



AI for Future Workforce

Module 17: 데이터
가져오기 및 처리

법률 고지사항

- Intel® 디지털 준비 프로그램 및 Intel® AI for Future Workfork 프로그램은 Intel Corporation에서 개발했습니다.
- © Intel Corporation. Intel, Intel 로고 및 기타 Intel 마크는 Intel Corporation 또는 자회사의 상표입니다. 다른 이름 및 브랜드는 다른 사람의 재산으로 주장될 수 있습니다. 프로그램 날짜와 수업 계획은 변경될 수 있습니다.
- Intel 기술에는 활성화된 하드웨어, 소프트웨어 또는 서비스 활성화가 필요할 수 있습니다.
- 모든 제품과 구성 요소는 안전을 보장 할 수 없습니다.
- 결과물은 추정되거나 시뮬레이션 되었습니다.
- Intel은 타사 데이터를 제어하거나 감사하지 않습니다. 정확성을 평가하려면 다른 출처를 참조해야 합니다.
- 당신이 투자한 비용과 그에 대한 결과물은 다를 수 있습니다.

학습 결과

이 모듈이 끝나면 다음을 수행할 수 있습니다:

- Python에서 데이터 가져오기 시연
- 데이터 다운로드 자동화 시연
- 백분위수 범위 사용 방법 알아보기
- 박스플롯 및 히스토그램 이해
- 오류와 이상값의 차이점 이해
- 오류 및 누락 데이터 조사 및 처리

데이터 수집 및 탐색

AI for Future Workforce

자기 주도 학습

노트북: 데이터 얻기

주요 내용

데이터 얻기

- 기계 학습 프로젝트를 위한 수많은 데이터 소스가 있습니다.
- 그 중 일부를 나열해보세요!
- 인터넷에서 컴퓨터로 데이터를 다운로드하는 몇 가지 방법을 나열해보세요.
- Pandas는 무엇에 사용할 수 있습니까?

시퀀셜코딩 게임

시퀀셜 코딩 노트북

시퀀셜 코딩 게임

- "Seq_Coding" 노트북 열기
- 4명씩 팀을 이루어 순차적으로 코딩하게 됩니다
- 1인당 4분의 코딩 시간이 있습니다.

시퀀셜 코딩 게임

- 첫 번째 구성원이 컴퓨터로 앞으로 호출됩니다.
- 작업 내용을 읽고 4분 동안 코딩을 시작합니다.
- 4분 후 첫 번째 구성원은 두 번째 구성원에게 30초 동안 앞서 작업한 코딩 내용을 설명합니다.
- 두 번째 구성원이 4분 동안 코딩합니다.
- 마지막 구성원은 4분 안에 코드를 완성합니다. 그 후 작성된 코드를 1분 동안 전체에게 설명합니다.

축하합니다!

자기 주도 학습

노트북 : 기본 데이터 처리 및 시각화

주요 내용

기본 데이터 처리 및 시각화

- describe()를 사용하여 데이터에 대해 자세히 알아볼 수 있습니다.
- describe()를 사용하여 데이터에 대해 무엇을 배울 수 있나요?
- 박스플롯은 데이터 분포를 시각화하는 데 유용합니다.
- 백분위수 값과 사분위수 범위는 무엇입니까?
- 히스토그램이란 무엇인가요?
- 산점도는 무엇을 위한 것인가요?

자기 주도 학습

노트북 : 오류 및 누락 데이터 처리

주요 내용

오류 및 누락 데이터 처리

- 잘못된 데이터와 이상값의 차이점은 무엇인가요?
- 잘못되거나 누락된 데이터를 처리하는 방법은 무엇인가요?

프로젝트(1/2)

프로젝트를 위해 4인 1팀을 구성해주세요.

프로젝트

프로젝트는 각 팀이 획득한 데이터를 수집, 처리 및 제시하는 것입니다.

코딩을 시작하기 전 계획하고 전략을 세우세요.

- 어떤 주제에 관심이 있나요?
- 어디에서 데이터를 얻을 계획인가요?
- 데이터를 처리하기 위해 무엇을 해야 하는지 어떻게 결정하시겠습니까?
- 주어진 시간 내에 작업을 완료할 수 있도록 팀의 작업을 어떻게 분할하시겠습니까?

