

전력 데이터 기반 객실 수요 예측 및 관리 솔루션

- 2020 꿈꾸는 아이 (AI) 본선 3라운드 기술 및 BM경쟁 -

2020.12.05

팀명: FE Lab.

발표자: 송준호

Contents

Part I . 기술 내용

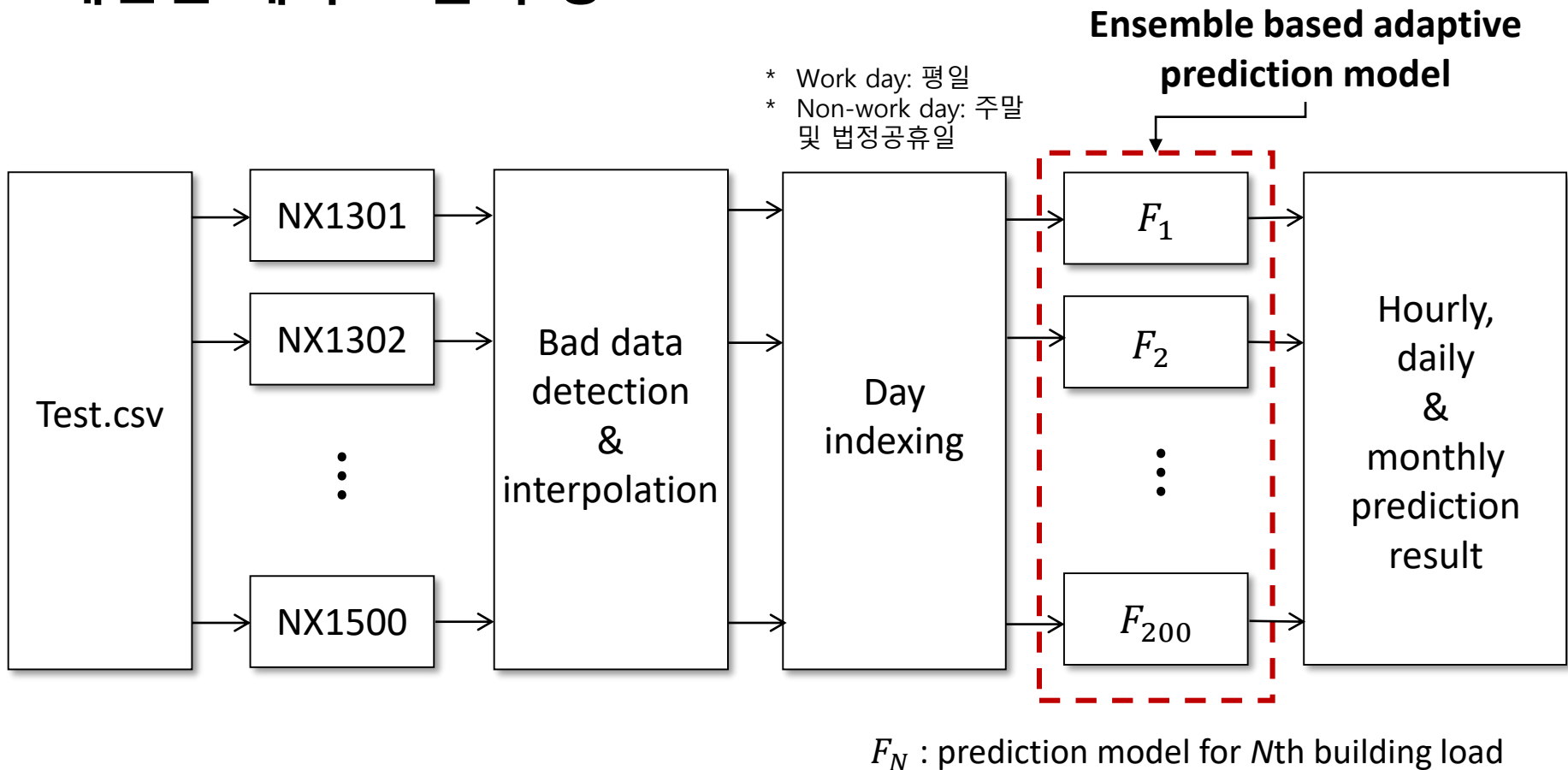
Part II . 비즈니스 모델

Part III . 참고문헌

Part IV . 부록

Part I . 기술 내용

제안한 예측 모델 구성도



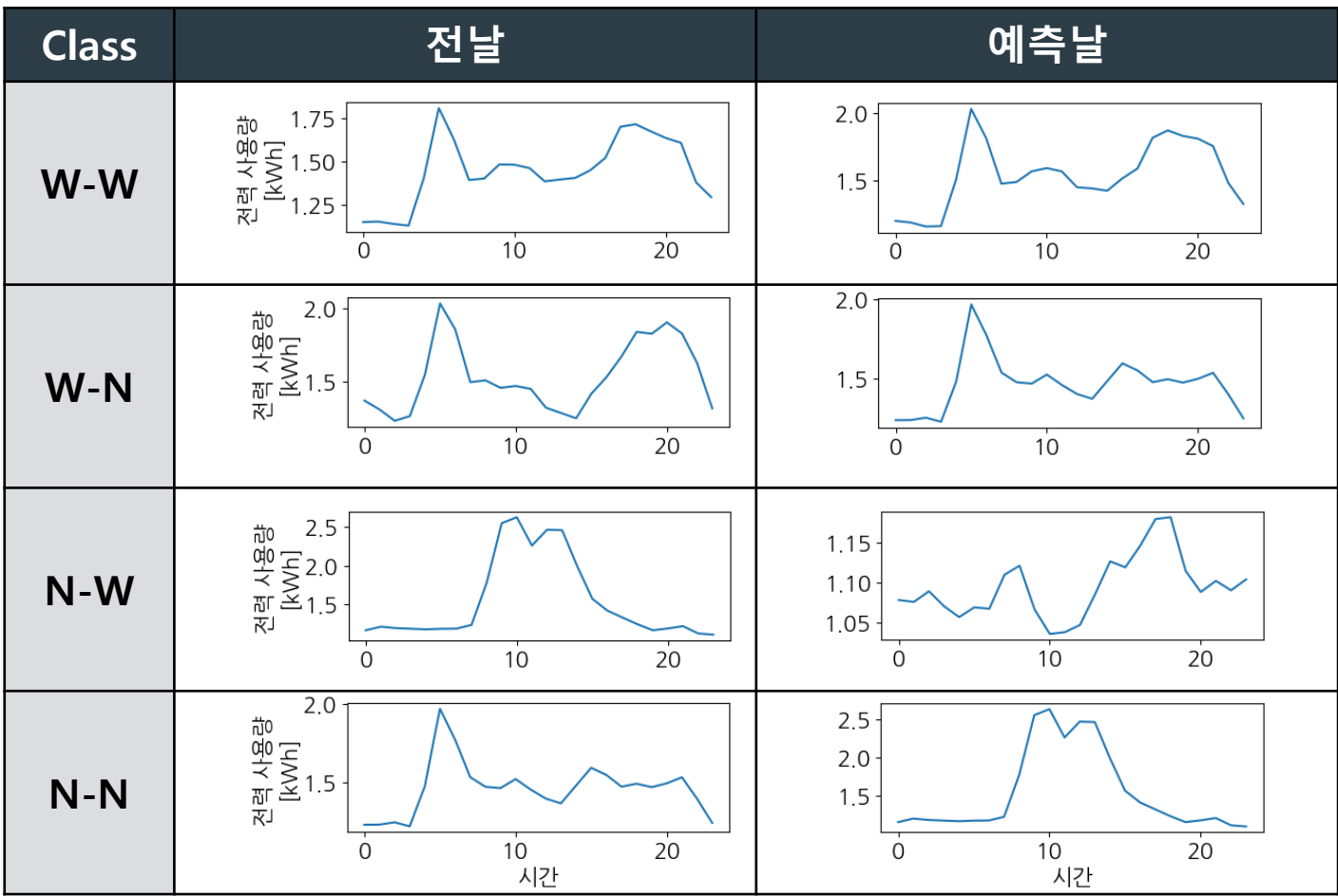
- **각 건물 부하**마다 적합한 예측 모델을 구성하여 시간별, 일별, 월별 예측을 수행

Bad data detection & interpolation

- 각 하루 전력 프로파일에 대해서 nan값 detection 수행
- 하루 전력 프로파일에서 nan 값이 특정 개수 (i.e., 4개) 이상을 넘어가면 그 날은 모델 학습에 사용하지 않음
- 하루 전력 프로파일에서 **nan 값이 특정 개수 이하**일 경우 nan 값이 있는 시간대를 기준으로 앞뒤로 20개 포인트를 이용해서 **spline interpolation**을 수행함

