**IS669 – Group Project:**

**Part 2:**  For a maximum of 75 points

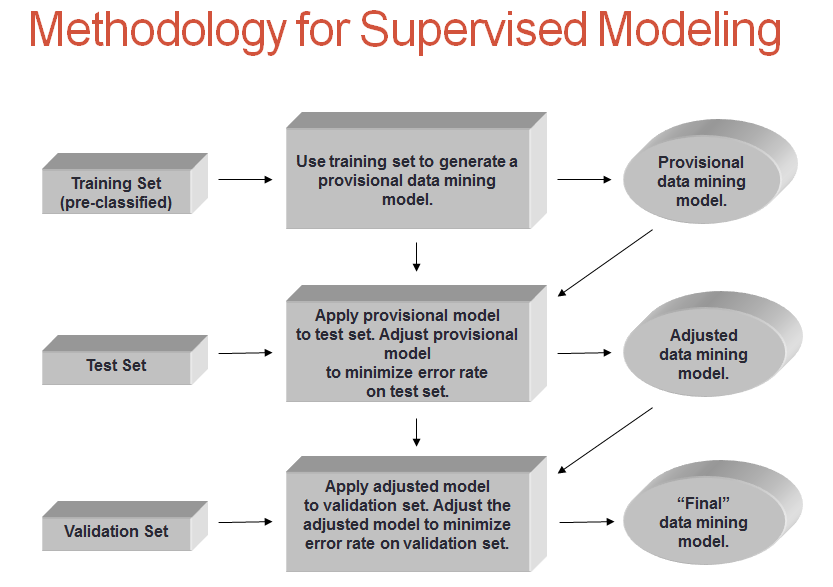
Important note: There is no need to use AWS in phase 2.

Each team will

* Select a sample of 1,000 records from the year that each member worked with in phase 1.
* Merge all the team’s files and load it into a new table in your machine as **Combined Years** table
* Add a new column to the **Combined Years** table and name it **Delays**
* Populate the **Delays** column as follows
  + Insert a “Y” if a record has a non-zero value in the ArrDelay or DepDelay columns
  + Insert a “N” in Delays column otherwise
* The team will choose a machine learning model that can predict which new records will have no delays
* Train, validate and test the analytics model with different segments of your sample data (see the methodology in page 2 of these instructions)
* The instructor will provide a new input file of airline flights records that your team will use to predict if it will have a delay or not.
* A representative will upload the presentation for the team.

**Upload to Blackboard:**

1. **A sample output result as well as captured screenshots of the successful completion of the preparation steps listed above.** Each individual member will upload the team’s power-point presentation to Blackboard. The presentation will contain sections of slides for each individual’s work, as well as  ending  slides describing the final combined years’ process and solution presented in a user-friendly way.
2. **Each member has to include a spreadsheet of the records selected from his or her year that will be included in the group’s total sample. (No more than a thousand per each year)**
3. **With the combined group sample (between 5000 to 6000 records), follow the training, validation and testing as described in class: 1/3 of the records for training the models, a different 1/3 for testing, and a different 1/3 for validation.**



1. **Upload the three data sets in spreadsheet format to Blackboard.**
2. **Include a matrix in the slides describing findings. See example below:**

