2/27 윤재욱 교수님 상담 내용

(원문)

플러그인 하기전에 지난번에 알 엑셀을 한 적이 있다 . 근데 그건 교수님이 실수 한 것이다. 왜냐면 더 이상 알엑셀은 가지 않는다. 기본적으로 우린 R을 편리하게 만들겠다는 것이다 R의 문제는 사람들이 접근하기 쉽지 않다. 그래서, 엑셀을 제일 먼저 생각 했었다. 엑셀은 VB를 통해 프로그램을 편하게 짤 수 있다. 해서, 실제론 한팀은 6시그마를, 한 팀은 R엑셀을 했었다. 6시그마는 잘한게 아닌데 내용이 좋은건 아닌데 대상을 받았다. 꼭 내용이 중요해야만 하는 건 아니다 남을 설득할 수 있느냐가 관건이다. 알 엑셀은 더 이상 연구가 잘 진행되지 않고 이제는 앞으론 알커멘더인거 같다.

앞으로의 데이터 분석은 R로 갈 것 같다. R이 중요하다는건 알겠고 사용하기 편한걸 생각 하자니 기존에 R커멘더가 있다. 그러면 여기서 우리가 해야할 생각은 무엇일까. 기능을 더 해 줄수 있는걸 생각해야된다. 항상 맨땅에서 내가 무언갈 할 수 있다고 보여주는건 우리가 보기엔 멋있어보일 수 있어도 딴 사람들이 보기엔 그렇지않다. 우리가 중요하게 생각한거라면 남들도 그렇게 생각 했을것이다. 해서, 남들이 무엇을 했는지 정리 해볼 필요가있다. 즉, 남들이 다 개발 했는데, 미쳐 깜빡 한 것, 혹은 미흡한 것을 “보완”한다는 개념으로 접근해야 한다. 분명 수 많은 사람들이 R을 쉽게 쓰게하려고 노력했을 것이다. 그러한 노력들의 산물이 무엇이 있는데 그게 R코멘더이고 우리는 이 R 코멘더로 갈것인데, 그 알코멘더는 무엇을 제공하고 해줄 수 있는데 이러한 부분이 미흡하고 부족한데 우리가 그것을 보완하겠다. 그리고 그것이 궁극적으로 CONTRIBUTION이 된다는 것을 보여주어야한다. 4학년 프로젝트로서, 남에게 물건을 파는 일이 아닌 우리가 이 프로젝트를 진행하며 어떤 공부를 했고, 그 과정들이 우리의 전공 심화에 어떻게 도움이 되었고 나아가서 고객들에게까지 도움이 될 것이다라는 것을 보여주는게 중요하다. 무엇이 있고 그 무엇을 개선이 된 무엇이 있으며 거기에 뭐가 더 필요하고 우리가 그것을 추가할 것이다.라는 이야기가 들어가야 한다.

알 플러그인이 뭐가 만들어져있고, 뭐가있는지 알아야하며 그걸 기반으로 우리가 무엇을 추가할것인지 즉 주위 환경 분석이 다 되고 우리가 무엇을 추가하며 그것이 어떻게 CONTRIBUTION이 되고 게다가 이러한 것에 대해 할 능력이 있다는 것을 입증해야 제안서가 완성이다.

CLASSFICATION을 하나만 해도 방법론이 여러가지가 있고 INFORMATION GAIN 하나에 대한 방법론만 해도 여러 개가 있다. 그 안에서도 여러 기법으로 나뉘기 까지 한다. 그래서 이러한 패키지가 여러 개가 담긴 알 커멘더를 만든다는 것은 불가능에 가깝다. 그래서 R패키지들은 세분화 되어 나뉘는 것이다. 알커멘더가 이미 최상위에 있는 존재이다. 모델에서는 리그래션으로 가고 나머지는 플러그인으로 해결 하겠다는 것이 R커멘더의 입장인 것이다.

중요한건 건더기 없이 껍데기만 크게 보이겠다는 것은 아무 의미가 없다. 그렇기에 뭔가 건더기는 하나 있고, 거기서 뻥을 치는 것이다. 잘 살려면 근거를 가지고 뻥을 처야 한다. 근거 하나 가지고 저 여기까지 했어요 하면 그냥 밥벌어 먹는덴 문제없는정도에 불과하다. 즉 어느 한가지 분야의 SPECIFIC한 한가지를 잡고 그 안으로 파고들면 어떨까 하는 것이다. 예를 들면 CLASSIFICATION 하나만 해도 관련된 패키지가 수십가지가 있을 것이다. 로지스틱, 판별 분석, 디시전 트리, 랜덤 포레스트 등 많은게 존재한다. 지금 이 플러그인 중에서 CLASSIFCIATION을 종합적으로 다루는 패키지는 없을 것이다. DOE만 해도 우리가 원하는 결과가 제대로 없다. 왜냐, 스플랏 디자인, 네스티 디자인도 잘 안되있다. 우리쪽에서 많이 쓰는 점성도에 의한 오쏘그널 어레이가 있는지 모르겠다. 실험계획에서 우리가 많이쓰는 분야가 많이 제공되는지 모르겠다. 실험계획은 많이 쓰는 분야다. 아니면 우리 품질관련해서도 컨트롤 차드, 심플릭 익스프레션, 공정능력관리, 게이지 R&R 이런 것들이 있는데 이런것들을 잘 모아서 뭔가 할 수 있는게 뭔가 있을까 고민하자는 거다. 지금 그런게 없으면 그런것들을 모아서 제공하면 어떨까 하는 것이다. 그런 점에서 너네가 5명이란 말이다. 그래서 그런것들을 두가지 하면 어떨까.