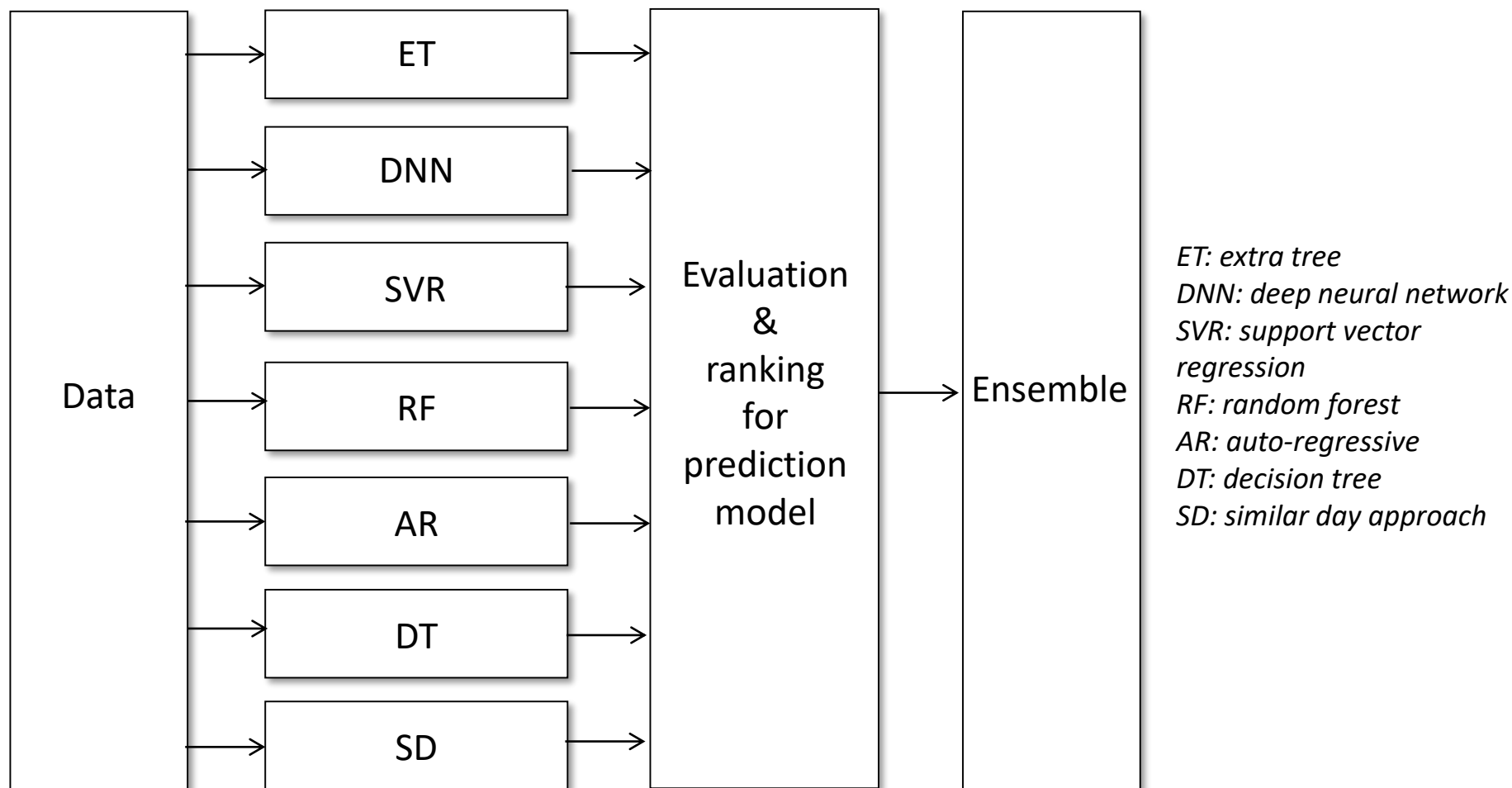
Day indexing

- Day type을 work day와 non-work day로 나눠 학습 및 테스트를 진행



* Work day: 평일
* Non-work day: 주말 및 법정공휴일

Ensemble based adaptive prediction model



- 각 건물 부하의 Validation test에서 7가지 모델을 통해 예측을 수행하고 SMAPE 값 비교를 통해 weight를 할당하여 **앙상블 수행**

일별, 월별 예측

- 일별 예측 모델

- 24시간 데이터를 합하여, 하루 사용량 데이터로 만들어서 예측 모델 구현
- **Similar day approach 예측 모델**로 예측

- 월별 예측 모델

- 예측한 Day 데이터의 합을 통해 결과 예측

* 일별, 월별 예측의 경우, **데이터 수가 부족**하여 예측모델을 충분하게 학습을 수행할 수 없으므로 'similar day approach'를 활용하여 예측을 진행

DACON 리더보드 결과

Public

Private

WINNER

1%

4%

10%

#	팀	팀 멤버	최종점수	제출수	등록일
1	FE lab.	<div> <div></div> <div></div> <div></div> <div></div> </div>	26.98438	31	16일 전
2	영듀	<div> <div></div> <div></div> </div>	27.3684	25	18일 전
3	shining_sunny	<div> <div></div> <div></div> <div></div> <div></div> </div>	27.79991	15	17일 전
4	Alhard	<div> <div></div> <div></div> <div></div> </div>	28.59153	25	16일 전
5	수요왕	<div> <div></div> <div></div> <div></div> </div>	28.63262	18	16일 전
6	ImSoPa	<div> <div></div> <div></div> <div></div> </div>	29.37608	10	16일 전
7	WooSeok Shin	<div> <div></div> </div>	29.51326	7	16일 전
8	디엔에이	<div> <div></div> <div></div> <div></div> <div></div> </div>	30.21613	30	16일 전

Part Ⅱ. 비즈니스 모델

배경

• My data 사업 [1]