각 레벨에 대해서

- **Level 1:** 최소 5,000행의 데이터세트 다운로드
- **Level 2:** 데이터 탐색 및 데이터 사전 처리 방법 결정
- **Level 3:** 데이터 전처리: 이상값, 누락된 데이터 등
- **Level 4:** 데이터 전처리 과정에서 사용한 기법과 그 결과
- **Level 5:** 선택한 산업에서 데이터를 어떻게 사용할 수 있는지
최소 5가지 방법을 제시하세요.

하프 타임!

각 팀은 진행 상황을 공유합니다

- 어느 수준에 도달했다고 생각하나요?
- 가장 높게 도달한 것에 대해 설명하세요.
- 프로젝트를 선보이기 전에 극복하고자 했던 가장 큰 애로사항을 설명하세요.
- 비슷한 애로사항이 있는 사람이 있나요?
- 도움을 주거나 조언을 하고 싶은 사람이 있나요?

프로젝트(2/2)

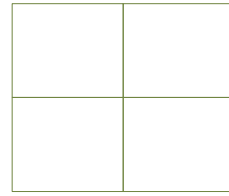
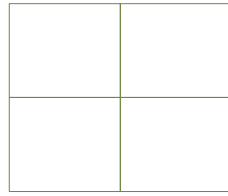
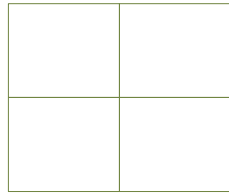
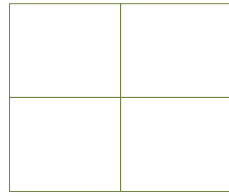
프로젝트

프로젝트는 각 팀이 획득한 데이터를 수집, 처리 및 제시하는 것입니다.

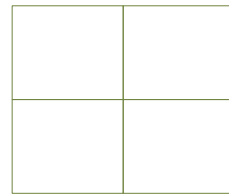
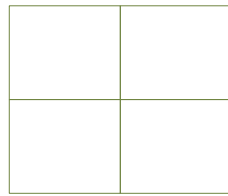
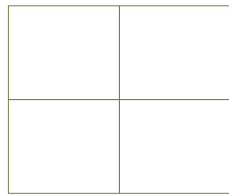
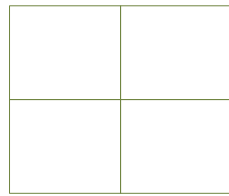
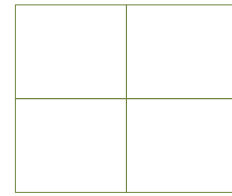
각 레벨에 대해서

- **Level 1:** 최소 5,000행의 데이터 세트 다운로드
- **Level 2:** 데이터 탐색 및 데이터 사전 처리 방법 결정
- **Level 3:** 데이터 전처리: 이상값, 누락된 데이터 등
- **Level 4:** 데이터 전처리 과정에서 사용한 기법과 그 결과
- **Level 5:** 선택한 산업에서 데이터를 어떻게 사용할 수 있는지
최소 5가지 방법을 제시하세요.

프로젝트 발표



1팀당 4명씩 10개조를
만드세요.



