

입사지원서

각오 한마디	기본에 충실한 자세로 꾸준한 성과를 내겠습니다.				
지원분야	데이터 분석	지원구분	신입	희망연봉	회사 내규에 따름

☐ 기본정보

	성명	권용진	성별	남
	생년월일	2001. 05. 02	나이	26 세(만 25 세)
	주소	강원도 횡성군 횡성읍 북천서로 19 번길 9		
	휴대번호	010-2380-4683	E-Mail	qxas4550@naver.com
	병역사항	병역必(2021.05.31~2022.11.30)	취미	풋살, 야구(관람), 게임, 헬스
포트폴리오		https://www.notion.so/YJ_-Dashboard-a3cba03e110049dcb987316f4fcc4f5b		

☐ 학력사항

기간	출신학교	전공	학위	학점
2020.03 ~ 2026.02	수원대학교	데이터 과학	학사	4.1/4.5
2017.03 ~ 2020.02	횡성고등학교	이과	-	-

☐ 보유기술

분야	보유기술	역량 소개
데이터베이스	SQL	MIMIC-IV 데이터에서 환자·입원·진단·ICU 테이블을 추출·결합하고 분석용 데이터셋 구성
프로그래밍	Python	pandas·NumPy 기반 데이터 전처리, 머신러닝 모델링, 웹 크롤링 및 시각화 수행
통계 분석	R	다중회귀분석, 상관관계 분석, 통계적 유의성 검정 및 지도 시각화 수행
공간 데이터 분석	QGIS	격자 생성, 좌표계 통일, 공간 조인 및 머신러닝용 공간 데이터셋 구축
데이터 시각화	Tableau	지역별 주요 지표와 분석 결과를 지도·차트 형태로 시각화
데이터 시각화	Power BI	이커머스 퍼널, 전환율, 매출·마진, 고객 세그먼트 성과 대시보드 구축
업무 자동화	Make	사용자 입력 기반 결과 생성·저장·전달 자동화 워크플로우 구축

☐ 교육이수내역

연수기간	연수기관	연수과정 및 내용
2025.08.26 ~ 2026.02.11 (6개월/920시간)	아이티윌	프로젝트 중심 빅데이터 융합 머신러닝 전문가 양성과정

□ 자격사항

취득일	종류 및 등급	발행처
2024.03.22	ADsP	한국데이터산업진흥원
2024.09.20	SQLD	한국데이터산업진흥원
2025.08.19	컴활 1 급	대한상공회의소
2026.06.12	정보처리기사	한국산업인력공단
2026.07.10	빅데이터분석기사	한국데이터산업진흥원
2026.06.30	Google Analytics Certification	Google

□ 경력사항

회사명	기간	소속부서	담당업무	퇴사사유

□ 수상내역

행사명	시상일	팀명 / 제목	상세 설명
제 2 회 AI/빅데이터 분석 경진대회 - 화성시연구원 주관	최우수상 2025/11/20	유방암 환자의 예후 예측을 위한 병리 이미지와 전사체 기반 멀티모달 딥러닝 모델	병리이미지(WSI)와 전사체(Transcriptome)를 통합한 멀티모달 딥러닝 기반 유방암 예후 예측 모델(VA-GATSurv)을 제안하여, 단일 모달 대비 생존 예측 성능(C-index)을 유의미하게 향상시킨 연구

□ 아르바이트 및 기타활동

활동기간	소속명	담당	상세 설명
2024.05 ~ 2024.11	DNA(학부 내 동아리)	C2 팀원	다양한 데이터셋을 다뤄보며 자신이 진출하고자 하는 분야 탐색 및 SQL, Tableau 같은 데이터 언어, BI 툴 학습 및 실습 (관련 블로그: https://memo-52.tistory.com)

□ 참여프로젝트

개발기간	구분	상세설명
24.03.15 ~ 24.06.05 (약 3 개월)	프로젝트명	▶ MIMIC-IV 데이터 분석을 통한 인사이트 도출 및 사망률 예측 / 5 명(팀프로젝트)
	개발개요	- MIMIC-IV 중환자 데이터를 활용해 환자의 사망 위험 요인을 분석하고, 의료 자원 배분과 환자 관리 의사결정을 지원할 수 있는 사망률 예측 모델을 구축했습니다.
	구현기술	· SQL 기반 환자·입원·진단·ICU 테이블 추출 및 결합 · Python 을 활용한 데이터 전처리 및 Feature Encoding · Logistic Regression, Random Forest, LightGBM 등 모델 비교

