

데이터 역량 개발을 위한 자생적 연구모임 스터디그룹 1회차

이 현 창
디지털혁신실 디지털신기술반

2022.10.19

Outline

1. **자생적 연구모임**
2. **데이터 전문가**
3. **디지털과 데이터**
4. **향후 계획**

자생적 연구모임

자생적 연구모임 운영

■ 스터디 그룹

- 2022년 1회(데이터 역량 개발)
- 2023년 상, 하반기 각 1회 예정

■ 데이터 분석 라이브러리 구축

■ 디지털·데이터 프로젝트 수행

2022년 스터디 그룹

- 한국은행이 데이터 전문가 집단으로 성장하기 위한 로드맵 제안
 - RAP(Reproducible Analysis Pipeline)
 - 데이터 분석 라이브러리
 - 통합 데이터 플랫폼
 - AI/ML
 - Best Practices

데이터 전문가

데이터 전문가

■ 한국은행은 데이터 전문가일까?

- 데이터사이언스, 빅데이터, AI/ML, 클라우드, 파이썬/R, 구글, 넷플릭스 등 뒤쳐지지 않기 위해서는 새로운 트렌드를 파악하고 공부해야할 것 같지만...
- 막상 관련 책이나 강좌, 유튜브 영상을 찾아보면 너무 기초적인 내용이거나 업무와는 동떨어진 내용,
중앙은행 특성상 우리가 따로 배울 건 없으니 하던 일이나 계속 하자

■ 한국은행은 다양한 경제·금융통계와 최신 분석도구를 이용하여 데이터 분석을 수행(이 만큼 하는 곳도 많지 않음)

데이터 전문가

- 다양한 데이터를 효율적으로 분석함으로써 데이터 기반 의사결정을 지원
 - [데이터] 여러 데이터를 빠르게 조회하고 입수하여 분석에 활용
 - [효율적 분석] 데이터 입수부터 전처리, 분석, 시각화, 보고서 작성 등 전 과정을 자동화
 - [의사결정 지원] 데이터 분석 결과가 새로운 데이터에 대해 얼마나 유효할 것인지 평가
 - 현재, 효율적인 분석 프로세스 공유 및 축적을 위한 여건이 마련되어 있지 않고, 데이터 분석은 주로 표본내 설명력을 높이는데 초점

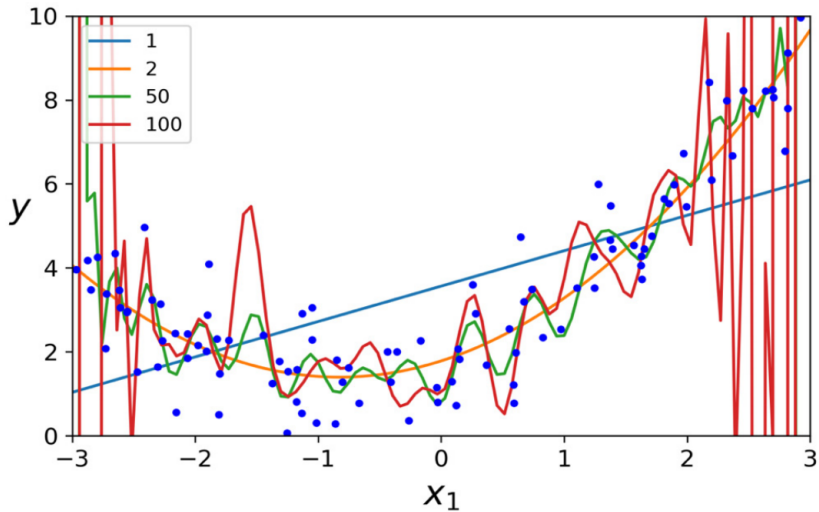
데이터

- 여러 정보 원천에 다양한 형태로 산재된 데이터를 빠르고 손쉽게 찾고 입수하여 분석에 활용
 - ECOS, IMF, BIS 등 당행에서 자주 사용하는 데이터는 BReiT를 통해 손쉽게 조회하고 입수하도록 지원하는 기능이 갖추어짐
 - [조회, 입수] 많은 경우 행내외 데이터를 탐색하고 (접근권한을 얻고) 입수하기 까지 많은 경험과 수고가 필요
- 또한, 비정형, 고빈도 등 빅데이터는 입수 절차 및 전처리가 까다로워 더 많은 시간과 노력을 들여야함

