



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

2022년 8월
석사학위 논문

인공지능 기반 소프트웨어 개발 비용 산정에 관한 소요공수 예측 모형

조선대학교 산업기술창업대학원

소프트웨어융합공학과

장 승 진

인공지능 기반 소프트웨어 개발 비용 산정에 관한 소요공수 예측 모형

Man-hours Prediction Model for Estimating
the Development Cost of AI-Based Software

2022년 8월 26일

조선대학교 산업기술창업대학원

소프트웨어융합공학과

장 승 진

인공지능 기반 소프트웨어 개발 비용 산정에 관한 소요공수 예측 모형

지도교수 신 주 현

이 논문을 공학석사학위신청 논문으로 제출함.

2022년 4월

조선대학교 산업기술창업대학원

소프트웨어융합공학과

장 승 진

장승진의 석사학위논문을 인준함

위원장 조선대학교 교수 김 판 구 (인)

위 원 조선대학교 교수 신 주 현 (인)

위 원 조선대학교 교수 최 준 호 (인)

2022년 6월

조선대학교 산업기술창업대학원

목 차

ABSTRACT

| | |
|---|----|
| I. 서론 | 1 |
| A. 연구 배경 및 목적 | 1 |
| B. 연구 내용 및 구성 | 3 |
| II. 관련 연구 | 4 |
| A. 소프트웨어 비용 산정 | 4 |
| B. 소프트웨어 개발 규모 예측 | 5 |
| 1. LOC 방식 | 5 |
| 2. 기능점수 산정 방식 | 6 |
| C. 빅 데이터 분석 사업비 산정 | 8 |
| 1. 빅 데이터 사업 추진단계 | 8 |
| 2. 투입공수 보정 | 9 |
| III. AI 기반 SW 개발 비용 산정에 관한 소요공수 예측 모형 | 10 |
| A. 시스템 구성도 | 10 |
| B. 사전 준비 | 12 |
| 1. AI 기반 SW 개발단계 정의 | 12 |
| 2. 설문대상 AI 기반 SW 정의 | 14 |

| | |
|---------------------------|----|
| C. 설문 설계 | 15 |
| 1. 설문 항목 설계 | 15 |
| 2. 설문 결과 수집 | 17 |
| D. 설문 분석 | 19 |
| 1. 데이터 전처리 | 19 |
| 2. Box Plot 분석 | 21 |
| 3. 상관 분석 | 26 |
| E. 개발단계별 소요공수 예측 모형 | 30 |
| 1. 회귀 분석 | 30 |
| 2. 소요공수 예측 회귀모형 | 31 |
| | |
| IV. 실험 및 결과 | 41 |
| A. 실험 대상 견적 | 41 |
| B. 실험 평가 및 분석 | 43 |
| 1. 실험 평가 방법 | 43 |
| 2. 실험 결과 분석 | 45 |
| | |
| V. 결론 및 향후 연구 | 48 |
| | |
| 참고문헌 | 49 |
| 부록 | 50 |

표 목 차

| | |
|--|----|
| [표 2-1] 빅 데이터 분석사업 보정요소 | 9 |
| [표 3-1] 빅 데이터 분석사업 수행 단계별 수행 내용 검토 | 12 |
| [표 3-2] 설문 데이터 분석 내용 | 19 |
| [표 3-3] 설문항목 통계처리를 위한 칼럼별 영문 명명 규칙 설계 | 20 |
| [표 3-4] Box Plot 난이도 분석을 위한 Python Code | 21 |
| [표 3-5] Box Plot 소요공수 분석을 위한 Python Code | 23 |
| [표 3-6] 단계별 소요공수 설문 결과 | 24 |
| [표 3-7] 상관분석을 위한 Python Code | 26 |
| [표 3-8] 개발 단계 간 소요공수의 강한 상관관계 목록 | 28 |
| [표 3-9] 회귀분석 시행목록 | 30 |
| [표 3-10] 개발 단계별 소요공수 예측 회귀 모형 | 40 |
| [표 4-1] 견적서 소요공수 환산 내역 | 42 |
| [표 4-2] 설문 결과의 중앙값 소요공수 | 43 |
| [표 4-3] 소요공수 예측비교 실험 결과 | 45 |
| [표 4-4] 소요공수 예측모형별 정확도 | 47 |

그림 목 차

| | |
|---|----|
| [그림 1-1] AI SW 시장 규모 예측 | 1 |
| [그림 1-2] AI SW 개발비용 신뢰도 | 2 |
| [그림 2-1] SW 비용 산정 가이드 대상 사업유형 | 4 |
| [그림 2-2] 기능점수 산정 절차 | 6 |
| [그림 2-3] 빅 데이터 분석 사업 수행 7단계 | 8 |
| [그림 3-1] 시스템 구성도 | 10 |
| [그림 3-2] 표준 품셈 예시 | 14 |
| [그림 3-3] 설문 의 구성 | 15 |
| [그림 3-4] 수행활동 난이도 설문 설계 예시 | 15 |
| [그림 3-5] 수행활동 소요공수 설문 설계 예시 | 16 |
| [그림 3-6] 설문 결과 | 17 |
| [그림 3-7] Box Plot 난이도 분석 결과 | 22 |
| [그림 3-8] Box Plot 소요공수 분석 결과 | 23 |
| [그림 3-9] 개발 단계 간 Pearson 상관분석 결과 | 27 |
| [그림 3-10] 개발 단계 간 Heat Map 상관분석 결과 | 27 |
| [그림 3-11] 개발 단계 간 소요공수 상관관계도(圖) | 28 |
| [그림 3-12] 회귀 식 도출을 위한 방법(1) | 31 |
| [그림 3-13] 회귀 식 도출을 위한 방법(2) | 31 |
| [그림 3-14] 요건정의와 데이터 수집의 소요공수 회귀분석 | 32 |
| [그림 3-15] 요건정의와 데이터 전처리의 소요공수 회귀분석 | 33 |
| [그림 3-16] 데이터 수집과 검증 및 안정화의 소요공수 회귀분석 | 34 |
| [그림 3-17] 요건정의와 검증 및 안정화의 소요공수 회귀분석 | 35 |
| [그림 3-18] 데이터 전처리와 모델 구현의 소요공수 회귀분석 | 36 |
| [그림 3-19] 데이터 전처리와 분석 모델링의 소요공수 회귀분석 | 37 |
| [그림 3-20] 분석 모델링과 시각화의 소요공수 회귀분석 | 38 |
| [그림 3-21] 모델 구현과 시각화의 소요공수 회귀분석 | 39 |
| [그림 4-1] AI SW 개발 견적 사례 | 41 |
| [그림 4-2] 소요공수 예측 모형 간 입출력 흐름도 | 44 |
| [그림 4-3] 예측소요공수와 설문소요공수 비교 | 46 |

ABSTRACT

Man-hours Prediction Model for Estimating the Development Cost of AI-Based Software

SeongJin Chang

Advisor : Prof. JuHyun Shin, Ph.D.

Department of Software Convergence

Engineering

Graduate School of Industrial Technology
and Entrepreneurship, Chosun University

The artificial intelligence software market is expected to grow sixfold from 2020 to 2025. However, because the software development procedure is not standardized and there is no cost calculation standard, there is growing distrust in the quality and development cost of the development software.

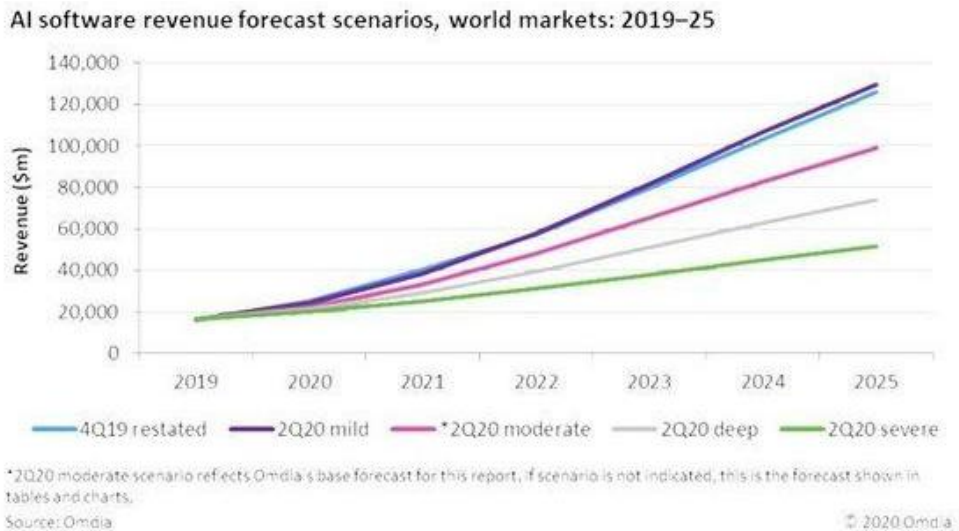
Therefore, in this paper, we define the development stage of “artificial intelligence-based software” that learns with a large amount of data, derives and applies algorithms, and randomly selects 27 people from 19 companies and conducts a questionnaire to collect input man-hours for each development stage. did.

By performing correlation analysis and regression analysis on the collected man-hours, a man-hour estimation model with a high accuracy of 92% was proposed, and the estimation model is simple to use and easy to measure. It is expected that research on the development cost of artificial intelligence software will become more active in the future.

I. 서론

A. 연구 배경 및 목적

4차 산업혁명의 여파로 인공지능 소프트웨어 개발은 지속적으로 증가하고 있다. 세계적인 기술 유력 연구기업인 Omdia에 따르면 2020년을 기준으로 2025년까지 6배 정도 인공지능 소프트웨어 시장 규모가 증가할 것으로 예측하고 있다. [그림 1-1]은 AI SW 시장 규모의 예측을 나타낸다.



[그림 1-1] AI SW 시장 규모 예측

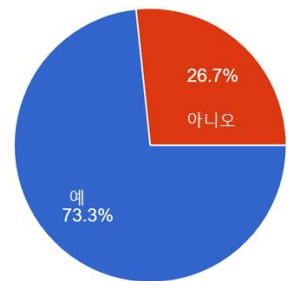
일반적으로 정보화 사업의 범위와 그에 상응하는 합리적인 비용을 정확히 예측하기 위해서는 사업의 규모를 가능한 한 정확히 파악하고, 충분한 예산을 투입해야 한다. 인력과 예산 투입은 정보시스템의 납기, 품질, 생산성에 결정적인 영향을 미치므로[1, 2], 수 발주 사업자 모두가 동의할 수 있는, 규모 예측과 예산 산출 방식이 필요하다. 본 논문에서 다룬 “인공지능 기반 소프트웨어”는 대량의 데이터를 통하여 학습을 진행하고 알고리즘을 도출하여 적용하는 소프트웨어로 한정하여 개발 비용에 대한 연구를 진행하고자 한다.

정보시스템 개발 프로젝트의 크기에 대해서 오래전부터 계산하려는 노력이 있었으나 유형자산과 달리 무형자산의 경우는 정확한 규모와 가치를 측정하는 것이 매우 어렵다. 소프트웨어도 무형자산이므로 소프트웨어의 정확한 규모와 가치를 측정하는 것이 어렵지만 눈에 보이지 않는 무형자산의 경우 정확한 원가와 공수 도출을 위한 규모 산정은 매우 중요한 일이다[3].

그러나 산업계 간의 연구를 통해 여러 분야의 산업에 융합된 새로운 제품군과 다양한 서비스가 나오고 있는 실정이다[4]. 이렇듯 AI, Big Data 등 신기술의 등장과 기존 산업과의 융합으로 인한 복잡도의 증가로 개발 공수와 일정 등을 예측하기가 어렵고 관련한 연구도 부족한 상황이다.

현재, 인공지능 소프트웨어의 개발비용은 투입공수 방식으로 산정되고 있으나 투입공수의 적정성을 확인하기 어렵고 투입공수 산정을 위한 개발단계는 국내에 정의되어 있지 않아 더더욱 신뢰하기 어렵다.

[그림 1-2]와 같이 본 논문을 위한 설문을 통하여, 인공지능 소프트웨어의 개발비용에 대한 불신이 73%에 달하고 있음이 확인되었다. 따라서 본 논문에서는 개발비 산정의 기준이 되는 소요공수 예측 모형을 연구하고자 한다.



[그림 1-2] AI SW 개발비용 신뢰도

B. 연구 내용 및 구성

인공지능 소프트웨어 관련 업계 담당자들로부터 인공지능 소프트웨어의 개발 단계와 단계별 소요공수를 수집하고 분석하여 개발 단계별 소요공수 예측 회귀모형을 제안하고자 한다. 본 논문의 구성은 다음과 같다.

본 장인 서론에 이어 2장 관련 연구에서는 행정안전부와 NIA가 공동 발간한 “공공 빅데이터 분석 사업비 대가 산정 가이드”에서 제시하는 프로젝트 수행 단계를 살펴봄으로써 시대적인 변화에 의한 신기술 기반 사업의 비용 산정에 대한 이해를 돕는다. 또한, 소프트웨어 개발대가 산정 가이드에서 제시하는 공수 산정방식과 기능점수 산정방식에 대해 검토 한다.

3장에서는 국내 인공지능 소프트웨어 관련 업계로부터 설문을 통해 개발 단계별 소요공수를 수집한 결과와 개발 단계 간 소요공수의 상관 분석을 수행하고 강한 상관관계를 보이는 개발 단계를 기반으로 회귀분석을 실시하고 최종적으로 회귀식을 도출한다. 도출된 개발 단계별 회귀 식을 소요공수 예측 모형으로 제안 한다.

4장에서는 개발 단계별 소요공수 예측 모형을 활용한 인공지능 기반 소프트웨어 개발 전체 소요공수 예측결과와 설문으로 수집된 소요공수, 업체로부터 견적을 받은 소요공수를 비교하여 예측 모형의 정확성을 검증 한다.

마지막으로 5장에서는 본 논문에 대해 전체적인 결과를 요약하고, 향후 연구를 제시하며 마무리 한다.

II. 관련 연구

A. 소프트웨어 비용 산정

소프트웨어 비용 산정에 대한 가이드는 한국 소프트웨어 산업협회에서 수행하고 있다. 또한, SW의 수명주기, 즉 기획단계, 구현단계, 운영단계에 따른 사업유형을 [그림 2-1]과 같이 분류하고 있으며 사업유형별 대가 산정 방법을 안내하고 있다.



