Things to do:

1. Preprocessing:

-We will need to preprocess our data, meaning we need to select attributes that we have chosen, see if those attributes have noise or missing values, fill those missing values with either N/A, mean/average of the whole instances of data( in the case of numerical attributes), or get rid of the instances.

-Ridwan and Hajun will meet on Monday to choose which attributes so that we could decide on which algorithm is the right one to approach.

-This would belong to the first section for our final project report.

1. Algorithms

-After having decided the attributes from the previous step, we will decide on what algorithms to use for our data set. I roughly think that since we are trying to classify someone as low-risk, medium-risk, or high-risk, it would be good for us to use (Any ideas). (Probably Apriori depending on the confidence and support for certain attributes)

-We would need to divide our data into training and testing datas. Create a model out of the training data and perform the testing using testing data. Like the professor said during the last lecture, I believe that an ensemble of models could increase our accuracy. I think this could also take a spot in the final report. We could say “our model had …% accuracy, and we tried to enhance our accuracy after doing …” Just for the purposes of learning and increasing our accuracy, I think we could divide the data into 4 datasets in which everyone gets to create a model and then we will combine these models to see what the majority of each classification/association model shows.

1. A When2meet link has been sent out to determine new meeting times.

Preprocessing:

1. Discretize data(transform)
2. Missing Value Detection
3. Use Mode to fill out the missing values (Object types)
4. Use Interpolate for Float Types

3. Check for outliers

4. Relation Between Variable using Covariance and Correlation

# **data collection**

# **Processing**

1. Discretize data(transform)
2. Missing Value Detection
3. Use Mode to fill out the missing values (Object types)
4. Use Interpolate for Float Types

# 

# **descriptive stats, and characteristics**

3. Check for outliers

4. Relation Between Variable using Covariance and Correlation

# **Justify your choice of techniques/design and describe them in detail**