

데이터셋 구축 프로젝트

---

**셀렉트스타**  
**인공지능**  
**데이터셋**  
**지원사업**

2020. 11.



**SELECTSTAR**

DATA CROWDSOURCING PLATFORM

# 셀렉트스타 인공지능 데이터셋 지원사업

## 1. 사업개요

### ○ 사업 목적

- AI 관련 사업을 진행 중인 기업(스타트업, 예비창업자) 또는 인공지능 관련 연구를 진행 중인 국내 우수 대학 연구실 혹은 정부출연 연구기관의 연구와 개발을 위한 인공지능 학습데이터셋 구축 지원. 국내/외에서 활용 가능한 완전히 새로운 데이터셋 제안을 권고하며, 기존 해외 데이터셋의 한국형 데이터셋 제안도 가능.

(예시 : KorQuad, ImageNet등)

- (연구실의 경우) 인공지능 연구개발에 널리 사용가능한 범용성이 있는 데이터셋 발굴 및 제작을 위한 연구 지원.

-> 대상: 국내 우수 대학 연구실

-> 세부 목표: 데이터셋 수집 및 가공을 통해 해외 유명 저널 및 학회 제출을 목적으로 한 데이터셋 개발 연구 또는 데이터셋 활용 연구 지원

- (기업의 경우) 국내 인공지능 기술 상용화를 위한 특수 목적의 범용성 있는 데이터셋 구축 지원 (예시: LiDAR, KorQuad 등)

-> 대상: 중소기업, 스타트업

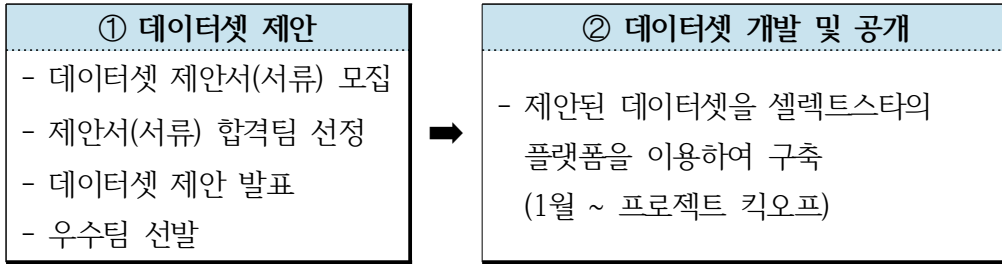
-> 세부 목표: 공공의 연구 또는 개발을 위한 전체 또는 일부 공개 가능한 범용성 있는 데이터셋 구축 지원

### ○ 사업 내용

#### 【 지원사업 개요 】

- 사업명 : 셀렉트스타 인공지능 데이터셋 지원사업 (2020 하반기 ~ )
- 사업기간 : 2020년 11월 ~ 2021년 데이터 구축 완료 시
- 지원대상 : 인공지능 관련 연구를 진행 중인 국내 우수 대학 연구실 또는 AI 관련 사업을 진행 중인 기업 (기존 ‘한국데이터산업진흥원’의 데이터 바우처 지원사업의 수혜를 받은 기업 제외)
- 지원내용 : - 총 10개 팀에게 각 7,000만원 상당의 데이터셋 구축 지원,  
- AWS 크레딧 등 1,000만원 상당의 부가 혜택지원  
(<https://selectstar.ai/benefit> 참고)  
- VC(현재 카카오벤처스 확정)에서 투자 검토
- 주 관 : 셀렉트스타 주식회사(SelectStar Inc.)
- 파 트 너 : SKTelecom True Innovation, 카카오벤처스, 한국인공지능협회

○ 추진 절차



# 셀렉트스타 인공지능 데이터셋 지원사업

## 2. 사업내용

### 1) 데이터셋 제안

제안서를 바탕으로 우수팀들을 1차로 선정하고, 2차 발표평가 결과에 따라 최종 평가 및 상위 10개 연구소 및 기업 선정

### (1) 프로젝트팀 발굴 절차

홍보 및 모집	서류심사	우수팀 선정 공지	발표평가 및 최종 선정
11월 2일 ~ 30일 (약 4주)	11월 30일 ~ 12월 4일 (1주)	12월 4일 (1차 심사결과 발표)	발표평가: 12월 9일 최종 발표: 12월 11일
온라인 홍보 (가능할 경우 오프라인 홍보) 온라인 접수	서류심사 진행 ↓ 우수팀 선발 (15개 팀 내외)	선정결과 공지 발표평가 안내	발표평가 및 프로젝트 진행할 최종 팀 선정 (연구실 00개 / 기업 00개)

### (2) 세부내용

- 모집
  - 제출 양식: 5페이지 이하 기획안 제출  
(dataset.or.kr 홈페이지에서 양식 다운 및 온라인 제출)
  - 기간: 11월 2일 ~ 11월 30일
- 서류심사(우수팀 선정)
  - 1차 심사 결과 발표: 12월 4일
  - 서류심사단은 셀렉트스타 주식회사 내부 심사로 자문 위원들 중에서 선정함
  - 서류심사를 통해 15개 내외의 우수팀 선정
- 발표평가 및 최종 선정
  - 발표평가: 12월 9일
  - 최종 발표: 12월 11일 공개
  - 우수팀은 발표평가에 참여
  - 제출 양식: PPT를 활용한 발표 평가
  - 발표평가를 통해 최종 선정 10개 팀 확정