- 개인정보 가명 처리, 동의 없이 정보 활용
- 제 3자에게 제공 가능 (상업 목적 포함)
- 데이터 거래소 및 고객 맞춤 서비스 활성화



2020



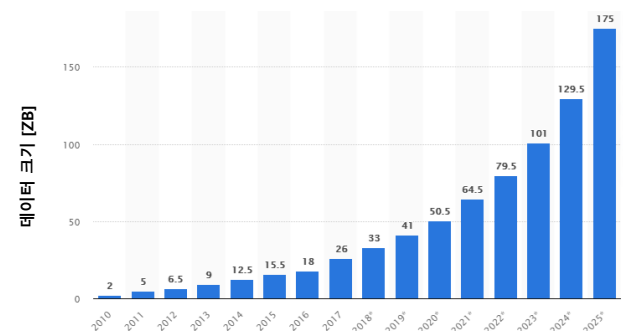
2022



2025

• 디지털 경제의 성장 [2]

- 25년도 175 ZB 규모의 데이터
- 21세기 원유로서 데이터의 가치
- 구글, 에어비엔비 등 데이터 관계 기업 성장



필요성

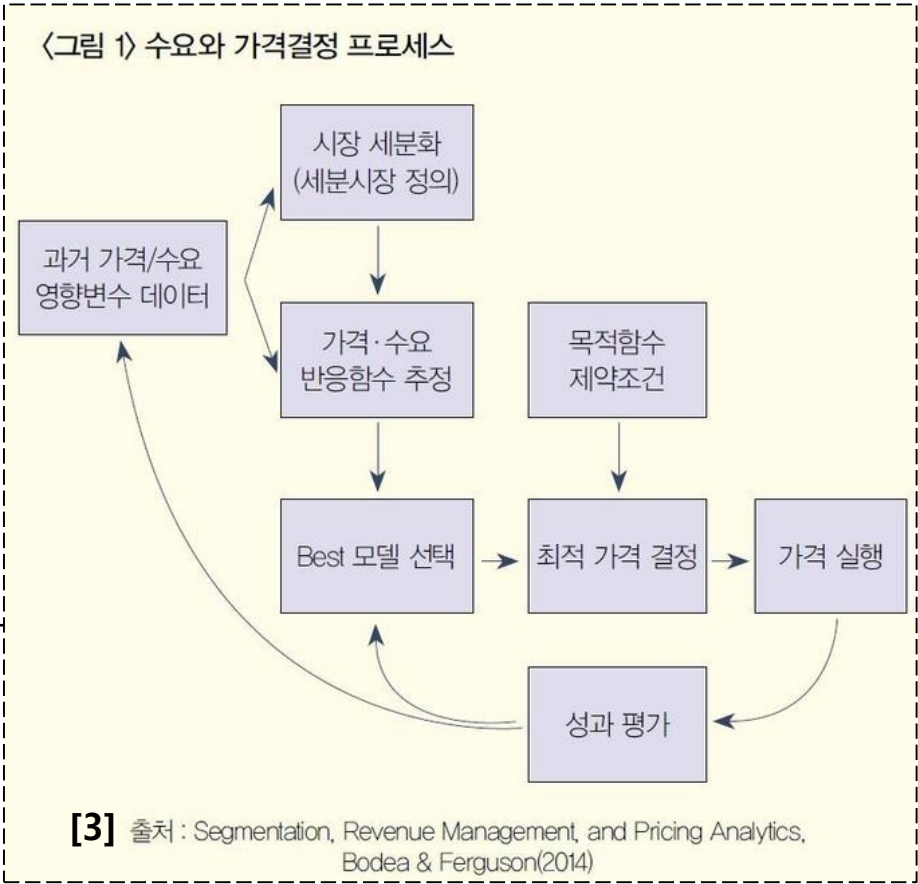
숙박 산업의 RM (Revenue Management)

