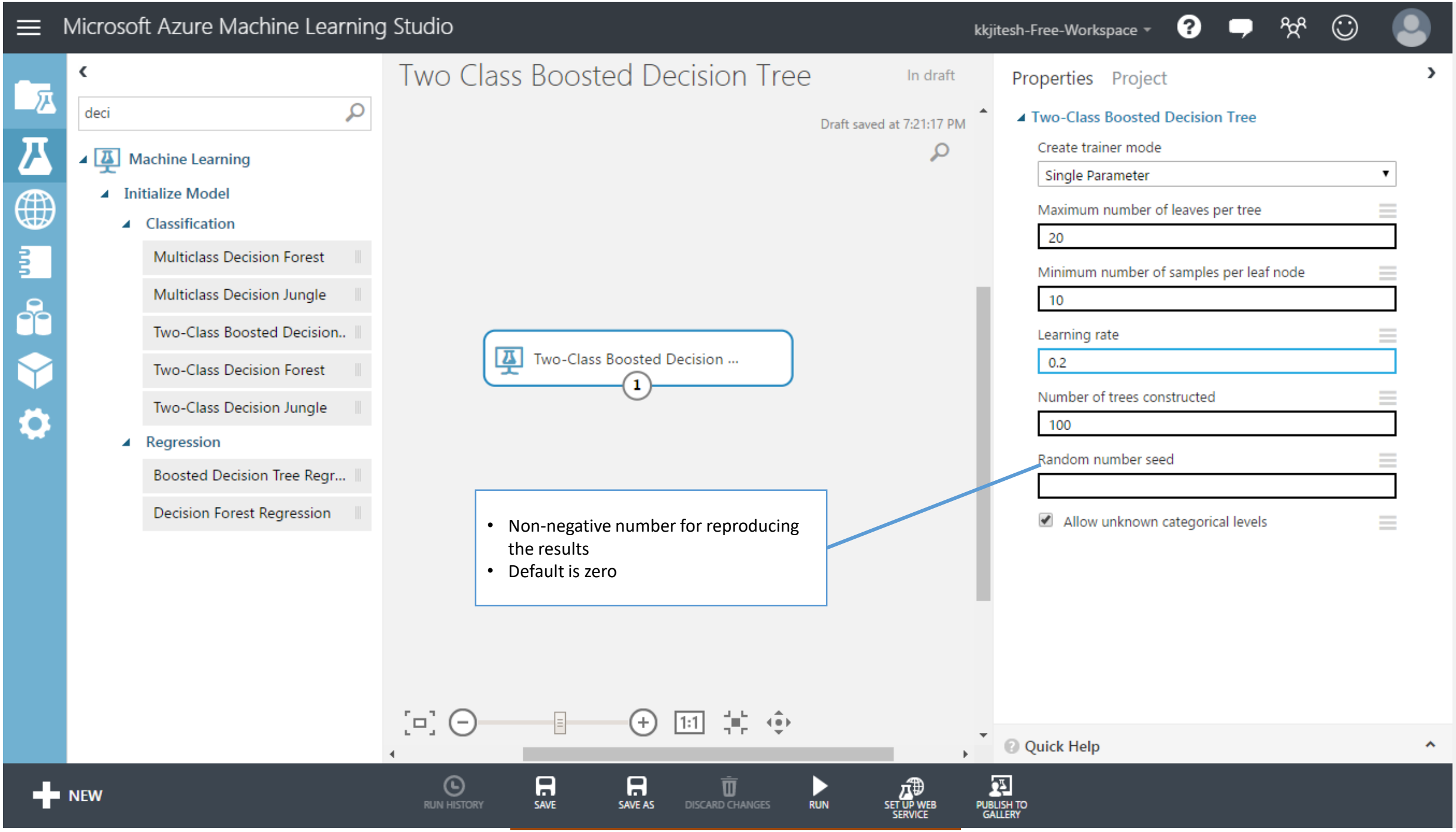

Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Boosted Decision Tree
In draft
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
How the model should be trained?
• Single Parameter
• Parameter Range
aft saved at 7:21:17 PM
Two-Class Boosted Decision ...
1
Properties
Project
Two-Class Boosted Decision Tree
Create trainer mode
Single Parameter
Single Parameter
Parameter Range
20
Minimum number of samples per leaf node
10
Learning rate
0.2
Number of trees constructed
100
Random number seed
Allow unknown categorical levels
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Boosted Decision Tree
In draft
at 7:18:27 PM
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision F...
Multiclass Decision J...
Two-Class Boosted D...
Two-Class Decision F...
Two-Class Decision J...
Regression
Boosted Decision Tre...
Decision Forest Regr...
Leaves/Terminal Nodes
• Increases the size of the tree
• Gives better Precision
• Risk of overfitting and longer training time
Two-Class Boosted Decision ...
1
Properties
Project
Two-Class Boosted Decision Tree
Create trainer mode
Single Parameter
Maximum number of leaves per tree
20
Minimum number of samples per leaf node
10
Learning rate
0.2
Number of trees constructed
100
Random number seed
Allow unknown categorical levels
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Boosted Decision Tree
In draft
Draft saved at 7:21:17 PM
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Number of cases required
• Increases/decreases the threshold for creating new node
Two-Class Boosted Decision ...
1
Properties
Project
Two-Class Boosted Decision Tree
Create trainer mode
Single Parameter
Maximum number of leaves per tree
20
Minimum number of samples per leaf node
10
Learning rate
0.2
Number of trees constructed
100
Random number seed
Allow unknown categorical levels
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Boosted Decision Tree
In draft
Draft saved at 7:21:17 PM
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Two-Class Boosted Decision ...
1
Number between 0 to 1
Defines the step size of learning
How fast or slow the learner reaches the optimal solution
Smaller the rate, longer time to reach solution but more accuracy
Properties
Project
Two-Class Boosted Decision Tree
Create trainer mode
Single Parameter
Maximum number of leaves per tree
20
Minimum number of samples per leaf node
10
Learning rate
0.2
Number of trees constructed
100
Random number seed
Allow unknown categorical levels
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

