

준비하면서 들었던 생각들..

- 어떤 사람을 뽑는가는 회사에서 가장 중요한 일중 하나라고 생각.
 - 까다롭게 뽑기 때문에 안에 계신분들도 좋은 분들일 것 같다.
 - 팀원 모두가 함께 일할 사람을 고르는 과정에 참가하는건 멋지다.
- 지원자에게도 만약에 입사하게 된다면, 함께 일하게될 사람을 미리 알수 있는 기회가 되기 때문에 좋은 기회다.
- 내가 가진 능력이나 할수 있는 것들 중에서 어떤 것들이 팀에서 중요하게 생각하는지 정확히 알수가 없기 때문에 일단은 중요하다고 생각하는 것들을 나열.
- 질문을 많이 해주시면, 많은 것들을 보여드릴수 있을것 같습니다.

소개

간단소개

- 제주도 제주시 거주
- 다음커뮤니케이션 데이터마이닝팀 (2010.3 ~ 현재)
- 송실대학교 학사(Computer Science) / 석사(Machine Learning)
 - 졸업논문은 Feature Selection 관련. Feature가 굉장히 많은 데이터에 대해서, Wrapper 방법으로 효과적으로 Feature를 추출하는 방법.
- 영화, 사진, 독서, 테니스, 피아노...

지원이유

- 이제 서울로 돌아오고 싶다.
- 업무내용**과 **요구조건**이 마음에 들었다.
 - 중요하다고 생각되는 부분들이 요구조건에 포함이 되어 있다.
- 변화와 발전

스킬

- 자주 사용하는 것: Python, Bash, Java (hadoop map-reduce)
- 가끔 사용하는 것: SAS, Hive, Pig, Weka, Javascript, C/C++, R, MongoDB, Redis

데이터분석에서 나의 장점이라고 생각하는 점

- 시스템, 데이터 전처리, 예측 모델링, 프로토타이핑, 대용량처리(hadoop) 개발이 전반적으로 가능한 점
- 잘 듣는 것, 정확하게 보려고 노력하는 것, 예술, 철학, 문학, Ways of Seeing

데이터 분석을 잘하기 위해서 중요하고 생각하는 것들

- 인문/철학, 수학(통계), 기술(CS)의 조화
- 이 일들을 도대체 왜 하는가? - 길을 잃지 않기 위해..
- 무엇이 좋은 건가? - ‘**질**’의 문제, 성능평가, 예측 모델의 방향성
- 좋은 눈 - 문제점 파악, 멀리서 바라보기
- 커뮤니케이션: 용어, 잘 전달하기, 꾸미지 않기
- 가능하다면 Case by Case
- Knowlege vs. Intuition: “거인의 어깨에 올라서서 더 넓은 세상을 바라보라.”
- ‘**기본 지키기**’와 ‘**관습의 파괴**’의 반복
- 빠른 프로토타입 & 반복된 실험

경험들

1. ECML PKDD Discovery Challenge 2008

- 소셜북마킹 서비스에서 스팸 찾는 문제, 3등
- Information Gain, Naive Bayes Classifier
- Naive Bayes classifier learning with feature selection for spam detection in social bookmarking*

2. 데이터 분석요청

- 기획자들의 수많은 요구사항
- 반복되는 일들을 어떻게 가치있게 만들것인가
- 커뮤니케이션의 중요성

3. 실시간뉴스 클러스터링

- 처음 담당하는 큰 서비스
- 속도 문제, C++

4. 블로그 스팸분류

- 학교에서 실험으로 사용하던 데이터와 다르더라..
- 학습데이터 수집의 문제
- 실제 적용에 대한 고려와 설득의 중요성

5. 관련검색어

- 관련검색어 개발 및 유지
- 관련검색어는 무엇인가? 어떻게 해야 잘하는 건가?
- 로직을 크게 개편: 실시간, 자동화
 - 개인적인 실험이나 시도가 많았다
 - 프로토타입을 이용한 커뮤니케이션을 적극적으로 활용
- 연관성분석, 키워드 파싱, 로직보다 많은 필터링, 어뷰징 찾기
- 기획, 개발과 함께 빠르게 많은 것들을 진행

6. 동일의도 검색어 추출

- 동일의도 검색어 추출 모델 개발
- 문제 정의의 어려움
 - 어디까지 동일한 의도로 볼것인가?
 - 효과적인 대화, 진행하면서 계속해서 이야기하기
 - 다양한 사람의 데이터에 기반하여 학습
- 학습데이터 안의 오분류 데이터
- Supervised-Learning, Feature Extraction, Feature Selection, Ensemble Model

7. 광고타겟팅 연구 개발 TFT

- Display Ad 관련 연구
 - targeting, real-time bidding, auction, optimization
 - 타겟팅로직 연구 및 개발
- 새로운 비즈니스, 큰 조직
 - 성능평가의 기준의 다양함
- 신규타겟팅 모델 개발
 - hierarchical classification

감사합니다.