

졸업 프로젝트 주제 : 축구 승부예측 모델 개발

디자인 스프린트 5일차 –survey, pitching

나 혼자 한다조-201302366 김규태

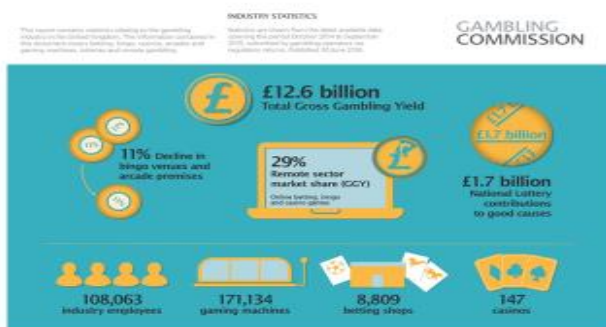
Github: <https://github.com/kimtuyer/Football-Predict-service>

Youtube: [https://www.youtube.com/watch?v=8NMbTl-qn\\_0&feature=youtu.be](https://www.youtube.com/watch?v=8NMbTl-qn_0&feature=youtu.be)

## 시장조사

- 승부예측 모델과 적중률 및 선정한 데이터를 비교할 만한 타 사이트의 적중률 비공개 및 공개된 승부예측 모델이 없음
- 따라서 승부 예측을 통해 실질적으로 비즈니스 모델을 창출하는 스포츠 베팅 시장에 대해 조사

## 시장조사



- 승부예측은 스포츠베팅 시장과 밀접한 관련,
- 축구의 나라 영국 베팅시장 약 10조원 규모, 10년전 베팅산업규모 10조규모, 영국 gdp 0.5, 연간 10만개 일자리창출,
- 1조원대 세수효과, 현재 중계권료 14조, 선수들의 총합 시장가치 12조에 달하는 프리미어 리그 20개구단중 절반이 베팅업체 스폰

## 국내 스포츠베팅시장 매출 현황



- 2001년 축구에서 처음 시작, 18년까지 7조9082억 기금 편입,
- 고객환급금,사업운영비제외, 국내 문화체육산업 지원

### 설문조사

> 승부예측 사이트를 이용해본 경험이 있는 분, 없는 분 모두 포함시킨 3분에게 조사

승부예측 모델에 대해 궁금한 점이 있나요?

- 왜 기존의 승부 예측 제공하는 사이트에서 과거예측 내역 및 적중률을 공개하지 않는가?, 공개하지 않는다면 기존의 사이트들과 어떻게 적중률을 비교할 수 있는가

답변: 이에 대해선 별다른 정보가 없어서 답을 하기 어렵지만,

추측해보자면 승부예측 스코어를 제공하는 사이트의 성격상

승부예측 자체가 사이트운영의 주 요소가 아닌 단지 부수적인 콘텐츠중 하나로

인식하는것 같습니다. 그리고 조사해보니 공개적으로 공개된 승부예측 프로그램이나

모델의 예가 없어서 구체적으로 다른 모델과 적중률을 비교해 보긴 어려울거 같습니다.

- 승부예측 모델 개발 및 구현을 통해서 사용자들이 기존과 다르게 얻을 수 있는 것은 무엇인가

답변: 승부예측 프로그램에 관심있어하는 사용자는 대개 앞서 말씀드렸다시피

비즈니스 수익으로 연결시킬 수 있는 스포츠베팅 유저분들일거라 생각합니다

다만 관련 정보를 공유하거나 의견을 나누는 사이트에 가보면 생각보다

스포츠베팅에 대해 데이터보단 자신의 감에 따르거나 주관적 해석에 의존하시는분들이 많은것 같습니다.

하지만 승부예측 모델을 개발하다보니 주관적인 판단 보단 객관적인 데이터에 근거해 확률적으로 접근하는 것이 승부예측을 적중하는데 보다 도움이 될거라는 확신이 듭니다.

- 어떤 경로로 데이터를 수집했는지, 해당 데이터를 수집한 근거, 선정 기준은 무엇인가

답변:

데이터를 수집한 경로,선정한기준에 대해선,인터넷보다

축구 승부예측에 관해 다룬 책에서 데이터 선정에 관한 많은 힌트를 얻은거같습니다.

