Notes on the slide

1. Random forest to predict decision of men

We used similar predictors above

We use several parameters to tune the process, the final model we select is the model with accuracy of 0.72 and kappa statistic of 0.42.

Variable importance concluded from Mean Decrease Gini statistics (from high to low ).

* Int\_corr
* Tuition.y
* Gaming
* Fun1\_1.x
* Attr1\_1.x

Recommendation: Male participants tend to assign more weights to variables above when they decide if someone matches or not.

1. Random forest to predict match

We used the person’s stated preferences, his demographical backgrounds and responses in the survey before the date.

We use several parameters to tune the process, the final model we select is the model with accuracy of 0.84 and kappa statistic of 0.054.

Variable importance concluded from Mean Decrease Gini statistics (from high to low ).

* Int\_corr
* Tuition.y
* Income.y
* Attr2\_1.x
* Clubbing.x

Recommendation: When the company tries to predict whether two people will match or not. It is important for them to assign more weights to factors above.

1. Logit model to see the correlation between economic power of one people and the match decision.

The model concludes income from both sides have some positive impact on match, but the effect is not significant.