1. 높은 목표설정

자발적으로 최고 수준의 목표를 세우고 끈질기게 성취한 경험에 대해 서술해 주십시오. (700자 10단락 이내)

(본인이 설정한 목표/목표의 수립과정/처음에 생각했던 목표 달성 가능성/수행 과정에서 부딪힌 장애물 및 그 때의 감정(생각)/목표 달성을 위한 구체적 노력/실제 결과/경험의 진실성을 증명할 수 있는 근거가 잘 드러나도록 기술)

데이터 분석

실제 데이터를 활용한 해커톤 참가 수료

목표가 무엇이 였을까?

데이터 과학자를 위한 1000시간의 마라톤

---한

최고 수준의 목표 : 데이터 분석 경진대회에서 상위 5%안에 들기

수립과정 : 데이터 분석 전문가가 되기 위한 1000시간의 교육과정 이행 중 학우들과 프로젝트 경험을 쌓기 위해 경진대회에 참가

달성 가능성 : 데이터 분석의 기초단계에서 수상권을 노리기보단 참가에 의의를 두자는 분위기

수행 중 장애물과 감정 : 수학적 지식, 하드웨어 성능의 부재

달성을 위한 구체적 노력 : aws를 활용한 하드웨어 성능 해결, 매일 점심시간이 끝나고 추가적인 데이터 과학 통계 책을 활용한 예/복습, 다른 개발자의 코드리뷰를 통한 간접경험 쌓기

실제 결과 : 다들 포기한 상태에서 제출 이틀을 남겨놓고 상위 10%대에 진입

경험의 진실성 증명 근거 : 끝까지 포기하지 않고 도전한 결과 처음 참가한 데이터 경진대회에서 1030팀중 26위를 달성

그 중 데이터 분석역량을 확인하기 위해 나간 첫 데이터 분석 경진대회에서 1030팀 중 26위를 달성하였습니다.

가슴에 새긴 ‘Nothing is impossible’

어떤 일이든 한계를 생각하는 순간이 발전을 멈추게 한다고 생각합니다. 가능하다는 생각과 믿음이 실제로 가능케 하는 행동의 밑거름이 되는 것을 경험하며 살아왔습니다.

‘DACON’의 원자력발전소 상태판단 대회에 출전한 적이 있습니다. 원자력 상태를 측정하는 5,120개의 센서 데이터를 학습해 상태 변화를 예측하는 알고리즘을 구축하는 대회였습니다. 팀원들 모두 처음 경험하는 대회여서 중간 등수만 해도 성공이라는 말들이 오고 갔지만 큰 꿈일수록 깨지는 조각도 크듯 상위 5% 안에 들자는 목표를 세웠습니다.

분석 처음부터 24gigabyte가 넘는 데이터 크기로 인해 로컬 환경에 불러오기조차 하지 못하였습니다. 저는 이 문제를 해결하기 위해 분석 가능한 환경을 찾던 중 가상환경인 AWS의 클라우드 서비스를 통해 해결 가능하단 것을 알게 되어 분석을 시작했습니다.

그리고 5,120개의 column이 오히려 모델의 정확도를 떨어뜨리는 점을 발견하여 통계지식을 활용한 feature engineering을 통해 정확도를 올리고자 매일 점심 직후 40분간 ‘데이터 과학을 위한 통계’ 책을 가지고 팀원들과 학습을 진행하여 분석 역량과 모델의 정확도를 함께 발전시켜나갔습니다.

그 결과 첫 대회에서 1,030팀 중 26위를 달성하여 목표보다 더 높은 성과를 낸 뜻깊은 경험이었습니다.

