|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| C:\Users\JIWON\AppData\Local\Microsoft\Windows\INetCache\Content.Word\로고.jpg | 27기 통계분석학회 P-SAT 입회지원서 | | | |
|  | | | 잔여학기 | 4 |
|  | 성 명 | 이정우 | 생년월일 | 1998.11.03 |
| 학 번 | 2017313195 | 원전공/복수전공 | 통계학과 |
| 휴대전화 | 010-4926-3143 | 학년 | 3학년 |
| 주소 | 서울특별시 종로구 명륜3가 153-7번지 201호 | | |

|  |
| --- |
| **Q1. 자기 소개** (간단히 자신을 소개해 주세요. 동아리, 기타 외부활동 및 특기사항 등을 자세히 소개해주세요.) |
| 저는 통계학과 17학번 성실의 아이콘 이정우입니다. 항상 후회없이 현재에 최선을 다하는 삶을 살려고 노력하고 있고, 맡은 활동에 대해서는 선택에 대한 책임감을 느끼고 열심히 합니다. 그래서인지 항상 열심히 한다는 칭찬을 지금까지 해왔던 동아리, 학회에서 많이 받아왔습니다.  저는 입학 후 노래패와 사회복지인권학회 활동을 했습니다. 17년 2학기엔 사회복지학과 소속 사회복지인권학회 “새봄”에서 활동하면서 인권, 빈곤 등 사회복지와 관련된 주제들에 대해 세미나를 하고 당시 사회적 이슈나 다양한 논의점에 대해서 의견을 주고받는 활동을 했습니다. 그리고 17~18년 2년 간은 사회과학대 노래패 “아우성”에서 1년은 노래패원으로, 1년은 사무팀장으로 활동했으며, 19~20년 2년 간은 강원도 태백에서 의경으로서 군복무를 하고 21년 1학기에 복학을 했습니다.  3학년에 진입하며, 통계학과 전공핵심 6과목을 모두 들었습니다. 통계 전공평점은 4.08이고 전역 후 전공 과목을 세세하게 복습했기 때문에 핵심 과목과 관련된 기본은 다져져있다고 생각합니다.  저의 특기는 선형대수적 직관력과 이해력, 그리고 그것에 기반해 통계 이론의 깊은 이해가 가능하다는 점이라고 생각합니다. 저는 전공 과목인 “행렬대수학”을 수강했지만 지식이 많이 얕다고 생각되었습니다. 그래서 MIT Gilbert Strang 교수의 “linear algebra”를 온라인 수강했고, 해당 강의에서 제공하는 과제를 거의 다 풀어봤습니다. 강의 내용의 수준에서 멈추지 않았습니다. 강의에서 생략하거나 대충 지나가는 SVD의 증명과 다양한 특징들, 라그랑주 승수법을 통한 PCA 증명 등을 검색해가며 자세히 공부했고, Haar wavelet transformation 또는 PCA, 기저변환을 활용한 다양한 데이터 압축 기법들, FFT를 활용해 연산량 줄이는 방법 등을 구글링을 통해 더 깊이 공부했습니다. 강의 내용 외의 선형대수 내용에도 관심이 많아 Cauchy-Binet formula의 증명이나 Hessian matrix의 의미, 숄레스키 분해 등 다양하게 공부했고, 이에 따라 선형대수 전반에 대해 폭넓게 알고 있습니다. 선형대수를 공부할 땐 증명실력과 응용실력 모두 잡으려고 노력했고, 이러한 수학적 기초를 토대로 통계분석에 대한 깊이있는 이해와 응용이 가능할 것이라고 생각합니다. 