▶ 빅데이터 적용 산업 리스트

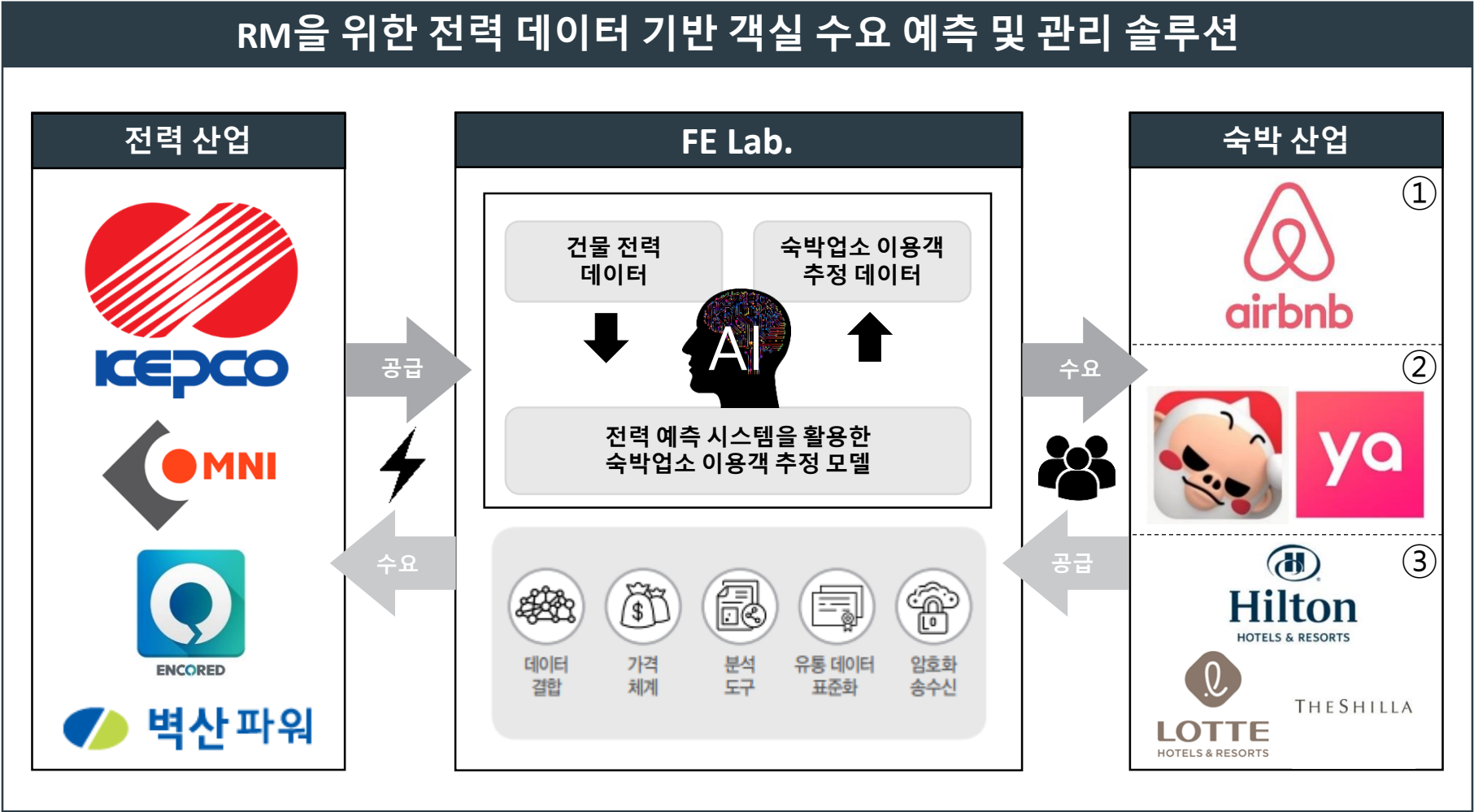
- 금융 산업
- 제조 산업
- ...
- **숙박 산업**

▶ 성과 창출 및 경쟁 우위 확보 목적

- 1) **객실 수요 예측**
- 2) 숙박 가격 결정
- 3) 비즈니스 운영 관련 의사결정



제안하는 비즈니스 모델



※ RM: Revenue management

차별성

구분	기존 숙박업 관련 서비스	제안하는 서비스
기술적 부문	<ul style="list-style-type: none"> • 객실 예약 서비스 제공 • AI 기반 챗봇 운영 	<ul style="list-style-type: none"> • 전력 데이터 기반 객실 수요 예측 서비스 제공 • 숙박시설 운영 솔루션 제공
경제적 부문	<ul style="list-style-type: none"> • 객실 예약 및 광고 수수료를 통한 이득 창출 - 숙박 업체 및 일반 소비자 	<ul style="list-style-type: none"> • 솔루션 서비스 이용료 - 숙박 업체
사회적 부문	<ul style="list-style-type: none"> • 항공업 및 숙박업소 인근 관광지 활성화 	<ul style="list-style-type: none"> • 효율적인 객실 운영을 통한 환경 보전 및 자원 절약 • 전력 사용량 및 투숙객 수 및 예측을 통한 숙박 업체의 DR (demand response) 참여 유도

기술적 실현 가능성

숙박 산업 관련 기술 동향 및 분석

- 스마트 프런트 개발 [4]
- 숙박 업소 재실 여부 판단 기술 개발 [5]

