



# Predicting Facebook Check-Ins — Notebook Critiques

Team 9: Erin DiFabio, Chris Farrell, Kris Hooper, and Jared Shawver



## “3rd Place Solution”

### Good

- Separates everything into steps
- Exports intermediate steps to CSVs
- Creates distributions of the data to understand time, accuracy
- Batch processing of data

### Bad

- Doesn't explain what the steps do
- Very confusing to read through
- No comments on the code
- Cannot tell what modeling technique they are using



## “Random Forest and KNN on a Few Blocks”

Good:

- Winner
- Clear and Easy to follow
- Dives into data before starting model
- Manipulation of Time
- Ranger package instead of Random Forest
- Good analysis of what happened after running the model
- Random Forest does well with common check ins

Bad:

- Too much data for a random Forest
- Model doesn't do well with seemingly random check in points



# “FBOOST”

- XGBoost -- Extreme Gradient Boosting
  - Pushes limits of computing power for boosted tree algorithm
  - -can better model performance and computational speed
- Goal is to use XGBoost to obtain variable importance
  - Convert x,y coordinates to spatial points and find spatial relationships
  - Cluster places together
- Comments:
  - Model took too long so they failed
  - Didn't sample data
  - No comments or documentation throughout notebook