의사결정

- 데이터 분석의 표본외 예측력(out-of-sample predictability)을 제고함으로써 데이터 기반 의사결정을 지원
 - [과적합] 기존 분석은 대체로 관측된 변수의 상호관계에 대한 추론, 즉 표본내 설명(in-sample description)에 초점,
분석 결과가 표본외 데이터에서 얼마나 유효한지에 대한 평가없이 표본외 예측력을 개선하기 어렵고 의사결정에도 활용하기 어려움
 - [비선형, 상호의존성] 대부분 계량경제모형은 변수 간 선형 관계를 가정함에 따라, 결과를 설명하기는 좋지만 예측력이 낮아질 가능성

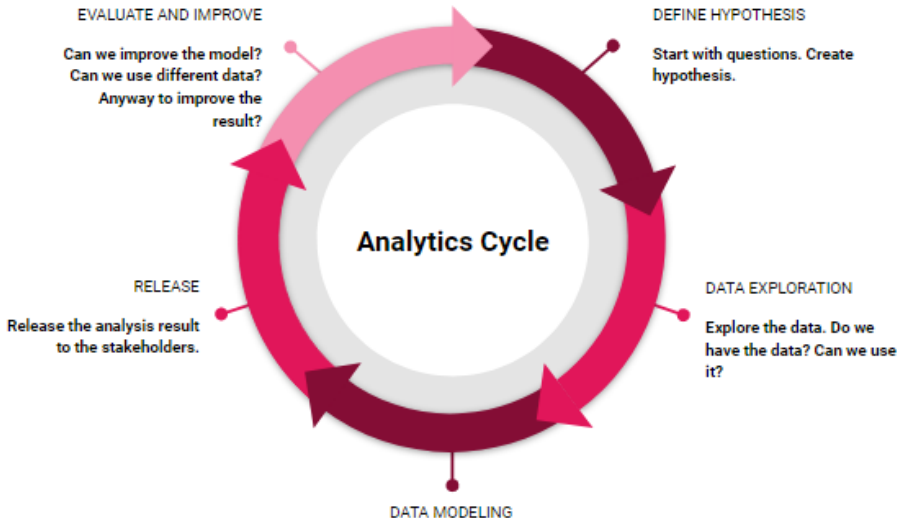
의사결정 - 과적합 (박기영·고정원, 2019)



효율적 분석

- 분석 프로세스의 모듈화, 자동화, 문서화를 통해 데이터 분석을 효율화
 - 현재 데이터 분석 업무는 개인 역량과 스타일, 업무 관행에 크게 의존
 - [모듈화] 전체 프로세스를 입력과 출력이 구분되는 단계들로 구분하고, 각 단계를 수행하는 코드(모듈)를 작성하거나 라이브러리에서 호출
 - [자동화] 데이터 입수부터 보고서 작성까지의 각 모듈을 독립적으로 수정하고, 전체 프로세스를 일괄 실행할 수 있는 분석 환경 활용
 - [문서화] 각 모듈에 대한 설명과 함께 작성된 분석 파일(코드, 데이터)은 손쉽게 공유되고 업데이트되며 한국은행 지적 자산으로 축적

효율적 분석 - Agile Analytics



디지털과 데이터

데이터 전문가를 위한 디지털 기술

	데이터 입수	효율적 분석	의사결정
1. 데이터 플랫폼 (BReiT)	●	◐	○
2. RAP	○	●	◐
3. 데이터 분석 라이브러리	○	●	○
4. AI/ML	◐	○	●

