Senior Project – Computer Science –2014

Machine Learning Examination of NBA Defense

Alex Block

Advisors – Chris Fernandes & Nick Webb

Objective

Identify the significant attributes that contribute to defensive effectiveness in the NBA

Attribute Selection & Removal

The data is gathered for each NBA regular season from 1996-1997 to 2013-2014

Attributes were removed for the following reasons
- Attribute is not an indicator of defense
- Attribute is an indicator of defensive effectiveness, but not a cause of it
- Attribute is highly correlated with another attribute
- Attribute is a composite of other attributes
- Attribute had minimal effect on the correlation of the model

Final Attribute Set (Ranked)

1. Opponent Turnover Ratio
2. Restricted Area Field Goal %
3. Defensive Rebound %
4. Mid-Range Field Goal %
5. Opponent Free Throw Attempt Rate
6. In The Paint (Non-RA) Field Goal %
7. Above the Break 3 Field Goal %
8. In The Paint (Non-RA) Field Goal Attempts
9. Mid-Range Field Goal Attempts
10. Restricted Area Field Goal Attempts
11. Opponent Free Throw %
12. Left Corner 3 Field Goal %
13. Right Corner 3 Field Goal%
14. Above the Break 3 Field Goal Attempts
15. Right Corner 3 Field Goal Attempts
16. Left Corner 3 Field Goal Attempts

Notice how each above attribute fits into one of Dean Oliver's "Four Factors" (defined top right)

Background

Basketball statistician Dean Oliver highlighted **four key factors to basketball success**.
1. Shooting
2. Turnovers
3. Rebounding
4. Free Throws

