

기업의 도산 가능성 분석 및 예측

(중소기업 데이터 활용)

학 기
강 의
지 도 교
조 수
조 장
 원

2020-1
산학 캡스톤디자인-1
정현숙
임형열
김동진
조재혁
김태완

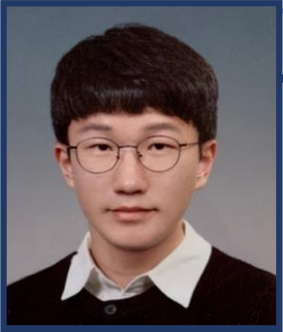
발표일자

2020. 4. 15. (수)

목 차

1. 소 개
2. 목 적
3. 개 발 환 경
4. 설계 및 구현
5. 의 견
6. 참 고 자 료

1. 소개



임형열

컴퓨터공학과 13
@doodleima

Big Data, Python



김동진

컴퓨터공학과 14
@dori.yeni

Big Data, JAVA



조재혁

컴퓨터공학과 14
@jhst_c

Big Data, Open CV, R



김태완

컴퓨터공학과 14
@tae._.wan

Big Data, Web Programming

2. 목 적

‘기업의 도산 가능성 분석 및 예측’

코로나 19

- 단기 유동성 문제 등
- 도산 기업의 증가

관련 정책

- 정부의 기업 맞춤형 정책 부족

기업과 투자자

- 문제 직관적 파악 어려움
- 기업에 투자해도 괜찮을지?

데이터 분석 & 시각화 - 문제를 직관적으로 파악
기존 값을 활용한 미래 예측 - 정책 수립, 컨설팅 자료로 활용

3. 개발 환경

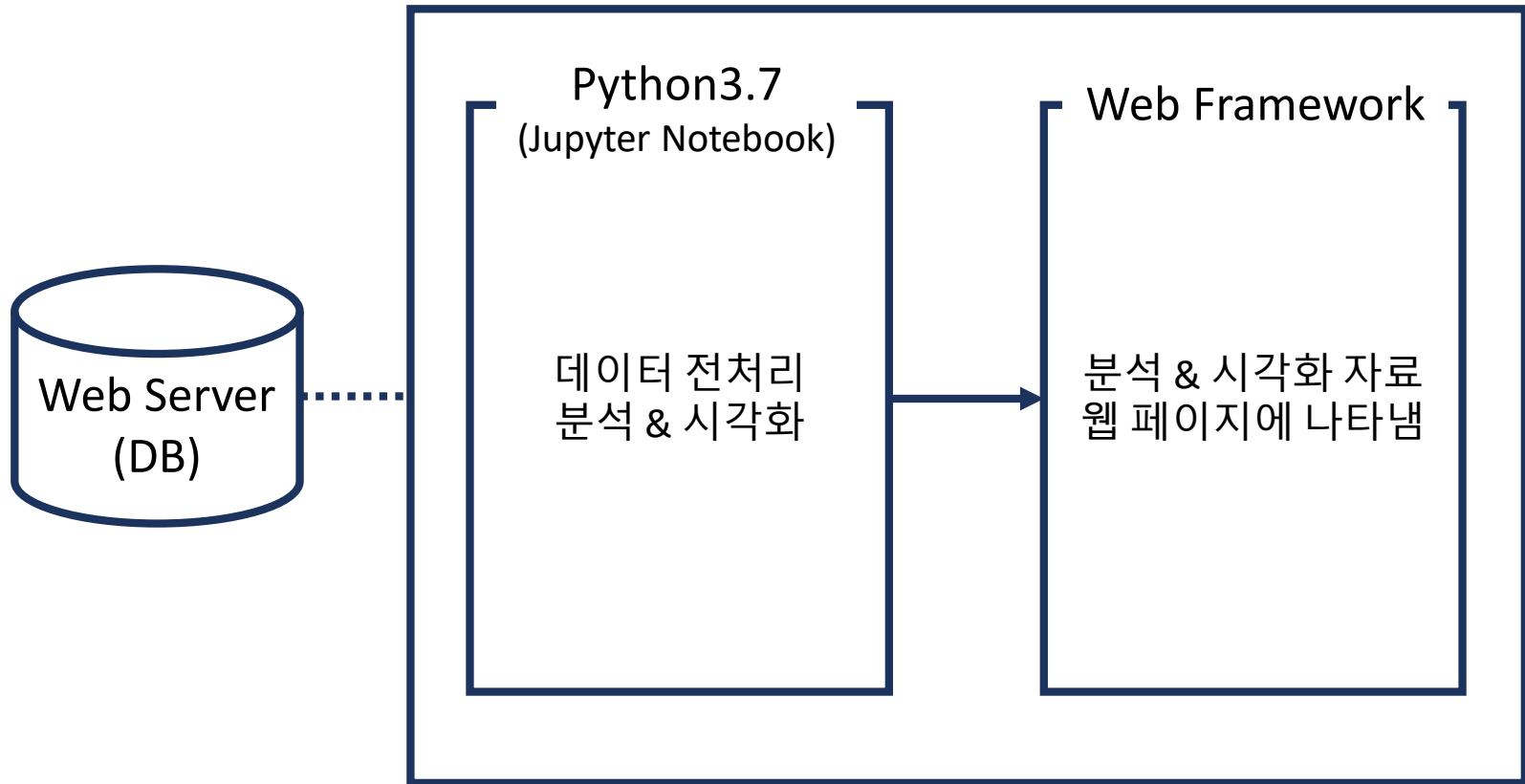
Web Framework (Django)
- 웹 개발 : 설계 & 구현
(일반 사용자가 사용하기 쉽도록..)

```
#Adj Close : 마감을 이동하여 히스토그램 작성  
sam_g = df_sam["Adj Close"]##삼성전자의 평균 마감가  
ax1 = sam_g.plot()  
lg_g = df_lg["Adj Close"]##lg전자의 평균마감가  
ax2 = lg_g.plot(grid=True)  
plt.show()
```