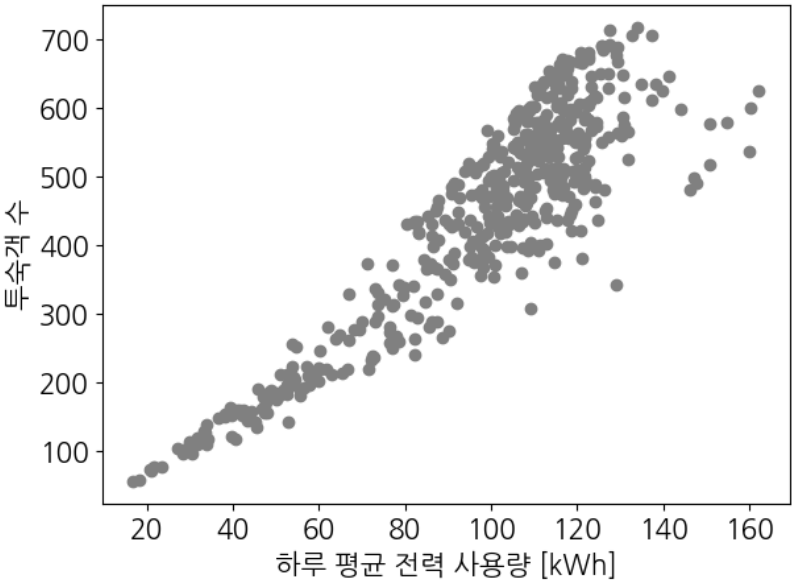


- 관련 기술 연구 동향 [6-7]

연도	저널	논문명	주요내용
2015	Energy and buildings	Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia	가구원 수 및 가구 특성을 이용한 선형 회귀 모델을 통해 연간 전력 수요를 예측
2020	IEEE Trans. on industry applications	Time-Frequency Feature Combination Based Household Characteristic Identification Approach Using Smart Meter Data	전력 사용량을 기반으로 가구원 수를 예측

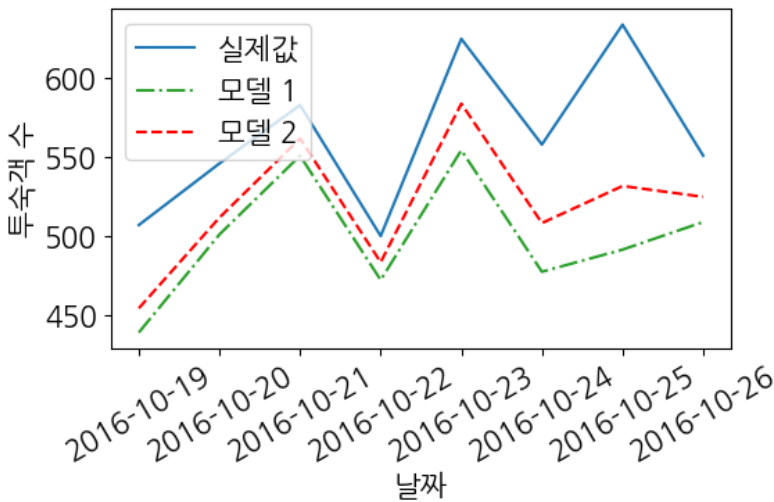
시뮬레이션

전력 데이터와 투숙객 수의 상관관계



- 상관 계수가 0.9로, 높은 상관관계를 보임
- 투숙객 수 예측에 있어 전력 사용량이 중요한 특징으로 사용될 수 있음

예측 결과



모델 1: 투숙객 수 데이터 기반 예측 모델
모델 2: 제안한 전력 데이터 기반 예측 모델

- 투숙객 수 데이터 기반 예측 모델의 SMAPE는 20.16
- 제안하는 방법의 SMAPE는 15.79로, **21.67%의 성능 향상**을 보임

모니터링 솔루션 시안

숙박업소 객실 수요 추정 모델

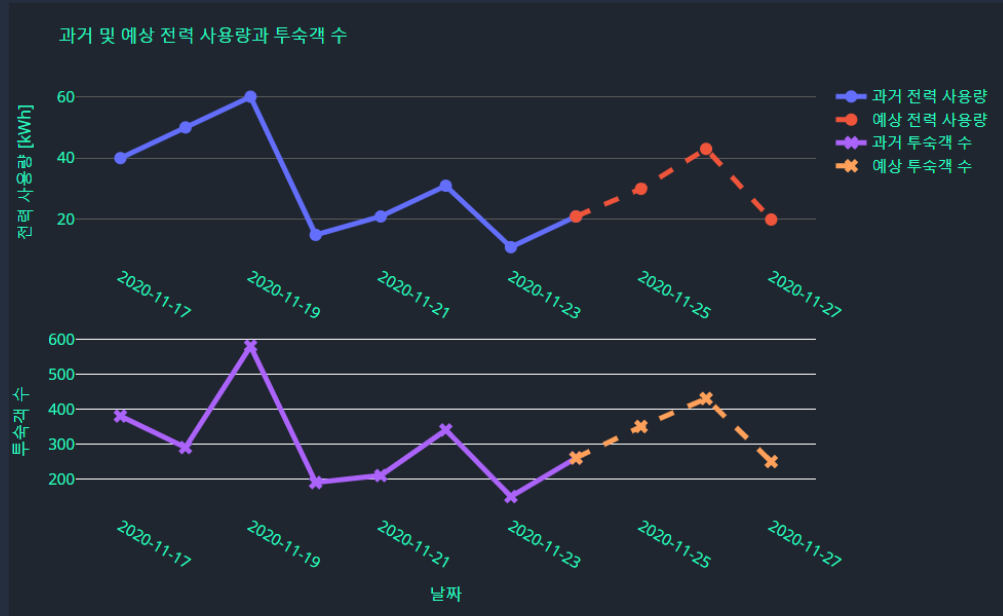
↑ 본 대시보드는 전력 사용량을 토대로 숙박업소의 객실 수요량을 추정하고 예상 수입을 산출하는 모델입니다.

