AdmissionBot (working title): Scope and Design
Daniel Ray, CS-195, Spring 2014

AdmissionBot is intended to assist Westmont College’s admission staff in deciding which prospective students to admit to the college. AdmissionBot should recommend for admission students who are not only desirable students (as determined by GPA, SAT/ACT scores, etc.), but will also matriculate in response to the offer of admission with high probability.

AdmissionBot will be a machine-learning algorithm. While the goal of the project is to implement and deliver the software artifact, research elements are also inherent in the project to guide the design of the algorithm. Machine Learning by Tom Mitchell will be the main source of guidance for the specific machine learning algorithm to be used, while James Scannell’s The Effect of Financial Aid Policies on Admission and Enrollment and A New Formula for Enrollment Management by John Maguire, et. al. may provide insight into the normal techniques and approaches of admissions staffs. As the project progresses, Westmont’s admissions department may also need to be consulted for insight.

After formulating an initial approach based on the results of a preliminary research phase, the first goal will be to tune the algorithm to differentiate between those prospective students who have desirable qualities and those who do not. Following this, the next goal will be to tune the algorithm to offer acceptance only to those who will matriculate with high probability. Multiple approaches may need to be tried to reach the desired high level of accuracy. Once this is achieved, additional modifications can be made to better meet the needs and desires of the admission staff.

Training will be accomplished using admissions data from the past five or so years. The task is essentially a classification problem, so the application of machine learning techniques should be straight-forward. However, many subjective factors – which are not represented in the available data – influence admissions, so some massaging of the data may need to take place to overcome this. It is expected that among the many factors available in the data, some of the more important ones will be GPA, SAT/ACTION score, state of residence, family income bracket, financial aid award, and legacy (family history with the college).