[그림 2-1] SW 비용 산정 가이드 대상 사업유형

각 사업유형별 비용 산정을 위한 핵심요소는 업무량, 투입공수, 기능점수, 유지관리 총 점수, 등급별 효율, 재개발 기능점수 등 다양한 요소들로 구성되어 있다[5]. 한국 소프트웨어 산업협회는 매년 변동요인을 반영하여 개정사항을 발표하고 있다. 그러나 인공지능 기반 소프트웨어의 개발 비용 산정 방법에 해답 언급은 없다.

단, “소프트웨어 개발” 사업의 대가 산정방법 중 하나인 “공수 산정 방식”을 참고할 수 있다. “공수 산정 방식”은 개발을 위해 특정 기술등급의 인력이 얼마의 기간 동안 투입되어야 하는지를 산정하는 방식으로 기술등급별 인력에 대해서는 산업협회가 기술등급별 단가를 기준 데이터로 보유하고 있어 이를 활용하면 개발비용으로의 산정이 가능하다.

B. 소프트웨어 개발 규모 예측

1. LOC(Line Of Code, 라인수) 방식

LOC 방식은 Barry Boehm이 1981년에 발표한 회귀식 모델로 공수는 규모와 비례하고 기간은 공수에 비례한다는 기법이다. 공수와 규모, 기간과 공수는 서로 독립적이라는 가정을 기반으로 정의하였으며 회귀식은 161개 프로젝트의 과거 경험 데이터에 기반 하였고 최초 모델은 라인수 기반이고 개발 단계에만 적용하였으나 후속 모델은 기능 점수와 라인수를 기반으로 전 공정에 적용하였다[6]. LOC 기반 모형을 살펴보면, Putnam이 고안한 Software Lifecycle Model(SLIM)은 매개변수를 이용하여 다음의 식 (1)과 같이 규모를 산정한다[7].

$$L = C_k \cdot K^{-1/3} \cdot T^{3/4}, K = D_0 \cdot t^3 \quad (1)$$

LOC 방식에서 제안하고 있는 회귀 식을 인공지능 기반 소프트웨어의 개발단계별 소요공수 예측에도 적용하고자 한다. 회귀 식을 도출하기 위해서는 어떤 개발단계가 인과(因果)관계에 있는지 분석하여야 한다. 그러므로 인공지능 기반 소프트웨어의 개발단계간의 상관관계의 정도를 설문 데이터 분석을 통하여 확인하고, 상관관계가 확인된 개발단계간의 회귀분석을 통하여 회귀 식을 도출 한다.

인공지능 기반 소프트웨어는 아직 개발단계, 적정 소요기간, 비용 등 어떤 것도 정립이 되지 않았기에 추후 설문을 통해 데이터 분석을 진행하고 분석된 데이터로부터 시사점을 도출할 수 있을 것이다.

2. 기능점수 산정 방식(Function Point)

기능점수 산정 방식은 1979년 Allan Albrecht에 의해 소프트웨어 개발 프로젝트의 산출물을 측정하는 개념으로 정의된 기법이다. 소프트웨어 기능 크기의 측정단위로써 FP는 세계적으로 관심이 높아지고 사용이 증가하였다. IFPUG(International Function Point User Group) Counting Practice Manual에서 지속적으로 개정해오고 있다[8].



[그림 2-2] 기능점수 산정 절차

기능점수 모형은 먼저 트랜잭션 기능 점수와 데이터 기능 점수의 합계를 산정한 후 시스템이 가지고 있는 특성에 따라 그 값을 보정하여 최종 소프트웨어 기능점수 계수를 산정한다. 보정하기 전의 기능점수는 데이터와 트랜잭션이라는 두 가지 기능 유형을 측정한다.

사업초기에는 기능의 복잡도를 확인하기 어려워 평균복잡도를 활용하는 간이법을 활용하고 사업종료 후에는 기능의 복잡도를 일일이 확인하여 정통법을 적용하여 기능점수를 계산한다. 이에 따라 기능점수의 정확성을 향상시키기 위한 연구들이 진행되어 왔다.

UML을 활용한 기능식별[9], 객체지향 개발방법론으로 개발된 소프트웨어의 유스 케이스와 클래스를 활용한 기능식별[10], 정통법을 개선하기 위하여 기존 복잡도가중치 체계를 두배로 확장시켜 더욱 세밀하게 측정하도록 한 Micro-FP 모형에

대한 연구[11], EI 기능점수 회귀 식을 기반으로 한 전체 기능점수 예측을 위한 모델에 대한 연구[12]에 이어 현재는 소프트웨어 사업정보 저장소의 사업초기 기능점수와 사업종료 기능점수 데이터를 기반으로 랜덤포레스트라는 기계학습 알고리즘을 활용하여 기능점수를 예측하는 연구까지 이르고 있다[13].

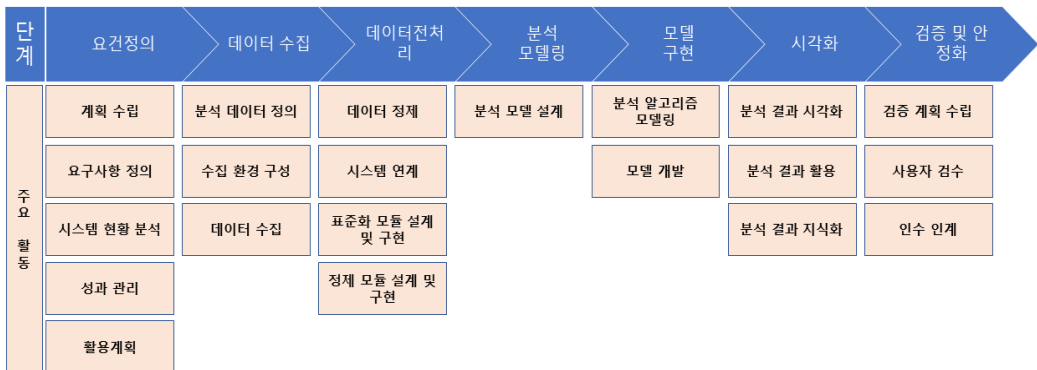
그러므로 인공지능 기반 소프트웨어의 기능점수 산정은 Data Base를 기반으로 입력, 수정, 삭제, 조회기능을 대상으로 한다. 그러나 Data Base가 아닌 데이터 수집, 정제, 분석, 학습, 알고리즘 모델 적용 등의 작업에 대하여는 “공수 산정 방식”이 타당하므로 인공지능 기반 소프트웨어에 대한 개발비용산정은 기능점수 산정방식과 공수 산정방식이 혼합되어 적용되는 것이 합리적이라고 말할 수 있다.

C. 빅 데이터 분석 사업비 산정

행안부와 NIA(한국지능정보사회진흥원)는 2017년 “공공 빅데이터 분석 사업비 산정 가이드”를 발표하였다. 빅데이터 분석 사업이 확대되고 있었지만, 분석 사업비 방안이 부재하여 걱정하고 합리적인 사업비 산정 방안을 마련하기 위해서였다. 이를 위해 현장의 요구사항을 반영하여 해당 사업의 WBS를 확정하고 보정계수를 적용한 투입공수 기반의 비용 산정 방안을 제시하였다.

1. 빅데이터 사업 추진단계

기 추진된 빅데이터 분석 사업의 산출 내역서를 취합하고 분석하여 아래의 사업 수행단계를 정의하였다.



[그림 2-3] 빅데이터 분석 사업 수행 추진 7단계

‘공공 빅데이터 업무적용 가이드’(2016, 행정자치부)의 부록 1. ‘사업추진 단계별 점검항목’ 내용을 참조하여 분석 사업에 필요한 태스크와 세부 태스크를 도출할 수 있도록 하였다[9]. 사업비용은 위의 수행 7단계의 세부 TASK별로 기술등급별 투입공수를 산정하는 방식으로 수행된다.

2. 투입공수 보정

공공 빅 데이터 분석 사업비 대가 산정 가이드는 공공 빅 데이터 협의회 구성을 시작으로 가이드를 공유하기까지 15개월이라는 장기간이 소요되었으며 기 추진 빅 데이터 분석 사업의 산출내역서 취합 및 분석을 통하여 보정요소에 대한 초안을 마련하였으며 이후에 전문가 의견수렴을 실시하여 적정성을 검토하였다.

빅 데이터 분석 사업은 데이터 규모에 따라, 데이터 전처리 기술의 종류에 따라, 분석 모델 수립의 난이도 등에 따라 투입공수를 보정한다. [표 2-1]은 빅 데이터 분석 사업의 보정요소를 나타낸다.

| 보정계수의 종류 | 설명 | | 적용 단계 |
|--------------------------|--|--|---------|
| (1) 데이터 규모에 따른 보정계수 | 분석에 사용되는 원천 데이터의 종수와 양, 보유기관 수 | | 데이터 수집 |
| (2) 데이터 처리기술 적용에 따른 보정계수 | 데이터 전처리 단계의 요소 기술로 축소, 여과, 변환, 정제, 통합, 개인정보 비식별화 등 | | 데이터 전처리 |
| (3) 분석모델 수립의 난이도 보정계수 | 업무 분야 특성 | 분석모델 수립 시 일반 분야, 전문 분야, 현안의 다양성이 큰 분야 등 업무 분야의 특성에 따라 측정 | 분석 모델링 |
| | 사용 데이터 종류 | 사용하는 데이터가 공공분야인지 민간분야인지, 정형인지 비정형인지에 따라 복잡도 측정 | |
| (4) 분석기술 적용에 따른 보정계수 | 분석 목적을 달성하기 위해 적용할 빅데이터 분석 기술에 따른 복잡도 측정 | | 모델 구현 |

[표 2-1] 빅 데이터 분석 사업 보정요소

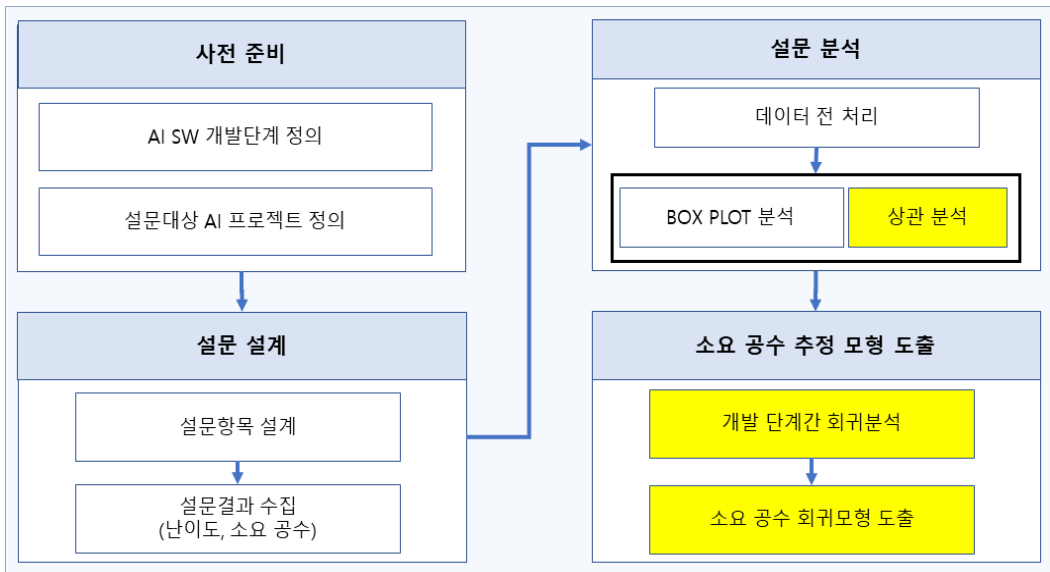
인공지능 기반 소프트웨어 개발 소요공수 산정을 위해서도 위와 같이 보정요소를 발견하고 전문가 의견을 수렴하여 확정하여야 했으나 시간과 인력의 한계로 본 논문에서는 보정요소에 대한 부분은 제외하기로 한다.

Ⅲ. AI 기반 SW 개발 비용 산정에 관한 소요공수 예측 모형

본 장에서는 인공지능 기반 소프트웨어의 개발단계별 소요공수에 대한 현장의 설문을 수행하고 수집된 설문 결과 분석하여 개발 단계별 소요공수 예측 모형을 제안하고자 한다.

A. 시스템 구성도

[그림 3-1]은 제안하는 인공지능 기반 소프트웨어 개발 소요공수 예측 모형의 전체 시스템 구성도이다. 현장의 설문을 수행하기 위한 사전 준비단계와 설문 진행, 설문결과를 분석하는 분석단계, 분석 결과를 바탕으로 소요공수 예측 모형을 도출하는 단계와 해당 모형을 검증하는 단계로 나뉜다.



[그림 3-1] 시스템 구성도

사전 준비 과정은 인공지능 기반 소프트웨어 소요공수 산출을 위한 세부 개발단계를 사전에 정의하며 설문자 입장에서 개발 단계별 소요공수 산정이 좀 더 실제

적일 수 있도록 설문 대상 인공지능 기반 소프트웨어 개발 프로젝트를 사전에 정의한다. 설문 설계 과정은 설문을 위한 항목을 설계하며 결과를 수집한다. 결과 수집의 용이성과 신속성을 위해 Google 설문지를 활용하였다.

설문 분석 과정은 수집된 데이터에 대해 무응답 항목 등에 대하여 삭제하고 응답결과가 TEXT로 처리된 건에 대하여는 수치화한다. 상관분석을 위한 데이터의 포맷 변환 등 전처리 과정을 거치고 시각화된 Box Plot 분석을 통해 데이터의 분포 및 중위값을 확인한다. 다음으로는 기초통계량 분석 및 개발단계간의 소요공수에 대한 상관 분석을 실시한다.

소요공수 예측 모형 도출 과정은 개발단계간의 상관분석 결과를 토대로 상관계수가 0.7이상 높게 나오는 개발단계를 도출하고 상호간의 회귀분석을 통해 회귀모형을 도출하게 된다.

마지막으로 모형 검증 과정은 도출된 회귀모형으로 개발단계별 소요공수를 예측하고 수집된 데이터 및 업체로부터 견적 받은 소요공수를 비교 검증하는 과정을 거친다.

B. 사전 준비

1. AI 기반 SW 개발단계 정의

본 절에서는 인공지능 기반 소프트웨어의 소요공수를 산정하기 위한 개발단계에 관하여 기술한다. 현재까지 인공지능 기반 소프트웨어의 개발절차에 대하여는 어디에서도 정의되어 있지 않다. 그래서 유사한 성격의 빅데이터 분야가 보유하고 있는 개발단계와 업무 내용을 분석하여 활용하고자 한다.