Note: ●는 큰 도움, ○는 작은 도움

1. 데이터 플랫폼

- 한국은행이 보유한 다양한 정형·비정형 데이터를 빠르게 탐색 및 입수하고, 효율적으로 분석하도록 지원하는 통합 플랫폼
 - BReiT(2019~)
 - 데이터 카탈로그(2022)
 - 데이터 레이크(2023)
 - 데이터 랩(2024)
 - 데이터 플랫폼을 100% 활용하기 위해서는 디지털·데이터 역량(RAP, AI/ML, 라이브러리 등)을 개발할 필요

1. 데이터 플랫폼 - BReiT

Model Hub

My Hub

Timeseries Hub

Microdata Hub

Bigdata Hub

Graph Hub

Model Hub

Department Hub

Reference Hub

검색

고급 데이터 분석환경

GDP nowcasting

주목시장 실거래가격지수

AI Transcribe

데이터 분석 참고자료 (10)

R (4)

데이터 입수 및 전처리

데이터 분석 시각화

데이터 모형분석

데이터 분석 참고자료

Python (4)

데이터 입수 및 전처리

데이터 분석 시각화

데이터 모형분석

데이터 분석 참고자료

Stata

EViews

기본 데이터 분석환경

안내사항

- 개인(행번) 및 팀(팀명) 작업 공간이 제공됩니다.
- 기본 분석환경에서는 16/CPU, 64GB RAM, 개인별 10GB, 팀별 20GB 저장 공간이 제공됩니다.
- 동 환경은 최신 Edge Browser에서 작동 합니다. 최초 사용 시 아래 버튼을 눌러 설치해 주시기 바랍니다.
- Matlab 사용 시 행번 및 원도우 비밀번호를 입력하여 로그인 하 주십시오.

Edge Browser 업데이트 >

사용자 설정 데이터 분석환경

안내사항

- 추가 연산자원(GPU포함) 및 저장 공간을 추가 설정하거나, 소속팀 외 직원간의 협력에 필요한 작업 공간이 제공됩니다.
- 필요시 데이터서비스팀에 문의하여 주시기 바랍니다.

데이터서비스팀 : 4600

BADA(BReiT Advanced Data Analytics) 소개

01

서버 자원을 활용한 분석환경

모든 작업이 개인 PC 성능의 한계를 넘어 고성능의 컴퓨팅 환경에 편리하게 접근하여 작업 수행이 가능합니다.

02

어느 PC에서나 동일한 환경

재무, 금융 등을 포함하여 업무에 연결된 어떤 PC에서도 동일한 분석환경을 이용할 수 있습니다.

03

추가 기능 설치 지원

로딩...
오픈소스 소프트웨어인 R, Python에서 사용하는 다양한 필수 패키지를 설치 가능합니다.
(CRAN/PyPI 지원)

04

분석용 데이터 제공

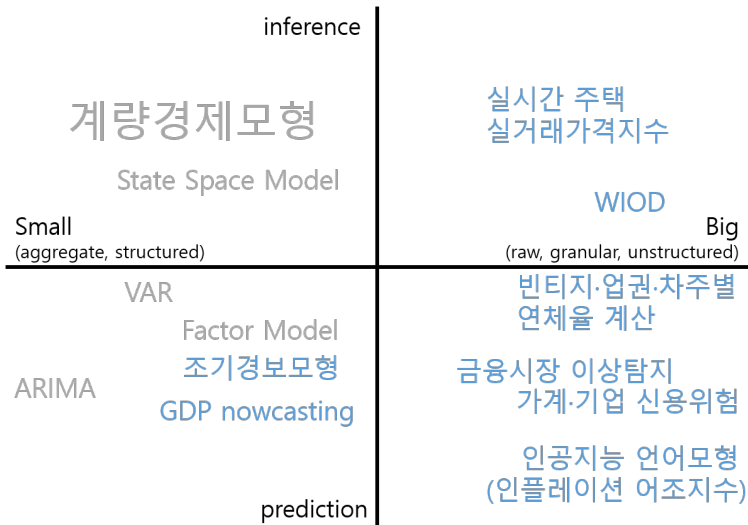
BReiT 시계열, 부동산 실거래가, 기업별 수출입 데이터 등 분석에 필요한 데이터를 사용할 수 있습니다.

