LEAGUE OF LEGEND OUTCOME PREDICTION

Yunxuan Li, Jie Chen

CEE 690-3 Final Project

Introductions

Research Qs and Hypotheses

- Predictions based on
  - *Pre-match information?*
  - *Early-game (10-minute benchmark) information?*
  - *Full-game information?*

Data Sampling and Preprocessing

- Source: Riot API
- Retrieval and preprocessing via Rstudio:
  - *Over 5000 matches distributed across all ranks*
  - *Retrieve match by match ID, and calculate features of interest*
  - *Deal with NA values, unclear player-role assignments*
- Other preprocessing in Python:
  - *Normalizing, feature selection*

Core Methods:

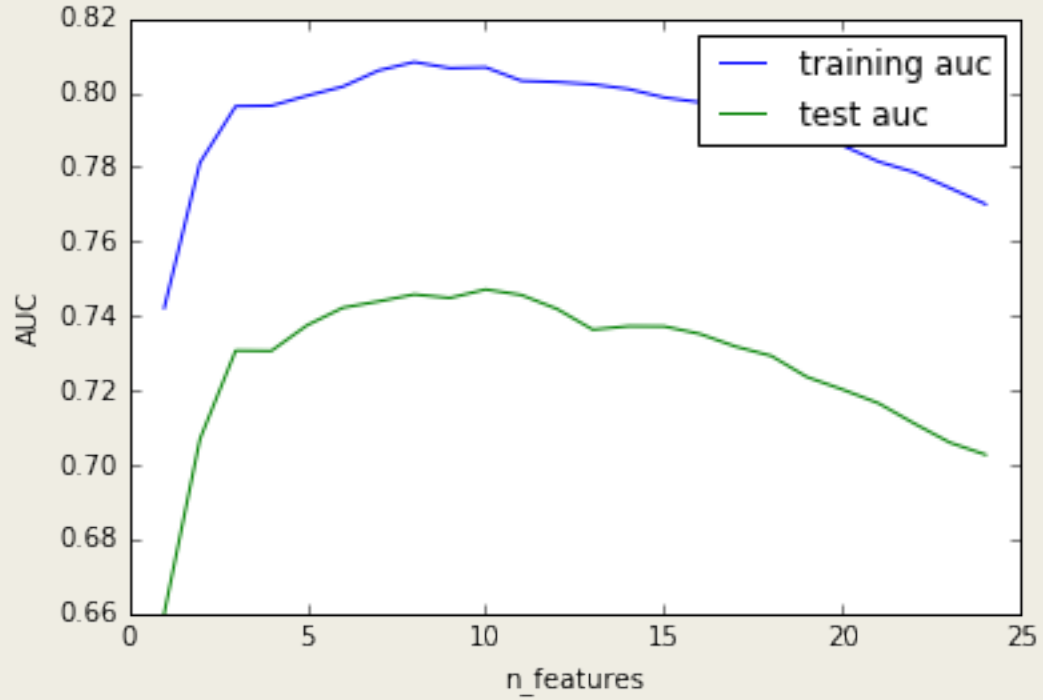
- Three-stage model:
  - *Pre-match*
  - *Early-game*
  - *Full-game*
- Classifiers investigated:
  - *Logistic Regression, K-Nearest Neighbors, Support Vector Machines, Random Forests, Gradiant Boosting, XGBoosting, Multi-layer Perceptron*
- Feature selection via cross validation; parameters of classifiers tuned through cross-validation
  - *Train, test, and validation set*