이는 회귀분석입문을 복습하는 과정에서도 느꼈습니다. 회귀분석입문을 수강할 당시에는 제대로 이해하지 못했던 Principal Component Regression에 대해서, PCA의 증명과정을 통해서는 X와 Y를 Centering해야하는 이유를, 기저변환을 통해서는 데이터 행렬 X의 표준화 행렬 에 의 고유벡터 행렬인 P의 곱 가 주성분 행렬이 되는 이유를 쉽게 이해할 수 있었습니다.  저는 또한, 학구열과 호기심이 뛰어납니다. 어떤 이론을 배울 때면 항상 왜 그렇게 해야만 하는지를 궁금해하고 알고자 하며, 무언가를 깨닫고 알아가는 걸 좋아합니다. 그래서 배우는 이론들의 증명에도 관심이 많은 편이고, 다양한 이론들의 작동 원리를 이해하고 있습니다. 예를 들어, 회귀분석의 여러 전제들이 회귀분석에서 왜 필요한지, Ridge regression이 설명변수들의 다중공선성을 어떤 방식으로 해결하는지 등을 이해하고 있습니다. 저는 이러한 기초들이 이론적인 측면에서 끝나지 않고 응용에도 도움을 주리라 생각합니다. 어떤 문제를 해결하는 데에는 여러가지 방법이 있습니다. 예를 들어, 설명변수들의 다중공선성을 해결하는 방법으로 PCA, Ridge regression 등이 있고 잔차의 독립성을 확인하는 방법으로 Run test, Durbin-Watson test, ACF plot 등이 있습니다. 각 방법마다 장단점이 있고, 그 장단점에 따라 가지고 있는 데이터에 어떤 분석법이나 진단법을 적용하는 것이 최적일지를 판단할 수 있습니다. 따라서 각 방법의 장단점을 아는 것은 중요하다고 생각하며, 각 방법에 대한 직관을 갖고 장단점을 제대로 알기 위해서는 그 이론의 증명과 작동 방식을 아는 것이 도움이 됩니다. 저는 이론들의 증명과 작동 방식을 이해할 수 있는 수준에서 공부해왔고, 앞으로 피셋에서 배울 새로운 이론들에 대해서도 그렇게 공부할 것입니다. 따라서 이러한 저의 학구열과 호기심은 학회에서 진행할 클린업과 주제분석, 공모전 모두에서 도움이 될 것이라고 생각합니다. |
| **Q2. 지원동기 및 학회활동을 통해 얻고 싶은 것을 서술해주세요.** (지원하게 된 동기와 학회원이 되어야 하는 이유를 세부적으로 서술해주세요.) |
| 저는 대학교 1학년 때부터 통계학과 선배들에게 피셋과 관련된 이야기를 줄곧 들어왔고, 데이터분석을 많이 배울 수 있다는 이야기에 항상 들어가고 싶어했습니다. 통계학을 이론적으론 깊이있게, 응용적으론 능수능란하게 다루고싶다는 욕심이 있었고 피셋은 그런 저의 욕심에 어울리는 학회였기 때문입니다. 그리고 피셋의 활동과 제가 공부하고 싶은 점이 부합했습니다. 학회에 들어간다는 것은 책임감을 갖고 선택해야 할 문제라고 생각하기 때문에 피셋에서 공유하는 깃헙에 들어가서 피셋의 활동 내용과 진행 방식이 저의 목표와 부합할지에 대해 살펴보았습니다. 저는 분석과 관련된 이론적인 지식을 쌓는 것에 그치지 않고 실제로 데이터를 가지고 분석을 하며 응용 실력을 기르고 싶었습니다. 이론적인 증명과 공부는 학교 수업과 개인 공부로 충분히 가능하다고 생각합니다. 하지만, 데이터 분석 실력과 관련해선 혼자하는 것보다 팀을 이루어 서로 피드백을 주며 공부하는 것이 효과적이라고 생각했습니다. 