모델의 정확도를 향상시키기 위해 기존 학습내용으로 부족하여 매일 점심시간 직후를 활용하여 데이터 과학자를 위한 통계책을 매일 14page씩

으로는 더 발전시키지 못하여 다른 개발자들의 코드리뷰를 통해 저의 데이터에 맞게 변환하는 과정과 부족한

다른 개발자들의 코드리뷰와 개인적인 데이터 통계지식 예복습을 통한 상위2.6%

1. 기존의 틀을 깨는 과감한 실행

새로운 것을 접목하거나 남다른 아이디어를 통해 문제를 개선했던 경험에 대해 서술해 주십시오. (700자 10단락 이내)

(기존 방식과 본인이 시도한 방식의 차이 / 새로운 시도를 하게 된 계기 / 새로운 시도를 했을 때 주변 반응 / 새로운 시도를 위해 감수해야 했던 점 / 구체적인 실행 과정 및 결과 / 경험의 진실성을 증명할 수 있는 근거가 잘 드러나도록 기술)

22시간 안에 문제해결과 발표자료 작성을 위한 전략을 새우기로 하였습니다.

산업 AI 해커톤 ‘Factory HACK’에 참가한 적이 있습니다. 산업데이터에 관심이 많은 후배와 참가하여 실제 제조환경에서 나온 데이터를 가지고 문제 해결을 위한 알고리즘을 22시간 안에 개발하는 해커톤이었습니다.

설비 센서 데이터를 활용해 생산품의 양/불량품을 예측하고 예측정확도를 시간에 따라 보전하는 알고리즘을 개발하는 것이었습니다. 하지만 실제 설비 데이터라 불량품이 0.6%밖에 없어 기존 방법으로는 모델을 학습시키기에 부적절하였습니다. 게다가 보안으로 인해 데이터 익명화가 되어 있어 산업공학도로서 도메인 지식을 활용할 수 없었습니다.

그러다 EDA를 진행하던 중 데이터가 생산주기로 일정하게 패턴을 보이는 것을 발견하였습니다. 그래서 대회 참가 전 학습한 CNN 알고리즘을 적용해 보기로 하였습니다. 원래 이미지, 주파수 데이터 학습에 최적화된 알고리즘이지만 센서데이터를 합치면 패턴을 시각화하여 문제를 해결할 수도 있다는 생각이 문뜩 들었습니다. 다른 팀 대부분 업샘플링을 통한 데이터 모델을 구축하였는데 체리픽킹이라는 생각이 들어 실제 제조환경에 맞지 않는다고 생각했기 때문입니다.

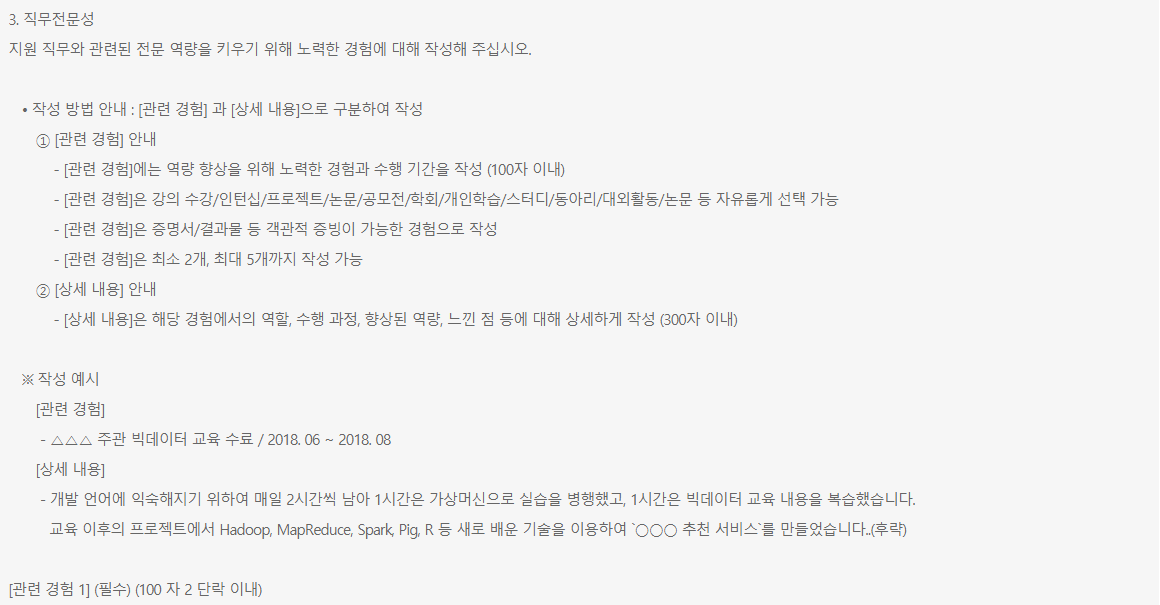
그리하여 생산 cycle time에 맞춰 모든 데이터의 시각화를 통해 새로운 방법론으로 접근할 수 있다는 결과를 발표하게 되었습니다. 비록 수상은 못 하였지만 참신한 방법론으로 아이디어 칭찬을 받은 도전적인 경험이었습니다.