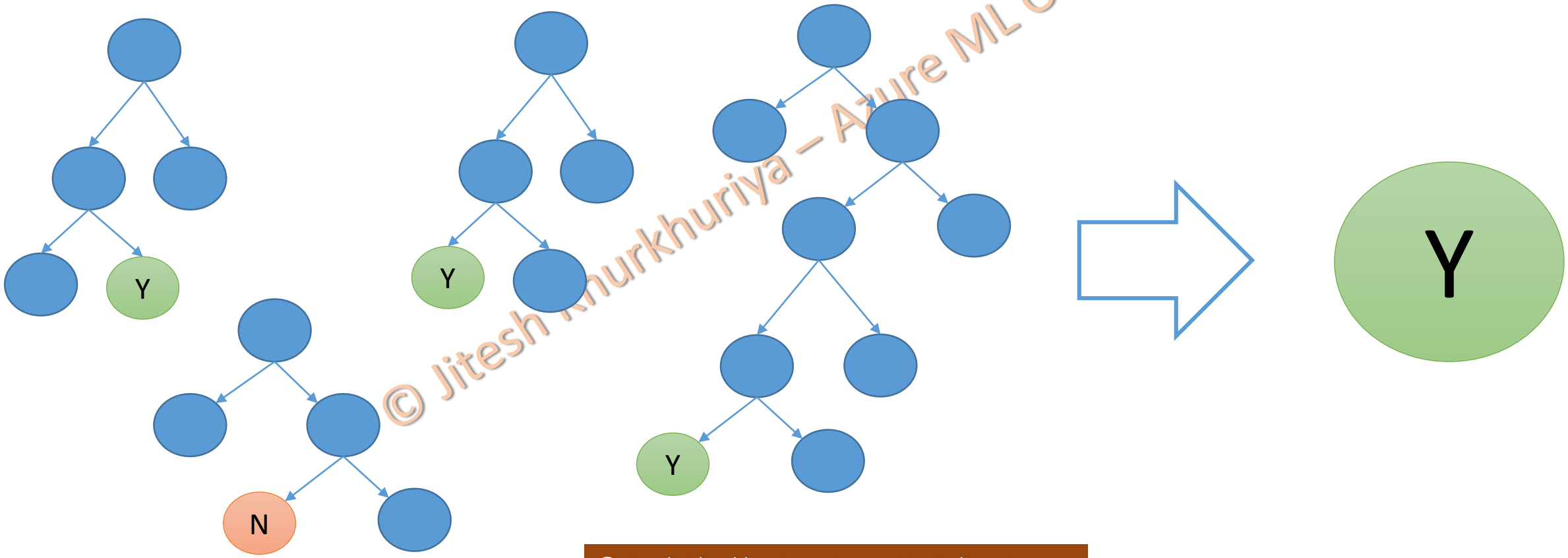

Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Boosted Decision Tree
In draft
Draft saved at 7:21:17 PM
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Two-Class Boosted Decision ...
1
Any integer value
Higher the number, better accuracy and more time
Properties
Project
Two-Class Boosted Decision Tree
Create trainer mode
Single Parameter
Maximum number of leaves per tree
20
Minimum number of samples per leaf node
10
Learning rate
0.2
Number of trees constructed
100
Random number seed
Allow unknown categorical levels
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Boosted Decision Tree
In draft
Draft saved at 7:21:17 PM
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Two-Class Boosted Decision ...
1
Properties
Project
Two-Class Boosted Decision Tree
Create trainer mode
Single Parameter
Maximum number of leaves per tree
20
Minimum number of samples per leaf node
10
Learning rate
0.2
Number of trees constructed
100
Random number seed
Allow unknown categorical levels
Non-negative number for reproducing the results
Default is zero
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Decision Forest

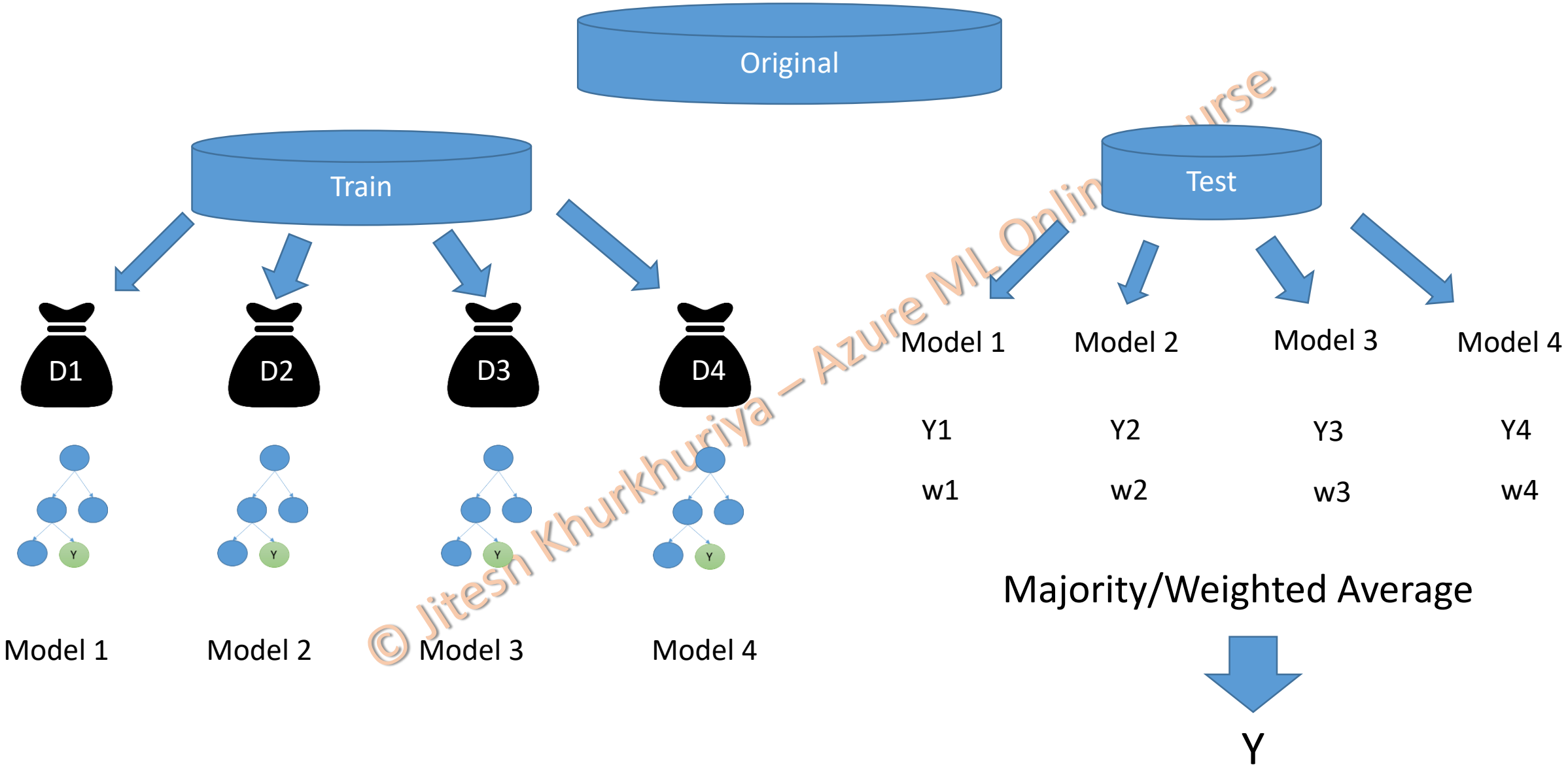
Bagging

- Various models are built in parallel
- All models vote to give the final prediction