피셋에서는 “클린업”으로 3주간 이론 공부를 하고 중간고사 이후 4주간 팀별로 “주제분석”으로 데이터분석을 진행하는 것으로 알고 있습니다. 피셋에서 공유하는 깃헙에 들어가서 각종 분석 기록들을 본 결과, “지지 정당 예측”, “서울시 살 곳 추천” 등 다양한 주제를 선정해 분석해왔다는 것을 알 수 있었습니다. 또한, 팀을 이루어 다양한 데이터 공모전에도 참가하는 걸로 알고 있습니다. 이러한 피셋의 활동들은 제가 공부하고 경험하고 싶은 점과 정확히 일치했습니다. 피셋을 통해 다양한 데이터 분석 경험을 쌓고 싶습니다. 이론 위주의 학교 공부에 머물지 않고 실제 데이터를 가지고 팀을 이루어 피드백을 주고 받으며 분석을 해보고 싶습니다. 이러한 점들이 제가 피셋에 지원한 이유이며, 다른 학회가 아닌 피셋의 학회원이 되어야 하는 이유입니다. |
| **Q3. 자신의 장·단점을 소개하고, 단점을 극복하기 위해서 어떤 노력을 했는지 서술해주세요.** |
| 저의 가장 큰 장점은 책임감과 열정이라고 할 수 있습니다. 먼저, 저는 책임감이 뛰어납니다. 제가 선택하고 맡은 바에 대해서는 책임감을 가지고 최선을 다합니다. 이러한 책임감은 제가 지금껏 해왔던 활동들에서 두드러지게 나타났습니다. 제가 노래패 “아우성”에서 사무팀장으로 활동할 때에는 사무일의 특성상 노래패에 필요한 것들을 미리 준비해야했고 돈과 관련된 만큼 최대한 합리적으로 일을 처리해야 했습니다. 제대로 하지 않으면 노래패 활동에 직접적인 타격을 줄 수 있는 일이었기 때문에 계획적이고 책임감있는 일처리가 필요했습니다. 학기 시작 전에는 노래패의 크고 작은 행사들을 정리하고 각 행사별 마감 기한을 정해놓고 그 기한 안에 모든 일을 처리했고, 정기모임 진행에 필요한 사무물품으로 무엇이 부족한지 수시로 체크하고 의견을 받았으며, 매달 회비는 얼마정도 걷는 것이 합리적일지, MT, TR, 그리고 각종 노래 관련 행사들에 쓰이는 돈들에 대한 점들을 항상 사무팀원들과 의논했습니다. 1년 간의 활동인 만큼, 사무팀의 일에 문제가 생긴 경우도 있었지만, 그런 경우에는 나를 믿고 맡겨준 일인 만큼 책임지고 문제의 대안을 찾아야된다는 생각으로 노력했고, 결국 큰 문제없이 일이 진행이 될 수 있었습니다.  둘째로, 저는 제가 하는 활동들에 대해서 열정적으로 참여합니다. 이런 특성 역시 제가 해왔던 활동들에서 눈에 띄었고, 저와 같이 활동하는 사람들에게 ‘정말 열심히 한다.’, ‘성실하다’라는 평을 많이 받아왔습니다. 제가 사회복지인권학회 “새봄”에서 학회원으로 활동할 때에는 매주 평균 100쪽 분량의 커리를 읽어와야 했습니다. 저는 100쪽 분량을 아무런 비판이나 궁금증 없이 받아들이지 않고 시간이 더 많이 걸리더라도 읽으면서 스스로 생각하고 궁금한 점들을 정리해서 세미나때 질문했습니다. 적극적인 태도로 임하는 것이 저의 발전에도 큰 도움을 주고 학회 차원에서도 풍부한 의논을 가능하게 하는 일이라고 생각했기 때문입니다. 이러한 저의 장점은 노래패 “아우성”에서도 발휘가 되었습니다. “아우성”에서는 집행기수는 주마다 정기모임을 2회 가지고 1학년들은 정기모임을 1회 가집니다. 