

DATA ANALYTICS IN SPORTS: IMPROVING THE ACCURACY OF NFL DRAFT SELECTION USING SUPERVISED LEARNING

Gary McKenzie

Dr. Dmitri Korin, Thesis Supervisor

ABSTRACT

Machine learning methodologies have been widely accepted as successful data mining techniques. In recent years these methods have been applied to sports data sets with some marginal success. The NFL is a highly competitive billion dollar industry. Creating a successful machine learning classifier to aid in the selection of college players as they transition into the NFL via the NFL Draft would not only offer a competitive advantage for any team who used such a successful classifier, but also increases the quality of the players in the league which would in turn increase revenue. However this is no easy task. The NFL prospect data sets are small and have varying feature set data which is difficult for machine learning algorithms to classify successfully. This thesis includes a new methodology for building successful classifiers with small datasets and varying feature sets. A multilayered, random sliding feature count, iterative genetic algorithm feature selection method coupled with several machine learning classifiers is used to attempt to successfully select players in the NFL draft as well as build a larger classification set that can be used to aid overall decision making in the NFL draft.