**분석 계열 보고서**

천재교육 K-Digital Training 빅데이터 7기

송지환, 최종은, 김수진, 이민아

**요약**

새로운 AWS 교육 서비스를 진행하기 위한 의사결정 근거 자료 산출을 위해 2월 27일부터 3월 5일까지, 영업일 기준으로 5일 간 웹 스크래핑, 전처리, 시각화, 기초통계분석을 이용한 가설검정을 진행하였다. 웹 스크래핑을 위해 python의 selenium, beautifulsoup, request, newspaper 패키지를 이용하여 랠릿, 프로그래머스, 원티드의 AWS 관련 채용정보를 검색한 후 웹 스크래핑을 진행하였다. 이후 python 환경에서 대부분의 전처리를, 일부 결측치와 문장 형식의 자료는 수작업으로 전처리를 진행하였다. 전처리 후 각 자료형에 대해 그래프로 시각화를 진행하였다. 이후 세 가지 가설검정을 진행하였다. 첫 번째 가설검정에서는 AWS 기술스택을 요구하는 채용업체들의 업종 분류에 따라, docker, 쿠버네티스, 컨테이너, Amazon EKS, Amazon K8S, Dev 기술스택에 대해 채용공고에서 추가로 언급하는 빈도의 차이가 있을 것이라고 초기 가설을 세웠고, 카이제곱 검정 결과 유의하지 않은 것으로 나타났다. 두 번째 가설검정에서는 AWS 기술스택을 요구하는 채용업체들이 구직자들에게 요구하는 경력(신입~시니어)에 따라 요구하는 기술스택도 달라질 것이라는 초기 가설을 세웠고, 결과는 신입, 주니어, 미들, 시니어 네 경력 분류에 대해 모든 비교 결과가 95% 유의수준에서 유의하였다. 즉, 요구하는 경력 별로 요구하는 기술 스택의 차이가 있는 것으로 밝혀졌다. 따라서, 새로 시행하는 교육 서비스의 대상자의 경력이 상이한 경우, 다른 강의 커리큘럼을 구성하는 것이 타당하다는 결론을 내렸다. 마지막 가설검정에서는 기업 형태(중소기업~대기업)에 따라, 언급되는 빈도수가 높은 기술 스택을 추출하여 각 기업별로 요구하는 기술스택의 차이가 있는지를 알아보고자 카이제곱 검정을 사용하였다. 결과는 대기업과 중견기업의 요구 기술스택 차이만이 95% 유의수준에서 유의하지 않았고, 중소기업과 대기업, 중소기업과 중견기업의 차이는 유의한 것으로 나타나, 중소기업에서 대기업이나 중견기업으로 이직 혹은 취직을 노리는 교육생에게 교육을 진행할 기술스택의 우선순위를 달리 하는 것이 타당하다는 결론을 내렸다.

**서론**

새로운 AWS 교육 서비스를 진행하기 위한 자료 수집과 분석을 진행하였다. AWS 관련 업체들의 수요 및 요구 조건과 기업 정보 등을 수집 및 분석하여 자료를 산출함으로써, 새로운 교육 서비스 전략의 수립 및 활용을 위한 의사결정 과정에 이용할 수 있는 근거를 제시하는 것이 목적이다.

새로운 AWS 교육 서비스를 시작하는데 있어, AWS 기술스택을 요구하는 업체들의 기업 정보를 파악하고, 관련 채용 시장에서 주로 요구하거나 우대하는 기술스택을 파악하는 등 향후 교육 서비스 전략의 수립 과정에 도움이 될 수 있는 자료를 산출하고자 한다.

이에 따라, AWS 관련 채용공고를 낸 업체들의 업종분류, 사원 수, 설립연도, 기업형태, 연매출, 요구 경력, 주소, 직무, 우대사항, 자격요건, 요구 기술스택에 대해 시각화를 진행하여 관련 업체들의 수요를 알아보고자 한다. 다음으로, docker, 쿠버네티스, 컨테이너, Amazon EKS, Amazon K8S, Dev 여섯 가지 기술스택에 대한 수요를 알아보기 위해 전체 채용공고에서 단어가 언급된 빈도 수를 구하였다. 마지막으로, 웹 스크래핑한 자료를 근거로 가설을 세운 뒤 검정하였다. 본 프로젝트에서는 두 가지 가설을 제안하였으며 그 내용은 아래와 같다.

가설 1: AWS 기술스택을 요구하는 채용업체들의 업종 분류에 따라, docker, 쿠버네티스, 컨테이너, Amazon EKS, Amazon K8S, Dev 기술스택에 대해 채용공고에서 추가로 언급하는 빈도의 차이가 있을 것이다.

가설 2: AWS 기술스택을 요구하는 채용업체들이 구직자들에게 요구하는 경력에 따라 요구하는 기술스택도 달라질 것이다.

가설 3: AWS 기술스택을 요구하는 채용 업체들의 기업형태에 따라 구직자들에게 요구하는 기술스택도 달라질 것이다.

**방법**

2월 27일부터 3월 5일까지, 영업일 기준으로 5일 간 프로젝트를 진행하였다. python의 selenium, beautifulsoup, request, newspaper 패키지를 이용하여 랠릿, 프로그래머스, 원티드의 AWS 관련 채용정보를 검색한 후 웹 스크래핑을 진행하였다. 전처리 과정을 거친 뒤, 각 데이터의 형태에 맞도록 시각화 자료를 생성하였다. 그 후 초기 가설에 근거하여 기초통계분석을 진행하였다.