행안부와 NIA에서는 빅데이터 사업비 산정에 대한 표준을 제공하고자 2017년에 공공 빅데이터 분석 사업비 대가 산정가이드를 발간하였다. 이는 2년여 기간에 걸쳐 산업계의 의견을 반영한 결과물이다. 해당 가이드에는 빅데이터 분석 사업에 대한 단계를 정의하고 사업비 산정을 위한 기준과 절차를 제공하고 있다. 본 논문에서는 빅데이터 분석 사업 수행 단계를 통해 인공지능 기반 소프트웨어 개발사업의 개발단계로 활용하고자 한다. 공공부문에서 기 추진된 빅데이터 분석 사업의 산출 내역서 및 작업명세서(WBS : Work Breakdown Structure)를 분석하여 표준화한 공공 빅데이터 분석 사업의 추진 단계는 [그림 2-3]과 같다[14].

[그림 2-3]을 참고하여 빅데이터 분석 사업의 단계별 업무 수행내용을 파악하고 AI SW 개발단계의 수행내용으로 적용가능한지 검토하였다. [표 3-1]은 빅데이터 분석 사업의 단계별 수행 내용을 정리한 것이다.

| 빅 데이터 분석 사업 수행 단계 | 수행 내용 | AI SW 개발 단계 적용 |
|-------------------|--|----------------|
| 요건 정의 | 빅데이터 분석과 활용을 통해 해결하고자 하는 현안에 대해 명확히 이해하고 정의하며, 사업 추진을 위해 필요한 환경을 분석 하고 사업의 추진 계획을 수립하는 과정을 포함함 | 적용 가능 |
| 데이터 수집 | 분석 데이터를 정의하고 수집 환경을 구성하여 데이터를 수집함 | 적용 가능 |

| | | |
|----------|--|-------|
| 데이터 전처리 | 수집된 데이터를 파악하여 데이터의 오류를 점검하고, 데이터를 정제함과 동시에 매칭 key 값을 기준으로 여러 데이터를 융합하고 기초 통계 분석을 통해 데이터의 고유 패턴을 파악함. 분석 대상 업무가 1회성이 아닌, 지속적으로 서비스를 수행해야 하는 경우 빅데이터 포털이나 기관 자체 인프라를 활용하여 데이터 수집, 정제, 표준화, 융합이 자동화 되도록 시스템 연계를 추진하고, 모듈을 설계하고 구현 | 적용 가능 |
| 분석 모델링 | 분석 대상 및 범위를 정하여 분석 목적을 구현하기 위한 분석 방법론을 설계하며 분석모델 설계가 타당한지 자문위원회 회의를 거쳐 논의하고, 내용을 추가 반영하여 분석모델 설계 완성 | 적용 가능 |
| 모델 구현 | 분석모델링 설계를 알고리즘으로 구현하기 위해 알고리즘 모델링을 설계하고 모듈을 개발하여 분석 알고리즘을 구현함 | 적용 가능 |
| 시각화 | 분석가 이외의 사람들이 이해하기 쉽도록 그림 및 도표를 활용하여 표현하며 수집한 데이터 및 분석 결과를 담당자 및 관련 기관들과 공유하기 위한 목적으로 필요하다면 Open-API 형태로 제공하도록 개발함 | 적용 가능 |
| 검증 및 안정화 | 사업 전반에 대한 결과물(시각화, 분석 프로세스 자동화 등)을 통합 테스트하여 검증하고, 사용자 테스트 후 검수를 거쳐 분석 결과를 사용할 실무자를 위하여 교육을 실시하고 인수인계를 함 | 적용 가능 |

[표 3-1] 빅데이터 분석 사업 단계별 수행 내용 검토

[표 3-1]과 같이 검토한 결과 빅데이터 분석 사업 수행의 7단계와 AI SW 개발 수행단계가 유사하여 소요공수를 파악하기 위한 AI SW 개발단계로 정의하였다.

2. 설문대상 AI 기반 SW 정의

설문의 신뢰도를 높이기 위해서는 특정 AI 기반 SW에 한정하여 소요 공수를 답변할 수 있도록 설문의 편의성을 고려하는 것이 필요하여 아래와 같이 정의하였다.

- 설문대상 AI SW : 품셈¹⁾ 공량 예측 인공지능 기반 소프트웨어
- 필요 기능 : 신규로 등록한 품셈의 내용과 기구축된 품셈의 내용을 비교하여 소요 공량을 예측한다. 비교하기 위해 자연어 처리를 해야 함

| 품셈 이름 | | | | |
|--------------------------|----------------------------|--------|--------|--------|
| 2-8-13 안전난간대 설치 및 해체(토목) | | | | |
| 공량 (10m당) | | | | |
| 구분 | 단위 | 2단 | 3단 | |
| 비계공 | 인 | 0.62 | 0.67 | |
| 비교 | - 난간기둥 간격에 따라 다음 효율을 적용한다. | | | |
| | 설치간격 | 1.0m이하 | 1.5m이하 | 1.5m초과 |
| | 효율 | 110% | 100% | 90% |

[그림 3-2] 표준 품셈 예시

[그림 3-2]와 같이 “안전난간대 설치 및 해체”라는 품셈이며 소요 공량은 10m당 2단일 경우에는 0.62인, 3단일 경우에는 0.67인의 “비계공” 인력이 소요된다는 것을 알 수 있다. 또한, 설치간격이 1m이하일 경우에는 110%, 1.5m 이하일 경우에는 100%, 1.5m 초과일 경우에는 90%의 효율을 가진다.

새로운 품셈을 생성 등록할 경우, 기존 품셈의 업무내용과 소요공량을 분석 비교하여 소요공량을 제시하는 것이 해당 AI SW의 요구 기능이다.

1) 품셈 : 인력이나 기계로 어떤 물체를 만드는 데 필요한, 단위당 노력과 능률 및 재료 등을 수량으로 나타낸 것

C. 설문 설계

1. 설문 항목 설계

AI SW의 개발단계별 난이도와 소요공수를 좀 더 정확히 파악하고자 발주자와 수주자에게 설문을 수행하였으며 설문자의 역할도 단순 개발자가 아닌 영업대표, 기획자, PM, 기타 등 AI SW와 관련 업무를 수행하는 모든 인력들에 대해 포괄적으로 설문을 진행하고자 노력하였다. [그림 3-3]은 설문 of 구성을 나타낸다.



[그림 3-3] 설문 of 구성

“개발 단계별 난이도”에 대하여는 개발 단계별 세부 TASK에 [그림 3-4]와 같이 설문을 설계하고 수행하였다.

| 수행 활동 | 수행활동별 난이도 | | | | |
|--------|-----------|----|----|----|----|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 |
| WBS 작성 | | | | | |

[그림 3-4] 수행활동 난이도 설문 설계 예시

- 전문 : 업무의 난이도가 특급기술자가 수행할 정도의 업무
- 복잡 : 업무의 난이도가 고급기술자가 수행할 정도의 업무
- 보통 : 업무의 난이도가 중급기술자가 수행할 정도의 업무
- 단순 : 업무의 난이도가 초급기술자가 수행할 정도의 업무
- 필요없음 : AI SW 개발에 필요 없는 업무

“개발 단계별 소요공수”에 대하여는 개발 단계별 세부 TASK에 대하여 적정 소요공수에 대한 설문을 받고자 하였다. 제작한 설문지는 [그림 3-5]와 같다.

또한, 추가 작업이 필요한 경우에는 해당 작업을 공백란에 기재할 수 있도록 구성하였다.

[세부 설문 내용]

※ 작성방법 : 단계별 수행활동에서 난이도와 소요공수를 기재해 주시고
 추가 작업 필요시 공백란에 기재하셔서 소요공수를 표시해 주세요

[난이도 선택 기준]

- 전문 : 특급기술자 투입이 필요한 작업, 복잡 : 고급기술자 투입이 필요한 작업
- 보통 : 중급기술자 투입이 필요한 작업, 단순 : 초급기술자 투입이 필요한 작업

1. 요구 정의 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

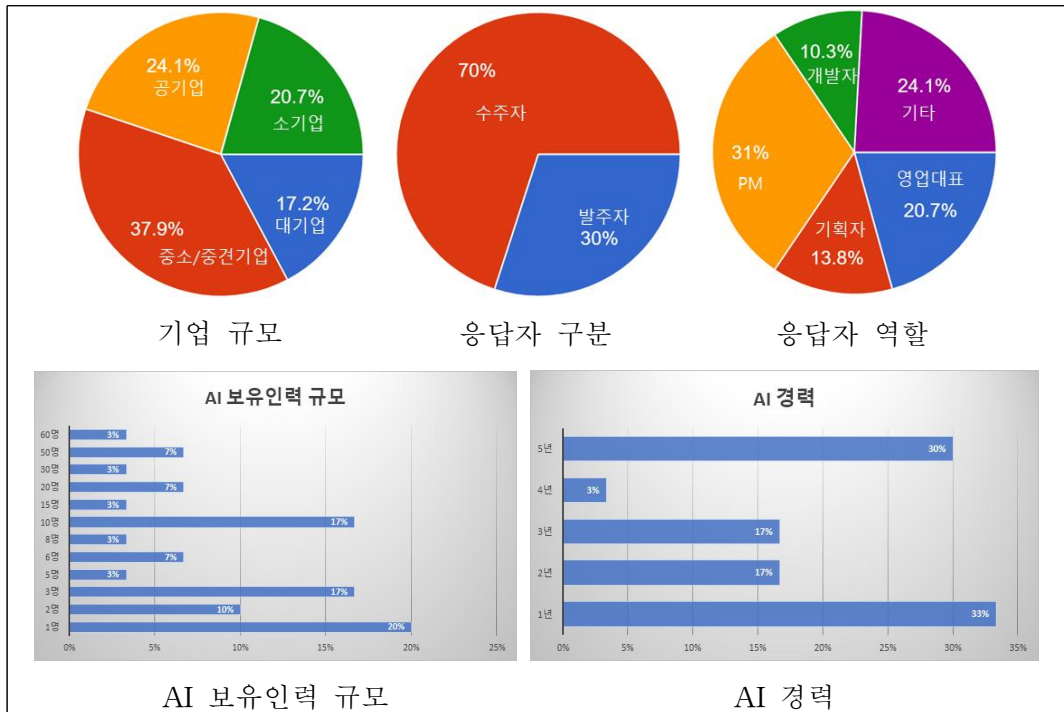
| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) 예시) 2명 x 10일 = 20 |
|----------------------------|-----------|----|----|----|----|--|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| WBS 작성 | | | | | | |
| 실무자 인터뷰 계획 | | | | | | |
| 실무자 인터뷰 질의서 작성 | | | | | | |
| 요구사항정의서 작성 | | | | | | |
| 관리자급 인터뷰 | | | | | | |
| 선행 사업 및 자료 분석 | | | | | | |
| 공통기반 플랫폼 인프라 및 소프트웨어 현황분석 | | | | | | |
| 수집대상 데이터 현황분석 | | | | | | |
| 데이터 수집 환경 분석 | | | | | | |
| 성과목표 및 성과 측정시기별 계획안 마련 | | | | | | |
| 성과관리를 위한 지표 및 운영방안 마련 | | | | | | |
| 과제 수행 결과 활용 계획 및 확산 계획안 작성 | | | | | | |
| | | | | | | |

[그림 3-5] 수행활동 소요공수 설문 설계 예시

2. 설문 결과 수집

설문 수집의 신속성을 고려하여 Google 설문지를 활용하였으며 대략 60여 명의 AI업계 관련 담당자들에게 설문 Link 및 설문 양식지를 발송하였으나 응답자는 30명으로 50%의 저조한 응답률을 보인다. 그러나 응답자는 고르게 분포되어 있다. [그림 3-7]은 설문 결과를 나타낸다.

- 설문 대상 : AI관련 19개 업체 249명 중의 27명을 통한 FGI²⁾(집단심층면접) 설문
- 설문 기간 : 4월 25일 ~ 4월 29일 (5일간)
- 설문 소속사 보유인력규모 : 1명~60명
- 설문 방식 : 인터넷 Google 설문



[그림 3-6] 설문 결과

2) FGI(Focus Group Interview) : 동질적인 특성을 지닌 소수의 조사 대상자로부터 다양하고 심층적인 의견을 청취하는 방법

[그림 3-7]의 설문 결과를 응답자 비율 순서대로 정리해 보면 다음과 같다.

- 기업규모 : 중소/중견기업 > 공기업 > 대기업 > 소기업
- 응답자 구분 : 수주자 > 발주자
- 응답자 역할 : PM > 영업대표 > 기획자 > 개발자 > 기타
- AI 보유인력 규모 : 1명 > 3명, 10명 > 2명 > 6명, 20명, 50명 > 5명, 8명, 15명, 30명, 60명
- AI 경력 : 1년 > 5년 > 2년,3년 > 4년

D. 설문 분석

수집된 설문 데이터에 대해 분석 환경을 기반으로 분석 단계를 실행하였다. [표 3-2]는 설문 데이터의 분석 내용을 나타낸다.

- 분석 환경 : Jupyter Notebook을 활용하여 Python 프로그램 활용

| 분석 단계 | 수행 내용 | 필요성 |
|-------------|--|-----------------|
| 데이터 전처리 | <ul style="list-style-type: none"> • AI SW 수행단계 이름의 코드화 • 난이도 이름의 코드화 등 | Python 분석 |
| Box Plot 분석 | <ul style="list-style-type: none"> • 난이도 분석, 소요공수 분석 • 세부 수행 단계별 설문결과의 데이터의 분포 및 Median-Value 확인 | 중간 값 확인 |
| 상관 분석 | <ul style="list-style-type: none"> • 소요공수 분석 • 개발 단계별 소요공수 확인 및 상관 분석 수행 | 개발 단계 간 상관관계 확인 |

[표 3-2] 설문 데이터 분석 내용

1. 데이터 전처리

엑셀 파일 형태인 수집된 설문 결과를 효과적으로 분석하기 위해 설문 결과 항목에 대한 영문 코드화 작업을 진행하였다. 난이도 설문결과에 대하여는 숫자로 변환하여 전처리 작업을 진행하였으며 설문의 결과 칼럼에 대하여는 Python 프로그램으로의 효과적인 통계처리를 위해 [표 3-3]과 같이 칼럼에 대한 영문 명명 규칙을 설계하고 적용하여 CSV(comma-separated values) 파일 형태로 변환하였다.

