빅데이터 프로그래밍

프로젝트 제안서

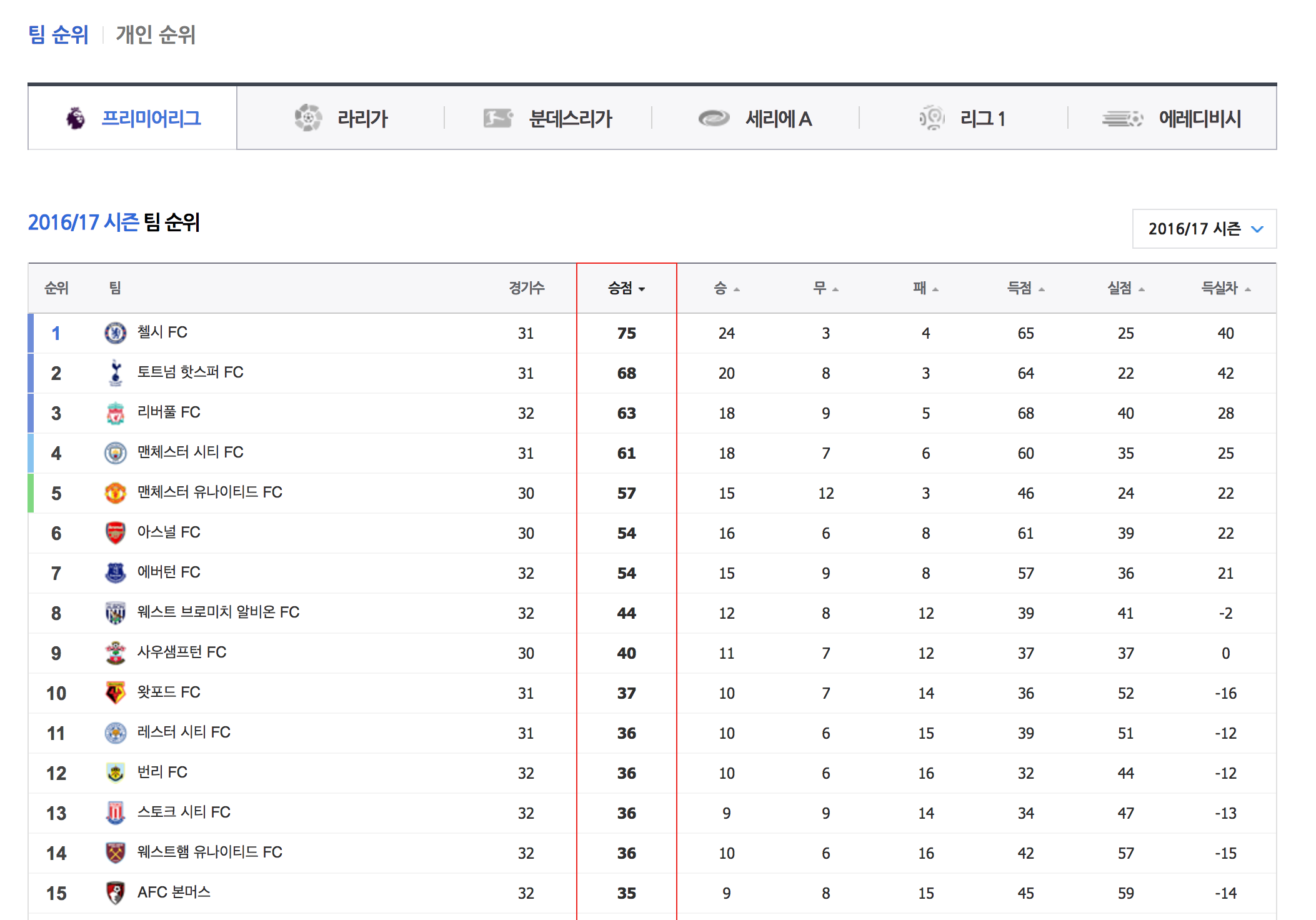
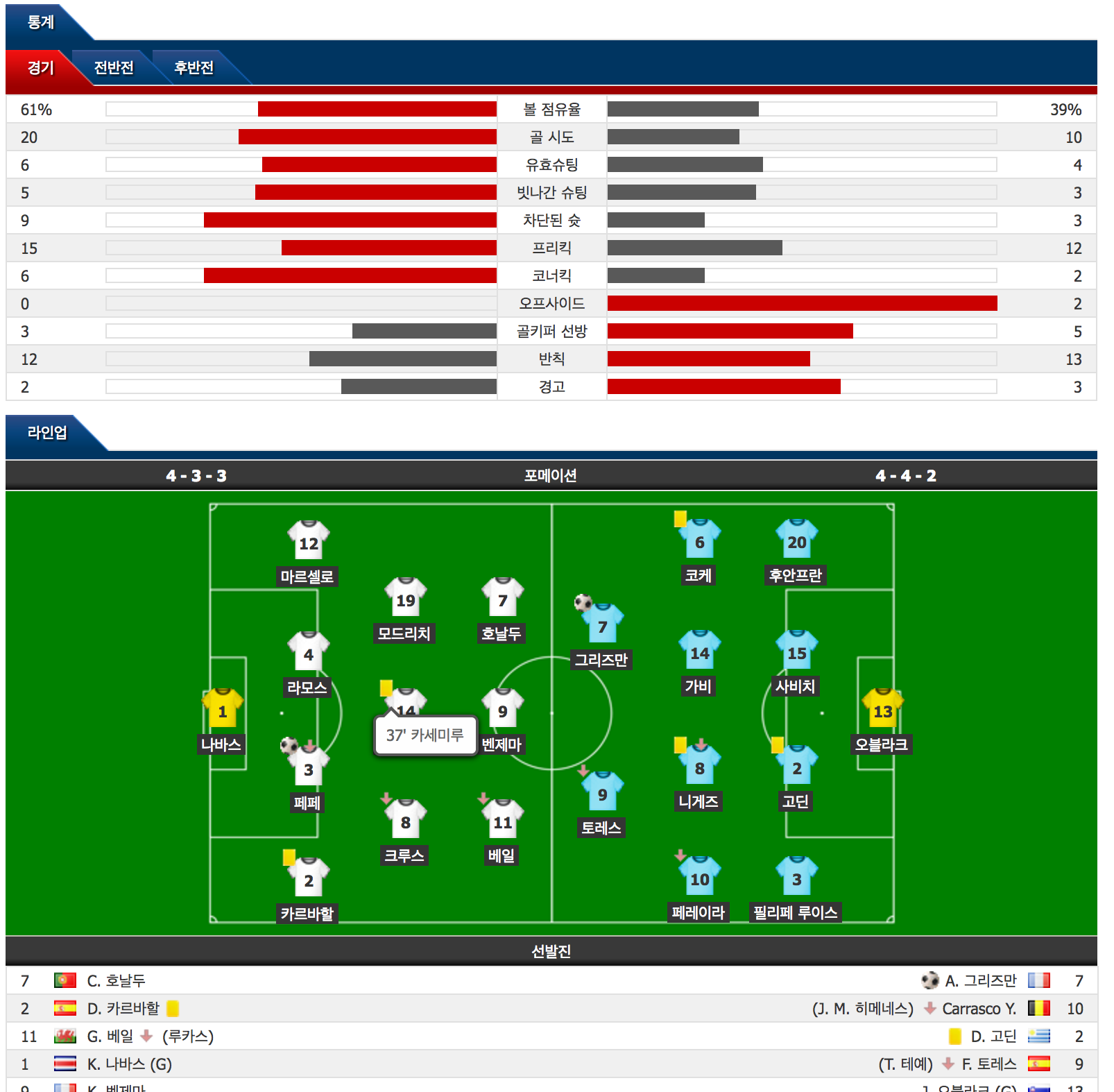
pasted-image.tiff