17 / 30

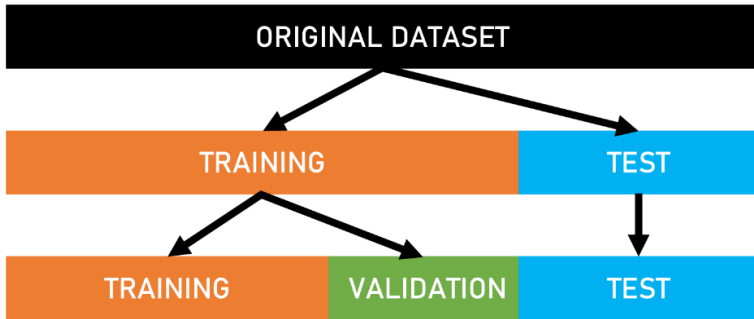
4. AI/ML

- 머신러닝 알고리즘, 모형검증 기법 등을 이용하여 변수 간 비선형, 상호의존적 관계를 포착함으로써 데이터 기반 의사결정을 지원
 - 인공지능 언어모형(AI transcribe, 인플레이션 어조지수, BERT/GPT)
 - 기계학습 알고리즘(penalized regressions, decision trees, Ensemble, Deep Neural Network)
 - 모형검증(model validation) 기법을 이용하여 모형의 표본외 예측력을 평가하고 개선
 - 분석사례 공유, 인재개발원 연수, 자생적 연구모임을 통해 지속적으로 학습

4. AI/ML - 계량경제모형과 머신러닝



4. AI/ML - Validation



2. RAP (Reproducible Analysis Pipeline)

- 데이터 분석의 전 과정을 하나의 파이프라인처럼 구축하여 분석 프로세스 실행과 업데이트, 공유/협업을 효율적으로 수행

- Replication crisis

Ioannidis, 2005, "Why Most Published Research Findings Are False"

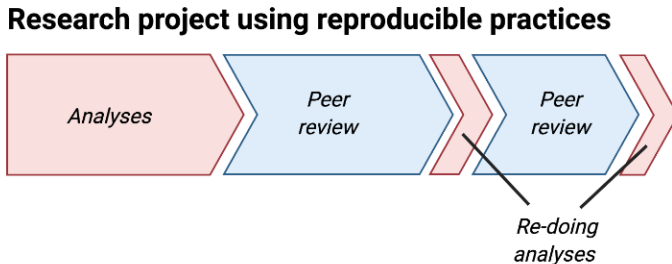
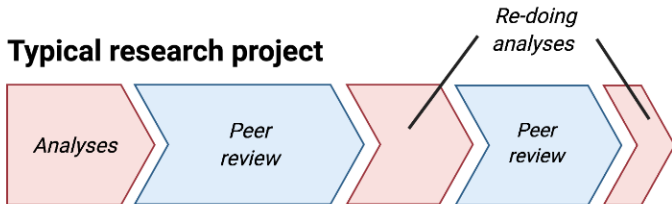
Bohannon, 2016, "About 40% of economics experiments fail replication survey"

- 모듈화, 자동화, 문서화

주피터, BReiT 모델허브, 구글 Colab, 파이썬(매틀랩, 스타타 연계)