• 난이도 설문결과 전처리

- 필요없음 : 0, 단순 : 1, 보통 : 2, 복잡 : 3, 전문 : 4

| 단계 | 난이도 칼럼 이름 | 공수 칼럼 이름 | 공수합계 칼럼 추가 | 비고 |
|----------|--------------|-------------|---------------|--|
| 요건 정의 | Rn | Rn_MD | RSUM_MD | R : Request n : 1 ~ 12 |
| 데이터 수집 | DCn | DCn_md | DCSUM_MD | DC: Data Collection n = 1 ~ 3 |
| 데이터 전처리 | DPn | DPn_md | DPSUM_MD | DP : Data Pre-processing n = 1 ~ 7 |
| 분석 모델링 | AMn | AMn_md | AMSUM_MD | AM : Analysis Modeling n = 1 ~ 3 |
| 모델 구현 | ADn | ADn_md | ADSUM_MD | AD : Analysis model Development n = 1 ~ 5 |
| 시각화 | Vn | Vn_md | VSUM_MD | V : Visualization n = 1 ~ 5 |
| 검증 및 안정화 | Sn | Sn_md | SSUM_MD | S : verification and Stabilization n = 1 ~ 4 |

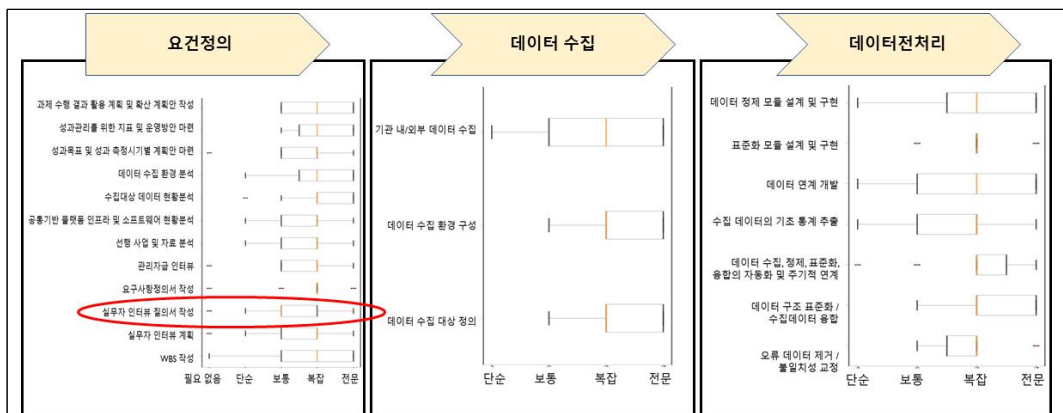
[표 3-3] 설문항목 통계처리를 위한 칼럼별 영문 명명 규칙 설계

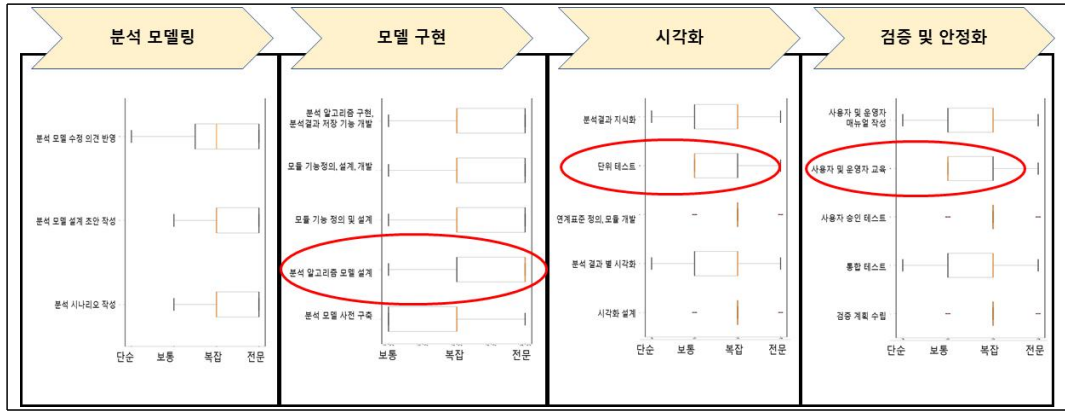
2. Box Plot 분석

전처리된 CSV 데이터 파일을 이용하여 개발 단계별 난이도와 단계별 소요공수를 데이터 분포와 중위값 확인을 위해 Box Plot으로 분석하였다. [표 3-3]은 Box Plot 난이도 분석을 위한 Python Code를 나타내고, [그림 3-7]은 Box Plot 난이도 분석 결과를 나타낸다.

| | |
|-------------------|---|
| 난이도 분석 | <pre> pd.set_option('display.max_columns', None) data2=pd.read_csv("diff_cd.csv") # 단계별 난이도 설문 결과 box_data = [data2['r1'], data2['r2'], data2['r3'], data2['r4'], data2['r5'], data2['r6'], data2['r7'], data2['r8'], data2['r9'], data2['r10'], data2['r11'], data2['r12']] #요건정의 단계 난이도 green_diamond = dict(markerfacecolor='r', marker='s') plt.boxplot(box_data,vert=False, flierprops=green_diamond) plt.show() # 요구정의 단계 난이도에 대한 Box Plot 출력 </pre> |
|-------------------|---|

[표 3-4] Box Plot 난이도 분석을 위한 Python Code





[그림 3-7] Box Plot 난이도 분석 결과

- 난이도의 설문 데이터의 분포가 가장 넓은 업무는 “WBS 작성단계”로서 필요 없다에서 전문까지 가장 넓게 응답결과가 분포되고 있으며 난이도가 “복잡”하다는 답변이 중앙값을 차지하고 있다.
- 반대로, 설문 데이터의 분포가 가장 좁은 업무는 “복잡”하다는 난이도에 수렴되는 아래의 업무이며, 현업 담당자들의 의견이 일치한다고 볼 수 있다.

- 요건정의 단계의 “요구사항 정의서 작성”
- 데이터 전처리 단계의 “표준화 모듈 설계 구현”
- 시각화 단계의 “시각화 설계”와 “연계표준 정의, 모듈 개발”
- 검증 및 안정화 단계의 “검증계획수립”과 “사용자 승인 테스트”

업무 난이도의 최종 분석결과와는 [그림 3-8]의 붉은색 타원형 외에는 “복잡”, 필요 없는 업무는 없다고 집계되었으며, 추가되어야 하는 업무도 거의 없다고 집계 되었다.


```

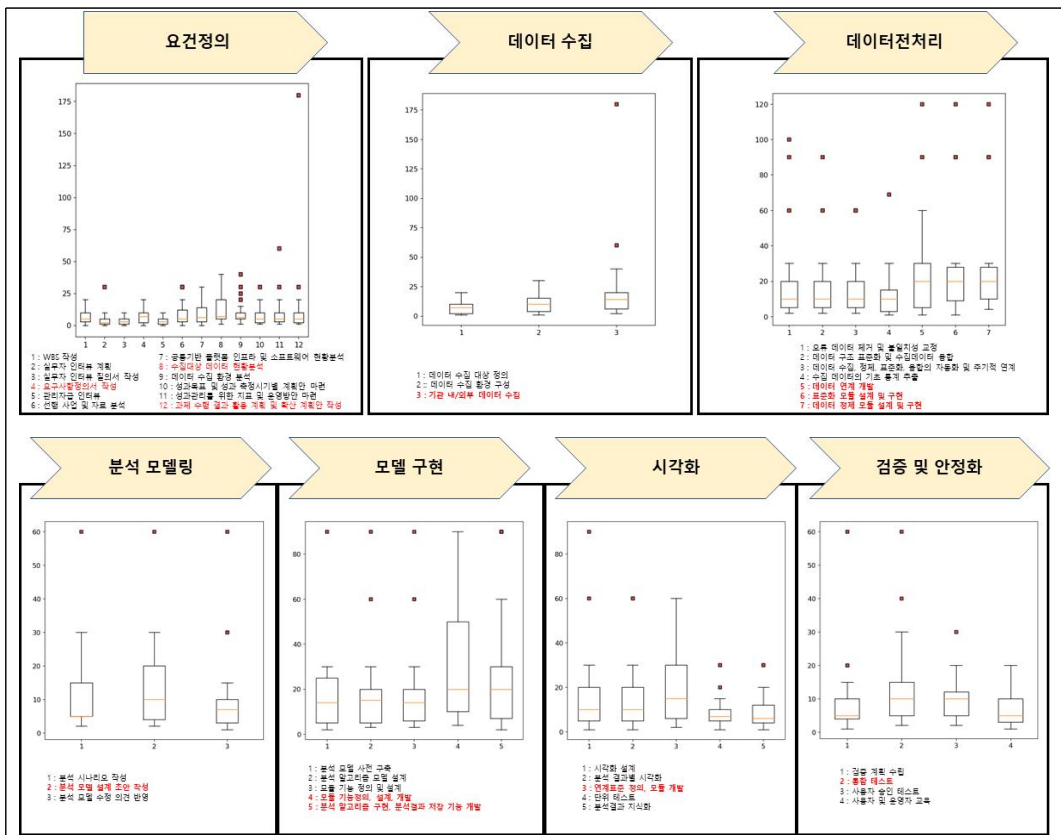
pd.set_option('display.max_columns', None)
data4=pd.read_csv("md.csv")# 단계별 소요공수 설문 결과

box_data4 = [ data4['r1_md'],data4['r2_md'],data4['r3_md'],
data4['r4_md'],data4['r5_md'],data4['r6_md'],data4['r7_md'],
data4['r8_md'],data4['r9_md'],data4['r10_md'],data4['r11_md'],
data4['r12_md']] #요건정의 단계 소요공수

green_diamond = dict(markerfacecolor='r', marker='s')
plt.boxplot(box_data4,flierprops=green_diamond)
plt.show() # 요구정의 단계 소요공수에 대한 Box Plot 출력
    
```

**소요공수
분석**

[표 3-5] Box Plot 소요공수 분석을 위한 Python Code



[그림 3-8] Box Plot 소요공수 분석 결과

난이도에 대한 설문결과와는 달리, 소요공수의 설문결과는 응답 범위가 상당히 넓게 나타나는 것이 특징이며 가장 넓은 응답 범위를 보이는 것은 “데이터 전처리” 단계임을 확인할 수 있다. [표 3-6]은 각 단계별 세부 업무 소요공수의 합을 기준으로 최소, 최대, 중앙값을 조사한 표이다.

| 단계 | 최소 | 최대 | 설문 응답 범위 | 중앙값 | 비중 |
|----------|----|-------|----------|-----|------|
| 요건정의 | 16 | 310 | 294 | 75 | 18% |
| 데이터 수집 | 4 | 220 | 216 | 30 | 7% |
| 데이터 전처리 | 21 | 609 | 588 | 108 | 27% |
| 분석 모델링 | 6 | 180 | 174 | 24 | 6% |
| 모델 구현 | 19 | 450 | 431 | 85 | 21% |
| 시각화 | 13 | 240 | 227 | 50 | 12% |
| 검증 및 안정화 | 8 | 160 | 152 | 34 | 8% |
| 합계 | 87 | 2,169 | 2,082 | 406 | 100% |

[표 3-6] 단계별 소요공수 설문 결과

[표 3-6]에서 알 수 있듯이

- “비중” 항목을 토대로 작업량이 많은 순서대로 개발 단계를 나열하면 데이터 전처리(27%) > 모델 구현(21%) > 요건정의(18%) > 시각화(12%)로 이어짐을 확인할 수 있다. 개발 7 단계 중 “데이터 전처리”와 “모델 구현”의 합계 비중이 48% 수준으로 다른 개발 단계에 비해 많은 비중을 차지하는 것을 확인할 수 있다.
- “설문 응답 범위” 항목은 변동성이 많은 순서대로 개발 단계를 나열하면 데이터 전처리(588) > 모델 구현(431) > 요건정의(294) > 시각화(227)로 작업량의 순서와 동일하다.

소요공수에 대한 최종 분석 결과, 데이터 전처리 단계와 모델 구현 단계가 변동성이 크고 AI SW의 개발단계의 가장 많은 비중을 차지하는 것을 확인할 수 있다. 이는 해당 단계의 소요공수 관리를 얼마나 잘하느냐가 개발비의 적정성에 영향을 미침을 알 수 있다.

3. 상관 분석

요르돈의 법칙, 일명 Snow Ball Effects에 따르면 소프트웨어 개발 초기에 체계적인 분석 및 설계가 수행되지 못하면 그 결과가 이후 단계에 영향을 미쳐 비용이 커진다는 법칙이다. 이 법칙은 개발 초기 단계와 이후 단계의 상관관계에 대해 언급한다고 할 수 있다.

이에 따라 AI SW 개발의 요건정의 초기 단계의 소요공수가 이후 단계의 소요공수에 영향을 미친다고 가정해 본다면 초기 “요건정의” 단계와 이후 개발단계와의 상관성을 분석하여 상관관계가 큰 개발단계들을 추출하여 상관성을 분석하고자 한다.