© Jitesh Khurkhuriya – Azure ML Online Course

Bagging

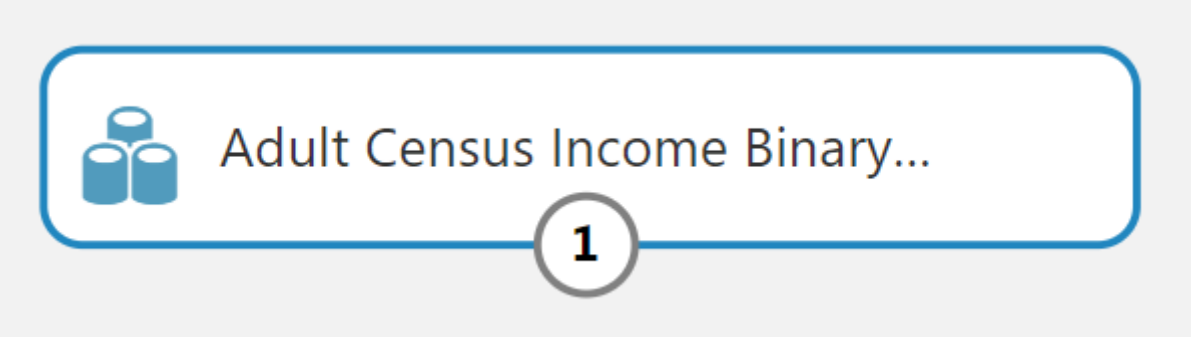


© Jitesh Khurkhuriya – Azure ML Online Course

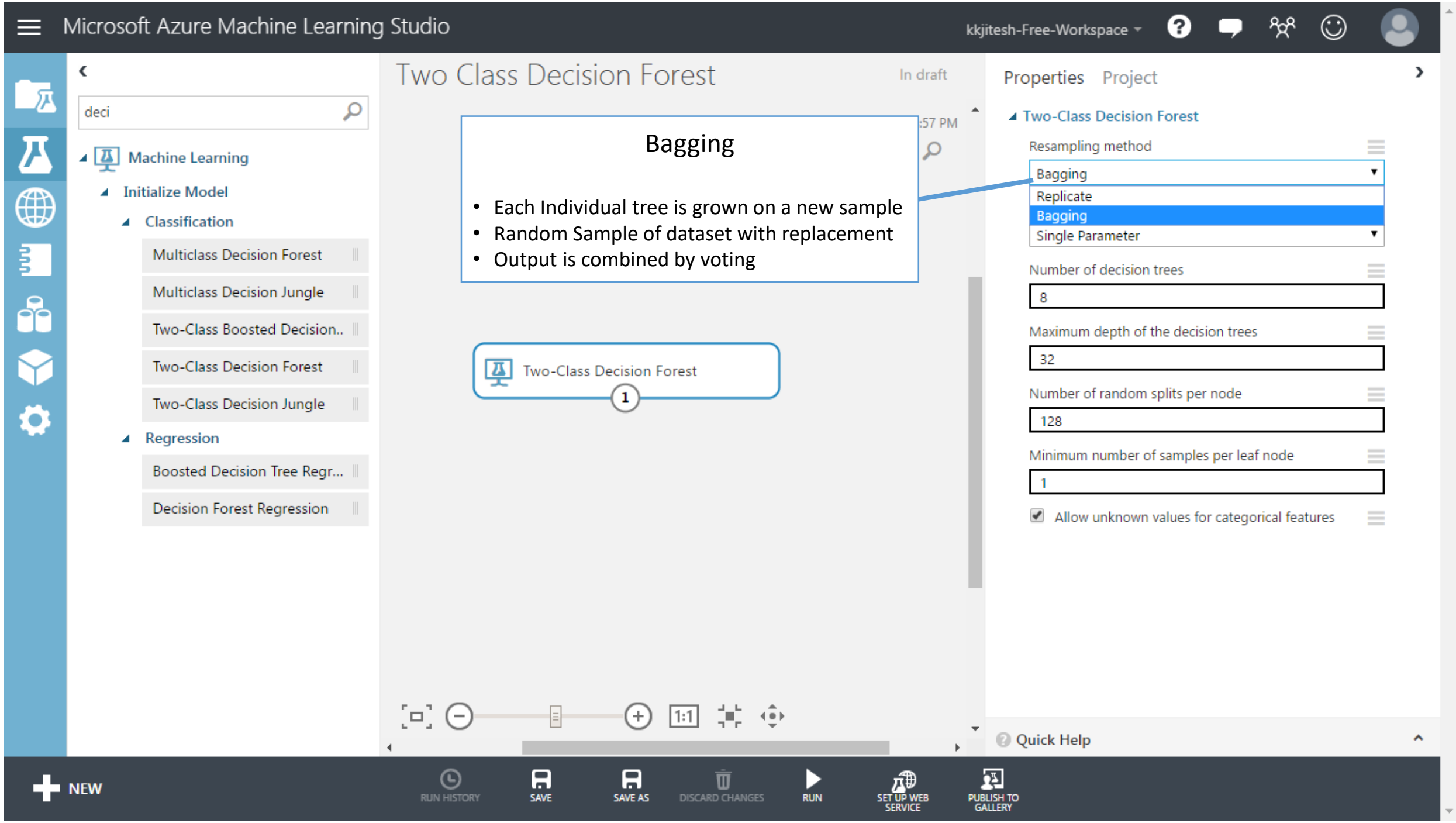
Two Class Decision Forest

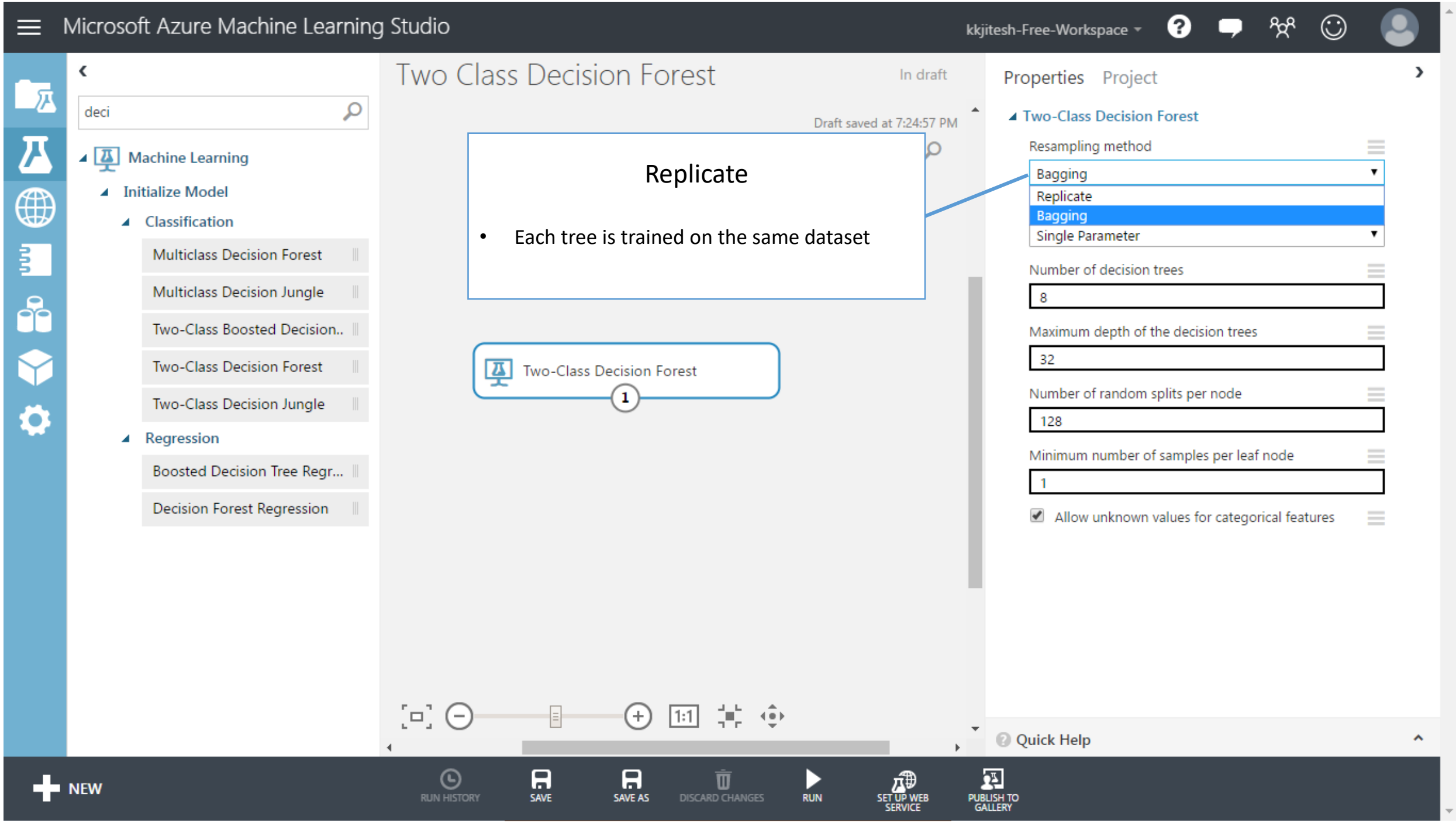
Adult Census Data

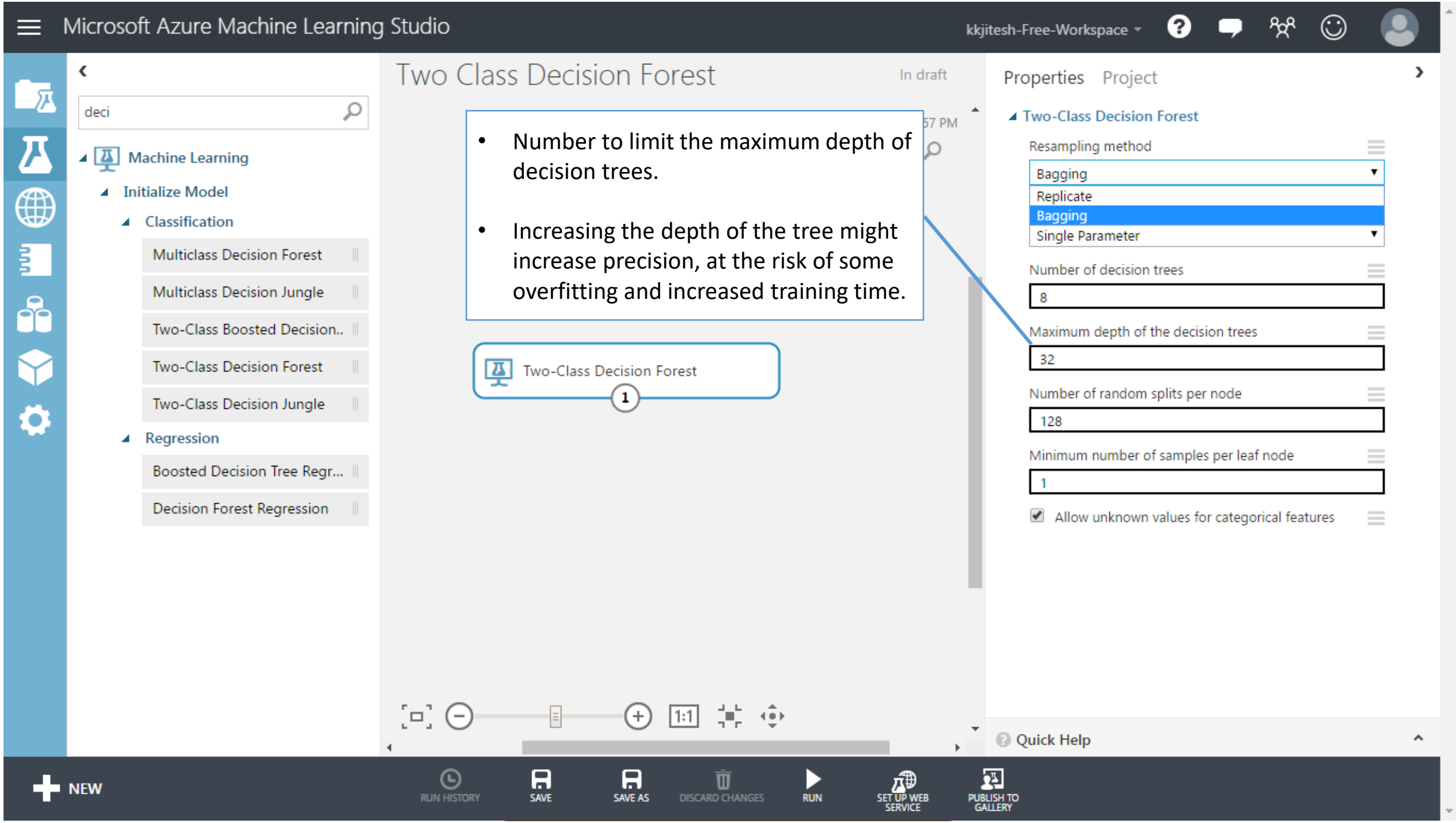
- Problem statement: Predict whether income exceeds $50K/yr based on census data.