저는 “아우성”에서 더 열심히 활동하고 즐기고 싶다는 생각이 있었고, 그래서 1학년이지만 선배들과 함께 주 2회 정기모임을 하는 날이 많았습니다. 또한, 악기를 배우고 무대에 서는 것은 필수가 아닌 선택사항이었지만 열심히 활동하고 싶다는 생각으로 노래패 내에서 기타를 배우고 무대에도 섰습니다. 이처럼 제가 선택한 활동들에 대해 열정적으로 참여해왔고, 그 덕분에 많이 배우고 성장하고 즐길 수 있었을 뿐만 아니라 제가 속한 단체들에도 긍정적인 영향을 줄 수 있었다고 생각합니다.  저의 이러한 장점들은 피셋에서 활동을 할 때 빛을 보리라 생각합니다. 이러한 장점들과 학구열을 기반으로 저에게 주어진 과제들에 대해서는 책임감있게 처리할 것이며, 개개인에게 주어진 과제가 아닌 공동의 논의나 피드백을 주고받는 자리에서도 적극적이고 열정적으로 임할 것입니다.  반면에 저의 단점은 다른 사람의 기분을 지나치게 헤아린 나머지 다른 사람이 귀찮아 할 만한 일을 잘 부탁하지 못한다는 것입니다. 제가 “아우성”에서 사무팀장으로서 일을 할 때, 팀원들에게 어떤 일을 해달라고 부탁을 잘 못했었습니다. 많은 일을 팀원들에게 넘기고 싶지 않았고 그래서 많은 일들을 스스로 처리하려 했었습니다. 하지만 그러다보니 할 일이 너무 많아지고, 시간에 쫓겨 점점 지쳐갔습니다. 하지만 시간이 지나고, 제가 혼자 많은 일들을 하려고 할 때 마다 팀원들이 자기가 하겠다고 하면서 적극적으로 나서줘서 집행기수 활동의 중반부터는 공평하게 일을 나눠서 하게 되었습니다. 그때 전 훨씬 편해지기도 했고 전체적으로 좀더 활동을 열심히 꼼꼼히 할 수 있게 되었습니다. 그때, 이렇게 부탁하지 못하는 태도는 좋지 않다고 느꼈습니다. 물론 자기 일은 열심히 해야되지만 그렇다고 모든 일을 혼자서 떠맡을 필요는 없고, 그렇게 혼자서 많은 일을 떠맡아버리면 저 자신도 지칠 뿐만 아니라 효율이 떨어져서 전체적인 성과도 떨어진다는 것을 깨달았습니다. 또한, 그렇게 혼자 일을 하는 것은 다른 사람을 배려하는 게 아니라 오히려 배려하지 못하는 것일 수도 있음을 깨달았습니다. 이 점은 군대에서 많이 연습이 되었습니다. 군대 특성상 함께 해야 하는 일이 많고 분담해야 하는 일이 많은데, 계급이 오르면서 직접 후임들을 관리하고 일을 분담하는 주체가 되면서 이러한 점에서 많이 발전할 수 있었습니다.  다른 사람에게 부탁을 잘 하지 못하는 단점은 이러한 깨달음과 실천으로 상당부분 해결되었고, 분담해야 할 일에 대해서는 분담하여 일을 효과적으로 처리할 수 있게 되었습니다. |
| **Q4. 학회원이 되어 해보고 싶은 실제 데이터분석에 대해서 설명해주세요.** (주제 및 데이터수집, 분석방법에 대해 구체적으로 설명해주세요.) |
| 데이터분석의 주제로는 공공기관이 들어서야 할 위치나 화재나 범죄 등 위험성에 대한 예측, 아니면 어느 사회경제적 현상에 대해 설명변수를 설정해서 설명하는 것 등 다양한 주제를 생각해볼 수 있을 것입니다. 저는 이들 중 범죄와 관련된 데이터분석을 해보고 싶습니다. 범죄는 사람들의 일상에 치명적인 영향을 미칠 수 있으므로 중요한 사회적 현상이라고 할 수 있습니다. 