

우상욱

Data Engineer

Birthday 1996.12.24

Email wjddm3@naver.com

Mobile 010-6659-6977

Address 서울시 중랑구

병역사항 현역필

소개 / About Me

- 메인 언어 Python 으로 데이터와 데이터엔지니어링을 사랑하는 주니어 개발자
- 데이터 엔지니어 인턴 경험과 팀 프로젝트, 그리고 꾸준한 학습으로 이론과 실무 경험을 쌓아가는 개발자
- 지식 기록을 위한 블로그 운영과, 기록 공유를 위한 스터디 운영으로 모두와 함께 성장하려는 개발자
- 창업보육 관련 인턴 경험을 통해 기업의 더 나은 방향을 끊임없이 고민하고 제안하는 개발자

기술 스택 / Skill Set

기능 구현 등의 사용 경험이 있는 Skill Set

구분	Skill
Programing Languages	Python, Java, HTML5/CSS3, JavaScript
Framework/ Library	Airflow, Superset, Spark, Pandas, Scikit-Learn, Tensorflow, FastAPI, Django, ReactJS, Hadoop, Hive, Pig
Database	MySQL, PostgreSQL, AWS Redshift
Tooling/ DevOps	Docker, GitHub, Gitkraken
Cloud	AWS(EC2, S3, RDS, Redshift, ECR, Lambda)
Environment	Linux(Ubuntu)
ETC	Notion, Slack

사용경험은 없으나, 이론적 지식이 있는 Skill Set

구분	Skill
Framework/ Library	Kafka, Presto, Sqoop, DBT, Kubeflow, Flask, Spring, Spring boot
Database	MongoDB, Redis
Tooling/ DevOps	Kubernetes
Cloud	AWS(Athena, Glue, EMR, EKS)

인턴 업무 경험 / Work Experience

(주)데이터누리 | 데이터 엔지니어 | 2022.04 ~ 2022.12(8 개월)

1. 라벨링 데이터 AI 모델 맞춤 가공

- 차아 이미지 데이터 딥러닝 모델에 맞춰, 라벨링 데이터 가공
- 데이터 라벨링 클라우드 워커 전용 보조 프로그램 개발(작업 속도 57% 향상)

2. 빅데이터 전처리 자동화 및 처리 내역 레포트 추출 프로그램 개발

- 고객사 데이터 정의서 기반, 전처리 자동화 프로그램 개발
- 가공 로그 기반 전처리 레포트 추출 프로그램 개발

3. 공공 데이터 DB 적재 업무

- 반정형 데이터 가공 및 DB 적재 업무

4. 고객사 요청 데이터 크롤링 프로그램 개발

- 웹 크롤링 프로그램 설계 및 개발

(재)서울경제진흥원(SBA) | 창업육성팀(비개발, 장기현장실습) | 2021.07 ~ 2021.12(6 개월)