		· 교차검증과 ROC-AUC 기반 모델 성능 평가 · Voting Classifier 기반 최종 예측 모델 구축
	담당업무	· 데이터 수집 및 전처리 · 머신러닝 모델링 및 시각화 · 분석 결과 정리 및 PPT 제작
24.06.17 ~ 24.08.30 (약 2 개월)	프로젝트명	▶ 경제 수준 차이에 따른 도서관 이에 미치는 영향 분석 / 4 명(팀프로젝트)
	개발개요	- 지역별 경제 수준과 도서관 운영 환경이 회원등록자 수에 미치는 영향을 분석하고, 지역 간 도서관 이용 격차를 완화하기 위한 정책 방향을 제안했습니다.
	구현기술	· 국가도서관통계시스템, KOSIS, 복지포 등 공공데이터 수집 · Python·Excel 기반 데이터 정제 및 지역 단위 통합 · R 기반 다중회귀분석 및 통계적 유의성 검정 · 회귀계수, p-value, Adjusted R² 기반 결과 해석 · 지역별 주요 지표 지도 시각화
	담당업무	· 공공데이터 수집 및 전처리 · 회귀분석 및 시각화 · 결과 해석 및 보고서 작성
24.08. ~ 24.10. (약 2 개월)	프로젝트명	▶ 화성시 노인 교통사고 위험 지역 분석 및 예방 정책 제안 / 4 명(팀프로젝트)
	개발개요	- 화성시 지역별 노인 교통사고 영향 요인을 분석하고, 사고 위험지역을 예측해 노인보호구역 확대와 단속카메라 재배치 등 예방 정책을 제안했습니다.
	구현기술	· Python·R 기반 데이터 정제 및 탐색적 분석 · 후진제거법을 활용한 유의 변수 선별 · Random Forest 기반 변수 중요도 분석 · XGBoost, Lasso, Ridge 모델 비교 및 5-Fold 교차검증 · 계층적 군집분석 및 Tableau 지도 시각화
	담당업무	· 데이터 수집 및 전처리 · 예측 모델링 및 군집분석 · 분석 결과 시각화 및 포스터 제작
24.09. ~ 24.12. (약 3 개월)	프로젝트명	▶ 지역별 생활자원순환지수 개발 및 군집 분석 / 5 명(팀프로젝트)
	개발개요	- 전국 시·군·구의 재활용 효율과 폐기물 처리 특성을 종합적으로 평가할 수 있는 생활자원순환지수를 개발하고, 군집별 특성에 맞는 개선 방향을 제안했습니다.

	구현기술	· 2019~2022 년 시·군·구 단위 공공데이터 통합 · EDA 및 Pearson 상관계수 기반 주요 변수 선정 · 가중합 및 정규화를 통한 생활자원순환지수 개발 · Elbow Method 기반 최적 군집 수 결정 · 계층적 군집분석 및 Tableau 지도 시각화
	담당업무	· 공공데이터 수집 및 전처리 · 지수 개발 및 군집 분석 수행 · 분석 결과 해석 및 시각화 · 정책 제안 및 결과 정리
24.09. ~ 24.12. (약 3 개월)	프로젝트명	▶ AI 해외여행 플래너 서비스 기획 및 자동화 구현 / 5 명(팀프로젝트)
	개발개요	- 사용자의 여행 목적, 일정, 예산, 선호도를 기반으로 맞춤형 해외여행 일정을 생성하고 저장·전달하는 AI 자동화 서비스를 구현했습니다.
	구현기술	· Google Form 을 활용한 사용자 여행 설문 입력 수집 · Google Sheets 를 통한 설문 응답 데이터 저장 및 관리 · ChatGPT API(OpenAI)를 활용한 사용자 맞춤형 여행 일정 자동 생성 · Make 노코드 자동화 플랫폼을 활용한 전체 워크플로우 구성 · Discord 및 Gmail 을 통한 여행 계획 결과 자동 전달 · Notion Database 에 여행 계획 데이터 저장 및 관리
	담당업무	· 서비스 기획 및 자동화 시나리오 설계 · Make 모듈 연동 및 워크플로우 구현 · OpenAI API 프롬프트 설계 · 결과 전달 및 데이터 저장 구조 설계 · 프로젝트 결과 정리 및 PPT 제작