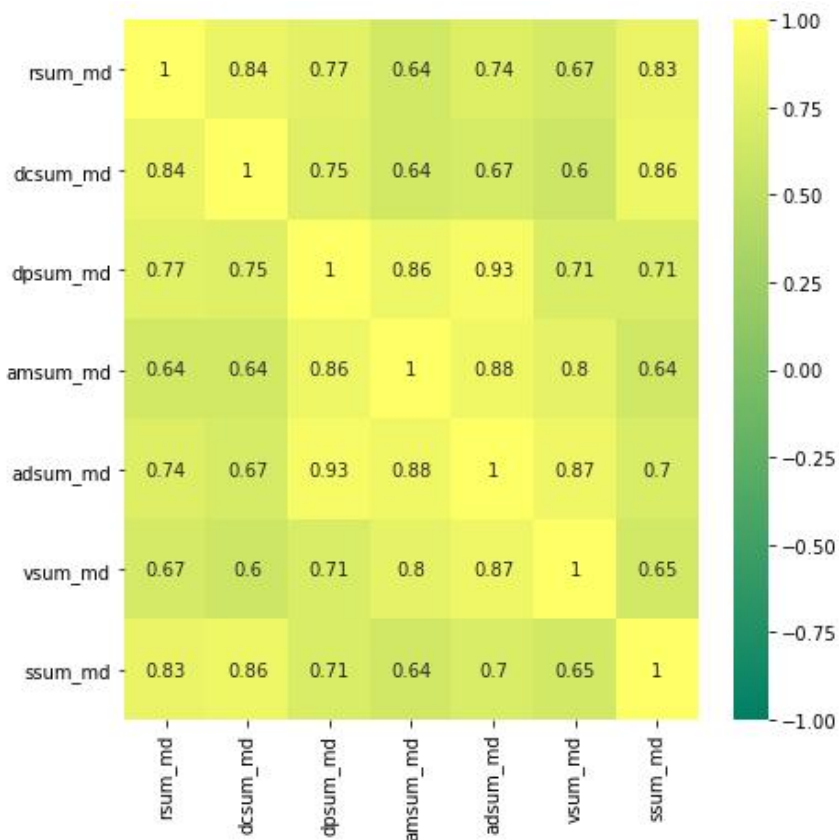
아래는, 개발 단계별 소요공수 합을 입력받아 개발 단계 상호간의 상관분석을 실시하고 HEAT MAP 출력을 통해 상관관계를 시각화 한다.

| | |
|------|--|
| 상관분석 | <pre> import pandas as pd import seaborn as sb import matplotlib.pyplot as plt data=pd.read_csv("md_sum.csv") # 단계별 소요공수 합계 lst=[data.rsum_md,data.dcsmd,data.dpsum_md,data.amsum_md, data.adsum_md,data.vsum_md,data.ssum_md] df=pd.DataFrame(lst).T corr=df.corr(method='pearson') # pearson 상관 분석 corr # 상관분석 결과 출력 # 상관분석 결과 HEAT MAP 출력 plt.rcParams["figure.figsize"] = (7,7) sb.heatmap(df.corr(), annot = True, #실제 값 화면에 나타내기 cmap = 'Blues', #색상 vmin = -1, vmax=1 , #컬러차트 영역 -1 ~ +1) </pre> |
|------|--|

[표 3-7] 상관분석을 위한 Python Code

| | rsum_md | dcsum_md | dpsum_md | amsum_md | adsum_md | vsum_md | ssum_md |
|----------|----------|----------|----------|----------|----------|----------|----------|
| rsum_md | 1.000000 | 0.841357 | 0.773467 | 0.641090 | 0.740947 | 0.671045 | 0.833465 |
| dcsum_md | 0.841357 | 1.000000 | 0.749297 | 0.642153 | 0.673877 | 0.604615 | 0.860543 |
| dpsum_md | 0.773467 | 0.749297 | 1.000000 | 0.862848 | 0.931280 | 0.706339 | 0.707481 |
| amsum_md | 0.641090 | 0.642153 | 0.862848 | 1.000000 | 0.884530 | 0.799490 | 0.637878 |
| adsum_md | 0.740947 | 0.673877 | 0.931280 | 0.884530 | 1.000000 | 0.874300 | 0.704025 |
| vsum_md | 0.671045 | 0.604615 | 0.706339 | 0.799490 | 0.874300 | 1.000000 | 0.645485 |
| ssum_md | 0.833465 | 0.860543 | 0.707481 | 0.637878 | 0.704025 | 0.645485 | 1.000000 |

[그림3-9] 개발 단계간 Pearson 상관분석 결과



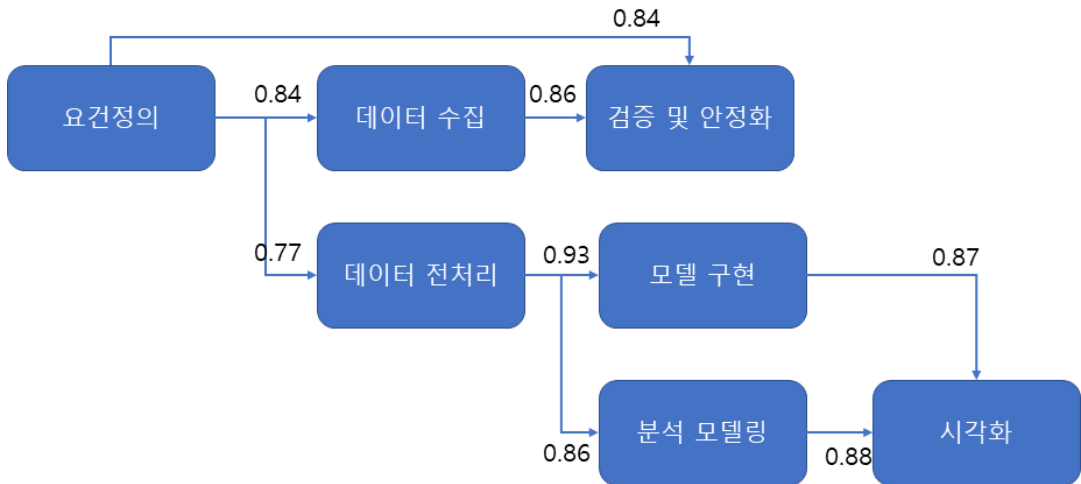
[그림 3-10] 개발 단계 간 Heat Map 상관분석 결과

[표 3-8]은 [그림 3-10]에서 분석한 상관관계 중 상관계수가 0.77 이상의 강한 상관관계를 갖는 개발 단계를 나타낸다.

| 순위 | R(상관계수) | 개발 단계 1 | 개발 단계 2 |
|----|---------|---------|----------|
| 1 | 0.93 | 데이터 전처리 | 모델 구현 |
| 2 | 0.88 | 분석 모델링 | 시각화 |
| 3 | 0.87 | 모델 구현 | 시각화 |
| 4 | 0.84 | 요건 정의 | 데이터 수집 |
| 5 | 0.86 | 데이터 수집 | 검증 및 안정화 |
| 5 | 0.86 | 데이터 전처리 | 분석 모델링 |
| 6 | 0.83 | 요건 정의 | 검증 및 안정화 |
| 7 | 0.77 | 요건 정의 | 데이터 전처리 |

[표 3-8] 개발 단계 간 소요공수의 강한 상관관계 목록

[표 3-8]을 기반으로 개발단계간의 상관관계를 [그림 3-10]과 같이 표현하였다.



[그림 3-11] 개발 단계 간 소요공수 상관관계도(圖)

[그림 3-11]과 같이 요건정의 단계의 소요공수는 데이터 수집과 데이터 전처리에 영향을 미치며, 데이터 수집은 검증 및 안정화에 데이터 전처리는 모델구현과 분석 모델링에, 분석 모델링은 시각화 소요공수에 영향을 미친다고 할 수 있다.

이 후 단계에서는, 상관관계도(圖)를 기반으로 개발 단계간의 회귀 분석을 시행하고 단계별 소요공수 예측 회귀 모델을 도출하고 검증하고자 한다.

E. 개발단계별 소요공수 예측 모형

1. 회귀 분석

개발단계간의 상관분석을 기반으로 본 단락에서는 회귀분석을 실시하였다. 회귀 분석은 독립변수가 종속변수에 어떤 영향을 미치는지 알아보고자 할 때 실시하는 분석방법으로서 3)상관계수가 높은 개발 단계 간에 [표 3-9]와 같이 8종류의 단순 선형 회귀분석을 시행하여 개발 단계별 소요공수를 예측하는 모형을 도출하였다.

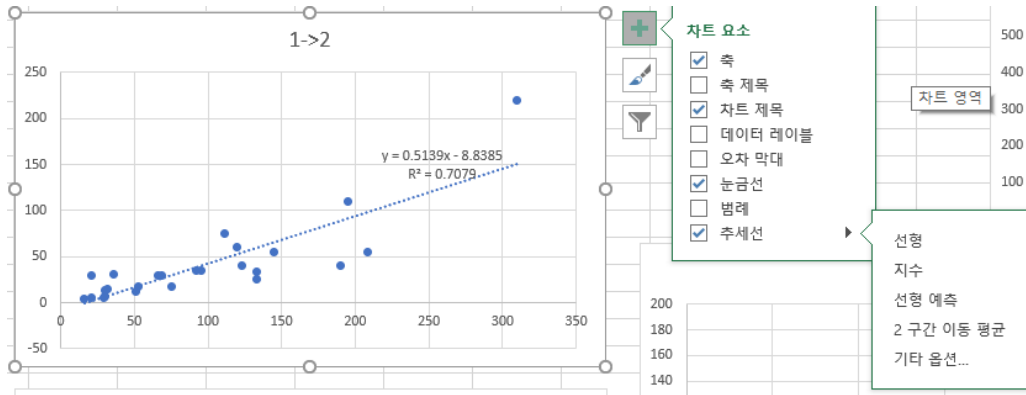
| 독립 변수 | 종속 변수 | 회귀 모형 도출 |
|--------------|---------------|------------------------|
| 요건 정의 소요공수 | 데이터 수집 소요공수 | 데이터 수집 소요공수 예측 모형 |
| 요건 정의 소요공수 | 데이터 전처리 소요공수 | 데이터 전처리 소요공수 예측 모형 |
| 데이터 수집 소요공수 | 검증 및 안정화 소요공수 | 검증 및 안정화 소요공수 예측 모형 |
| 요건 정의 소요공수 | 검증 및 안정화 소요공수 | |
| 데이터 전처리 소요공수 | 모델 구현 소요공수 | 모델 구현 소요공수 예측 모형 |
| 데이터 전처리 소요공수 | 분석 모델링 소요공수 | 분석 모델링 소요공수 예측 모형 |
| 분석 모델링 소요공수 | 시각화 소요공수 | 시각화 소요공수 예측 모형 |
| 모델 구현 소요공수 | 시각화 소요공수 | |

[표 3-9] 회귀분석 시행 목록

3) 상관계수 : 두 변수 x, y 사이의 상관관계의 정도를 나타내는 수치

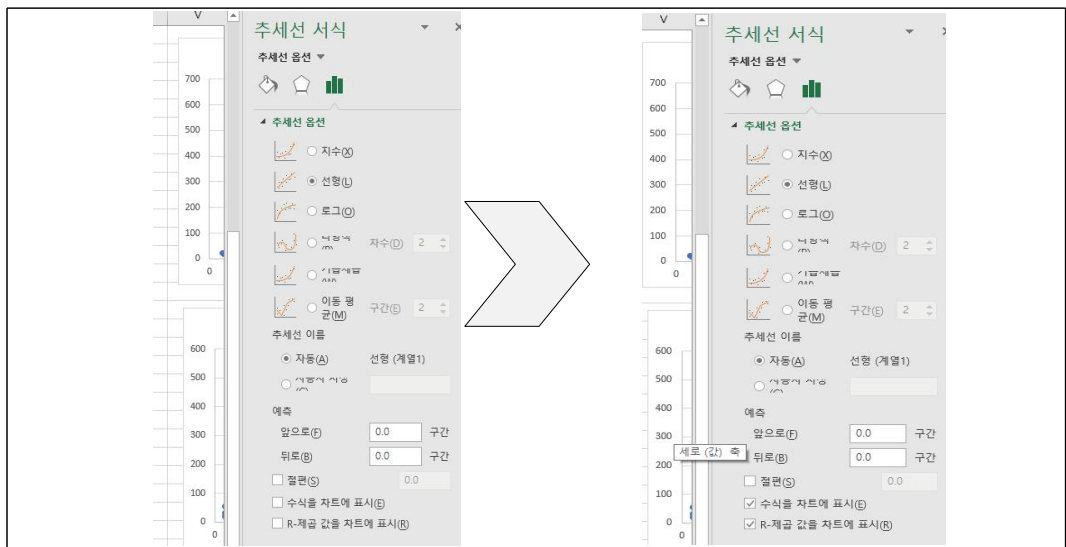
2. 소요공수 예측 회귀모형

[표 3-9]의 회귀분석 시행 목록에 의거하여 개발 단계 간 소요공수에 대한 단일 선형 회귀분석을 시행하였고 소요공수 예측 회귀모형을 도출하였다.



[그림 3-12] 회귀식 도출을 위한 방법(1)

x축, y축에 대한 그래프를 엑셀에서 삽입한 후에, [그림 3-12]와 같이 그래프의 우측 상단에 **+** 버튼을 클릭하고 메뉴중에서 “추세선”을 클릭하고 우측 화살표를 클릭, “기타옵션”이라는 메뉴를 클릭한다. [그림 3-13]과 같이 추세선에 대한 서식을 설정할수 있다. 여기서 “선형”에 클릭, “수식을 차트에 표시” 클릭, “R-제공 값을 차트에 표시” 클릭한다.

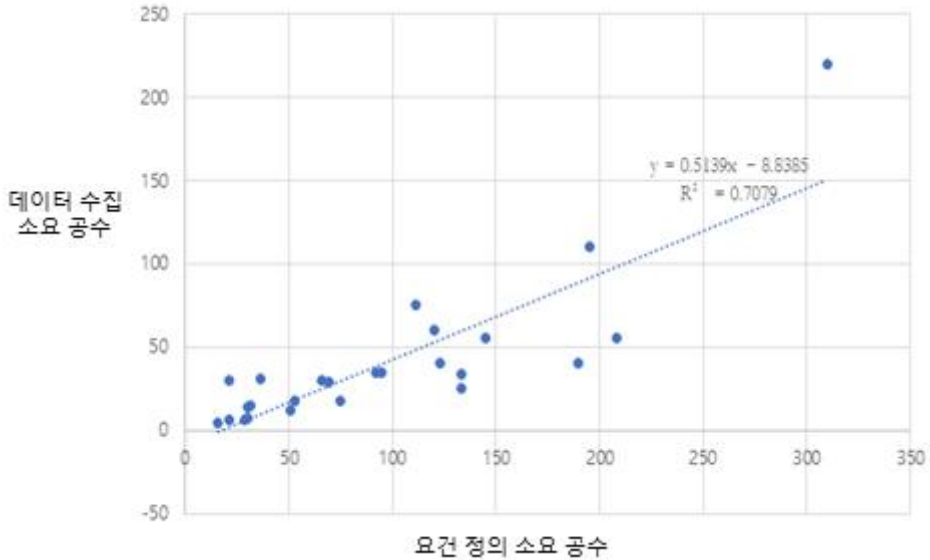


[그림 3-13] 회귀식 도출을 위한 방법(2)

- 데이터 수집 소요공수 예측 회귀모형

요건정의 소요공수와 데이터 수집의 소요공수에 인과 관계를 갖는다고 가정하고 회귀분석을 시행한 결과 [그림 3-14]와 같은 회귀선과 회귀식이 도출되었다.