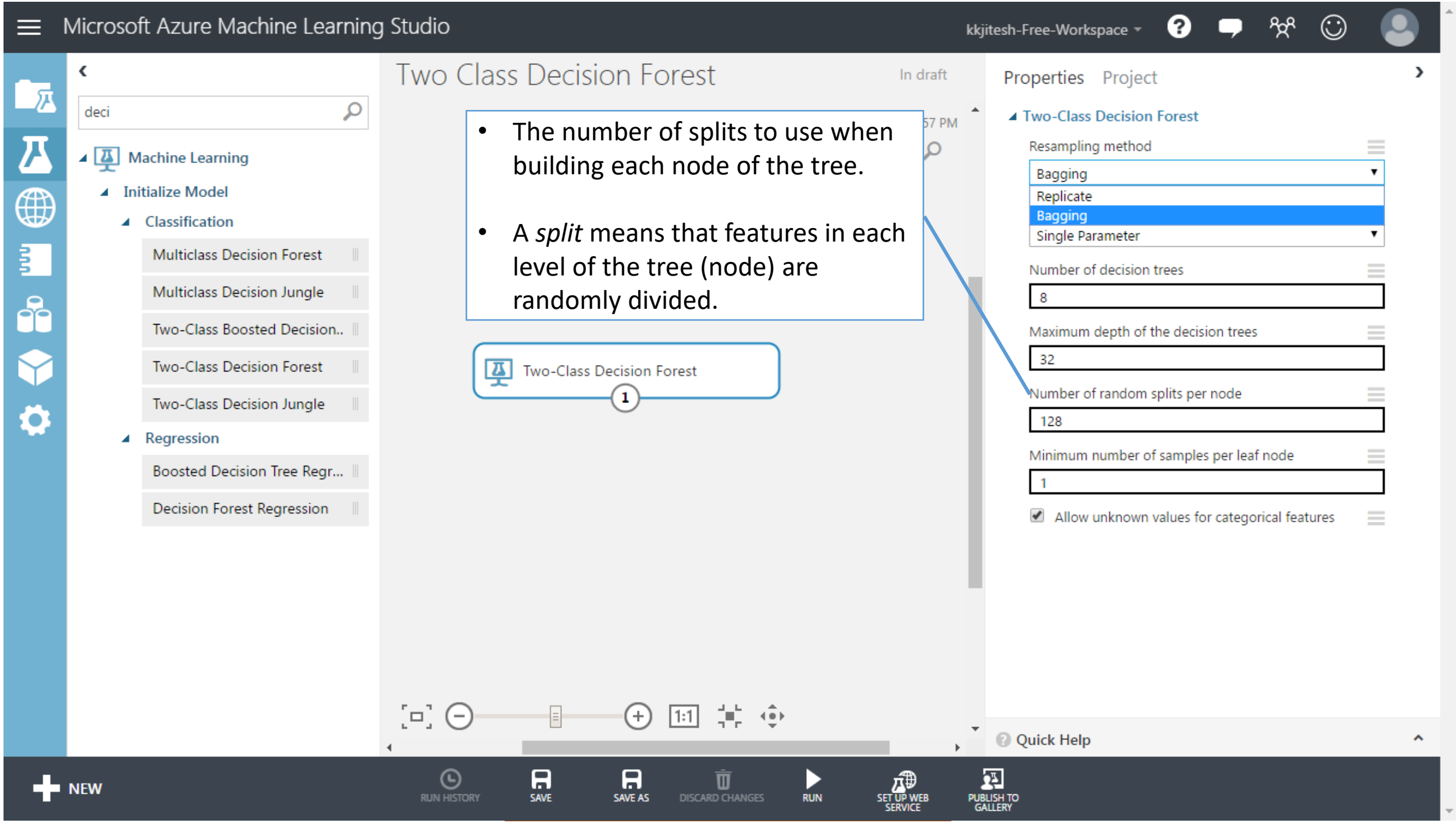


1. Age
2. Workclass
3. Fnlwgt
4. Education
5. Education-Num
6. Marital Status
7. Occupation
8. Relationship
9. Race
10. Sex
11. Capital Gains
12. Capital Losses
13. Hours per week
14. Native Country
15. Income


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Bagging
Each Individual tree is grown on a new sample
Random Sample of dataset with replacement
Output is combined by voting
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
Draft saved at 7:24:57 PM
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Replicate
• Each tree is trained on the same dataset
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

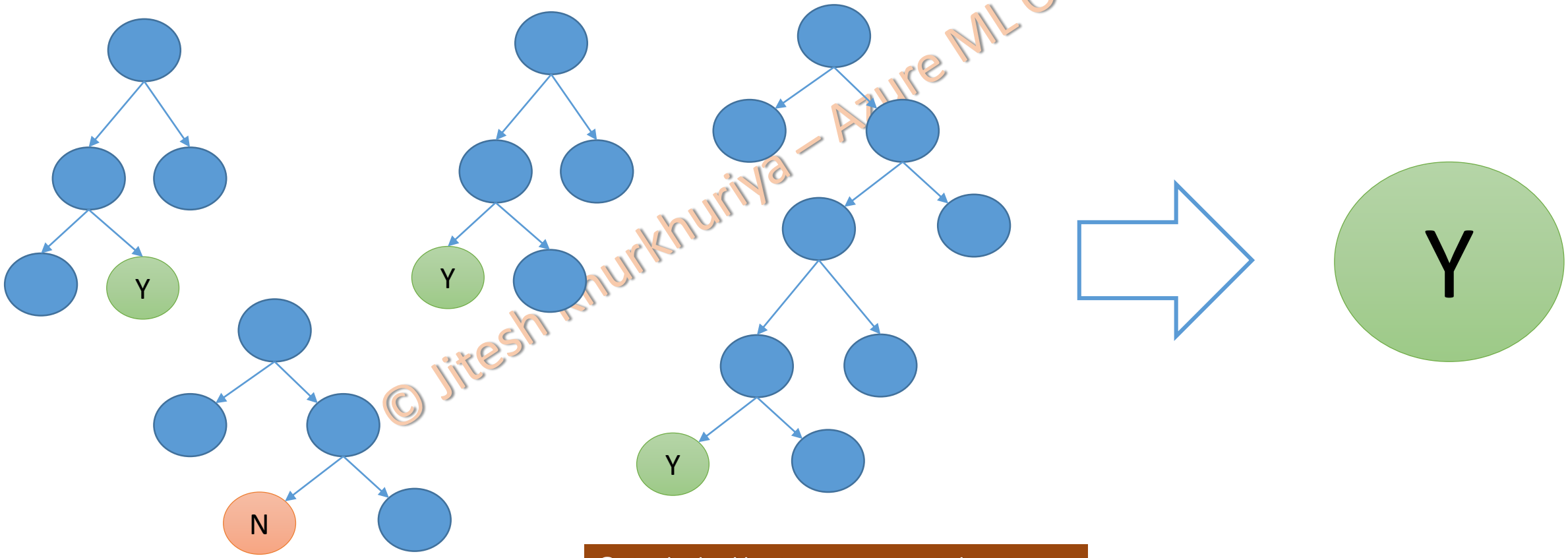

Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
• Number to limit the maximum depth of decision trees.
• Increasing the depth of the tree might increase precision, at the risk of some overfitting and increased training time.
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
• The number of splits to use when building each node of the tree.
• A split means that features in each level of the tree (node) are randomly divided.
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Decision Forest