25.03. ~ 25.06. (약 3 개월)	프로젝트명	▶ 유방암 환자 예후 예측을 위한 병리 이미지·전사체 기반 멀티모달 딥러닝 모델 / 3 명(팀프로젝트)
	개발개요	- 병리 이미지와 전사체 데이터를 결합한 멀티모달 딥러닝 모델을 구축하여 유방암 환자의 생존 위험을 예측하고, 단일 데이터 기반 모델의 한계를 개선했습니다.
	구현기술	· TCGA-BRCA 병리 이미지·전사체·생존 데이터 활용 · WSI 패치 분할 및 GAT 기반 이미지 표현 학습 · VAE Encoder 기반 전사체 잠재 표현 생성 · Cross-Attention 기반 모달리티 정보 융합 · Contrastive Loss 및 CoxPH Loss 기반 생존 분석
	담당업무	· 전사체 표현 학습을 위한 VAE 기반 선행 논문 스터디 및 모델 설계 방향 정리 · VAE Encoder 구조 구현 및 전사체 잠재 표현 생성 · 생성된 전사체 표현을 멀티모달 모델 입력으로 연계 · 연구 결과 정리 및 포스터 제작

25.10.13 ~ 25.10.22 (10 일)	프로젝트명	▶ 무신사 상품 데이터 자동 수집 및 이미지 포함 엑셀 보고서 생성 크롤러 개발 / 개인 프로젝트
	개발개요	- 사용자가 선택한 무신사 카테고리의 상품 정보와 이미지를 자동으로 수집하고, 서식이 적용된 Excel 보고서로 생성하는 크롤러를 개발했습니다.
	구현기술	· Selenium 기반 로그인 및 브라우저 자동화 · BeautifulSoup 기반 상품 정보 파싱 · urllib 을 활용한 상품 이미지 다운로드 · pandas 기반 CSV·XLSX 파일 생성 · openpyxl·Pillow 기반 이미지 삽입 및 엑셀 서식 자동화
	담당업무	· Selenium 기반 로그인 및 카테고리 선택 자동화 구현 · 상품 데이터 및 이미지 크롤링 로직 설계 및 구현 · CSV 및 XLSX 파일 저장 로직 구현 · Excel 이미지 삽입 및 서식 자동 조정 기능 개발
25.10.13 ~ 25.10.22 (10 일)	프로젝트명	▶ FotMob 리그·시즌·카테고리 기반 축구 데이터 자동 수집 크롤러 개발 / 개인 프로젝트
	개발개요	- FotMob 에서 리그, 시즌, 카테고리, 세부 섹션을 사용자가 선택하고 경기·순위·선수·팀·뉴스 데이터를 수집할 수 있는 크롤러를 개발했습니다.
	구현기술	· Selenium 기반 브라우저 자동 제어 · WebDriverWait 기반 DOM 상태 감지 · 사용자 입력 기반 리그·시즌·카테고리 선택 · 경기, 순위, 선수·팀 통계 및 뉴스 데이터 파싱 · 섹션별 Excel 파일 자동 생성
	담당업무	· FotMob 크롤러 전체 구조 설계 및 구현 · 리그·시즌·카테고리·섹션 선택 인터페이스 로직 구현 · DOM 기반 섹션 감지 및 데이터 수집 함수 개발 · 경기, 순위, 선수·팀 통계, 뉴스 데이터 파싱 및 저장 · 크롤링 결과 엑셀 파일 자동 생성 및 관리

25.11.17 ~ 25.12.19 (약 1 개월)	프로젝트명	▶ 머신러닝을 활용한 경기도 싱크홀 위험도 예측 및 위험지도 구축 / 6 명(팀프로젝트)
	개발개요	- 경기도 전역을 250m×250m 격자로 세분화하고 지반·지질·인구·교통·건축·하수도 데이터를 통합하여 싱크홀 위험도를 예측했습니다. 예측 결과는 공간 위험지도와 반응형 대시보드로 구현하여 위험지역을 직관적으로 확인할 수 있도록 구성했습니다.