- 회귀 분석 대상 : X축 요건정의 소요공수, Y축 데이터 수집 소요공수



[그림 3-14] 요건정의와 데이터 수집의 소요공수 회귀분석

- 데이터 수집 소요공수 예측 회귀식

$$= (0.5139 * \text{요건정의 소요공수}) - 8.8385$$

- 회귀식의 4)R²(결정계수) = **0.7079**

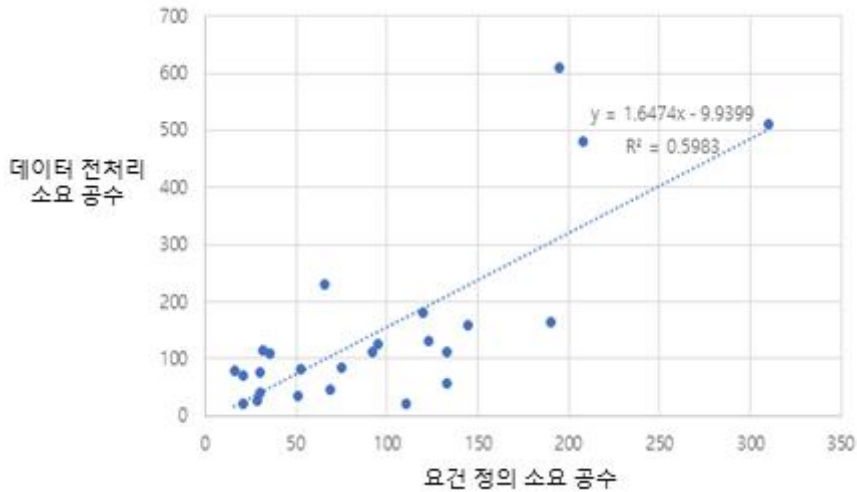
예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.7079로 확인 되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 높다고 할 수 있다.

4) R²(결정계수) : 표본자료로부터 예측된 회귀식이 그 측정 자료에 어느 정도 적합한가를 측정하는 척도

- 데이터 전처리 소요공수 예측 회귀 모형

요건정의 소요공수와 데이터 전처리 소요공수가 인과 관계를 갖는다고 가정하고 회귀분석을 시행한 결과 [그림 3-15]와 같은 회귀선과 회귀식이 도출되었다.

- 회귀 분석 대상 : X축 요건정의 소요공수, Y축 데이터 전처리 소요공수



[그림 3-15] 요건정의와 데이터 전처리의 소요공수 회귀분석

- 데이터 전처리 소요공수 예측 회귀식

$$= (1.6474 * \text{요건정의 소요공수}) - 9.9399$$

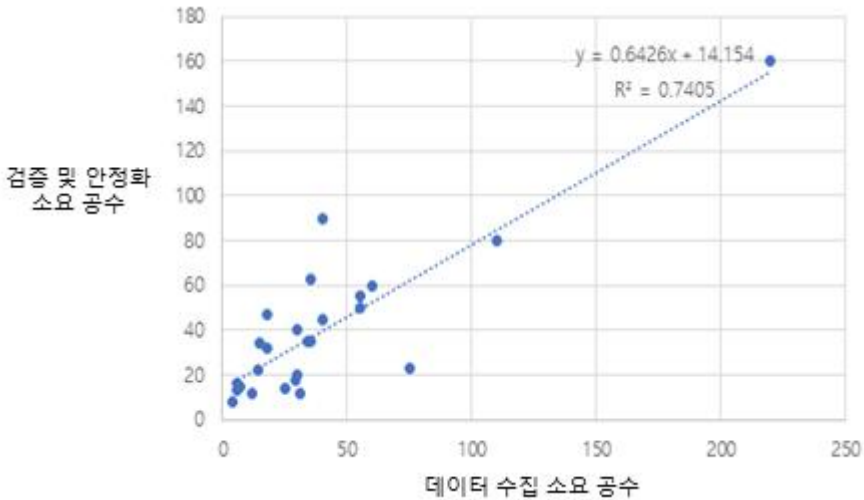
- 회귀식의 R^2 (결정계수) = **0.5983**

예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.5983으로 확인 되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 보통이라 할 수 있다.

- 검증 및 안정화 소요공수 예측 회귀 모형

데이터 수집 소요공수와 검증 및 안정화 소요공수가 인과 관계를 갖는다고 가정하고 회귀분석을 시행한 결과 [그림 3-16]과 같은 회귀선과 회귀식이 도출되었다.

1) 회귀 분석 대상 : X축 데이터 수집 소요공수, Y축 검증 및 안정화 소요공수



[그림 3-16] 데이터 수집과 검증 및 안정화의 소요공수 회귀분석

- 검증 및 안정화 소요공수 예측 회귀식

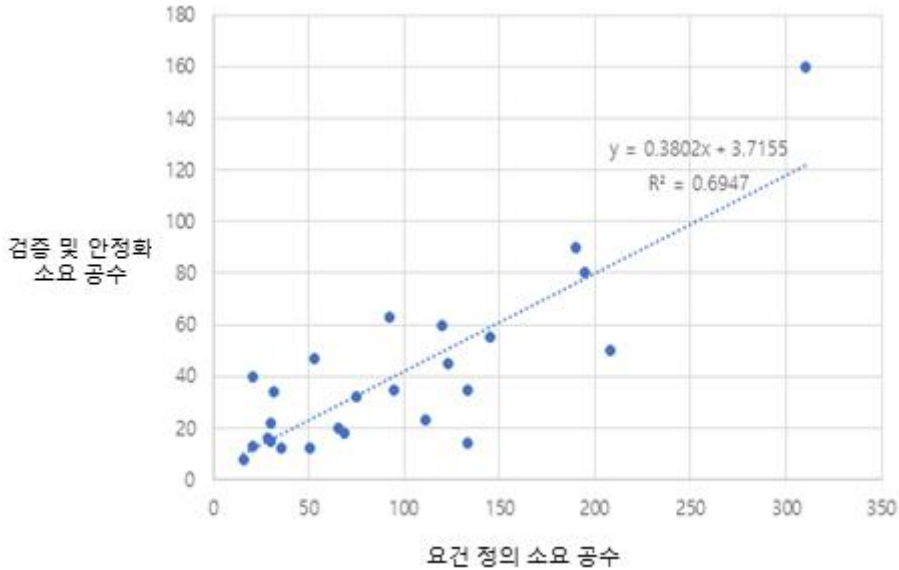
$$= (0.6426 * \text{데이터 수집 소요공수}) + 14.154$$

- 회귀식의 R^2 (결정계수) = **0.7405**

예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.7405으로 확인 되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 높다고 할 수 있다.

요건 정의 소요공수와 검증 및 안정화 소요공수가 인과 관계를 갖는다고 가정하고 회귀분석을 시행한 결과 [그림 3-17]과 같은 회귀선과 회귀식이 도출되었다.

2) 회귀 분석 대상 : X축 요건 정의 소요공수, Y축 검증 및 안정화 소요공수



[그림 3-17] 요건정의와 검증 및 안정화의 소요공수 회귀분석

- 검증 및 안정화 소요공수 예측 회귀식

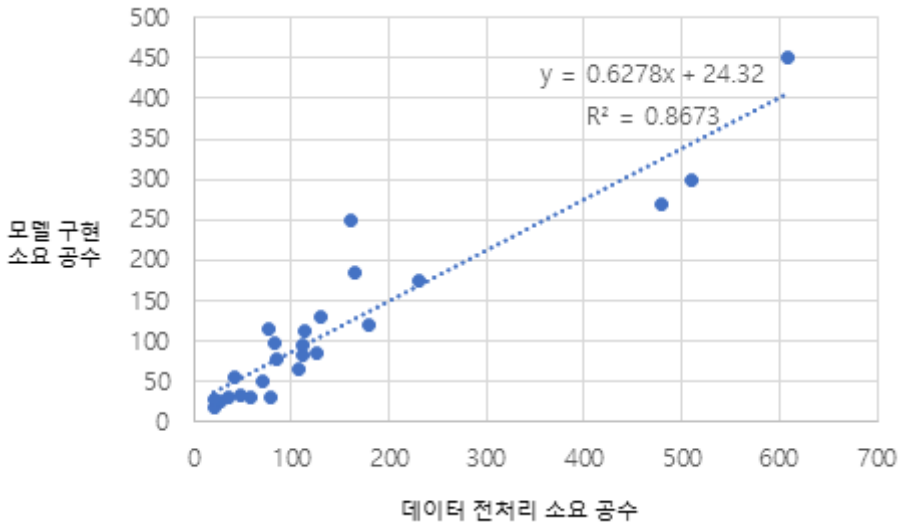
$$= (0.3802 * \text{요건 정의 소요공수}) + 3.7155$$
- 회귀식의 R^2 (결정계수) = 0.6947

예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.6947로 확인되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 보통이라고 할 수 있다. 검증 및 안정화 소요공수를 예측하는 선형 모형 [그림 3-13]과 [그림 3-14]를 비교해 볼 때 검증 및 안정화 소요공수를 예측하기에 적합한 것은 결정계수가 높은 [그림 3-13]의 선형 모형이다.

- 모델 구현 소요공수 예측 회귀 모형

데이터 전처리 소요공수와 모델 구현 소요공수가 인과관계를 갖는다고 가정하고 회귀분석을 시행한 결과 [그림 3-18]과 같은 회귀선과 회귀식이 도출되었다.

- 회귀 분석 대상 : X축 데이터 전처리 소요공수, Y축 모델 구현 소요공수



[그림 3-18] 데이터 전처리와 모델 구현의 소요공수 회귀분석

- 검증 및 안정화 소요공수 예측 회귀식

$$= (0.6278 * \text{데이터 전처리 소요공수}) + 24.32$$

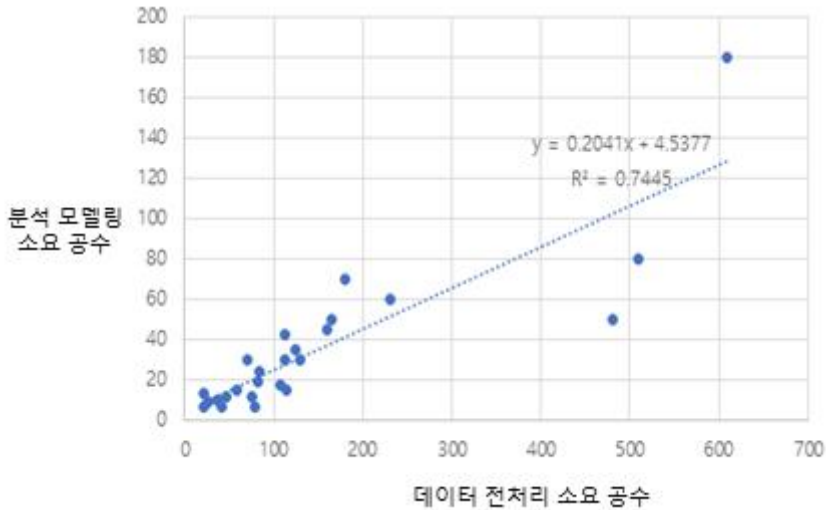
- 회귀식의 R²(결정계수) = 0.8673

예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.8673으로 확인되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 높다고 할 수 있다.

- 분석 모델링 소요공수 예측 회귀 모형

데이터 전처리 소요공수와 분석 모델링 소요공수가 인과 관계를 갖는다고 가정하고 회귀분석을 시행한 결과 [그림 3-19]와 같은 회귀선과 회귀식이 도출되었다.

- 회귀 분석 대상 : X축 데이터 전처리 소요공수, Y축 분석 모델링 소요공수



[그림 3-19] 데이터 전처리와 분석 모델링의 소요공수 회귀분석

- 분석 모델링 소요공수 예측 회귀식

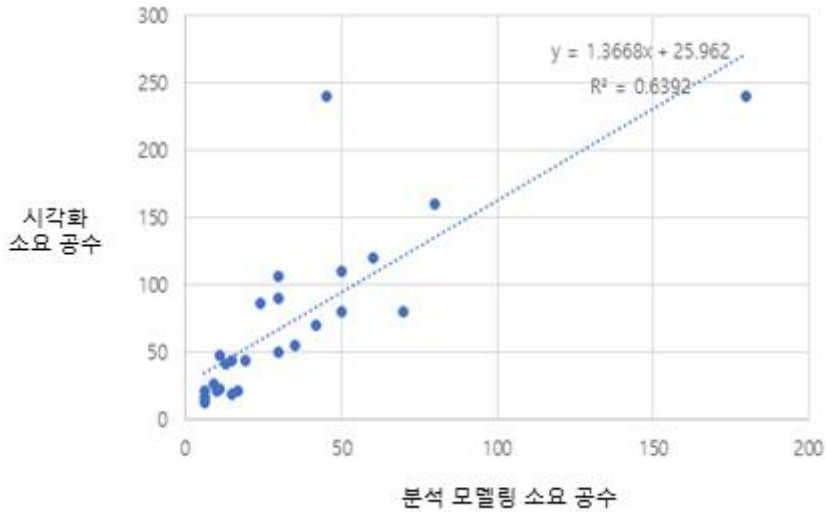
$$= (0.2041 * \text{데이터 전처리 소요공수}) + 4.5377$$
- 회귀식의 R^2 (결정계수) = **0.7445**

예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.7445로 확인되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 높다고 할 수 있다.

- 시각화 소요공수 예측 회귀 모형

분석 모델링 소요공수와 시각화 소요공수가 인과 관계를 갖는다고 가정하고 회귀 분석을 시행한 결과 [그림 3-20]과 같은 회귀선과 회귀식이 도출되었다.

1) 회귀 분석 대상 : X축 분석 모델링 소요공수, Y축 시각화 소요공수



[그림 3-20] 분석 모델링과 시각화의 소요공수 회귀분석

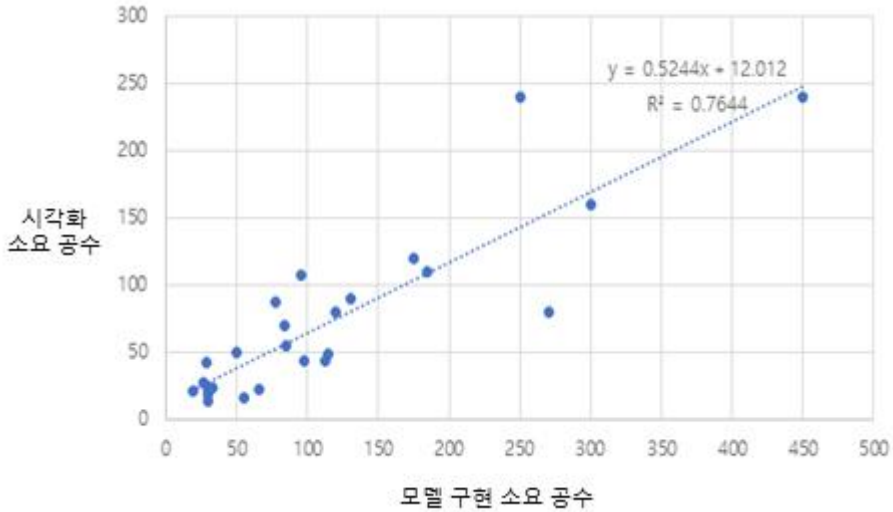
- 시각화 소요공수 예측 회귀식

$$= (1.3668 * \text{분석 모델링 소요공수}) + 25.962$$
- 회귀식의 R²(결정계수) = 0.6392

예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.6392로 확인되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 보통이라고 할 수 있다.