승부예측 역시 어떤 감에 따른 도박이 아닌 통계,확률적인 영역에서 접근하면 충분히 좋은 모델 개발 가능하다 여겨, 통계나 확률적으로 축구를 다룬 자료들에서

근거를 얻어 데이터 수집함

머신러닝 모델의 경우 승부예측모델이 기존의 결과가 주어진 경기들을 대상으로

분류하는 것이기 때문에 지도학습에 속하는 모델들을 위주로 실험했는데

우선 기본적인 모델들로 서포트 벡터 머신, 의사결정트리를 적용해봤는데 의사결정트리

약간 더 좋은 적중률을 갖는것 같습니다. 아직 몇가지 실험중인 feature들이 남아있어서

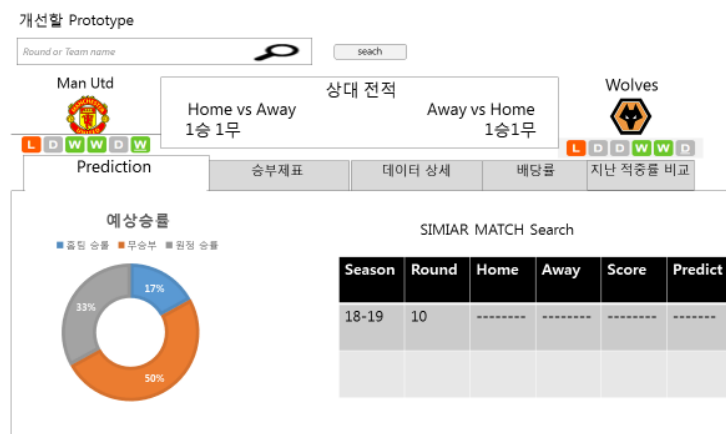
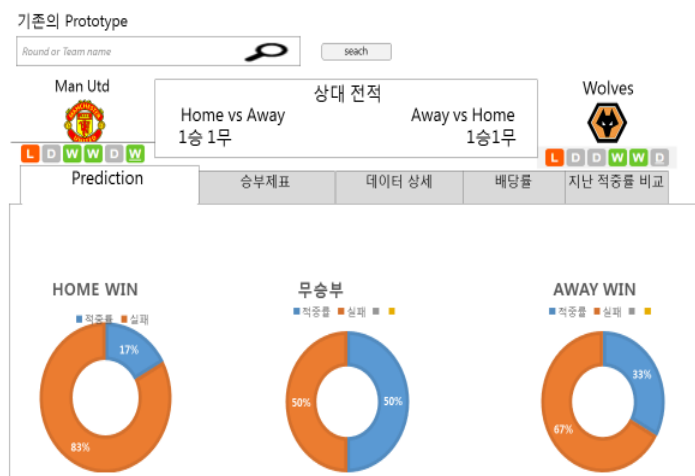
좀더 데이터를 정제화한 후에 신경망 모델도 적용해볼 계획입니다.

## 프로토 타입에 대한 피드백

- 경기 선택할시 나오는 Prediction 탭에서 홈,원정,무승부의 예상 승률을 나타 낸 그래프를 하나로 줄이고, 해당 경기와 유사한 데이터 패턴을 지닌 기존의 다른 경기결과를 같이 제시해 비교해주는 것도 좋겠다

답변: 저도 좋은 의견이라 생각해서 기존의 프로토타입이 다음과 같이 승률그래프만 있었는데.

다음과 같이 그래프를 하나로 합치고, 옆에 다른 유사한 경기결과를 제공해주는 란도 반영할 계획입니다.



- 데이터상세'란 탭에서 예시란을 보니 어떤 데이터지표에서 차이가난다는 건 알겠는데 그것이 확률적으로 두 팀의 승률에 어떻게 차이가 나는지 알려 주면 좋겠다

답변: 데이터들 마다 승률에 영향을 끼치는 정도가 다르고, 어떤 지표의 어떤 구간내에서 영향을 끼치는지 다 달라서 풀어서 설명하기 복잡해서 보기 좋게 보여드리기가 쉽지 않을 거 같습니다. 이 부분은 어떻게 설명드리는게 좋을지 더 고민해보고 결정해야될거 같습니다.