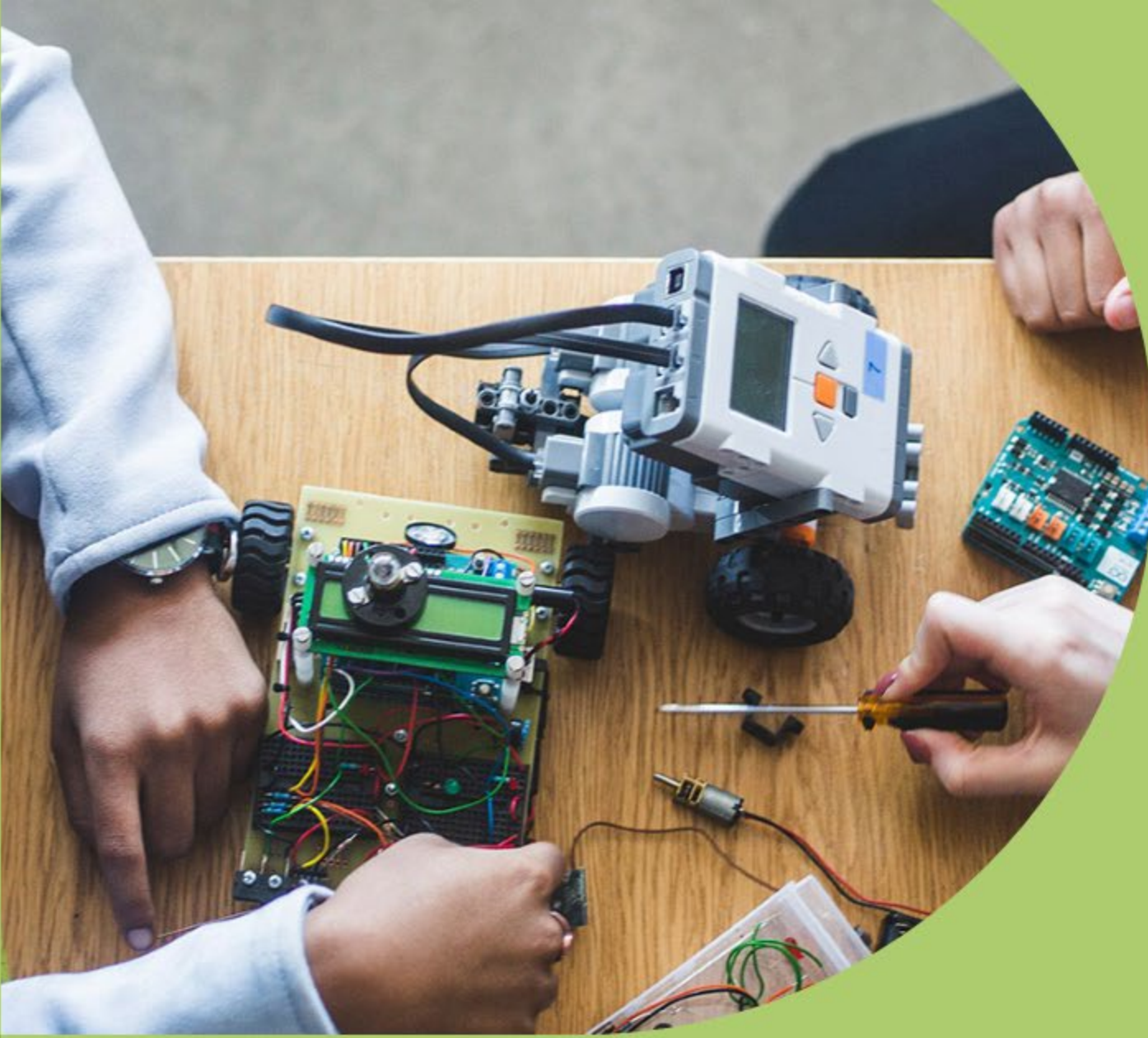
팀별 발표 순서

1. 어떻게 데이터 세트를 다운로드 했나요?
2. 특정 데이터 세트를 다운로드 하게 된 이유는 무엇인가요?
3. 데이터 세트를 어떤 애플리케이션에 사용할 수 있다고 생각 하나요?
4. 업무를 어떻게 정하였나요?
5. 데이터를 다운로드, 처리 및 분석하는 데 어떤 기술을 사용했나요?

모두 수고하셨습니다!

프로젝트에 대해 토론해봅시다.

- 접근 방식은 무엇이었나요?
- 어떤 도전에 직면했나요?
- 어떻게 극복했나요?
- 프로세스를 어떻게 개선하면 좋을까요?



요약

학습 결과

이 모듈이 끝나면 다음을 수행할 수 있습니다:

- Python에서 데이터 가져오기 시연
- 데이터 다운로드 자동화 시연
- 백분위수 범위 사용 방법 알아보기
- 박스플롯 및 히스토그램 이해
- 오류와 이상값의 차이점 이해
- 오류 및 누락 데이터 조사 및 처리

퀴즈

Link [here](#)

적용

- 오늘 배운 내용을 수업을 넘어서 어떻게 적용할 것인가요?
- 인공지능 애플리케이션을 구축할 때 주의해야 할 사항은 무엇인가요?
- 개인 정보 보호에 대한 고려 사항이 있나요?
- 오늘 배운 내용이 실제 산업에 도움이 되는 내용이라고 생각하나요?

A young man with glasses is shown in profile, looking intently at a computer screen. The background is a blurred classroom with other students. The text 'intel digital readiness' is overlaid on the left side of the image. A white curved line separates the text area from the man's face. A solid green rectangle is in the bottom-left corner.

intel. digital readiness