- 효율적인 분석 프로세스 구축은 빠른실행/공유/협업뿐만 아니라 분석결과 검증에도 도움

2. RAP - Coding for Economists



2. RAP - Colab



EWS_v5.ipynb ☆

PRO

파일 수정 보기 삽입 런타임 도구 도움말 7월 22일에 마지막으로 수정됨



목차



+ 코드 + 텍스트



Early Warning Indicators for Financial Stress

{x}

0 - Design

Object

Methodology

Consideration

1- Packages

2 - Data

2.1 Data load

2.2 Data comparison across sources

2.3. Exploratory Data Analysis

Financial distress events

금융외환시장

부채와 자산가격

3 - Crises dating

Financial Stress Indices

Identifying crises

4 - Classifiers

5 - Tuning, Training & Test

5.1 Experiments with CV out-of-sample

Models & Hyperparameters

Data range

Cross-validation

Early Warning Indicators for Financial Stress

- Tae Wan, Kim (kimtwan@bok.or.kr)

0 - Design

4 숨겨진 셀 3개

1- Packages

[] 4 숨겨진 셀 4개

2 - Data

2.1 Data load

[] 4 숨겨진 셀 9개

2.2 Data comparison across sources

[] 4 숨겨진 셀 18개

3. 라이브러리

- 한국은행에서 자주 사용되는 데이터 분석 모듈(기능)을 중심으로 데이터 분석 라이브러리를 구축하여 필요한 기능을 간편하게 사용
 - quantecon, pandas, statsmodels, fredapi, tensorflow/pytorch
 - 기존 분석 코드를 모듈화하고, 이외 활용도가 높은 기능은 내부 프로젝트 또는 외부 위탁으로 개발하는 방식으로 구축
 - 계량경제모형(DFM, UCSV 등), 전망 평가모형(DM test, bootstrap 등), 데이터 입수 및 전처리(가계부채, 기업재무 데이터), 시각화 기법 등
 - 핵심 기능을 모은 '한국은행 데이터 분석 라이브러리'를 통해 당행 데이터 분석 노하우를 축적(대외 공개도 검토)

3. 라이브러리 - 파이썬 연수 자료

github.com/hyunchangyi/python101

Product Solutions Open Source Pricing

Search / Sign in Sign up

hyunchangyi / python101 Public

Notifications Fork 0 Star 3

<> Code Issues Pull requests Actions Projects Security Insights

main 1 branch 0 tags

Go to file Code

About

hyunchangyi Delete GDP nowcasting_202107.pdf #26836f on 22 Jul 116 commits

input	Delete open_in_colab.png	11 months ago
script	Update plots.py	11 months ago
EWS_slides_20220718.pdf	Add files via upload	3 months ago
Ensemble.ipynb	Colaboratory를 통해 생성됨	3 months ago
Python_for_BigData.pdf	Add files via upload	3 months ago
README.md	Update README.md	3 months ago
RF.ipynb	Colaboratory를 통해 생성됨	3 months ago
intro.ipynb	Rename python_intro.ipynb to intro.ipynb	3 months ago
intro_new.ipynb	Colaboratory를 통해 생성됨	3 months ago
lecture_note.pdf	Add files via upload	3 months ago
pandas.ipynb	Colaboratory를 통해 update	11 months ago
preprocess.ipynb	Colaboratory를 통해 생성됨	3 months ago
test.ipynb	Update test.ipynb	11 months ago
wiot.ipynb	Colaboratory를 통해 update	11 months ago
파이썬 신입직원연수(2022).pdf	Add files via upload	3 months ago

README.md

No description, website, or topics provided.

Readme 3 stars 2 watching 0 forks

Releases

No releases published

Packages

No packages published

Languages

Jupyter Notebook 100.0%

향후 계획

- 2회차(11.9), RAP 활용 방법
- 3회차(11.23), 라이브러리 개발 계획
- 4회차(12.7), 통합 데이터 플랫폼 구축 현황 및 방향
- 5회차(12.14), AI/ML과 모형검증
- 6회차(12.21), 데이터 분석 사례

■ AI/ML

- Penalized regression, Ensemble, DNN, AI language model
- 주요 모형검증 기법

■ 데이터 분석 라이브러리

- 전망 평가모형(DM test, bootstrap 등)
- 계량경제모형(DFM, UCSV 등)
- 데이터 입수 및 전처리(가계부채, 기업재무 데이터), 시각화 기법 등

감사합니다.