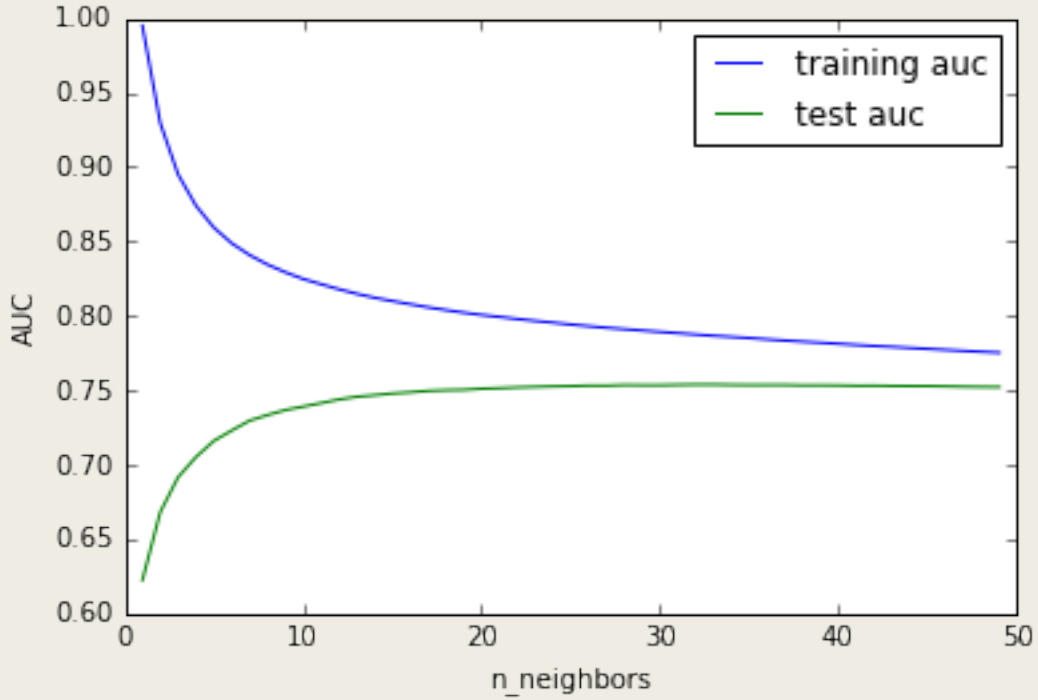
Results: Prediction Accuracy on Validation Set

Method	Pre-match	Early-game	Full-game
Logistic Regression	49.09%	71.48%	82.445%
KNN	54.94%	**79.49%**	99.49%
SVM	55.44%	63.32%	99.61%
Random Forest	53.4%	78.03%	99.89%
GBM	61.24%	75.15%	99.72%
XGBoosting	56.16%	65.32%	97.99%
MLP	56.73%	69.67%	96.27%

Best Model: KNN

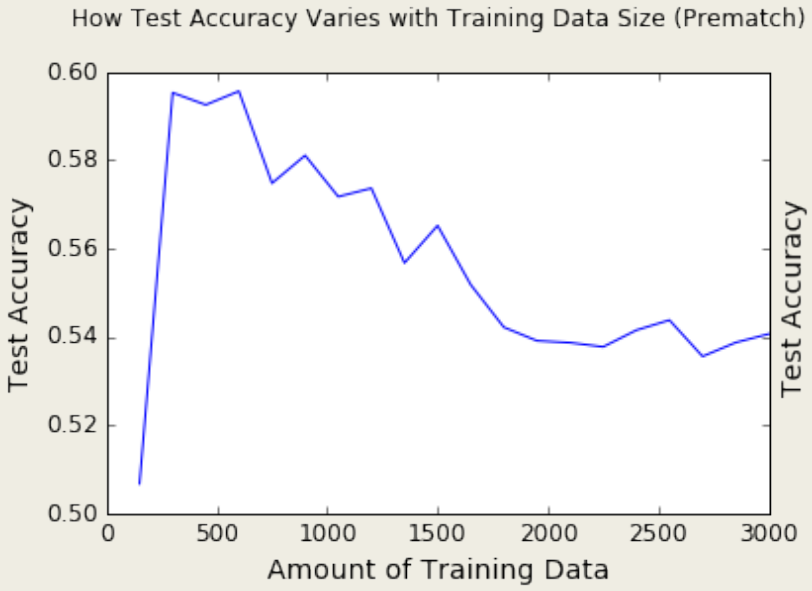
- Select n_features and n_neighbors using Cross-validation