	구현기술	· 싱크홀 발생 이력과 인구·지반·지질·교통·건축물·하수도 관련 공공데이터 수집 · QGIS 기반 250m×250m 격자 생성 및 좌표계(EPSC:5181) 통일 · Spatial Join 을 활용한 격자 단위 통합 데이터셋 구축 · 47 개 독립변수를 인구, 지반·지질, 교통·지하구조물, 건물, 하수도 인프라 등 5 개 범주로 구성 · 상관계수와 VIF 를 활용한 다중공선성 검토 · SHAP Feature Importance 와 PR-AUC 변화 기반 단계적 변수 선택 · Balanced Random Forest 와 Logistic Regression 모델 비교 · 클래스 불균형을 고려한 Recall·PR-AUC 중심 성능 평가 · Streamlit 기반 반응형 싱크홀 위험도 대시보드 및 로드뷰 기능 구현
	담당업무	· 공공기관 싱크홀 관련 데이터 수집 및 전처리 · QGIS 기반 격자 생성 및 공간 조인 수행 · 머신러닝 모델 비교·평가 및 결과 해석 · ppt 제작 및 발표
26.01. ~ 26.02. (약 1 개월)	프로젝트명	▶ Willpang Growth Analytics: AARRR 기반 이커머스 성장 전략 설계 및 시뮬레이션 / 5명(팀프로젝트)
	개발개요	- 가상 이커머스 플랫폼 Willpang 데이터를 기반으로 AARRR 프레임워크를 적용해 고객 유입부터 구매·재구매까지의 성장 구조와 퍼널 병목을 분석했습니다. - Power BI 로 단계별 KPI 와 구매 퍼널을 시각화하고, 트래픽·전환율·리텐션 개선에 따른 매출 및 마진 변화를 시뮬레이션했습니다. 또한 분석 결과를 활용할 수 있도록 RAG 기반 사내 데이터 분석 챗봇 ‘Willbot’을 구축했습니다.
	구현기술	· Mockaroo 기반 고객·세션·이벤트·주문·주문상품 이커머스 데이터셋 설계 · 원가·마진 컬럼 설계 및 Gross Margin 계산 구조 구현 · Python(Pandas, NumPy) 기반 AARRR 단계별 KPI 정의 및 구매 퍼널 구성 · Activation 전환율과 병목 구간, Retention 및 고객 매출 집중도 분석 · TF-IDF 기반 리뷰 감성 분석을 통한 정성 데이터 정량화 · Traffic·Funnel·Retention 개선 시나리오별 매출 및 마진 시뮬레이션 · Power BI 기반 AARRR 5 단계 대시보드 구축 · RAG 기반 사내 데이터 분석 챗봇 ‘Willbot’ 구현
	담당업무	· Mockaroo 를 활용한 이벤트 기반 이커머스 데이터 직접 설계 및 생성 · AARRR 분석 구조 설계 및 단계별 KPI 정의 · Power BI 대시보드 설계 및 시각화 · ppt 제작 및 발표

자기소개서

지원동기 및 포부

데이터분석 직무를 선택하게 된 이유는 데이터를 통해 막연한 현상을 구체적인 문제로 정의하고, 그 원인을 근거 있게 설명하는 과정에 매력을 느꼈기 때문입니다. 처음에는 스포츠 기록과 경기 데이터를 보며 숫자로 흐름을 해석하는 것에 관심을 가졌고, 이를 체계적으로 배우기 위해 데이터과학을 전공하게 되었습니다. 이후 전공 수업과 프로젝트를 경험하며 데이터 분석은 단순히 데이터를 정리하거나 모델을 만드는 일이 아니라, 문제를 구조화하고 의사결정에 필요한 방향을 제시하는 일이라는 점을 깨닫게 되었습니다.

이러한 관점이 가장 구체화된 경험은 Willpang 이커머스 성장 분석 프로젝트였습니다. 해당 프로젝트에서는 단순히 매출이 높고 낮음을 확인하는 것이 아니라, 고객의 유입부터 구매, 재구매까지 이어지는 성장 구조를 분석하는 것을 목표로 삼았습니다. 이를 위해 고객, 세션, 이벤트, 주문 데이터를 직접 설계하고, AARRR 프레임워크를 기준으로 Acquisition, Activation, Retention, Revenue 단계별 지표를 정의했습니다. 이후 퍼널 전환율, 재구매율, 매출 집중도를 분석하며 장바구니 이후 결제 단계와 재구매 전환 구조에서 병목이 발생하고 있음을 확인했습니다.

분석 결과를 단순히 보고서로 정리하는 데 그치지 않고, 2026 년 Baseline 과 개선 시나리오를 비교하는 성장 시뮬레이션을 수행했습니다. 이를 통해 단순히 유입을 늘리는 전략보다 퍼널 전환율 개선과 리텐션 강화가 더 큰 마진 상승 효과를 만들 수 있음을 수치로 확인했습니다. 또한 Power BI 대시보드와 RAG 기반 분석 챗봇을 구현하며, 분석 결과가 일회성 결과물이 아니라 운영 의사결정에 반복적으로 활용될 수 있는 구조로 확장될 수 있음을 경험했습니다.