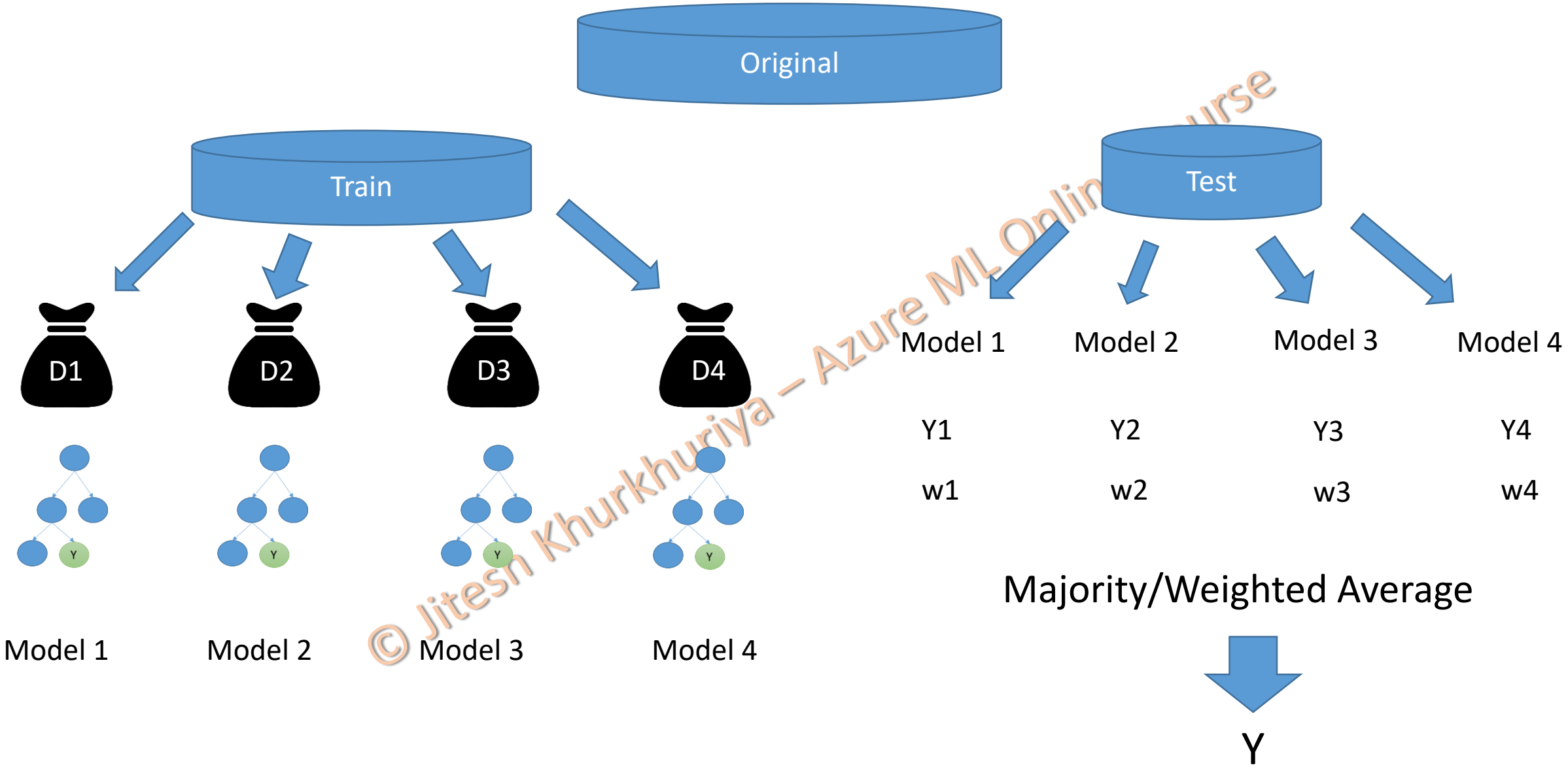
Bagging

- Various models are built in parallel
- All models vote to give the final prediction