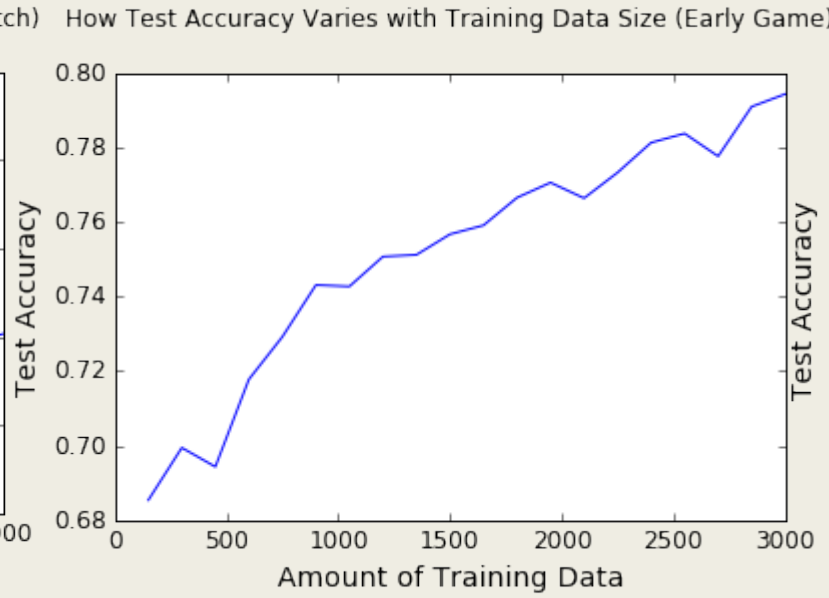


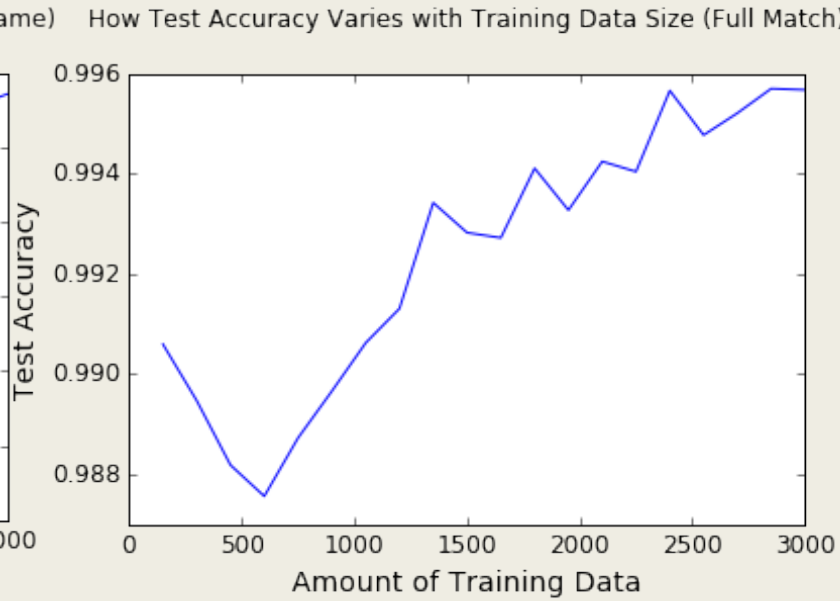


Has the Model Achieved its Limit?

- Researched on how the validation prediction accuracy changes as the amount of train data we feed into the model changes.
- %figures of three-stage of KNN







Conclusions and Implications

- Early-game prediction accuracy is beyond expectation!
  - *Our unique set of variables may play a role in achieving this*

- Can confirm that matchmaking system is… pretty fair!
- Early-game advantage is important – snowballing effect