이 경험을 통해 데이터분석 직무가 제게 잘 맞다고 느꼈습니다. 저는 단순히 정해진 데이터를 처리하는 것보다, 데이터 안에서 문제의 흐름을 파악하고, 어떤 지표를 봐야 하는지 정의하며, 분석 결과를 실제 개선 방향으로 연결하는 과정에서 가장 큰 흥미와 보람을 느꼈습니다. 또한 의료, 공공, 공간 데이터 프로젝트를 수행하며 도메인에 따라 문제 정의와 평가 기준이 달라져야 한다는 점도 배웠습니다. MIMIC-IV 사망률 예측 프로젝트에서는 환자 단위 데이터를 구성하고 예측 모델을 구축했으며, 싱크홀 위험도 예측 프로젝트에서는 단순 정확도보다 실제 위험지역을 놓치지 않는 평가 기준이 중요하다는 점을 경험했습니다.

입사 후에는 Python, SQL, R 을 활용한 데이터 전처리와 분석 역량을 바탕으로 실무 데이터에 빠르게 적응하겠습니다. 단기적으로는 주어진 데이터를 정확하게 이해하고, 분석 목적에 맞는 지표를 설정하며, 결과를 실무자가 이해할 수 있는 형태로 전달하는 분석가가 되고 싶습니다. 장기적으로는 반복적으로 활용 가능한 분석 체계와 대시보드를 구축하고, 데이터 기반 의사결정이 조직 안에서 자연스럽게 이루어질 수 있도록 기여하는 데이터 분석가로 성장하겠습니다.

직무역량 성장과정

데이터 분석 역량은 실제 데이터를 다루는 프로젝트를 반복하며 가장 크게 성장했습니다. 데이터과학을 전공하며 Python, SQL, R 을 중심으로 데이터 전처리, 통계 분석, 머신러닝의 기초를 학습했고, 이후 프로젝트 중심의 빅데이터 교육과정을 통해 의료, 공공, 공간, 이커머스 등 다양한 도메인의 데이터를 다루며 분석 전 과정을 경험했습니다.

초기 프로젝트에서는 분석 이전 단계인 데이터 수집과 전처리의 중요성을 배웠습니다. MIMIC-IV 의료 데이터 분석

프로젝트에서는 환자 정보, 입원 정보, 진단 코드, ICU 관련 데이터를 SQL 로 추출하고 조인하여 환자 단위 데이터셋을 구성했습니다. 이후 Python 을 활용해 결측치 처리, 범주형 변수 인코딩, 파생변수 생성을 수행하고, Logistic Regression, RandomForest, LightGBM 등 여러 모델을 비교하며 사망률 예측 모델을 구축했습니다. 이 경험을 통해 모델 성능은 단순히 알고리즘 선택만으로 결정되는 것이 아니라, 분석 목적에 맞게 데이터를 설계하고 정제하는 과정에서 출발한다는 점을 배웠습니다.

이후 공공데이터 프로젝트를 수행하며 데이터 분석의 범위를 통계적 검증과 결과 해석으로 확장했습니다. 도서관 이용 영향 분석 프로젝트에서는 지역 경제지표와 도서관 운영 데이터를 결합해 다중회귀분석을 수행했고, 변수별 유의성과 설명력을 바탕으로 도서관 이용에 영향을 미치는 요인을 해석했습니다. 생활자원순환지수 개발 프로젝트에서는 재활용률, 소각·매립 비율 등 여러 지표를 정리하고, 상관관계를 반영한 가중합 방식으로 지역별 지수를 산출했습니다. 이후 군집 분석과 지도 시각화를 통해 지역별 특성을 구분하고, 분석 결과를 정책 제언 방향으로 연결했습니다. 이를 통해 저는 단순한 수치 분석보다 결과가 어떤 의미를 가지는지 해석하고 전달하는 역량을 키울 수 있었습니다.