첫 번째 통계분석의 초기 가설 및 분석 방법은 다음과 같다.

* 가설: AWS 기술스택을 요구하는 채용업체들의 업종 분류에 따라, docker, 쿠버네티스, 컨테이너, Amazon EKS, Amazon K8S, Dev 기술스택에 대해 채용공고에서 추가로 언급하는 빈도의 차이가 있을 것이다.
* 가설 설정 근거: AWS 관련 구인시장에서도 업종에 따라 해당 6가지 기술스택을 요구하는 정도의 차이가 있는지 알게 된다면, 향후 교육 커리큘럼을 구성할 때 목표로 하는 업종에 따라 특정 기술스택에 더 무게를 두고 교육을 진행하는 것이 타당한지 의사결정을 하는 것에 도움이 될 것이다.
* 검정방법: 카이제곱(Chi-square) 검정

두 번째 통계분석의 초기 가설 및 분석 방법은 다음과 같다.

* 가설: AWS 기술스택을 요구하는 채용업체들이 구직자들에게 요구하는 경력에 따라 요구하는 기술스택도 달라질 것이다.
* 가설 설정 근거: 구직자의 경력에 따라 업체가 요구하는 기술스택이 다르다면, 향후 교육 커리큘럼을 구성할 때 교육생의 경력에 맞춰 어느 기술스택에 더 무게를 두고 교육을 시켜야 해당 업계의 취업시장에서 경쟁력을 갖게 할 수 있는지 판단하는데 도움이 될 것이다.
* 검정방법: 카이제곱(Chi-square) 검정

세 번째 통계분석의 초기 가설 및 분석 방법은 다음과 같다.

* 가설: AWS 기술스택을 요구하는 채용업체들의 기업형태에 따라 구직자들에게 요구하는 기술스택도 달라질 것이다.
* 가설 설정 근거: 업체의 기업형태 별로 요구하는 기술스택이 다르다면, 향후 교육 커리큘럼을 구성할 때 교육생들이 특정 기업형태의 업체에 취직을 목표로 하는 경우 그 목표에 부합하여 어느 기술스택에 더 무게를 두고 교육을 제공하는 것이 타당한지 의사결정하는데 도움이 될 것이다.
* 검정방법: 카이제곱(Chi-square) 검정

**결과 및 토론**

첫 번째 분석에서는 AWS 기술스택을 요구하는 채용업체들의 업종 분류에 따라서 여섯 가지 키워드(docker, 쿠버네티스, 컨테이너, Amazon EKS, Amazon K8S, Dev)에 대해 채용공고에서 추가로 언급하는 빈도의 차이가 있는지를 알아보고자 하였다. 이를 위해 범주형 데이터 간 빈도 차이를 알 아볼 수 있는 카이제곱 검정을 시행하였다.

결과는 95% 유의수준에서 유의하지 않았다. 따라서 AWS 기술스택을 요구하는 업체들에서 여섯 가지 키워드에 대해 추가적으로 언급하는 빈도가 업체의 업종분류에 따라 상이하지 않음이 밝혀졌다.

따라서, 향후 AWS 교육 커리큘럼에 해당 키워드와 관련된 교육 과정을 수립할 때, 교육 주체가 관심을 갖고 있는 업종분류의 업체들이 있다면 해당 업종분류가 많이 언급하는 기술스택 키워드에 따라서 교육 커리큘럼을 추가하고 수정하는 것이 타당한지 재고해야 한다.

두 번째 분석에서는 AWS 기술스택을 요구하는 채용 업체들이 구직자에게 요구하는 경력(신입~시니어)에 따라서, 언급되는 빈도수가 높은 기술스택을 추출하여, 각 기업별로 요구하는 기술스택의 요구 빈도 차이가 있는지 알아보고자 하였다. 마찬가지로 범주형 데이터 간의 빈도 차이를 검정하기 위해 카이제곱 검정을 시행하였다.

결과는 신입, 주니어, 미들, 시니어 네 경력 분류에 대해 모든 비교 결과가 95% 유의수준에서 유의하였다. 즉, 요구하는 경력 별로 요구하는 기술 스택의 차이가 있는 것으로 밝혀졌다.

따라서, 새로 시행하는 교육 서비스의 대상자의 경력이 상이한 경우, 다른 강의 커리큘럼을 구성하는 것에는 통계적 근거가 있다.

마지막 분석에서는 AWS 기술스택을 요구하는 채용업체들의 기업 형태(중소기업~대기업)에 따라, 언급되는 빈도수가 높은 기술 스택을 추출하여 각 기업별로 요구하는 기술스택의 차이가 있는지를 알아보고자 하였다. 카이제곱 검정을 사용하였다.

결과는 대기업과 중견기업의 요구 기술스택의 차이만이 95% 수준에서 유의하지 않았으며, 중소기업과 대기업, 중소기업과 중견기업의 차이는 95% 수준에서 유의하였다. 따라서, 특정 기업형태의 소속, 혹은 특정 기업형태의 업체로의 취직을 노리는 교육생에게 교육 서비스를 제공할 때, 중견기업과 대기업을 구분하여 교육할 기술스택의 차이를 두는 것보다, 이외의 경우에서 중점적으로 교육을 진행할 기술스택의 우선순위를 달리 하는 것이 타당하다.

**깃허브 링크:**

https://github.com/6321martys/big-data-7-teamproject