모델 구현 소요공수와 시각화 소요공수가 인과 관계를 갖는다고 가정하고 회귀분석을 시행한 결과 [그림 3-21]과 같은 회귀선과 회귀식이 도출되었다.

2) 회귀 분석 대상 : X축 모델 구현 소요공수, Y축 시각화 소요공수



[그림 3-21] 모델 구현과 시각화의 소요공수 회귀분석

- 시각화 소요공수 예측 회귀식

$$= (0.5244 * \text{모델 구현 소요공수}) + 12.012$$
- 회귀식의 R^2 (결정계수) = **0.7644**

예측한 선형 모형이 주어진 자료에 적합한 정도를 재는 척도인 결정계수가 0.7644로 확인 되었다. 0과 1사이의 값을 갖는 결정계수를 고려하면 해당 선형 모형의 결정계수는 높다고 할 수 있다. 시각화 소요공수를 예측하는 선형 모형 [그림 3-17]과 [그림 3-18]을 비교해 볼 때 시각화 소요공수를 예측하기에 적합한 것은 결정계수가 높은 [그림 3-18]의 선형 모형이다.

이상과 같이 요건정의 이후 개발 단계들에 대한 소요공수 예측 모형을 도출하였으며 회귀모형의 결정계수와 상관계수는 [표 3-10]과 같이 정리하였다.

| 개발단계 | 소요공수 예측 회귀 모형 | 채택 여부 | 통계량 | | |
|----------|------------------------------|-------|----------------|------|-------|
| | | | R ² | R | F |
| 데이터 수집 | 0.5139 * 요건 정의 공수 - 8.8385 | ✓ | 0.7079 | 0.84 | 55.74 |
| 데이터 전처리 | 1.6474 * 요건 정의 공수 - 9.9399 | ✓ | 0.5983 | 0.77 | 34.25 |
| 분석 모델링 | 0.2041 * 데이터 전처리 공수 + 4.5377 | ✓ | 0.7445 | 0.86 | 67.02 |
| 모델 구현 | 0.6278 * 데이터 전처리 공수 + 24.32 | ✓ | 0.8673 | 0.93 | 150.3 |
| 시각화 | 1.3668 * 분석 모델링 공수 + 25.962 | | 0.6392 | 0.88 | 40.74 |
| | 0.5244 * 모델 구현 공수 + 12.012 | ✓ | 0.7644 | 0.87 | 74.62 |
| 검증 및 안정화 | 0.6426 * 데이터 수집 공수 + 14.154 | ✓ | 0.7405 | 0.86 | 65.64 |
| | 0.3802 * 요건 정의 공수 + 3.7155 | | 0.6947 | 0.84 | 52.33 |

[표 3-10] 개발단계별 소요공수 예측 회귀 모형

[표 3-10]과 같이 개발단계별 소요공수 예측 모형에 따르면 “데이터 전처리”의 소요공수 예측 회귀모형에 대한 결정계수 R²이 0.5983으로 다른 회귀모형의 결정계수에 비해 너무 낮게 나타나고 있다. 이에 따라, 데이터 전처리의 결정계수가 높게 나오는 기타 2개 모형을 비교 실험하였고 [표 3-10]의 모형에 따른 전체 소요공수 예측결과와 정확도가 가장 높게 나타났다.

4장에서는 [표 3-10]의 소요공수 예측 모형을 활용하여 전체 프로젝트의 소요공수를 예측하고 성능을 검증하고자 한다.

IV. 실험 및 결과

본 장에서는 [표 3-8]의 개발단계별 소요공수 예측 모형을 활용하여 적정성을 검증하고자 한다. 검증하는 방법은 설문에서 제시한 인공지능 기반 소프트웨어에 대한 업체 견적과 설문결과로 도출된 소요공수의 중앙값을 비교하는 것으로 시행 한다.

A. 실험 대상 견적

설문에서 제시한 인공지능 기반 소프트웨어에 대하여, AI 업체로부터 RFI⁵⁾ (Request For Information) 용도로 받은 견적서는 아래와 같다.

| 문서번호 202203-1 | | 용역견적서 | | | |
|------------------|--------------|-------|--|--|--|
| 귀중 | | 등록번호 | | | |
| 공급금액 | ₩291,908,587 | 상호 | | | |
| 부가세 | ₩29,190,359 | 소재지 | | | |
| 총합계 | ₩321,098,946 | 업태 | | | |
| 견적일 | 2022.04.18 | 전화 | | | |

* 사업명 : 디지털 공사비정보 시스템 구축

| 업무활동 | IT 직무별 투입공수(M/M) | | | | |
|-----------------------|---------------------|-------------|-------------|------------|-------------|
| | 직무 | 투입인원 (명) | 투입기간 (월) | 투입률 (%) | 월 평균임금(원) |
| 데이터 수집 | 데이터 추가 수집 | 15 | 3 | 100 | 2,000,000 |
| 데이터 정제 | 데이터 라벨링 | 6 | 3 | 100 | |
| 데이터 정제 | 데이터 보완 개선 | 4 | 3.5 | 100 | |
| 데이터 정제 | 데이터통합 | 4 | 7 | 100 | |
| 중 투입공수 | (투입인원 * 투입기간 * 투입률) | | | | 105 |
| 직접인건비 한계 | (직무별 투입공수 * 평균임금) | | | | 46,500,000원 |
| 제 경 비 기 율 | (110~120%) | | | | 4,750,000 |
| | (20~40%) | | | | 4,000,000 |
| 직접경비 | | | | | 56,000,000 |
| 합계 | | | | | 60,000,000 |

[그림 4-1] AI SW 개발 견적 사례

5) RFI(Request For Information) : 발주자가(화주)가 RFP를 작성하기 전에 프로젝트 계획 및 수행에 필요한 정보를 수집하기 위해 공급업체에 요청하는 정보요청서.

본 논문에서는 소요공수 비용 산정의 정확성을 높이기 위해 소요공수 기준을 M/D로 정의하였다. 따라서 견적서에 제시된 월간(M/M) 소요공수를 일간(M/D) 소요공수로 변환하여 개발단계별 소요공수를 [표 4-1]과 같이 정리하였다.

| 개발 단계 | 견적 소요공수 | 개발 단계 | 견적 소요공수 |
|---------|-----------------|-----------------|------------------|
| 요건 정의 | 0 | 모델 구현 | 0 |
| 데이터 수집 | 45 * 20 = 900 | 시각화 | 0 |
| 데이터 전처리 | 60 * 20 = 1,200 | 검증 및 안정화 | 0 |
| 분석 모델링 | 0 | 투입공수 합 계 | 2,100 m/d |

[표 4-1] 견적서 소요공수 환산 내역

이상과 같이 확인해 본 결과, [그림 2-3]과 같이 빅데이터 사업 수행단계에서 가장 중요한 모델링과 모델 구현에 대하여는 소요공수가 누락되어 개발비용에 대해 신뢰할 수 없다는 것을 확인할 수 있다.

B. 실험 평가 및 분석

1. 실험 평가 방법

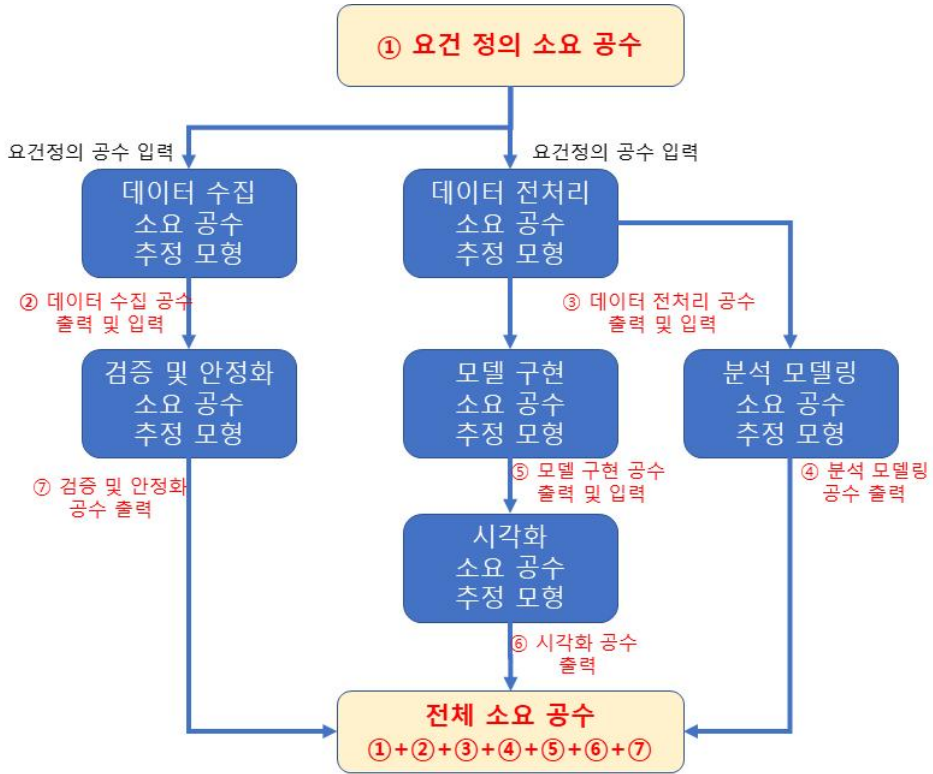
본 절에서는 소요공수 예측 모형의 성능을 검증하기 위해 예측모형을 통한 소요공수와 설문을 통한 소요공수를 비교 검증하여 정확도를 확인한다.

전체 개발단계별 소요공수의 예측모형은 요건정의 소요공수가 확인되면 이후 단계의 소요공수가 예측이 가능하도록 모형이 도출되었다. 실험에 앞서 설문결과를 통해 도출된 요건정의 단계의 중앙값 소요공수를 확인한다. 평균과 중앙값을 비교해보면 평균은 극단의 값에 영향을 받는 반면, 중앙값은 전체의 순서 중에 중앙의 위치의 값을 사용함으로써 극단의 값에 영향을 거의 받지 않는다. 따라서 중앙값은 소요공수에 대한 설문결과의 대표성을 부여할 수 있다.

| | 요건 정의 | 데이터 수집 | 데이터 전처리 | 분석 모델링 | 모델 구현 | 시각화 | 검증 및 안정화 |
|-----|-----------|-----------|------------|-----------|------------|-----------|-----------|
| 평균 | 95.34 | 40.16000 | 147.12 | 34.56 | 116.68 | 73.20 | 39.96 |
| 분산 | 73.088804 | 44.64553 | 155.668001 | 36.816074 | 104.937172 | 62.941772 | 33.337266 |
| 최소 | 16.00 | 4.0 | 21.00 | 6.00 | 19.00 | 13.00 | 8.00 |
| 25% | 31.50 | 15.0 | 58.00 | 11.00 | 33.00 | 23.00 | 16.00 |
| 50% | 75 | 30 | 108 | 24 | 85 | 50 | 34 |
| 75% | 133.00 | 40.0 | 160.00 | 45.00 | 130.00 | 90.00 | 50.00 |
| 최대 | 310.00 | 220.0 | 609.00 | 180.00 | 450.00 | 240.00 | 160.00 |

[표 4-2] 설문 결과의 중앙값 소요공수

요건정의 소요공수는 75 m/d로 확인되었다. 아래 [그림 4-2]에 따라 요건정의 소요공수를 입력으로 연쇄적인 단계별 소요공수를 예측하고 마지막으로 전체 소요공수를 예측하게 된다.



[그림 4-2] AI SW 개발 소요공수 예측 모형

위 모형에 따라 예측된 전체 소요공수와 설문으로 수집된 중앙값 소요공수, 견적을 통해 수집된 소요공수를 비교하여 성능을 검증한다.

2. 실험 결과 분석

[그림 4-2]에서 제안한 소요공수 예측모형을 통해 도출된 소요공수와 설문 의 소요공수, 견적을 통해 확보된 소요공수를 비교 검증하여 본 논문에서 제안한 소요공수 예측모형이 우수함을 확인한다.

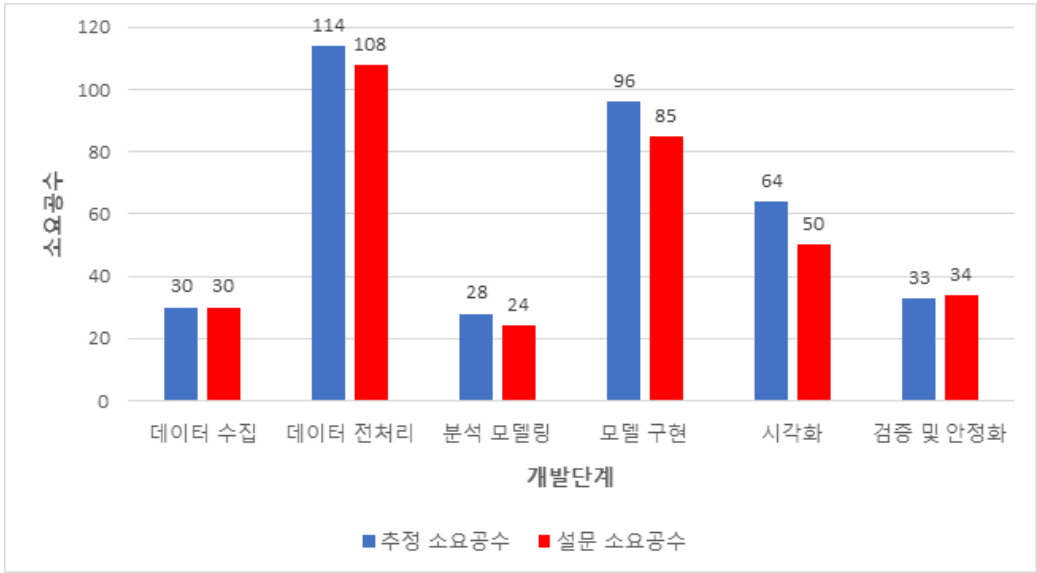
요건 정의 소요공수는 75m/d이며 개발단계별 소요공수 예측 회귀모형을 수행하게 되면 아래와 같이 개발단계별 소요공수가 도출된다.

| 개발단계 | 예측 소요공수 | 견적 소요공수 | 설문 소요공수 |
|----------|---------|---------|---------|
| 데이터 수집 | 30 | 900 | 30 |
| 데이터 전처리 | 114 | 1,200 | 108 |
| 분석 모델링 | 28 | | 24 |
| 모델 구현 | 96 | | 85 |
| 시각화 | 64 | | 50 |
| 검증 및 안정화 | 33 | | 34 |
| 합계 | 440 | 2,100 | 406 |

[표 4-3] 소요공수 예측 비교 실험 결과

[표 4-3]과 같이 2,100m/d를 필요로 하는 견적 소요공수는 예측 소요공수와 설문 소요공수에 비해 월등히 높다는 것을 확인할 수 있다. 이는 분석 모델링, 모델 구현, 시각화, 검증 및 안정화 단계에서 소요공수 산정이 누락되었고, 정확한 계산 방법이 아닌 주먹구구식으로 소요공수를 산정한 것으로 예측된다.