분석을 원하는 업체의 ID를 입력하십시오.

호텔 A

SUBMIT

모니터링 현황

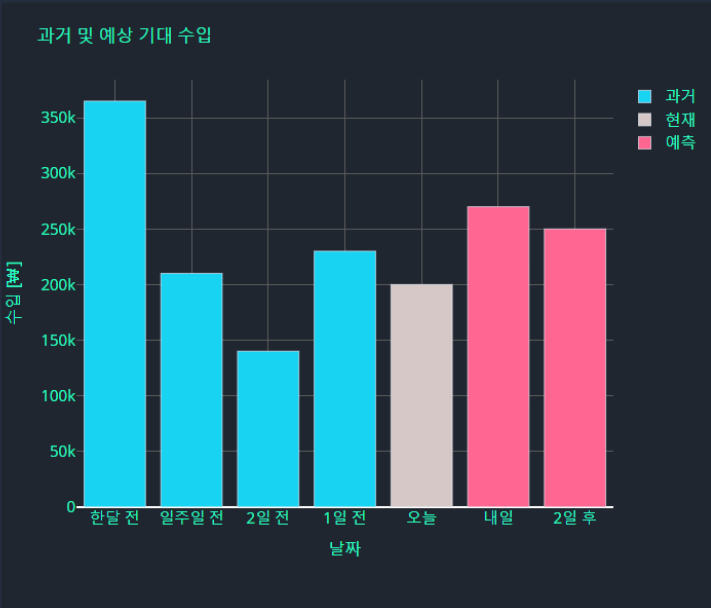


분석 결과

예상 수입

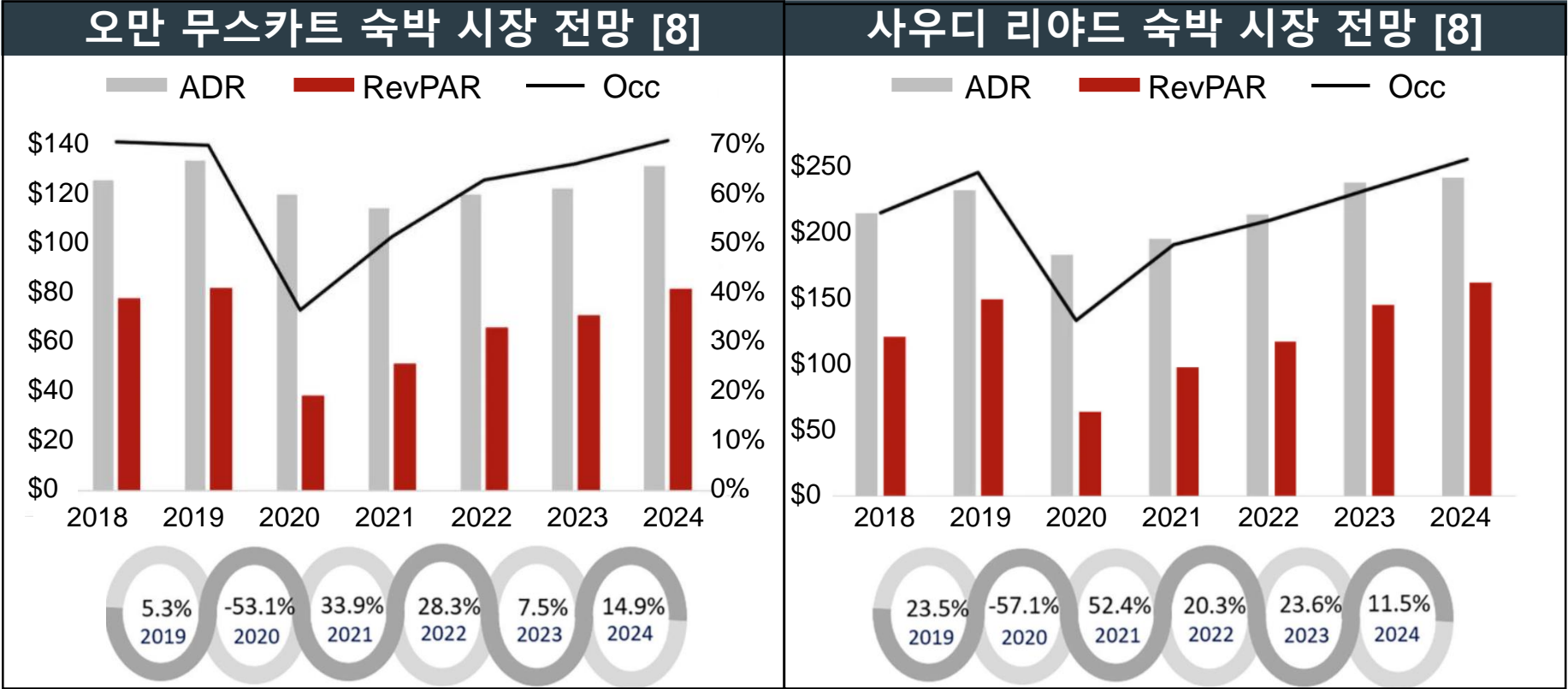
×

▼



시장진입

ADR: Average Daily Rate
RevPAR: Revenue Per Available Room
Occ: Occupancy



숙박업 시장	고객
코로나 영향에 의한 조정이 있었으나 지속적인 성장이 기대됨	숙박 사업자, 숙박시설 예약 업체

성과창출전략

■ 판매전략

- 숙박 산업 전문 **전시회 부스 운영**
- 숙박 사업자 대형 **커뮤니티 사이트 광고**
- 공공 숙박 시설(e.g. 유스호스텔)과 협업 및 시범 운영을 통한 홍보
- 숙박 시설 예약 업체 (e.g. 야놀자)와 협업을 통한 정보 공유 및 사업 확장

전문 전시회 부스 운영



숙박업 커뮤니티 광고

카페정보 나의활동

메니지 서울 1 미...

since 2012.12.09

카페소개

나루1단체

11,795 회

초대하기

* 즐겨찾는 멤버

1,088명

* 게시판 구독수

217회

* 우리카페알 수

80회

주제 친목/모임 > 업종/직종

지역 경기도

카페 가입하기

카페 체험

금액

회 전체클보기

41,552

TO문영진

처음오셨나요?

이벤트

자금소요 및 조달계획

■ 총 사업비 현황

총사업비	정부지원금	기업 대응자금	
		현금	현물
100백만원	70백만원	10백만원	20백만원
100%	70%	10%	20%

*정부지원금: 청년창업지원금

■ 조달계획

- 청년창업지원금 (7천만원)
- 예비/초기 창업 패키지 자금 투자유치
- 성장사다리펀드 투자 유치
- 기술보증기금/신용보증기금/중소벤처기업진흥공단 정책자금 투자유치

자금소요 및 조달계획

▪ 사업비 세부내역

비목	산출근거	금액(원)	
		정부지원금+대응자금 (현금)	대응자금 (현물)
재료비	• 데이터저장장치 (NAS): 500천원× 1개	5,000,000	-
기계장치 (공구, 가구, 비품, SW 등)	• 시뮬레이션용 서버 PC :300천원×4개	12,000,000	-
