

**ISM 6136 – Datamining/Predictive Analytics**

**Class Assignment 10**

**Dr. Bharti Sharma**

**5 points**

**TASK: Time series forecasting – Data Mining Task using XLMiner**

**Perform time series forecasting on the Average Income of Tax payers in US dataset determine the best model for any ‘two’ of the states. Provide screen shots with your explanation below.**

1. Perform Data Partitioning on the dataset (select Training set at least 60 % or higher)

2. Perform Lag analysis and explain the ACF and PACF plots. Take Training lag of 15 and Validation lag < 14 as the max lag (lag analysis window).

3. Develop 2 ARIMA models for each of the three states by changing the partitioning ‘or’ p,d,q parameters ‘or’ iterations. So total you should have four models (2 models for each State, choose the best model for each State). Create a table to show all the models.

4. Compare the models and determine the best model for each of the two states - based on the following – provide screen shots and your comments for the following model selection criteria:

a) Forecast and Actual plot for Training

b) Forecast and Actual plot for Validation (select any year and compare the results with respect to that in all the models)

c) MSE training and validation error

d) p-values of the coeff and AR1, AR2 etc