이에 따라 결과 분석에서는 견적 소요공수는 제외하도록 한다.



[그림 4-3] 예측 소요공수와 실문 소요공수 비교

인공지능 기반 소프트웨어 소요공수 예측을 위해 도출된 모형은 정확도를 검증하는 계산식은 다음과 같다. 식 (2)는 개발단계별 소요공수 예측모형에 대한 정확도 계산식이며, 식 (3)은 개발 전체 소요공수 예측모형에 대한 정확도 계산식이다.

$$a = \left(1 - \frac{d}{S}\right) * 100 \quad (2)$$

$$A = \left(1 - \frac{\sum_{n=1}^6 d_n}{\sum_{n=1}^6 S_n}\right) * 100 \quad (3)$$

- a : 개발단계별 소요공수 예측 모형 정확도
- d : 예측 소요공수와 실문 소요공수 오차의 절대값
- S : 개발단계별 실문 소요공수

도출된 식을 적용하여 전체 소요공수에 대한 정확도를 산정한 결과 92%로 계산되었다.

| 소요공수 예측 회귀모형 | 예측 소요공수(A) | 설문소요 공수(B) | A - B | 모형 정확도 |
|-----------------|---------------|---------------|-------|--------|
| 데이터 수집 | 30 | 30 | 0 | 100% |
| 데이터 전처리 | 114 | 108 | 6 | 94% |
| 분석 모델링 | 28 | 24 | 4 | 83% |
| 모델 구현 | 96 | 85 | 11 | 87% |
| 시각화 | 64 | 50 | 14 | 72% |
| 검증 및 안정화 | 33 | 34 | 1 | 97% |
| 합계 | 440 | 406 | 34 | 92% |

[표 4-4] 소요공수 예측모형별 정확도

[표 4-5]와 같이 “A - B”는 본 논문에서 제안한 모형에 따른 예측 소요공수에
 서 설문을 통해 산정된 소요공수와의 차이를 나타낸다.

개발단계별 예측 모형 정확도는 데이터 수집단계의 경우 100%, 검증 및 안정화
 97%, 데이터 전처리 94% 등으로 높게 나타났으며 전체 개발 소요공수에 대한 예
 측 정확도가 92%로 본 논문에서 제안하는 소요공수 예측에 관한 모형의 높은 신
 뢰도를 확인할 수 있다.

V. 결론 및 향후 연구

인공지능 기반 소프트웨어는 해를 거듭할수록 물량이 증대되고 있으나 소프트웨어의 개발 절차가 표준화되어 있지 않고 비용 산정 기준이 없어 소프트웨어의 품질과 개발 비용에 대한 불신이 커지고 있다. 업계에서의 인공지능 기반 소프트웨어의 투입공수 견적은 데이터 수집과 데이터 전처리 단계로 산정하고 있음이 견적서를 통해 확인되었다. 수주자 관점에서 작성하는 투입공수는 모호성과 주관성으로 근거가 부족하여 발주자와 수주자 모두에게 논란의 대상이 되고 있다.

이에 본 논문에서는 대량의 데이터로 학습을 진행하고 알고리즘을 도출하여 적용하는 “인공지능 기반 소프트웨어”의 개발 단계를 정의하고 19개 업체의 27명을 무작위 추출하여 설문을 진행하여 개발단계별 투입 공수를 수집하였다.

개발 단계 간 상관분석을 실시하여 0.77 이상의 상관계수가 높은 개발 단계에 대하여 회귀분석을 실시하였고 개발단계별 소요공수 예측 회귀 식을 도출 하였다.

또한 예측 모형의 성능을 평가하기 위해 설문으로 수집된 소요공수 중위 값을 기준으로 정확도 평가를 시행하였으며 그 결과 92%의 정확도를 나타내고 있음도 확인하였다. 향후, 모호하고 주관적인 투입공수 산정으로 수주자나 발주자가 산정하기 어려웠던 문제점을 해결할 수 있을 것으로 기대한다.

향후, 인공지능 소프트웨어의 종류별 개발단계 표준화 연구와 이를 기반으로 개발단계별 소요공수의 데이터 축적이 꾸준히 이루어진다면 실 데이터를 기반으로 하는 정확한 개발비용 산정에 대한 연구가 가능하리라 본다.

참고 문헌

- [1] C. Jones, and T. C. Jones, “Estimating Software Costs”, McGraw-Hill, 1998.
- [2] Y. Yokoyama, M. Kodaira, “Software cost and quality analysis by statistical approach”, Proceedings of the 1998 20th International Conference on Software Engineering, Kyoto, Japan, pp. 465-467, 1998.
- [3] J. S. Bae, J. Y. Jung, “Software Size Measurement from Information Strategy Planning with the Function Point Method”, The Journal of Society for e-Business Studies, Vol. 14, No. 3, pp. 153-168, 2009.
- [4] 광송해 (2017), “SI 프로젝트의 리스크 관리와 소요공수 산정을 위한 통합 모델 연구”, 공주대학교 컴퓨터공학과 박사학위 논문
- [5] 한국소프트웨어산업협회 (2022), “SW사업 대가 산정 가이드”
- [6] B. W. Boehm, et al., (2000) “Software cost estimation with COCOMO ”, PrenticeHall PTR
- [7] B. W. Boehm, “Software and Its Impact : A Quantitative Assessment”,
- [8] IFPUG, (2000) “Function Point Counting Practices Manual”, Release 4.1.1, International Function Point Users Group
- [9] 안계중, 이남용(2002), “UML기반 소프트웨어시스템의 규모측정을 위한 기능점수분 석기법의 응용에 관한 연구”, 한국전자거래학회지 7(2), 2002.8, 173-190(18 pages)
- [10] 국방대학교 전산정보학과 김동선, 윤희병(2008), “객체지향 기반 효율적인 기능점수 측정 프로세스 설계 및 사례연구”, 정보처리학회 논문지 d 제15-d권 제3호, 2008. 6.
- [11] 안연식(2009), “소프트웨어 규모산정을 위한 기능점수 개선 Micro-FP 모형의 제안”, 韓國컴퓨터情報學會 論文誌 第14卷 第12號, 2009. 12.
- [12] 유재상(2019), “소프트웨어 사업규모 추정 모델 개발”, 2019년 한성대학교 지식 서비스&컨설팅대학원 스마트융합컨설팅학과 석사학위 논문
- [13] 이지혜, 강동수(2021), “랜덤포레스트를 이용한 프로젝트 초기단계 소프트웨어 기능점수 추정”, 정보과학회 컴퓨팅의 실제 논문지 27(1), 2021.1, 41-47 (7 pages)
- [14] 행정안전부, 한국정보화진흥원 (2016) “공공 빅데이터 분석 사업비 대가 산정 가이드”

[부록] 인공지능 SW의 적정 개발비용 산정방안 연구를 위한 설문

인공지능 SW의 적정 개발비용 산정방안 연구를 위한 설문

안녕하십니까? 조선대 SW융합공학과 대학원에 재학 중인 장승진이라고 합니다

4차산업혁명 시대를 살아가고 있는 지금
과거와 달리 인공지능 소프트웨어의 개발사업이 증가하고 있습니다

그러나 합리적인 개발비용 산정기준이 없기에
발주자와 수주자 모두 어려움을 겪고 있습니다

이에 본 설문에서는
아래의 "인공지능 소프트웨어 사례"를 기반으로
개발 작업 단계별 소요인력의 기술등급과 공수를 취합하여
적정 개발비를 산정하기 위한 투입공수 기준을 연구하고자 합니다

[인공지능 소프트웨어 사례]

- 프로젝트명 : 신규 공사작업의 작업공수 예측 모델 구축
- 데이터 : 기존 공사작업별 작업공수
- 모델 기능 : 신규 작업 생성시 기존 작업의 유사성을 비교하여 개략 작업공수 제안
- * 기존 작업내용과 신규 작업내용의 비교를 위해 자연어 처리기술 활용

여러분들의 소중한 의견이
인공지능 소프트웨어의 적정 개발대가 산정기준 수립에 중요한 데이터로 활용될 수 있습니다

작성하신 설문은 아래 계정으로 회신해 주시면 감사하겠습니다
회신계정 및 기한 : sarafather@gmail.com / 2022.4.28.(목)

[기본 정보]

- ▶ 인공지능 소프트웨어 개발비용 설득에 어려움을 경험한 적이 있습니까? 있다(), 없다()
- ▶ 발주자 수주자 구분 : 발주자(), 수주자()
- ▶ 업 체 명 : _____
- ▶ 응답자 이름/연락처 : _____ / _____
* 작성해 주신 연락처는 소정의 선물을 발송하기 위한 용도로 활용됩니다
- ▶ 현재 직무 : 대표이사(), 영업대표(), 기획자(), PM(), 개발자(), 기타()
- ▶ 기업규모 : 대기업(), 중소/중견기업(), 공기업(), 소기업()
- ▶ AI 인력 보유규모 : () 명
- ▶ AI 소프트웨어 경력 : 1년미만(), 1년~3년미만(), 3년~5년미만(), 5년 초과()

[세부 설문 내용]

※ 작성방법 : 단계별 수행활동에서 난이도와 소요공수를 기재해 주시고
 추가 작업 필요시 공백란에 기재하셔서 소요공수를 표시해 주세요

[난이도 선택 기준]

- 전문 : 특급기술자 투입이 필요한 작업, 복잡 : 고급기술자 투입이 필요한 작업
- 보통 : 중급기술자 투입이 필요한 작업, 단순 : 초급기술자 투입이 필요한 작업

1. 요구 정의 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) <small>예시) 2명 x 10일 = 20</small> |
|----------------------------|-----------|----|----|----|----|---|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| WBS 작성 | | | | | | |
| 실무자 인터뷰 계획 | | | | | | |
| 실무자 인터뷰 질의서 작성 | | | | | | |
| 요구사항정의서 작성 | | | | | | |
| 관리자급 인터뷰 | | | | | | |
| 선행 사업 및 자료 분석 | | | | | | |
| 공통기반 플랫폼 인프라 및 소프트웨어 현황분석 | | | | | | |
| 수집대상 데이터 현황분석 | | | | | | |
| 데이터 수집 환경 분석 | | | | | | |
| 성과목표 및 성과 측정시기별 계획안 마련 | | | | | | |
| 성과관리를 위한 지표 및 운영방안 마련 | | | | | | |
| 과제 수행 결과 활용 계획 및 확산 계획안 작성 | | | | | | |
| | | | | | | |

2. 데이터 수집 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) <small>예시) 2명 x 10일 = 20</small> |
|----------------|-----------|----|----|----|----|---|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| 데이터 수집 대상 정의 | | | | | | |
| 수집 환경 구성 | | | | | | |
| 기관 내/외부 데이터 수집 | | | | | | |
| | | | | | | |

3. 데이터 전처리 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) 예시) 2명 x 10일 = 20 |
|-----------------------------------|-----------|----|----|----|----|--|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| 오류 데이터 제거 및 불일치성 교정 | | | | | | |
| 데이터 구조 표준화 및 수집데이터 융합 | | | | | | |
| 데이터 수집, 정제, 표준화, 융합의 자동화 및 주기적 연계 | | | | | | |
| 수집 데이터의 기초 통계 추출 | | | | | | |
| 데이터의 시스템 연계 개발 | | | | | | |
| 표준화 모듈 설계 및 구현 | | | | | | |
| 데이터 정제 모듈 설계 및 구현 | | | | | | |
| | | | | | | |

4. 분석 모델링 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) 예시) 2명 x 10일 = 20 |
|----------------|-----------|----|----|----|----|--|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| 분석 시나리오 작성 | | | | | | |
| 분석 모델 설계 초안 작성 | | | | | | |
| 분석 모델 수정 의견 반영 | | | | | | |
| | | | | | | |

5. 모델 구현 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) 예시) 2명 x 10일 = 20 |
|---------------------------|-----------|----|----|----|----|--|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| 사전 구축 | | | | | | |
| 분석 알고리즘 모델 설계 | | | | | | |
| 모듈 기능 정의 및 설계 | | | | | | |
| 모듈 기능정의, 설계, 개발 | | | | | | |
| 분석 알고리즘 구현, 분석결과 저장 기능 개발 | | | | | | |
| | | | | | | |

6. 시각화 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) 예시) 2명 x 10일 = 20 |
|----------------|-----------|----|----|----|----|--|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| 시각화 설계 | | | | | | |
| 분석 결과별 시각화 | | | | | | |
| 연계표준 정의, 모듈 개발 | | | | | | |
| 단위테스트 | | | | | | |
| 지식화 | | | | | | |
| | | | | | | |

7. 검증 및 안정화 단계의 세부작업별 필요인력 공수를 기재해 주세요

(*추가 작업 필요시 표의 맨아래 공백 란에 기재하시고 소요공수를 표시해 주세요)

| 수행 활동 | 수행활동별 난이도 | | | | | 소요공수 (Man Day) 예시) 2명 x 10일 = 20 |
|--------------|-----------|----|----|----|----|--|
| | 필요없음 | 전문 | 복잡 | 보통 | 단순 | |
| 검증 계획 수립 | | | | | | |
| 통합테스트 | | | | | | |
| 사용자 승인테스트 | | | | | | |
| 사용자 및 운영자 교육 | | | | | | |
| | | | | | | |

♡ 바쁘신 중에도 설문에 응해 주셔서 대단히 감사합니다 ♡