



THE UNIVERSITY OF  
**CHICAGO**

Department of Statistics  
**MASTER'S THESIS PRESENTATION**

---

**RISHI CHOPRA**

Department of Statistics  
The University of Chicago

**Clustering Professional Soccer Players**

**TUESDAY, July 30, 2013, at 11:00 AM**

117 Eckhart Hall, 5734 S. University Avenue

## **ABSTRACT**

A fundamental area of research in sports analytics is the discovery and description of player archetypes. Using play-by-play data furnished by StatDNA on Europe's Top 5 football leagues for the past two seasons, we attempt to discover particular playing styles among professional soccer players. The analysis proceeds in two steps: a supervised model selection followed by an unsupervised clustering. The feature set was extracted from the data by performance on classification of a player's traditional playing position using support vector machines. With this feature set clustering was performed using both the  $k$ -Means clustering algorithm and a Gaussian mixture fit with the EM algorithm. Each method was evaluated based on its year-over-year consistency as measured by the mutual information of overlapping players in each season. The plausibility of the make-up of each clustering was also considered given traditional position labels. Both algorithms performed remarkably similarly, showing the same consistency between seasons and largely agreeing on the type of playing styles in the data. The results give insight both into the rate of change of certain playing styles and tactics as well as directions for further study.

---

For information about building access for persons with disabilities, please contact Matt Johnston at 773.702-0541 or send an email to [mhj@galton.uchicago.edu](mailto:mhj@galton.uchicago.edu). If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.