너희들의 고객은 교수들이다. 교수님들에게는 우리가 큰 프로젝트로 패키지 하나 개발했습니다. 보다는 이런 기능을 가지는 이런기능을 가지는 패키지 두개를 개발 해 보았습니다. 그리고 이러한 것들이 통합할수도 있습니다. 두개를 합쳐서 알코멘더에 올릴 수잇습니다 등등 포장을 잘해라. 잘 모를땐 DIVIDE AND CONQUER다. 쪼개고 약한놈부터 하나씩 정복해야한다.

검증에 관해서는 아주 좋은 방법이 있다. 패키지를 만든 후 사용자들에게 동일 기능을 제공하는 기존 TOOL과 사용 후 비교를 권해보면 된다. 예를 들어 5~10명의 사용자를 모집하여 미니탭과 우리의 패키지가 추가된 R코멘더를 써보고 인터뷰라든지, 설문지를 통해 어느정도 검증을 받을 수 있을 것이다. 사실 그 이후엔 여러가지로 검증할 방법이 많다. 근데 ㅋ 버그만 안나도 ㅋ 너네들이 거기까지 할 수 있을까.. 할 수 있으면 좋지..ㅋ

아이디어가 좋은건 좋아. 근데, 기존의 분석을 안하고 덤벼드는건 아주 큰 낭패를 보게 될 것이다. 뭐가 했고 남들이 그걸 위해 무엇을 했고 난 거기에 뭘 더 추가할수 있을까 이게 중요하다.

AREA별로 물론 할 수 있다. 이런 서비스경영이나 등.. 그러나 그것은 다음 년도나 차기의 후배들 때를 생각하고 있다.

(정리)

-R commander를 생각 하게 된 배경

→ 앞으론 R이 데이터 분석의 선도적인 역할을 할 것이라는 데에는 많은 사람들이 공감 할 것이다. 그런데 그러한 R은 비 전문가가 사용하기에는 어려움이 있다. 해서, 처음 생각 한 것이 접근성이 용이한 R excel이었다. 그러나 이는 더 이상 연구가 활발하게 진행되지도 않고 많은 문제점이 있다. 때문에 이제는 R commander가 GUI를 갖춘 R통계 툴의 주역이 될 것이라 생각.

-제안서가 갖추어야 할 요소들

→ 확장적으로 생각 할 필요가 있다. 기존 R의 불편함 때문에 R코멘더가 있다면, 그러한 R코멘더를 추가적으로 편리하게 만들기 위한 수많은 노력들이 분명 있었을 것이다. 그러한 그 동안의 노력(현재까지의 산물)을 다 조사하고, 와중에 깜빡 놓치거나 미흡한 부분을 우리가 “보완” 할 것이라는 것을 강조해야 할 필요가 있다. 그러한 보완 내용이 어떻게 고객들에게 contribution되는지 설명 해야 할 것이다.

또한, 4학년 졸업프로젝트로서 우리가 이 패키지 개발 플젝을 진행하며 어떤 공부를 해야 했고 어떻게 배운 내용을 활용 했다는 설득이 이루어져야 할 것이다.

-패키지 내용에 대한 대략적인 예시

→ classification 하나만 해도 방법론이 여러 가지가 있다. 우리가 KDD와 같은 프로세스를 통합적으로 구축하겠다는 것은 그 안에 들어갈 수 많은 패키지를 생각하면 불가능에 가깝다고 생각한다. 때문에 좀더 특정화 하여 깊이 파고들 필요가 있다. 품질, 실험계획 등의 분야에서 중요한 이슈임에도 아직 R commander에서는 해당 분야에 대한 종합적인 패키지가 없는 경우가 있을 것이다. 그런 것을 찾고, 적어도 2개 정도는 만들어 보면 어떨까.

-검증에 관하여..

→ 기업을 통한(ex 실험계획법) 검증 또한 방법이 될 수 있으나, 가장 간단한 검증 방법으로는 개발한 패키지 기능을 제공하는 타 tool 사용자에게 동시사용을 권하여 인터뷰, 설문 등을 통해 우리 패키지의 성능을 검증 하는 방법이 있다.

-졸플에 대한 tip..(?)

→ 우리들의 고객은 교수님들이다. 교수님들에게는 우리가 큰 프로젝트로 패키지 하나 개발했습니다. 보다는 이런 기능을 가지는 패키지 저런 기능을 가지는 패키지, 합쳐서 두 개를 개발 해 보았습니다. 그리고 이러한 것들이 통합되어 함께 사용 될 수도 있습니다. 두 개를 합쳐서 R commander에 올릴 수도 있습니다 등등 포장을 잘해라. 잘 모를 땐 divide and conquer다. 쪼개고 약한 놈부터 하나씩 정복 해야 한다.

앞서 말한 KDD 처럼 area 단위로(ex 서비스 경영관련 패키지) 물론 생각 해 볼 수있다. 그러나 시간이 얼마 안 남은 시점에서 권하고 싶진 않다. 이는 차기 후배 대에서 추가 개발로 생각 해 볼 일인 것 같다.