컴퓨터과학과

201111235 신동규

1. 기획

빅데이터 분야에서 스포츠는 단연 빼 놓을 수없는 데이터 분석을 많이 사용하는 곳중 하나 입니다.

물론 이많은 데이터들을 분석하여 보여주는 사이트들이 많습니다.

그러나 이 데이터들은 각자의 팀의 승무패 의 횟수와 경기가 끝나고 난 후에 올라오는 데이터 분석 결과 밖에 볼수없습니다. 물론 팀마다의 선수 개개인의 능력의 차로인해 승수가 많은팀이 어느팀과 경기를 하던 강하다는것은 사실이지만 새로운 경기의 예측은 개개인 선수의 능력 및 관심있는 사람들이 따로 분석을해야합니다. 또한 이러한 분석결과를 자신이 프로토나 토토 같은 스포츠 게임에 적용 시키기도 합니다.

이미 분석되어있는 데이터는 많으므로 이 데이터들을 다시한번 분석하여 정확도 80%이상의 예측 시스템을 구성하는것이 이번 프로젝트의 목표입니다.

2. 개발

데이터를 총 3가지 분류로 가져올 예정입니다.

첫번째는 팀의 5년간 승률 및 평균스코어

두번째는 상대팀과의 최근 5년간의 전적 의 평균스코어와 승률 ,

세번째는 팀선수의 개개인의 상대팀 포메이션 별 평점을 가져올 예정입니다.

데이터는 <http://www.scoreboard.com> 혹은 naver sport 에서 팀별 데이터들을 가져올 예정입니다.

또한 세번째 데이터를 위해 예를들어 손흥민의 상대팀 포메이션 별 평점을 구하기위해 손흥민선수가 출전한 경기의 상대방 의 포메이션 과 그경기의 평점을 <https://www.premierleague.com/> 사이트에서 크롤링은해서 데이터 정리를 할예정입니다.

팀의 주력선수들의 상대팀 포메이션 별 평점이 높을 수록 상대팀의 포메이션에 강하다는것이 증명되므로 승부예측에 포함시킬수있다고 생각합니다.

처음 개발을 10팀을 기준으로 할예정입니다. 한팀당 9팀과 경기를 할수있으므로 각팀마다 세가지 분류의 데이터 분석값들의 합을 9개로 mongodb에 저장할 예정입니다. 자신이 선택한 팀과 상대팀을 정한다면 그 두팀 사이의 데이터들을 보여주므로서 예측 결과를 보여줄수있습니다.

3. 기대효과 및 경제성

이미 스포츠에 대한 데이터분석은 스포츠 팬들이 알아보기 쉽게 많이 정리되어있습니다.

그것들을 사람들이 일일이 분석을해서 예측하거나 중계해설자들이 해주는 해설로 승부예측을 하는것보다 ,

크롤링을 통하여 상대팀에 대한 승률을 분석하여 한번에 보여줄수 있는 시스템이있다면 보다 편하게 축구를 즐길수 있다고 봅니다. 또한 스포츠 토토와 프로토를 즐기기위해서는 자신이 응원하는 팀 뿐만아니라 다른 팀들의 정보들까지도 세세하게 분석을해야하는 등 손이 많이가서 접근성이 로또보다 떨어져있었습니다. 다음 프로젝트를 진행하여 승부예측이 80%이상 된다면 이 결과들로 접근성을 낮추고 많은사람들이 스포츠를 즐기고 스포츠 기금도 올라가는 효과를 볼수있다고 봅니다.