
FIT3152 데이터 분석 (Data analytics) - 2025년 1학기

과목 정보

환영합니다 FIT3152 데이터 분석에 오신 것을 환영합니다. 시작하는 데 필요한 주요 정보를 간략히 요약해 드립니다.

세미나 (Seminar): 매주 2시간씩 캠퍼스 내 세미나가 있습니다. 이는 강의 형태로 진행되며, 중간중간 토론과 질문의 기회가 있습니다.

- 시간: 화요일 오후 7:00 - 9:00
- 장소: 모든 학생은 C1 강의실(CL_Exh-25.FIT_C1 Lec (314))에 참석해야 합니다.
- 녹화: 세미나는 MULO에 녹화됩니다. C1 강의실이 꽉 찰 경우를 대비한 보조 강의실(CL_Rnf-11.SciN_S13 Lec (56))이 있지만, 이는 C1이 만석일 때만 사용됩니다.

응용 세션 (Applied Sessions): 캠퍼스에서 진행되며, 2주 차부터 시작됩니다. 세미나 주제를 한 주 뒤에 다룹니다. 본인의 세션 시간은 Allocate+에서 확인할 수 있습니다.

강의 노트 및 사전 읽기 자료: 강의 사전 읽기 슬라이드는 세미나 전 주에 제공됩니다. 단, 1주 차는 예외입니다. 업데이트나 주석이 포함된 강의 슬라이드는 1주 차를 포함하여 매 강의 후 가능한 한 빨리 게시될 것입니다.

자료 (Resources): 모든 소프트웨어는 오픈 소스이며 무료입니다. 대부분의 참고 자료는 온라인이나 모나시 도서관을 통해 무료로 이용할 수 있습니다.

포럼 (Forum): 과목 공지 사항 및 학생 토론을 위해 Ed Discussion을 사용합니다.

교육팀 (Teaching Team)

- Dr John Betts (공동 강사, 책임 교수): john.betts@monash.edu
- Dr Heshan Kumarage (공동 강사, 과목 코디네이터): heshan.kumarage@monash.edu

FIT3152 데이터 분석 - 개요 최근 몇 년 동안 연구 및 산업 분야에서 일상적으로 기록하고 분석하는 데 데이터의 양과 다양성이 폭발적으로 증가했습니다. 일부 사회 평론가들은 이 현상을 "빅 데이터"의 부상이라고 부르며, 데이터를 조사하는 분석가와 실무자를 "데이터 과학자"라고 부릅니다.

데이터는 과학 실험 및 측정, 인터넷 브라우징 데이터, 소셜 네트워크, 휴대전화 사용 또는 금융 거래와 같은 인간 상호 작용 등 다양한 소스에서 얻을 수 있습니다. 많은 기업들도 고객 행동 및 선호도 분석, 신용 카드 사용이나 보험 청구와 같은 행동 패턴 인식을 통한 사기 탐지, 그리고 더 정확한 위험 평가 및 수익 증대를 위해 데이터의 가치를 깨닫고 있습니다.

빅 데이터에서 통찰력을 얻기 위해 실무자들에게는 새로운 분석 기술이 필요합니다. 여기에는 시각화, 클러스터링, 데이터 마이닝과 같은 계산 집약적이고 대화형 접근 방식이 포함됩니다. 대규모 데이터 세트의 관리 및 처리는 향상된 계산 자원과 분산 컴퓨터에서 작동하는 새로운 알고리즘의 개발을 필요로 합니다.

이 과목은 최신 기술과 오픈 소스 및 독점 소프트웨어 도구를 사용하여 빅 데이터의 분석 및 관리를 학생들에게 소개합니다. 데이터 및 사례 연구는 보건 및 정보학, 생명 과학, 웹 트래픽 및 소셜 네트워킹, 거래를 포함한 비즈니스 데이터, 고객 트래픽, 과학 연구 및 실험 데이터 등 다양한 출처에서 가져옵니다. 분석, 조사 및 보고의 일반적인 원칙을 다루며, 학생들은 자신의 관심 분야 내에서 데이터 분석 과정을 비판적으로 성찰하도록 장려됩니다.

과목 자료

핸드북 및 과목 미리보기

이 과목의 강의 및 세미나 중 적어도 하나는 복습 목적 및 수업을 놓친 학생들에게 일회성 대안을 제공하기 위해 녹화되어 제공될 예정입니다. 녹화본은 이 과목에 등록된 학생들만 사용해야 합니다. 녹화본은 복습을 위한 소중한 자료이지만, 대면 수업에 참석하여 얻을 수 있는 학습 경험과 상호 작용을 대체할 수는 없습니다.