공간 데이터와 머신러닝을 함께 활용한 싱크홀 위험도 예측 프로젝트에서는 문제 특성에 맞는 평가 기준 설정의 중요성을 배웠습니다. 해당 프로젝트에서는 QGIS 를 활용해 경기도 전역을 250m × 250m 격자 단위로 나누고, 지반·지질, 인구, 교통, 건축, 하수도 인프라 데이터를 공간 조인하여 분석 데이터셋을 구축했습니다. 하지만 싱크홀 발생 데이터는 클래스 불균형이 매우 심했기 때문에 Accuracy 만으로 모델을 평가하면 실제 위험지역을 놓칠 수 있었습니다. 이에 Recall 과 PR-AUC 를 중심으로 평가 기준을 설정하고, 변수 중요도와 다중공선성을 검토하며 모델의 안정성과 해석 가능성을 함께 고려했습니다. 이 경험을 통해 분석가는 문제의 목적에 맞는 기준을 직접 설정할 수 있어야 한다는 점을 배웠습니다.

최근에는 Willpang 이커머스 성장 분석 프로젝트를 통해 분석 결과를 실제 운영 전략과 연결하는 경험을 했습니다. 고객, 세션, 이벤트, 주문 데이터를 설계하고 AARRR 프레임워크를 기준으로 유입, 활성화, 유지, 매출 지표를 정의했습니다. 이후 퍼널 전환율, 재구매율, 매출 집중도를 분석해 병목 구간을 파악하고, 전환율 개선 시나리오를 통해 주문 수와 마진 변화까지 검토했습니다. 또한 Power BI 대시보드와 RAG 기반 분석 챗봇을 구현하며 분석 결과를 반복적으로 활용 가능한 형태로 확장했습니다.

이러한 프로젝트 경험을 통해 저는 데이터 분석가에게 필요한 핵심 역량이 도구 사용 능력에만 있지 않다는 것을 배웠습니다. 중요한 것은 문제를 정확히 이해하고, 분석 목적에 맞게 데이터를 구성하며, 적절한 지표와 평가 기준을 설정하고, 결과를 실무자가 이해할 수 있는 언어로 전달하는 능력이라고 생각합니다. 입사 후에도 그동안 쌓아온 Python, SQL, R 기반 분석 역량과 프로젝트 경험을 바탕으로 새로운 데이터에 빠르게 적응하고, 문제 정의부터 결과 해석까지 책임감 있게 수행하는 데이터 분석가로 성장하겠습니다.

문제해결능력

저는 문제를 해결할 때 먼저 “프로젝트의 목적이 무엇인가”를 명확히 정의하는 것이 가장 중요하다고 생각합니다. 경기도 싱크홀 위험도 예측 프로젝트에서 이러한 방식으로 문제를 해결한 경험이 있습니다.

해당 프로젝트는 경기도 전역을 250m × 250m 격자 단위로 세분화하고, 지반·지질, 인구, 교통, 건축, 하수도 인프라 등 다양한 공간 데이터를 통합하여 싱크홀 발생 위험지역을 사전에 예측하는 것이 목표였습니다. 기존 싱크홀 관련 정보가 행정동이나 시군구 단위의 사후 관리 중심에 머무르는 한계가 있었기 때문에, 더 세밀한 공간 단위에서 행정적으로 활용

가능한 위험도 지도를 구축하는 것이 필요했습니다.

프로젝트를 진행하며 가장 큰 문제는 싱크홀 발생 데이터의 극단적인 클래스 불균형이었습니다. 전체 격자 중 실제 싱크홀 발생 비율이 약 0.2%에 불과했기 때문에, 단순히 Accuracy 를 기준으로 모델을 평가하면 대부분의 지역을 미발생으로 예측해도 높은 정확도가 나올 수 있었습니다. 하지만 이 경우 실제 위험지역을 놓칠 가능성이 높아, 사전 예방이라는 프로젝트 목적과 맞지 않는다고 판단했습니다.

이에 저는 먼저 분석 목표를 다시 정리했습니다. 이 프로젝트의 핵심은 전체 예측 정확도를 높이는 것이 아니라, 실제 발생 가능성이 있는 위험지역을 최대한 놓치지 않는 것이었습니다. 따라서 평가 기준을 Accuracy 중심에서 Recall 과 PR-AUC 중심으로 전환했습니다. 또한 QGIS 를 활용해 경기도 전역의 격자를 생성하고, EPSG:5181 좌표계 통일과 공간 조인을 통해 격자 단위 통합 데이터셋을 구축했습니다. 이후 인구, 지반·지질, 교통·지하구조물, 건물, 하수도 인프라 범주의 총 47 개 독립변수를 구성하고, 상관계수와 VIF 를 활용해 다중공선성을 검토했습니다. SHAP 기반 변수 중요도와 PR-AUC 변화 검증을 함께 활용하여 예측 기여도가 낮거나 모델 안정성을 떨어뜨릴 수 있는 변수를 단계적으로 제거했습니다.