서로 다른 범죄 유형이 어떤 설명변수들과 관련이 있는지를 밝혀낸다면 어떤 점을 보완해서 관심있는 범죄 유형의 발생률을 줄일 수 있을지 생각해볼 수 있을 것입니다.  저는 학회원이 되어 절도, 폭력 등 각 범죄 유형들의 발생 건수가 가로등의 개수, CCTV 개수, 해당 지역의 산업 등 다양한 설명변수들을 통해 어떤 식으로 설명이 되는지를 보고 싶습니다. 데이터는 “공공데이터포털”에서 얻을 수 있었습니다. “공공데이터포털”에 검색해본 결과, 서울특별시 자치구 별 범죄들의 발생건수를 찾을 수 있었고, 각 자치구 별 가로등 현황, CCTV 현황을 확인할 수 있었습니다. 하지만 가로등과 CCTV 개수를 숫자 그대로 설명변수의 값으로 넣기에는 무리가 있다고 생각합니다. 왜냐하면 가로등과 CCTV는 공간과 관련된 변수이기 때문입니다. 그래서 각 지역구별 면적으로 나눠서 면적별 가로등과 CCTV의 개수를 사용하는 것이 바람직하다고 생각합니다. 세번째 설명변수인 해당 지역의 산업과 범죄 유형들의 상관관계를 보기 위해서는 해당 지역의 산업을 대표할 수 있는 지표들을 선택해야 합니다. 무슨 지표를 선택해야 할지에 대해서는 실제로 이 주제로 분석을 할 때 학회원들과 자세히 의논을 해봐야겠지만, 저의 의견으로는 지역구 내 각 산업 분야들의 지역내 총부가가치를 모든 산업의 지역내 총부가가치 합으로 나눈 비율을 사용할 수 있을 것 같습니다. 각 자치구의 산업 분야별 지역내 총부가가치 데이터 역시 “공공데이터포털”에서 찾을 수 있었습니다.  위의 주제로 데이터 분석을 할 땐 반응변수와 설명변수가 모두 양적 변수인 회귀분석 모델을 설계할 수 있습니다. 여기서 반응변수는 각 범죄 유형들의 발생 건수로, 관심있는 범죄 유형의 개수만큼 변수가 많아지므로, 각 범죄 유형별로 따로 회귀분석을 진행할 수 있을 것 같습니다. 그런데 각 지역구의 산업 분야가 주어진 데이터에 16가지로, 가로등과 CCTV 변수까지 추가하면 설명변수가 총 18개가 되어 지나치게 많다고 생각합니다. 그래서 단계 선택법을 통해서 먼저 BIC가 작은 변수들을 선택해야 합니다. 이렇게 중요한 변수들을 뽑아낸 후엔, 회귀분석의 정확성을 보장하기 위해 회귀분석의 가정을 변수들이 따르는지를 확인해야 합니다. 먼저, 설명변수와 반응변수가 선형관계를 따르는지를 확인하기 위해 산점도를 확인합니다. 선형관계에서 벗어나있다고 생각될 경우 반응변수에 log를 취해서 선형관계에 가까워지도록 조정해줄 수 있습니다. 그 후, 오차에 대한 가정을 확인해야 합니다. 오차가 정규성을 따지는지를 확인하기 위해 Shapiro-Wilk test를 실시하고, 오차의 독립성을 확인하기 위해 ACF plot을 그려서 확인할 수 있습니다. 또한, Cook’s distance를 통해 극단적인 값을 가지는 데이터를 제거해줘야 합니다. 이렇게 회귀분석의 가정들을 변수들이 따르도록 조정한 후에는 마지막으로 설명변수들끼리 다중공선성이 발생하는지를 확인해줘야 합니다. 특히, 각 지역구의 산업 분야별 지역내 총부가가치 비율끼리는 다중공선성이 발생할 가능성이 클 수 있습니다. 그래서 먼저 설명변수들끼리 다중공선성이 발생하는지를 VIF를 통해 확인하고, 발생했다면 OLS가 아닌 Ridge regression을 통해 회귀분석을 실시해주면 됩니다.  