새로운 것

어떤 프로젝트에서 새로운 것을 시도했는지

주변 반응, 감수해야 했던 비용

등



1. 관련 경험1(필수 100자 2단락 이내)

- 4차 산업혁명 선도인력 양성과정

- 멀티캠퍼스 주관 서비스 산업 데이터를 활용한 머신러닝 분석 1000시간 교육과정 수료 / 2019.10 ~ 2020.04

1. 상세 내용1(필수 300자 7단락 이내)

데이터 분석전문가가 되기위한 파이썬, R을 활용한 탐색적 데이터분석기술 부터 머신러닝/딥러닝에 필요한 고급기술 교육을 학습하였습니다.

기초적인 전처리에 필요한 프로그래밍과 SQL, Flask Web framework, Javascript 학습을 하였고 project와 실무에 필요한 Github 특강을 통해 Git 사용법을 익혔습니다.

교육과정 마지막 Final Project를 통해 `주 52시간 근무제 웹 반응 감성분석 알고리즘`을 개발하였습니다.

1. 관련 경험2(필수 100자 2단락 이내)

- DACON 데이터 분석 경진대회 2회 참가

- 1. 원자력발전소 상태판단 대회 ~2020.02.12 / 2. 월간데이콘 2 천체 유형 분류 대회 ~2020.02.29

1. 상세 내용2(필수 300자 7단락 이내)

데이터 분석가로서 역량을 확인해보기 위해 팀원들과 함께 데이터 분석 경진대회 참가하였습니다.

학습용 Data set을 활용해 모델을 구축하여 Test set을 예측하여 Log loss 측정지표를 활용한 competition.

딥러닝 모델과 Lightgbm 모델을 활용해 예측 정확도가 높은 모델을 구축하기 위한 수많은 시도 끝에 참가한 두 대회 모두 상위 5% 안에 드는 성과를 달성하였습니다.

1. 관련 경험 3

- 데이터 과학자가 되기위한 필수 역량 스터디 / 2019.12 ~ 2020.03

1. 상세 경험 3

머신러닝 교육과정 중 데이터 분석가에 관심이 많은 학우들과 함께 스터디를 진행하였습니다.

매일 점심시간을 활용하여 데이터 과학에 필요한 통계지식을 책을 통해 단원별 학습 후 다른 학우들에게 당일 학습한 개념에 관해 설명하는 형식으로 진행하였습니다.

매주 화요일 개발자 및 코딩테스트에 필요한 역량을 쌓기 위해 주제를 선정하여 해당 내용을 강의하는 형태로 알고리즘 스터디도 하였습니다.

- 한국과학창의재단 주관 찾아가는 기술교육프로그램 활동 / 2017.10~2017.12

- 창의융합형 교육법 수료 / 초, 중학교 학급에 배치되어 4차산업기술 교육 재능기부

어린 학생들에게 재능기부 교육을 해주기 위해 교육법 및 4차산업 기술에 대한 기초 교육부터 실습에 대해 강의를 들었습니다.

교육을 해주기 위한 교안을 팀원들과 함께 만들고 학급당 2시간씩 교육 커리큘럼을 구성하여 수차례 예행연습을 통해 포항초등학교 6학년 학생들에게 교육 진행

어린 학생들에게 새로운 꿈과 희망을 심어준 활동으로 보람차고 재미있는 활동이었습니다. 그리고 4차산업 기술에 대한 넓은 안목을 가질 수 있었습니다.