모델링 단계에서는 클래스 불균형에 대응하기 위해 Balanced RandomForest 와 Logistic Regression 을 비교했습니다. Balanced RandomForest 는 소수 클래스 탐지에 강점이 있었지만, 최종적으로는 성능 안정성과 해석 가능성, 행정 활용 가능성을 함께 고려해 Logistic Regression 을 선정했습니다. 또한 Threshold 조정을 통해 탐지 민감도를 조정하며 실제 위험지역을 더 잘 식별할 수 있도록 모델을 개선했습니다.

그 결과 싱크홀 발생지역에 대한 높은 재현율을 확보할 수 있었고, 격자 기반 위험도 지도와 Streamlit 대시보드를 통해 분석 결과를 실제 의사결정에 활용 가능한 형태로 제시했습니다. 단순히 성능이 높은 모델을 선택한 것이 아니라, 문제의 목적에 맞는 평가 기준과 모델을 선택했다는 점에서 의미 있는 경험이었습니다.

이 경험을 통해 문제 해결은 기술을 많이 적용하는 것보다, 해결해야 할 문제의 본질을 정확히 정의하고 그에 맞는 기준을 설정하는 데서 시작된다는 것을 배웠습니다. 앞으로도 실무에서 데이터를 분석할 때 단순 수치나 모델 성능만을 기준으로 판단하지 않고, 실제 업무 목적과 활용 가능성을 기준으로 문제를 해결하는 데이터 분석가가 되겠습니다.

의사소통능력

저는 데이터 분석에서 의사소통능력은 분석 결과를 단순히 설명하는 것이 아니라, 상대방이 결과를 이해하고 실제 의사결정에 활용할 수 있도록 구조화해 전달하는 능력이라고 생각합니다. 팀 프로젝트를 수행하며 분석 결과가 아무리 의미 있더라도, 지표의 의미와 해석 방향이 명확히 공유되지 않으면 이후 개선 방향을 함께 논의하기 어렵다는 점을 경험했습니다.

이러한 점을 가장 크게 느낀 경험은 Willpang 이커머스 성장 분석 프로젝트였습니다. 해당 프로젝트는 고객의 유입부터 구매, 재구매까지 이어지는 흐름을 분석하고, 이커머스 서비스의 성장 전략을 제안하는 것이 목표였습니다. 그러나 프로젝트 초반에는 팀원마다 데이터를 바라보는 관점이 달랐습니다. 일부는 매출 규모나 주문 수에 집중했고, 일부는 방문 수와 전환율을 중요하게 보았습니다. 이처럼 같은 데이터를 보더라도 어떤 지표를 중심으로 해석하느냐에 따라 결론이 달라질 수 있었기 때문에, 먼저 공통된 분석 기준을 세우는 과정이 필요했습니다.

저는 팀원들이 전체 흐름을 같은 기준으로 이해할 수 있도록 AARRR 프레임워크를 기준으로 분석 구조를 정리했습니다. 고객 유입은 Acquisition, 상품 조회와 장바구니 행동은 Activation, 재방문과 재구매는 Retention, 주문과 마진은 Revenue 로 구분했습니다. 이후 page view, add to cart, checkout, purchase 로 이어지는 퍼널 구조를 시각화하고, 각 단계의 전환율과 이탈 구간을 함께 정리했습니다. 단순히 “전환율이 낮다”고 말하는 대신, 어느 단계에서 고객이 이탈하고 있으며 그 결과가 주문 수와 매출에 어떤 영향을 주는지 연결해 설명하려고 했습니다.

분석 결과를 전달할 때는 수치만 나열하지 않고, 팀원들이 의사결정 관점에서 이해할 수 있도록 시나리오 형태로 정리했습니다. 예를 들어 장바구니에서 결제로 넘어가는 단계의 전환율이 개선될 경우 주문 수와 마진이 어떻게 달라지는지 Baseline 과 개선안을 비교했습니다. 이를 통해 단순히 방문자를 늘리는 전략보다 퍼널 병목 구간을 개선하는 것이 더 효과적일 수 있다는 점을 팀원들과 함께 확인할 수 있었습니다.