## 2) 데이터셋 개발 및 공개

최종 선정된 팀들이 제안한 데이터셋들을 셀렉트스타의 플랫폼을 이용하여 구축 및 데이터셋 공개

### (1) 데이터셋 구축

- 발표평가를 통해 선정된 최종 선정된 팀들의 데이터셋을 구축
  - > 셀렉트스타가 보유중인 데이터 생산 클라우드소싱 플랫폼 캐시미션을 이용, 제안한 데이터셋을 구축
- 프로젝트 진행 일정
  - > 12월 말 ~ 1월 초: 데이터 구축 관련 킷오프
  - > 1월 초 ~ 데이터셋 생성 마감 까지: 캐시미션을 통한 데이터셋 구축 진행

### (2) 데이터셋 공개 일정

- 연구실: 해외 인공지능 관련 학회 또는 저널 논문 제작을 통한 데이터셋 공개
  - > 데이터셋 활용에 대한 홍보 진행
  - > 공개 일시: 논문 결과 이후 (논문 일정에 따라 빠르면 2021년 하반기, 늦으면 2022년 상반기)
- 기업: 인공지능 업계 공공의 목적을 위한 일부 또는 전체 데이터셋 공개
  - > 공개가 어려운 기업의 경우 협의를 통해 가능한 범위 내에서 공개
  - > 공개 일시: 데이터셋 구축이 완료 되는대로 공개 (2021년 3월 ~ 4월 예상)

## 셀렉트스타 인공지능 데이터셋 지원사업

### 3. 전체 일정

일시	내용	장소
11월 2일 ~ 11월 30일	홍보 및 모집	온라인*
11월 중순 ~ 12월 초	서류심사(우수팀 선정) 및 우수팀 선정 공지(12월 4일 발표)	온라인
12월 초 ~ 12월 중순	선정된 우수팀 대상 발표평가(12월 9일) 및 최종 진행할 팀 선정(12월 11일 발표)	온라인*
12월 말 ~ 1월 초	데이터셋 구축을 위한 킥오프	팀별 오프라인 대면 미팅
1월 중순 ~	데이터셋 구축 시작 및 완료된 데이터셋 공개	온라인

\* 국내 코로나 19 관련 상황에 맞춰 가능할 경우 오프라인으로도 진행 예정

### 4. 유의 사항

- 지원사업에 관련된 모든 사항은 추후 변동될 수 있으며 홈페이지를 통해 공지됩니다.
- 셀렉트스타의 웹/앱 플랫폼 '캐시미션'으로 수집 및 가공이 가능한 데이터셋을 제안해 주시기 바랍니다. 셀렉트스타의 '캐시미션' 플랫폼에 대한 안내는 <https://selectstar.ai> 와 <https://cashmission.com> 을 참고하시기 바랍니다. 필요시, 제안하실 데이터셋의 수집 및 가공 가능여부는 사전에 충분한 여유를 가지고 지원사업 공식 이메일(contest@selectstar.ai) 등으로 문의 주시기 바랍니다. (접수 마감 직전에는 문의가 많아 메일에 답변드릴 수 없을 가능성이 높으며, 셀렉트스타는 문의에 답변하지 못함으로 발생하는 사항에 책임지지 않습니다.)
- 2차 심사는 PT(발표평가)로 진행되며, 불참 시 심사에서 제외됩니다. (발표형식 자유, 발표시간 10분 이내로 제한 - 자세한 내용은 추후 제공 예정)
- 제안된 프로젝트에 대한 제3자의 저작권, 특허권, 초상권 등의 모든 지적 재산권 및 정보의 무단 사용 등으로 발생하는 법적 문제에 대한 책임은 참가팀에게 있으며 추후 문제 발생 시 접수 또는 지원이 취소됩니다.
- 심사 점수는 공개하지 않으며, 심사 이후 부적합 사유가 생길 경우 지원 대상에서 제외될 수 있습니다.
- 실제 데이터셋 구축과 관련된 세부 조율 및 협의는 최종 선정 후 진행됩니다.
- 기획안 제출 시 필요한 모든 사항을 숙지하시기 바라며, 숙지하지 못한 책임은 참가팀에게 있습니다.
- 구축된 데이터셋에 대한 소유권과 저작권은 선정된 기업 또는 연구실이 갖게 됩니다. (단, 데이터셋 전량 혹은 일부 공개 및 운영에 동의해야 하며, 개인정보가 포함된 데이터의 경우 AI 개발 목적으로만 활용할 수 있습니다.) 셀렉트스타는 데이터셋 공개 및 운영에 관한 권리를 갖게 됩니다. 셀렉트스타는 본 사업을 통해 개발한 데이터셋의 학술용 무료 공개를 지향하고 있습니다.
- 파트너 VC(카카오벤처스 등)나 파트너 기업(SKTelecom 등)은 선정된 기업/연구기관에 대해 투자나 사업 연계를 확정하지 않습니다.
- 파트너 VC의 경우 선정된 기업/연구기관을 대상으로 투자 검토를 진행하여 미팅 대상을 따로 선별하여 개별적으로 연락을 진행 합니다. 또한, 경우에 따라 투자 대상이 없을 수도 있습니다.