1. 입주기업 데이터베이스 관리 및 엑셀러레이팅

- 입주기업 데이터베이스, 관련 서류 관리 및 안내 메일링
- 입주기업 홍보 관련 기획

2. 네트워킹 행사 기획

- 입주기업 간 비대면 네트워킹 행사 기획 및 운영
- VC 네트워킹 행사 기획

3. 서울창업허브 센터 운영 관련 지원

- TryEverything 2021 행사 운영 지원
- SBA 창업본부 2021 년 하반기 사회공헌활동 기획 및 진행

과제 프로젝트 경험

서울시 혼잡도 데이터를 활용한 안전·문화생활 통합 웹서비스

작업 기간	2023.06. ~ 2023.07. (1 개월)	
인력 구성	데이터엔지니어 1 명, 데이터 분석가 1 명, 프론트 개발자 1 명, 백엔드 개발자 2 명	
프로젝트 목적	<ul style="list-style-type: none">- 다양한 소스로부터 온 데이터 통합 및 데이터베이스 구축- 클라우드 기반, 다양하고 실용적인 데이터 파이프라인 구현- 다양한 직군의 개발자들이 협업하는 과정 경험	
주요업무 및 상세역할	<ol style="list-style-type: none">1. 데이터베이스 설계 및 구축2. 맛집, 명소 데이터 크롤링3. Airflow 활용 데이터 수집 및 ETL 파이프라인 설계 및 구현4. Aws Lambda 활용, 딥러닝 모델(Tensorflow)을 활용한 별점 업데이트 파이프라인 설계 및 구현5. 데이터웨어하우스 설계 및 구축(Aws Redshift)6. Superset 데이터 시각화 서버 구축 및 DB, DW 연동7. 머신러닝 예측 페이지 전용 API 구현(FastAPI)8. Airflow 활용 ML 모델 자동화 배포 파이프라인 설계 및 구현9. 로그 데이터 수집 및 가공 파이프라인 설계 및 구현	
사용언어 및 개발환경	사용 언어	Python, Java, JavaScript, HTML, CSS
	프론트엔드	HTML, CSS, ReactJS
	백엔드	Spring, SpringBoot, FastAPI
	데이터	Tensorflow, Scikit-Learn, Airflow, Pandas, Superset
	데이터베이스	MySQL(AWS RDS), AWS Redshift
	테스팅 툴	Selenium, Postman
	클라우드	AWS(EC2, S3, RDS, Redshift, ECR, Lambda)
	데브옵스	Git, GitHub, GitKraken, Docker
	협업 툴	Notion, Slack, Canva
느낀 점	<ul style="list-style-type: none">- 기획과 설계, 그리고 CS 전반 지식에 대한 중요성을 깨닫게 된 계기가 되었습니다.- 다양한 직군의 개발자들과 협업하면서, 데이터 사용자가 쓰기 편한 데이터를 만들어주는 것이 데이터엔지니어의 중요한 역할 중 하나라는 점을 절실히 체감하였습니다.- 한정된 비용으로 시스템을 구성하면서, 각 클라우드 서버의 사양과 비용 등을 지속적으로 고민할 필요가 있다는 것을 알게 되었습니다.- 다양한 데이터파이프라인을 구성해보면서, 생각하지 못한 오류가 생길 때가 있었습니다. 이에 대한 대비책을 강구하는 일은 데이터 사용자와 함께 생각 해야 한다는 것을 깨닫게 되었습니다.	
참고자료	발표영상	https://www.youtube.com/watch?v=2tBL4qRNx7w
	시연영상	https://www.youtube.com/watch?v=xOqyAW4Hq0Y
	GitHub	https://github.com/sangwookWoo/Stand_Up_Seoul
	문서	https://url.kr/ihlpxz
	회고록	https://dataengineerstudy.tistory.com/196
	팀노션	https://url.kr/fjyvx9

SPOTIFY 데이터를 활용한 음악 평론 웹서비스(미니 프로젝트)

작업 기간	2023.05. ~ 2023.05. (4 일)	
인력 구성	데이터엔지니어 1 명, 백엔드 개발자 1 명	
프로젝트 목적	<ul style="list-style-type: none"> - Spotify 데이터 수집 파이프라인 구현 - DB 설계 및 웹 사이트 구현 	
주요업무 및 상세역할	<ol style="list-style-type: none"> 1. DB 설계 및 연동 2. 데이터 수집 파이프라인 구현 3. 가수 및 음악 검색 기능 및 페이지 프론트엔드, 백엔드 구현 4. 가수, 음악 상세 페이지 프론트엔드, 백엔드 구현 	
사용언어 및 개발환경	사용 언어	Python, HTML/CSS, JavaScript
	프론트엔드	HTML/CSS, JavaScript
	백엔드	Django
	데이터	Airflow, Pandas
	데이터베이스	MySQL
	클라우드	AWS(RDS)
	협업 툴	Notion, Slack
느낀 점	<ul style="list-style-type: none"> - Django 를 활용해 직접 백엔드와 프론트를 구현하면서, 백엔드 개발자와 프론트 개발자가 DB 의 데이터를 어떻게 활용하는 지를 제대로 알게 되었습니다. - 가벼운 데이터 파이프라인이었지만, DB 참조 관계 때문에 task 별 순서를 조정하는 과정에 많은 시간을 할애하였습니다. 데이터의 관계를 제대로 이해하는 것이 중요하다고 느꼈습니다. - 프론트엔드를 구성함에 있어서, 사용자가 사용하기 쉬운 프론트엔드에 대한 중요성을 절실히 깨닫게 되었습니다. 	
참고 자료	시연영상	https://www.youtube.com/watch?v=5Slg7gHpyv4
	회고록	https://dataengineerstudy.tistory.com/179

사용자 대화 기반 감성 분석 및 음악 추천 딥러닝 프로젝트(미니 프로젝트)