Notice how each of the statistics in the final attribute set fit into one of these four categories.
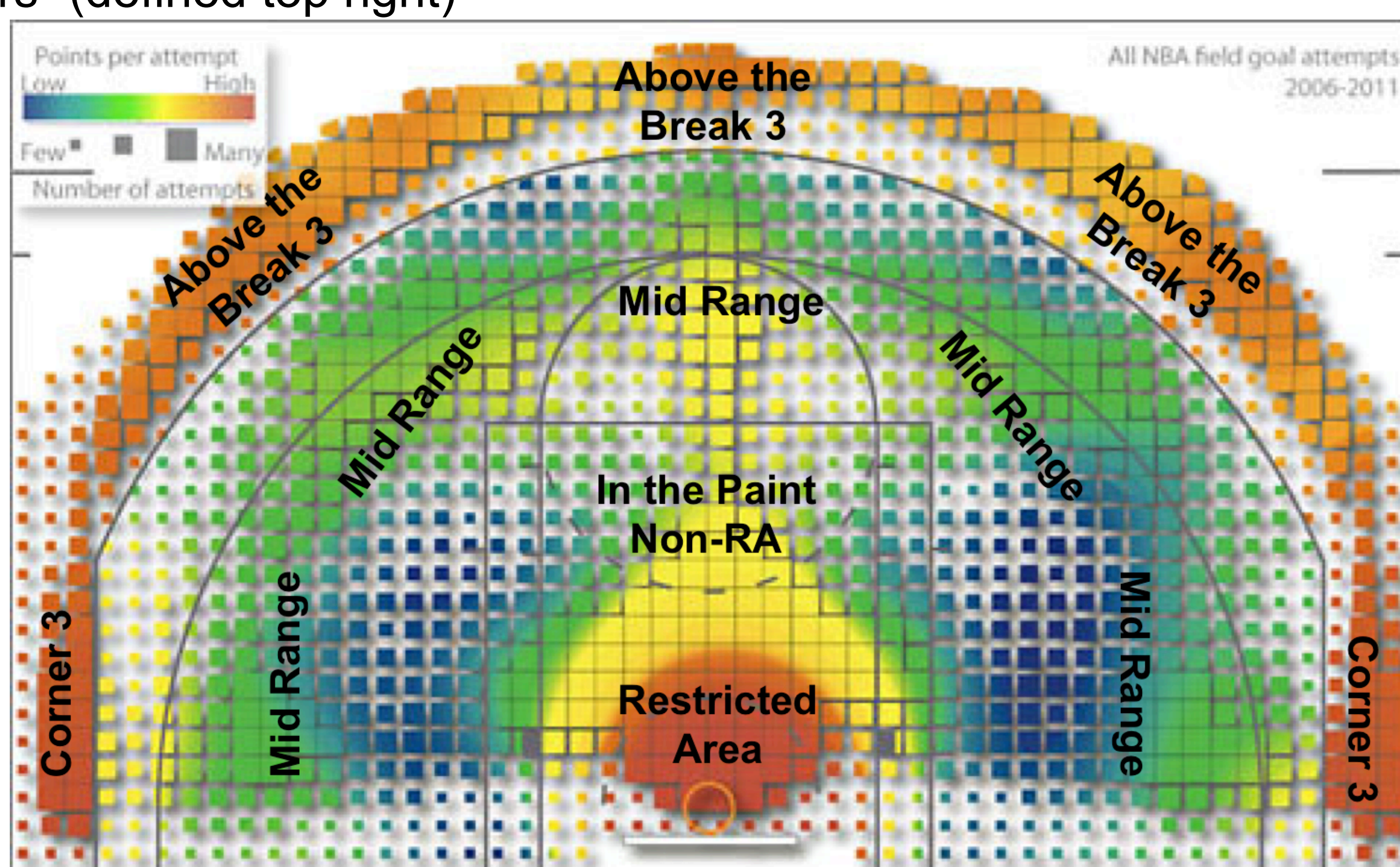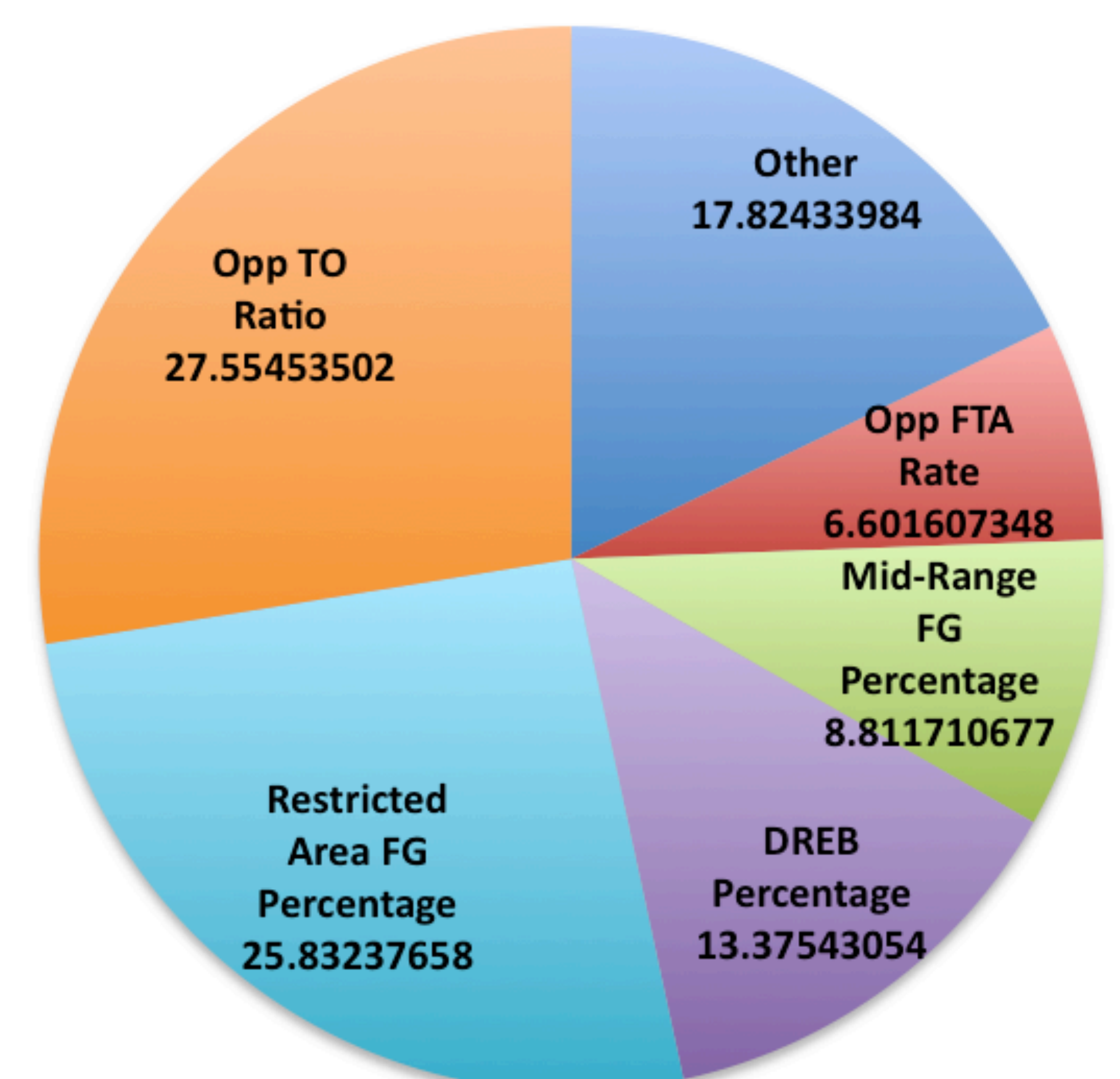

Defensive Efficiency

$$\frac{TotalPointsAllowed}{TotalPossessions} \times 100$$

Defensive Efficiency is used as a numeric proxy for defensive effectiveness in this project. The value is points normalized by 100 possessions.

Attribute Weights

By removing each remaining attribute and tracking its effect on the correlation of my model, I was able to calculate a relative weight for each attribute in the final attribute set. These weights convey the importance of each attribute in comparison to the other attributes.

Other 17.82433984
Opp TO Ratio 27.55453502
Opp FTA Rate 6.601607348
Mid-Range FG Percentage 8.811710677
Restricted Area FG Percentage 25.83237658
DREB Percentage 13.37543054

Final Model

- The best performing classifier for *Defensive Efficiency* is a multilayer perceptron
- For the final attribute set, the correlation is 0.9952
  - This high correlation value suggests that there is a statistically significant relationship between my attributes and predicted value