"Don't let the digital supply chain scare you."

# **Assignment 4**

Induce decision trees using three attribute selection measures. For each tree, plot test and training error. Plot at least 20 points for training error and 40 for test error. Clearly mention the protocol used for plotting the error curves.

27.03.2020

Nitesh Yadav(35), Rohit Rana(40), Rohit Shakya(41)

M.Sc. Computer Science
Department of Computer Science,
University of Delhi

# **O**verview

Predict the outcome of the game with the given positions of 2 kings and a rook.

# **About Data**

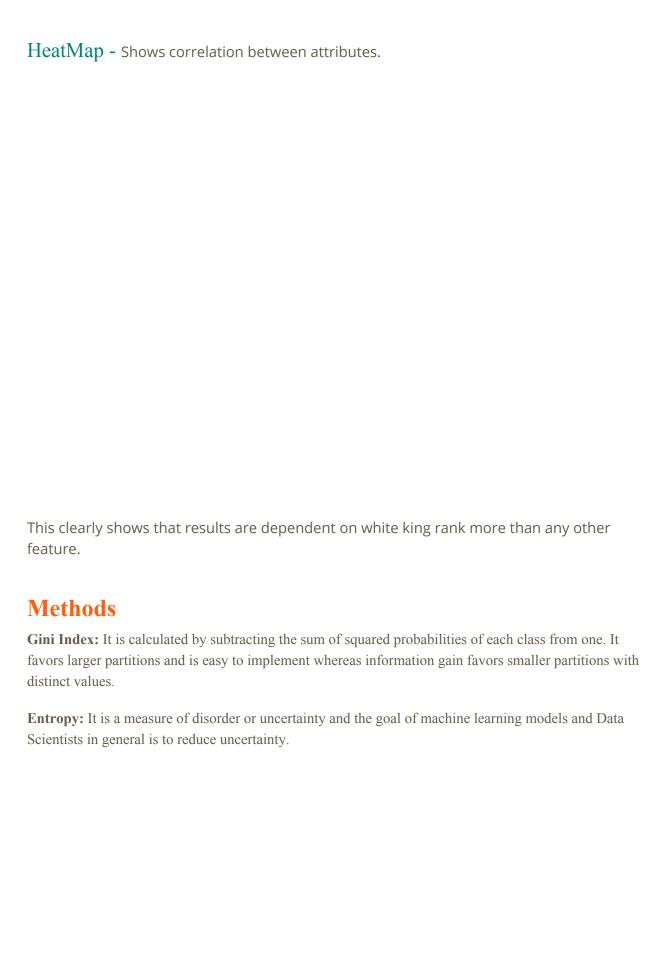
• Instances: 28056 Attributes: 7 and No Missing Values

# **Attribute Details:**

Name	Туре	Description
white_king_file	string	Column location on the chess board of the white king
white_king_rank	string	Row location on the chess board of the white king
white_rook_file	string	Column location on the chess board of the white rook
white_rook_rank	string	Row location on the chess board of the white rook
black_king_file	string	Column location on the chess board of the black king
black_king_rank	string	Row location on the chess board of the black king
result	string	Predictor Class. optimal depth-of-win for White in 0 to 16 moves, otherwise drawn Values: {draw, zero, one, two,, sixteen}

#### Sample Dataset

Used factorize function for handling Categorical Attributes.



# Train and Test error for different size of test tests and at different depth of Decision trees

Incremented test size every time by 2%.

**Gini Index** 

#### **Entropy**

#### **Results**

#### Gini

Mean Train\_error 0.098414Mean Test error 0.340853

#### **Entropy**

- Mean Train\_error 0.099461
- Mean Test error 0.338760

From here we can't conclude that which one is better both are almost equally effective.

#### **Nuts and Bolts**

- At max\_depth = 14 and test size of 20% gives the best results in both the scenarios (Gini Index and Entropy).
- As the max\_depth of the tree increases the accuracy of our model gets better.
- When test size is less than (50%-60%) then the relationship between training and testing error becomes linear.
- When the training error increases exponentially, The model is Under-Fitting.
- When test error starts increasing exponentially. The model is overfitting.