© Jitesh Khurkhuriya – Azure ML Online Course

Bagging

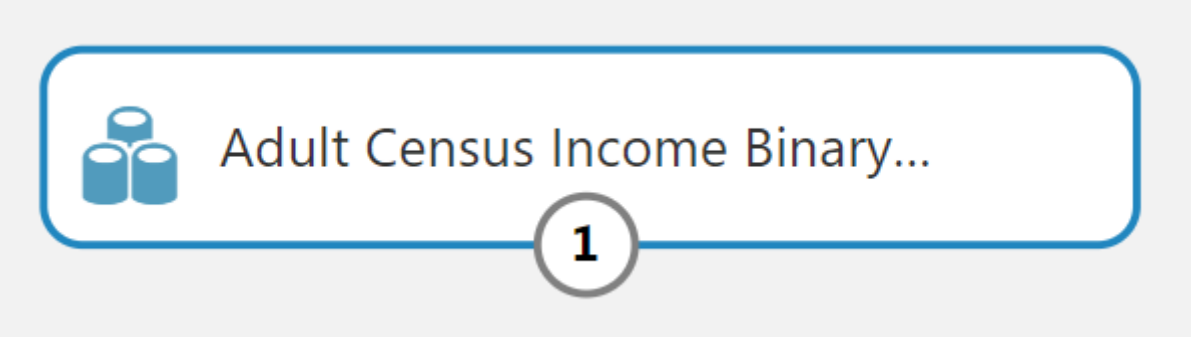


© Jitesh Khurkhuriya – Azure ML Online Course

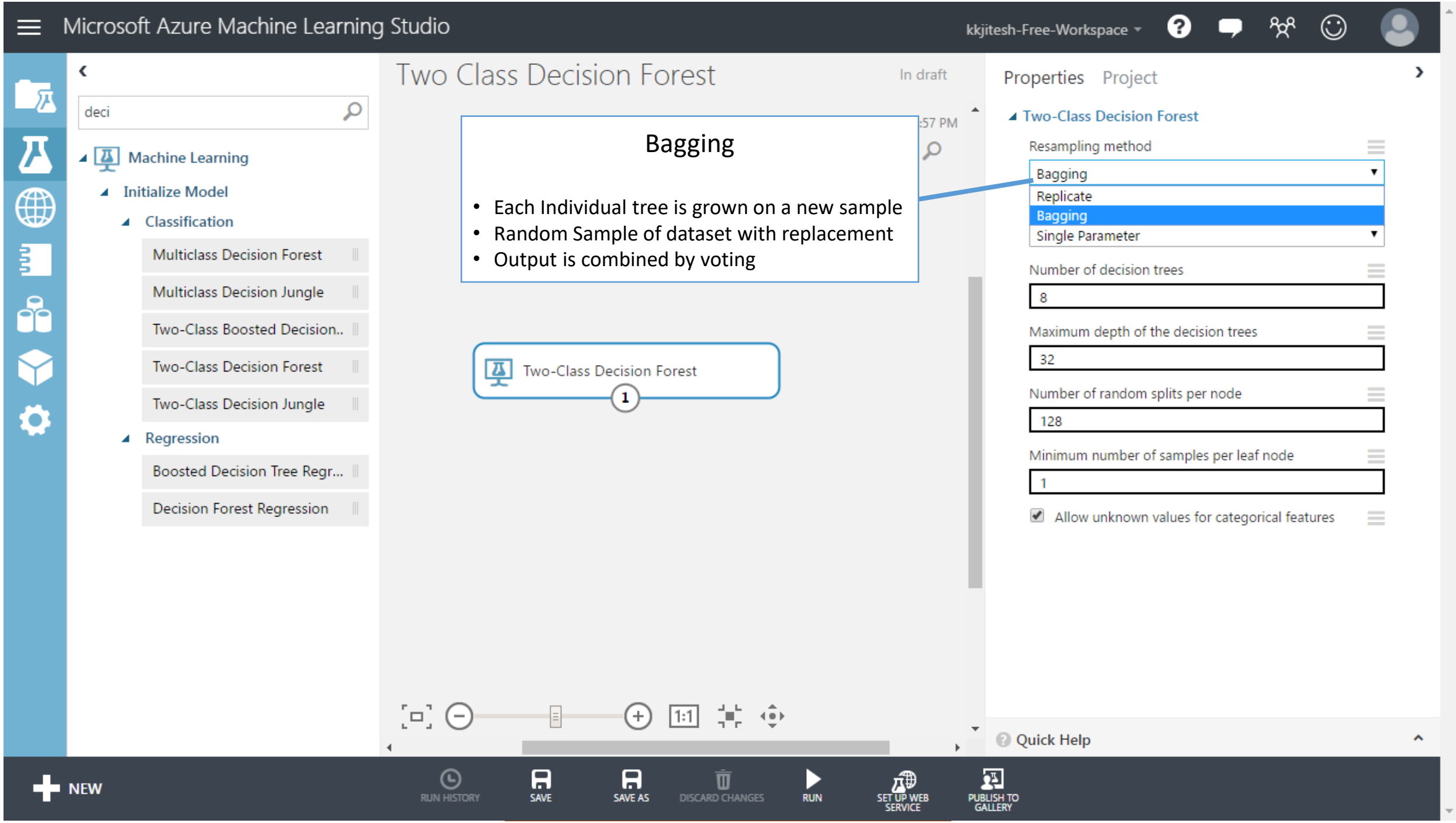