소프트웨어 (Software)

R 통계 프로그래밍 언어는 모든 강의 시연, 튜토리얼 연습 및 과제에 사용됩니다. 최근 몇 년 동안 R은 데이터 과학의 사실상 표준 플랫폼이 되었습니다. 따라서 이 과목에서 R을 배우고 사용하는 것은 이 환경에서 작업하는 경험을 제공할 것입니다. R은 무료 오픈 소스 소프트웨어이며 The Comprehensive R Archive Network에서 다운로드할 수 있습니다. R은 RStudio 환경에서 사용하는 것이 가장 쉽습니다. 각각의 최신 버전을 다운로드하십시오. 일반적으로 RStudio를 설치하기 전에 R을 먼저 설치하는 것이 좋습니다. 두 소프트웨어 사이트 모두 일반적인 컴퓨터 운영 체제(Mac, Windows, Linux)에 대한 설치 지침을 제공합니다.

주간 학습 (Study Each Week)

일반적인 모나시 과목(6학점)은 학생들이 12주 동안 매주 12시간씩 공부한다는 기초하에 설계되었습니다. FIT3152에서는 다음과 같이 시간을 할애할 것으로 가정합니다.

- 예정된 세미나 참석: 2시간
- 응용 세션 참석: 2시간
- 개인 학습: 8시간

개인 학습 시간은 응용 세션의 MULO 녹화본 시청, 강의 노트 복습, 응용 세션 준비(자료는 세션 최소 일주일 전에 게시됨), 과제 작업, 퀴즈/실습 활동 준비에 사용할 수 있습니다. 강의 슬라이드에는 많은 읽기 자료, 예제, 관련성 있고 흥미로운 배경 자료에 대한 참조가 포함되어 있습니다. 거의 모든 출판된(도서) 참고 자료는 모나시 도서관을 통해 무료로 얻을 수 있습니다. 또한 무료로 제공되는 온라인 지원 자료도 매우 많습니다. 이러한 링크를 따라가며 매우 흥미로운 데이터 과학의 세계를 탐험해 보시기를 권장합니다.

Leganto 독서 목록 Leganto 독서 목록 링크가 포함되어 있습니다. 독서 목록과 관련된 질문이나 문제는 일차적으로 담당 강사에게 문의하십시오.

과목 개요 (FIT3152 2025)

최근 몇 년 동안 정부, 기업 및 사회 전반에서 일상적으로 수집하고 분석하는 데이터의 양과 다양성이 폭발적으로 증가했습니다. 이 과목은 빅 데이터 분석과 데이터 과학자의 역할을 소개합니다. 다루는 기술에는 데이터 관리 및 변환, 시각적 분석, 소셜 네트워크 분석, 통계적 학습, 클러스터링 및 자연어 처리가 포함됩니다. 오픈 소스 산업 표준 소프트웨어를 사용하여 이러한 방법을 소개받게 됩니다. 데이터 및 사례 연구는 다양한 출처에서 가져옵니다. 분석, 조사 및 보고의 일반적인 원칙이 다루어집니다.

학습 성과 (FIT3152 2025)

- 현실 세계의 문제를 데이터 분석 기술을 사용하여 해결할 수 있는 문제로 변환하는 능력을 보여 준다.
- 분석을 위해 데이터를 정제하고 준비한다.
- 다양한 통계적, 그래픽적, 기계 학습 기술을 사용하여 대규모 데이터 세트를 분석한다.
- 분석 결과를 검증하고 비판적으로 평가한다.
- 분석 결과를 해석하고 이를 폭넓은 청중에게 전달한다.

교수법 (Teaching approach) 수업 자료는 세미나 중에 소개되며, 세미나는 강의 형태로 진행되면서 토론과 질문 기회가 주어집니다. 응용 세션은 학생들이 교육팀의 지원을 받아 개별적으로 또는 그룹으로 실무 및 이론 질문을 시도함으로써 지식을 더욱 발전시킬 기회를 제공합니다. 평가는 과제와 수업 중 퀴즈/실습 활동을 통해 이루어집니다.