하지만 위 방법과 주제에는 한계가 있었습니다. 첫째로, 지역구별 CCTV 개수, 경제활동 지표는 2018년 데이터였지만 가로등의 개수 데이터는 2021년 데이터밖에 찾을 수 없었습니다. 데이터를 수집한 날짜가 다르다면 회귀분석의 정확성이 떨어질 수 있습니다. 둘째로, 데이터의 수가 너무 적습니다. 이는 서울특별시로 제한하지 않고 지역의 수를 충분히 늘림으로써 해결할 수 있을 것입니다. 이러한 한계점들을 어떤 방식으로 해결해나갈지에 대해서도 피셋에 들어가서 배워가고 싶습니다. |
| **Q5. 본인이 희망하는 팀과 이유를 설명해주세요.** (복수선택이 가능하며, 참고사항으로만 활용될 뿐, **희망하는 팀으로 배정되는 것을 의미하지 않습니다**.) |
| 제가 가장 희망하는 팀은 데이터마이닝 팀입니다. 저는 기존의 데이터들 간에 관계를 찾아내어 새로운 데이터가 들어왔을 때 그 데이터를 분류하고 특징을 예측하는 일에 큰 매력을 느꼈습니다. 기존의 데이터를 분석하고 새로운 데이터를 예측하는 일은 지금까지 알 수 없었던 미지의 현상에 대해서 수학적으로 그 특징을 밝혀줄 뿐만 아니라, 굉장히 다양한 분야에서 활용될 수 있기 때문입니다. 실제로 넷플릭스의 영화, 드라마 추천 시스템이 데이터마이닝을 사용하는 것으로 알고 있고, 비즈니스적인 활용 외에도 다양한 분야에서 무궁무진하게 활용할 수 있습니다. 예를 들어, 사회복지적으로는 복지 사각지대에 있는 사람들을 찾아내어 도움을 주는 일을 할 수도 있고, 개인을 대상으로 해서는 같은 기질을 공유한 사람들이 어떤 방식으로 행복을 느끼는지를 분석해서 각 개인의 특성별로 행복을 위한 솔루션을 줄 수도 있을 것입니다. 저는 이처럼 기존에 해결하기 힘들었던 다양한 문제들에 대해서 해결책을 제시하거나, 새로운 데이터가 들어왔을 때 그 데이터의 특징들을 예측할 수 있게 해주는 데이터마이닝에 대해 깊이 공부하고 활용해보고 싶습니다. 그래서 데이터마이닝 팀을 희망합니다. |
| **참고사항** (해당과목 중에 수강하신 과목이 있다면 체크해 주세요) |
| ◦ 통계학원론(O) ◦ 행렬대수학/선형대수학(O) ◦ 회귀분석(O) ◦ 범주형자료분석( ) ◦ 시계열분석( ) ◦ 데이터마이닝( ) ◦딥러닝( )  기타 데이터분석 관련 과목을 적어주세요.   1. 통계프로그래밍입문 |
| **가능한 면접 시간을 말해주세요 [2월 27일(토) / 2월 28(일) 12~20시 中]**  **면접은 온라인으로 진행될 예정입니다**  토, 일 둘 다 전체 가능.  \* 지원 마감 후, 확정된 면접시간은 추후 문자(SMS)로 공지하겠습니다. |

본 지원서는 성균관대 통계분석학회 P-SAT **리크루팅 이외의 목적으로 사용되지 않음**을 밝힙니다.

본 지원서에 기재한 사실이 허위나 누락이 없음을 확인합니다.

2021년 2월 23일, 지원자 : 이정우 (인)

**<시간표 첨부>**

* **추가적으로 수업 중 실시간 및 오프라인 수업이 있는 경우 표기해주시면 감사합니다.**

****

온라인 사전 녹화 : 해석학1, 실험계획법, 통계적데이터마이닝, 생명의료윤리

온-오프 혼합 : 바이오통계입문

일부 실시간 강의 : 없음

오프라인 강의 : 없음