Two Class Decision Forest

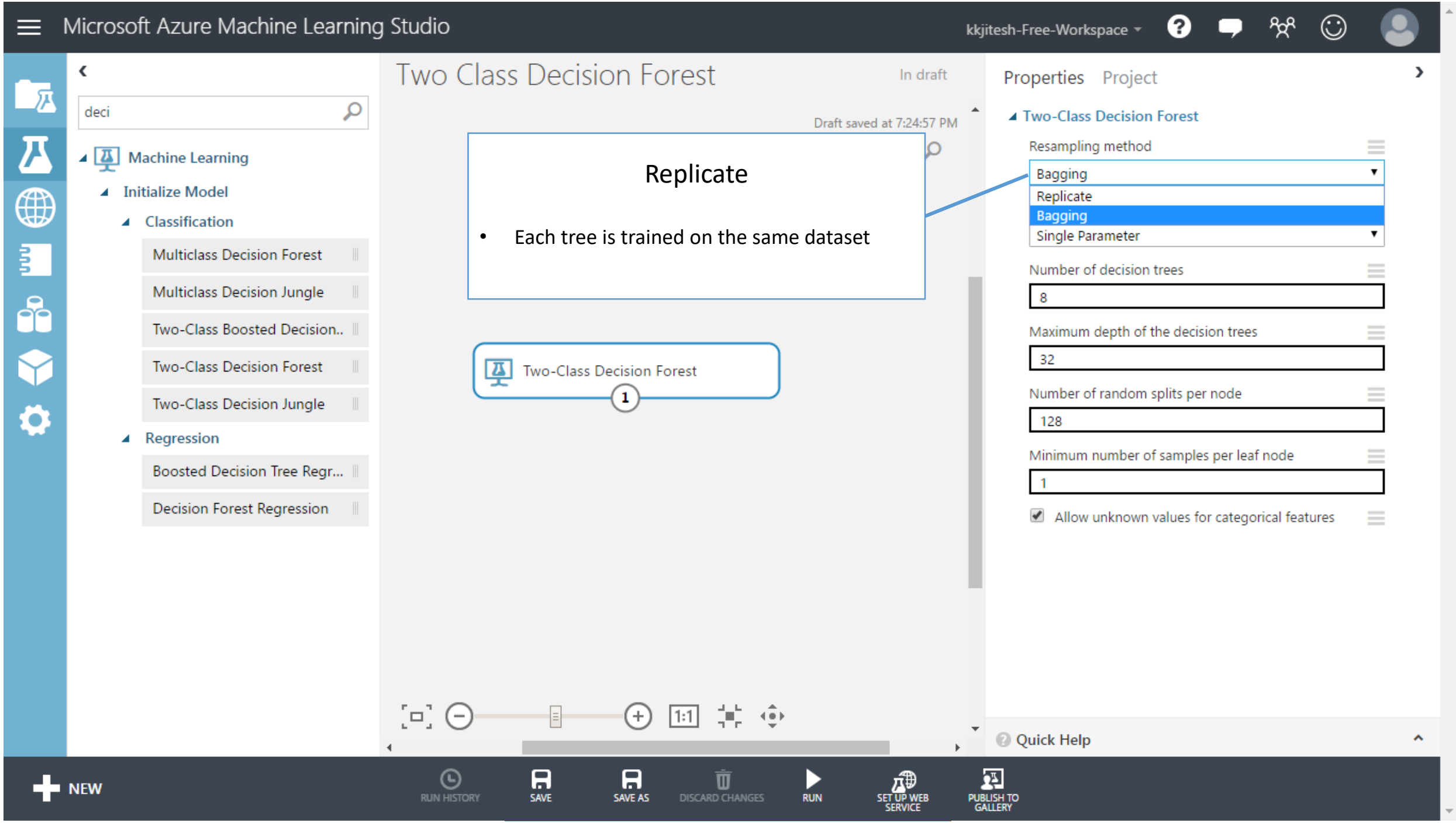
Adult Census Data

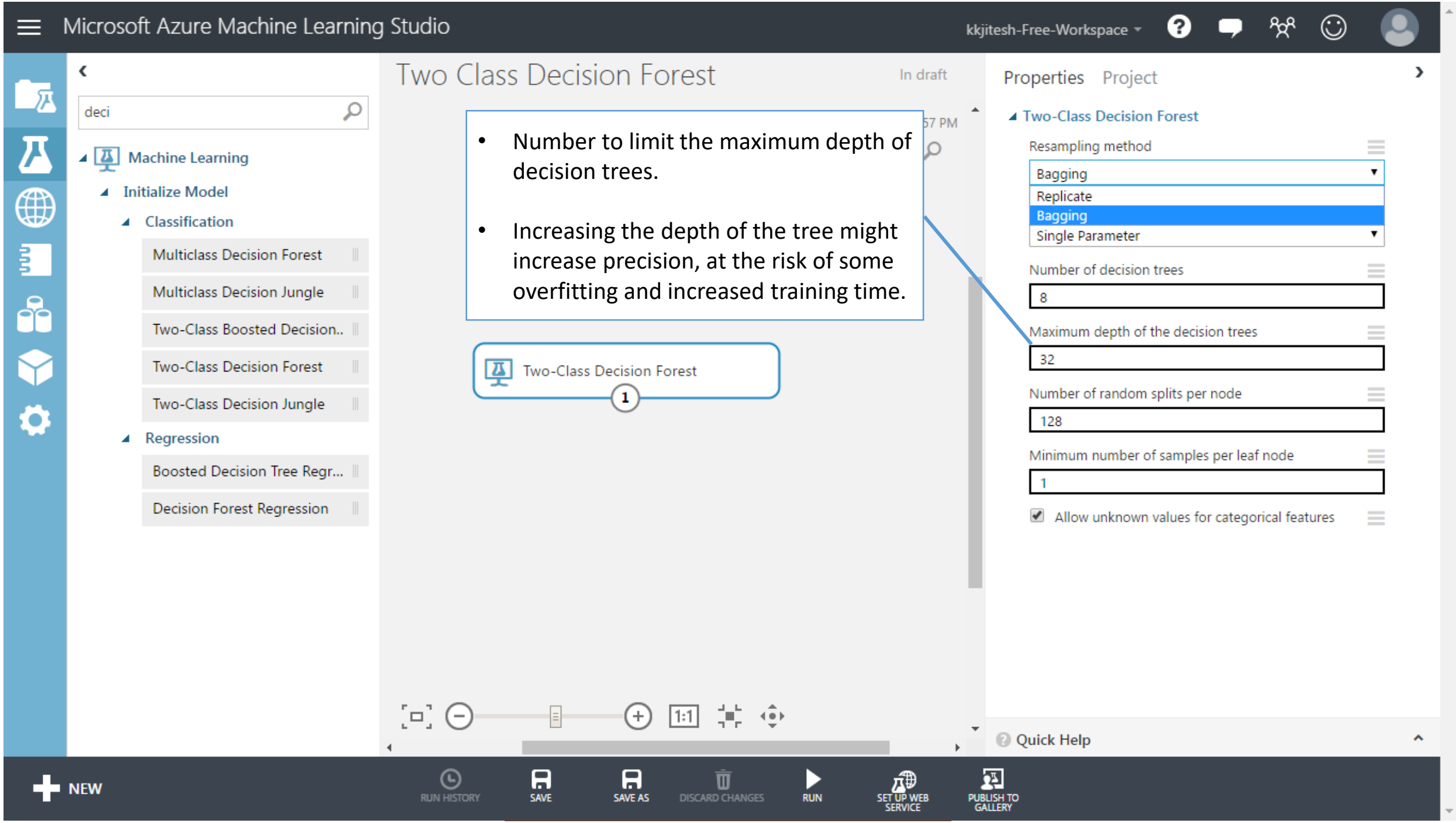
- Problem statement: Predict whether income exceeds $50K/yr based on census data.