또한 분석 결과를 반복적으로 확인할 수 있도록 Power BI 대시보드로 주요 지표를 시각화했습니다. 매출, 주문 수, 전환율, 재구매율, 고객 세그먼트별 성과를 한눈에 볼 수 있도록 구성했고, RAG 기반 분석 챗봇을 구현해 프로젝트 결과를 질문 형태로 확인할 수 있게 했습니다. 이를 통해 분석 결과가 일회성 발표 자료에 그치지 않고, 팀원들이 필요할 때 다시 확인하고 활용할 수 있는 형태로 확장되었습니다.

그 결과 팀 내에서 분석 결과에 대한 이해도가 높아졌고, 단순히 “매출을 높여야 한다”는 추상적인 결론이 아니라 “어떤 퍼널 단계와 고객 행동을 개선해야 하는지”에 대해 구체적으로 논의할 수 있었습니다. 최종적으로 분석 결과를 성장 전략, 개선 시나리오, 대시보드 형태로 정리하며 의사결정에 활용 가능한 결과물로 완성할 수 있었습니다.

이 경험을 통해 데이터 분석가의 의사소통은 분석 결과를 많이 설명하는 것이 아니라, 상대방이 같은 기준으로 문제를 이해하고 다음 행동을 결정할 수 있도록 돕는 과정이라는 점을 배웠습니다. 앞으로도 실무에서 지표 정의, 분석 결과, 개선 방향을 명확하게 전달하며 데이터 기반 의사결정에 기여하는 분석가가 되겠습니다.

성격 소개

제 성격의 강점은 꼼꼼함과 꾸준함입니다. 저는 새로운 내용을 배울 때 단순히 한 번 이해하고 넘어가기보다, 직접 정리하고 다시 확인하는 과정을 통해 제 것으로 만드는 편입니다. 특히 데이터 분석을 학습하고 프로젝트를 수행하는 과정에서 이러한 성향은 노선에 꾸준히 기록하는 습관으로 이어졌습니다.

데이터 분석은 수집, 전처리, 분석, 모델링, 시각화, 결과 해석까지 여러 단계가 연결되어 있습니다. 이 과정에서 데이터 출처, 변수 정의, 전처리 기준, 모델 선택 이유, 평가 지표를 명확히 기록하지 않으면 나중에 결과를 다시 설명하거나 수정할 때 어려움이 생길 수 있다고 느꼈습니다. 그래서 저는 프로젝트를 진행할 때마다 사용한 데이터의 구조, 분석 목적, 주요 코드, 오류 해결 과정, 결과 해석을 노선에 단계별로 정리했습니다.

이러한 기록 습관은 팀 프로젝트에서도 도움이 되었습니다. 예를 들어 분석 방향을 다시 점검해야 하거나, 모델 결과가 예상과 다르게 나왔을 때 이전에 정리해둔 전처리 방식과 변수 선택 기준을 확인하며 원인을 빠르게 찾을 수 있었습니다. 또한 팀원들과 결과를 공유할 때도 단순히 결론만 전달하는 것이 아니라, 어떤 과정과 기준으로 해당 결과가 나왔는지 설명할 수 있어 협업 과정의 혼선을 줄일 수 있었습니다.

프로젝트가 끝난 뒤에도 기록을 다시 보완하며 부족한 부분을 정리했습니다. 당시에는 이해했다고 생각했던 개념도 시간이

지나 다시 보면 보완할 부분이 있었고, 이를 다시 정리하는 과정에서 분석 흐름을 더 명확히 이해할 수 있었습니다. 이러한 꾸준한 정리 습관은 새로운 프로젝트를 시작할 때도 이전 경험을 빠르게 참고하고 적용할 수 있는 기반이 되었습니다.

물론 새로운 업무를 접할 때는 정확하게 이해하고 진행하려는 성향 때문에 초기 속도가 다소 느릴 때가 있습니다. 이를 보완하기 위해 먼저 전체 업무의 목적과 우선순위를 파악한 뒤, 작업을 작은 단위로 나누어 실행하려고 노력하고 있습니다. 또한 반복적으로 확인해야 하는 부분은 체크리스트로 정리하여 업무 속도와 정확성을 함께 높이하고자 합니다.

저는 노선에 꾸준히 기록하는 습관이 단순한 정리를 넘어, 분석의 신뢰성과 재현 가능성을 높이는 중요한 태도라고 생각합니다. 앞으로도 맡은 업무의 과정과 근거를 성실하게 기록하고 점검하며, 신뢰할 수 있는 결과를 만드는 데이터 분석가로 성장하겠습니다.