핸드북 정보

과목명: FIT3152 - 데이터 분석

개요 이 과목은 빅 데이터 분석과 데이터 과학자의 역할을 소개합니다. 다루는 기술에는 데이터 관리 및 변환, 시각적 분석, 소셜 네트워크 분석, 통계적 학습, 클러스터링 및 자연어 처리가 포함됩니다. 오픈 소스 산업 표준 소프트웨어를 사용하여 이러한 방법을 소개받게 됩니다. 데이터 및 사례 연구는 다양한 출처에서 가져옵니다. 분석, 조사 및 보고의 일반적인 원칙이 다루어집니다. 학생들은 자신의 관심 분야 내에서 데이터 분석 과정을 비판적으로 성찰하도록 장려됩니다.

개설 정보 (Offerings)

- **S1-01-CLAYTON-ON-CAMPUS:** Clayton, 1학기, 대면 수업
- **S1-01-MALAYSIA-ON-CAMPUS:** Malaysia, 1학기, 대면 수업
- **S2-01-CLAYTON-ON-CAMPUS:** Clayton, 2학기, 대면 수업

선수 과목 (Requisites)

- **금지 과목 (Prohibition):** ETX2250
- **필수 선수 과목 (Prerequisite):**
 - 다음 중 하나 이수 (통계 관련):
 - FIT1006 (비즈니스 정보 분석)
 - 또는 ETC1000 (비즈니스 및 경제 통계)
 - 또는 ETF1100 (비즈니스 통계)
 - 또는 FIT2086 (데이터 분석을 위한 모델링)
 - 또는 ETC1010 (데이터 분석 입문)
 - 또는 ETW1000 (비즈니스 및 경제 통계)
 - 또는 ETW1010 (데이터 모델링 및 컴퓨팅)
 - 또는 STA1010 (과학을 위한 통계적 방법)

- 그리고(AND) 다음 중 하나 이수 (데이터베이스 관련):
 - FIT2094 (데이터베이스)
 - 또는 FIT3171 (데이터베이스)

문의처

- 책임 교수 (Chief Examiner): Dr John Betts (John.Betts@monash.edu)
- 과목 코디네이터 (Unit Coordinator): Dr Bisan Alsalibi (Bisan.Alsalibi@monash.edu) - 말레이시아 캠퍼스 담당

평가 (Assessment)

1. 과제 1 (Assignment 1): 25%
2. 과제 2 (Assignment 2): 30%
3. 퀴즈 및 실습 활동 (Quiz and practical activity): 25%
4. 과제 3 (Assignment 3): 20%

수업 활동 시간

- 응용 세션: 총 22시간
- 세미나: 총 24시간

학습량 요구사항 응용 세션은 2주 차부터 12주 차까지 예정되어 있습니다. 이 과목의 학습 성과를 달성하기 위한 최소 예상 총 학습량은 학기당 144시간입니다.

학습 자원

- 권장 도서:
 - A Tiny Handbook of R (Allerhand, 2011)
 - An Introduction to Statistical Learning (James et al., 2021)
 - Data Science for Business (Provost & Fawcett, 2013)

- Introduction to Data Mining (Tan et al., 2006)
 - An Introduction to R (Venables & Smith, 2021)
 - R for Data Science (Wickham & Grolemund, 2017)
- **기술 자원:** R 통계 프로그래밍 언어 (RStudio 환경 권장).

학문 분야 가용성 고급 컴퓨터 과학, 비즈니스 정보 시스템, 계산 과학, 데이터 과학, 비즈니스용 IT

주차별 일정 (Weekly Schedule)

- **1주 차:** 데이터 과학 및 R 소개, 기초 통계 복습
- **2주 차:** 데이터 시각화 (Data Visualisation)
- **3주 차:** 데이터 조작 (Data Manipulation)
- **4주 차:** 회귀 모델링 (Regression Modelling)
- **5주 차:** 클러스터링 (Clustering)
- **6주 차:** 의사결정 트리를 이용한 분류 (Classification using Decision Trees)
- **7주 차:** 분류기 개선 및 평가, 나이브 베이즈 분류 (Improving and Evaluating Classifiers, Naïve Bayes Classification)
- **8주 차:** 앙상블 기법, 인공 신경망 소개 (Ensemble methods, Introduction to Artificial Neural Networks)
- **9주 차:** 네트워크 분석 (Network Analysis)
- **10주 차:** 텍스트 분석 소개 (Introduction to Text Analysis)
- **11주 차:** 텍스트 분석 응용 (Text Analysis Applications)
- **12주 차:** 과목 복습, 특별 주제 (Review of the unit, special topics)