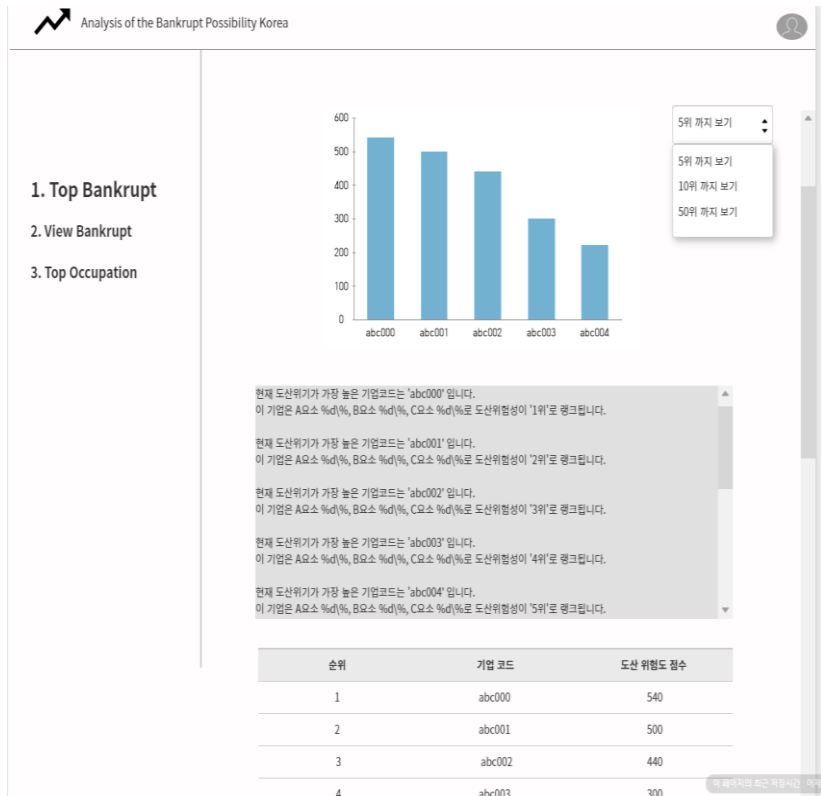


Python 3.7 (Jupyter Notebook)
- 데이터 분석, 시각화
(데이터 전처리 과정 필요)

4. 설계 및 구현_ 설계



4. 설계 및 구현_ 설계



4. 설계 및 구현_ 구현

데이터 전처리

- 광주 & 전남 기업 데이터 대상
- 많은 평가요소 존재

x20	경기실적	1: 매우나쁨 / 2: 다소나쁨 / 3: 동일 / 4: 다소좋음 / 5: 매우좋음
x21	경기전망	1: 매우나쁨 / 2: 다소나쁨 / 3: 동일 / 4: 다소좋음 / 5: 매우좋음
x30	생산실적(제조업)	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x31	생산전망(제조업)	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x32	내수실적	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x33	내수전망	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x34	수출실적	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x35	수출전망	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x36	영업이익실적	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x37	영업이익전망	1: 매우감소 / 2: 다소감소 / 3: 동일 / 4: 다소증가 / 5: 매우증가
x38	자금사정실적	1: 매우악화 / 2: 다소악화 / 3: 동일 / 4: 다소호전 / 5: 매우호전
x39	자금사정전망	1: 매우악화 / 2: 다소악화 / 3: 동일 / 4: 다소호전 / 5: 매우호전
x40	원자재조달사정실적(제조업)	1: 매우곤란 / 2: 다소곤란 / 3: 동일 / 4: 다소원활 / 5: 매우원활
x41	원자재조달사정전망(제조업)	1: 매우곤란 / 2: 다소곤란 / 3: 동일 / 4: 다소원활 / 5: 매우원활

전문가 자문을 통해
주요 요소 추출 (예정)

4. 설계 및 구현_ 구현

데이터셋 불러오기

- '19. 2. ~ '20. 1.
- 정제되지 않은 자료
(전처리 필요)

'19. 2. 데이터

2730	135972	25	1	71	1
2732	135973	25	6	26	1
2734	135975	25	4	582	2

128 rows · 46 columns

'19. 3. 데이터

2804	118159	25	7	11	1
2815	118218	25	6	20	2
2821	135925	25	1	56	1

135 rows · 46 columns

어떠한 기준으로 데이터를 처리할 것인지?

5. 의 견

기준 월 지정

기준 월에 존재하는 기업 코드만 분석

카운트

기업 코드를 카운트하여
일정 횟수 이상 존재할 경우만 분석

‘두 방식 모두 데이터 손실 가능성 존재’

두 번째 방식을 보완

일정 횟수 이상 존재할 경우 분석(횟수 높이기)
결측 데이터는 이전/이후 데이터의 평균값으로 채움

6. 참고자료

파이썬으로 데이터 주무르기 - 민형기, 비제이퍼블릭

파이썬 라이브러리를 활용한 데이터 분석 - 웨스 맥키니, 한빛미디어

Django로 배우는 쉽고 빠른 웹 개발 : 파이썬 웹 프로그래밍 - 김석훈, 한빛미디어