작업기간	2023.03. ~ 2023.03. (5 일)	
인력구성	데이터 엔지니어 1 명, 백엔드 개발자 1 명, 데이터분석 3 명,	
프로젝트 목적	<ul style="list-style-type: none"> - 자연어 처리를 통한 딥러닝 모델 활용 및 사용경험 - 유사도 기반 추천 시스템 이해 - 자연어 데이터 가공 숙련도 향상 	
주요업무 및 상세역할	<ol style="list-style-type: none"> 1. Konlpy 활용 한국어 데이터 가공 2. Streamlit 프레임워크 활용 간단한 웹 구현 3. 사용자와 대화하는 AI 챗봇 개발 4. ML/DL 모델링 	
사용언어 및 개발환경	사용 언어	Python
	웹 프레임워크	Streamlit

	데이터	Konlpy, Pandas, Torch ,Scikit-Learn
	데이터베이스	PostgreSQL
	데브옵스	Git, GitHub
	협업 툴	Notion, Slack
느낀 점	<ul style="list-style-type: none"> - 추천 시스템을 간단히 구현해보는 과정을 통해, DB 에 모인 데이터가 강력한 추천 알고리즘을 만들 수 있다는 것을 깨닫게 되었습니다. - 조금 더 성능이 좋은 DL 모델이 있었으나, 웹에서의 사용제약으로 인해 다른 ML 모델로 감성분석을 진행하였습니다. 시스템의 제약에 대한 이해 또한 개발 과정에서 중요하다고 생각하였습니다. 	
참고 자료	시연영상	https://youtu.be/4qhXDmpotuw
	GitHub	https://github.com/sangwookWoo/music_resting_place
	회고록	https://dataengineerstudy.tistory.com/79

심장질환 AI 예측 머신러닝 프로젝트(미니 프로젝트)

작업기간	2023.02. ~ 2023.02. (3 일)	
인력구성	데이터 엔지니어 1 명, 데이터 분석가 2 명	
프로젝트 목적	<ul style="list-style-type: none"> - EDA 를 통한 데이터 탐색 및 시각화 - 머신러닝에 관한 이해 및 성능 개선 - 머신러닝 활용 방안 모색 및 구현 	
주요업무 및 상세역할	<ol style="list-style-type: none"> 1. 심장질환 분류 데이터 EDA 및 전처리 2. 심장질환 분류 ML 모델링 3. Streamlit 프레임워크 활용 간단한 웹 구현 4. 심장질환 분류 확률 및 독립변수 조작을 통한 심장질환 개선 방향 로직 구현 	
사용언어 및 개발환경	사용 언어	Python
	웹 프레임워크	Streamlit
	데이터	Pandas, Scikit-Learn
	데브옵스	Git, GitHub
	협업 툴	Notion, Slack
느낀 점	<ul style="list-style-type: none"> - 팀원이 함께 고민하고 계속 검토한 결과, 기존엔 심장질환의 유무를 진단만 하는 AI 모델이었지만, 사용자에게 받은 독립변수를 백엔드 단에서 조작하면서, 독립변수의 변화에 따른 확률을 비교하는 로직을 구현할 수 있었습니다. 이 과정에서 열려 있는 팀원 간 토론을 통해 좋은 아이디어가 나올 수 있다는 점을 직접 경험하였습니다.. 	
참고자료	시연영상	https://www.youtube.com/watch?v=r16EihZyOk
	GitHub	https://github.com/sangwookWoo/ML_project_predict_heart_disease
	회고록	https://dataengineerstudy.tistory.com/61

학력

- 2015.03 ~ 2023.02 동국대학교 경영학부(경영정보학과) 졸업 | 학점 3.73/4.5
- 2012.03 ~ 2015.02 서울 대진고등학교 졸업

자격증

- 정보처리기사 (2023. 06 취득)
- ADSP(데이터분석준전문가) (2022. 06 취득)
- TOEIC(850) (2022. 03 취득)
- 창업보육전문매니저 (2021. 12 취득)
- 컴퓨터활용능력 1 급 (2021.04 취득)
- 한국사능력검정시험 1 급 (2020.10 취득)
- 사회조사분석사 2 급 (2019.11 취득)

교육 내용

- 2023.01 ~ 2023.06 플레이데이터 20 기 데이터엔지니어링 트랙 | 엔코아
- 2023.03 ~ 2023.05 실리콘밸리에서 날아온 데이터엔지니어링 스타터 키트 | 프로그래머스

수상 내역

- 2022 POSTECH OIBC CHALLENGE 태양광 발전량 예측 경진대회 | 참가상
- 2021 동국대학교 기업연계형 장기현장실습 우수사례 경진 대회 | 최우수상