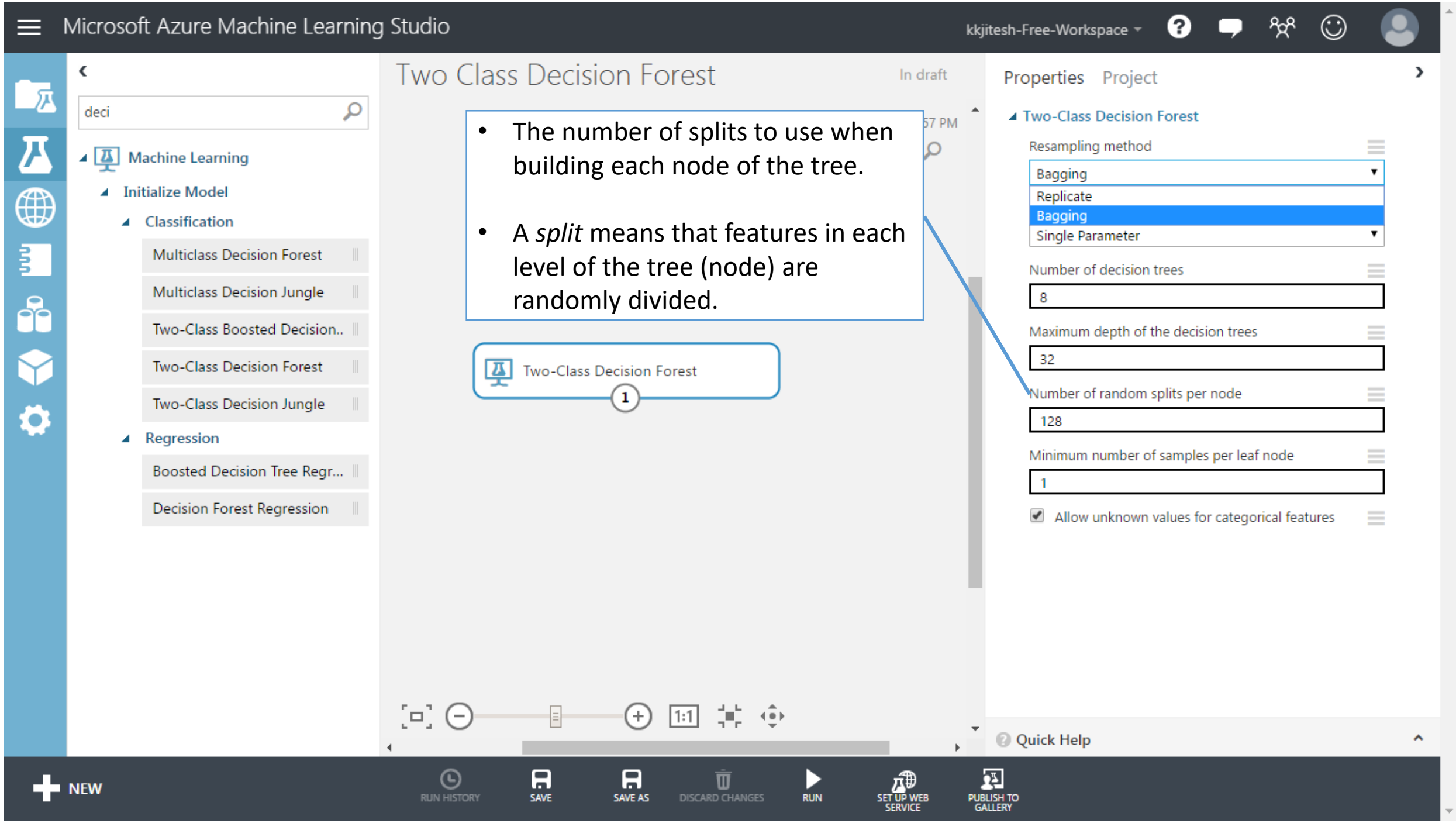


1. Age
2. Workclass
3. Fnlwgt
4. Education
5. Education-Num
6. Marital Status
7. Occupation
8. Relationship
9. Race
10. Sex
11. Capital Gains
12. Capital Losses
13. Hours per week
14. Native Country
15. Income


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Bagging
Each Individual tree is grown on a new sample
Random Sample of dataset with replacement
Output is combined by voting
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
Draft saved at 7:24:57 PM
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
Replicate
• Each tree is trained on the same dataset
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
• Number to limit the maximum depth of decision trees.
• Increasing the depth of the tree might increase precision, at the risk of some overfitting and increased training time.
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY


Microsoft Azure Machine Learning Studio
kkjitesh-Free-Workspace
Two Class Decision Forest
In draft
deci
Machine Learning
Initialize Model
Classification
Multiclass Decision Forest
Multiclass Decision Jungle
Two-Class Boosted Decision..
Two-Class Decision Forest
Two-Class Decision Jungle
Regression
Boosted Decision Tree Regr...
Decision Forest Regression
• The number of splits to use when building each node of the tree.
• A split means that features in each level of the tree (node) are randomly divided.
Two-Class Decision Forest
1
Properties
Project
Two-Class Decision Forest
Resampling method
Bagging
Replicate
Bagging
Single Parameter
Number of decision trees
8
Maximum depth of the decision trees
32
Number of random splits per node
128
Minimum number of samples per leaf node
1
Allow unknown values for categorical features
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Thank You…!