	• 시뮬레이션용 SW (MATLAB): 1000천원× 1개	10,000,000	-
인건비	• 홈페이지 및 홍보영상 제작비	3,000,000	-
	• 내부인건비: 5000천원×22%×12개월×4명	32,800,000	20,000,000
특허권 등 무형자산 취득비	• 특허출원비 및 관납수수료	2,000,000	-
	• 상표출원비 및 관납수수료	1,500,000	-
광고선전비	• 홍보물 제작비: 카달로그, 브로셔	1,000,000	-
교육훈련비	• SW 툴 및 알고리즘 관련 교육	1,200,000	-
여비	• 국내 출장비: 300천원×10번	3,000,000	-
	• 해외 출장비: 2000천원×2번	4,000,000	-
지급수수료	• 전시회(박람회 참가비): 2000천원×2	4,000,000	-
	• 회계감사비	500,000	-
합계		80,000,000	20,000,000

팀 구성

성명	학력	주요경력	담당업무
송준호	GIST 기전공학부 박사과정	<ul style="list-style-type: none"> ▪ 데이콘 전력 수요 예측 부문 2등 ▪ 전력 데이터 활용 신서비스 개발 경진대회 우수상 	<ul style="list-style-type: none"> - 팀 리더 - 예측 모델 설계
이지훈	GIST 전기전자컴퓨터공학부 석사과정	<ul style="list-style-type: none"> ▪ 전력 데이터 활용 신서비스 개발 경진대회 최우수상 ▪ 대구 빅데이터 분석 경진대회 장려상 	<ul style="list-style-type: none"> - 데이터 수집 및 분석 - 비즈니스 모델 설계
김명선	GIST 전기전자컴퓨터공학부 석사과정	<ul style="list-style-type: none"> ▪ 전력 데이터 활용 신서비스 개발 경진대회 최우수상 ▪ 대구 빅데이터 분석 경진대회 장려상 	<ul style="list-style-type: none"> - 변수 추출 - GUI 개발
서기업	GIST 전기전자컴퓨터공학부 석사과정	<ul style="list-style-type: none"> ▪ KSC 2019 우수발표논문상 	<ul style="list-style-type: none"> - 비즈니스 모델 설계 - 예측 모델 지원

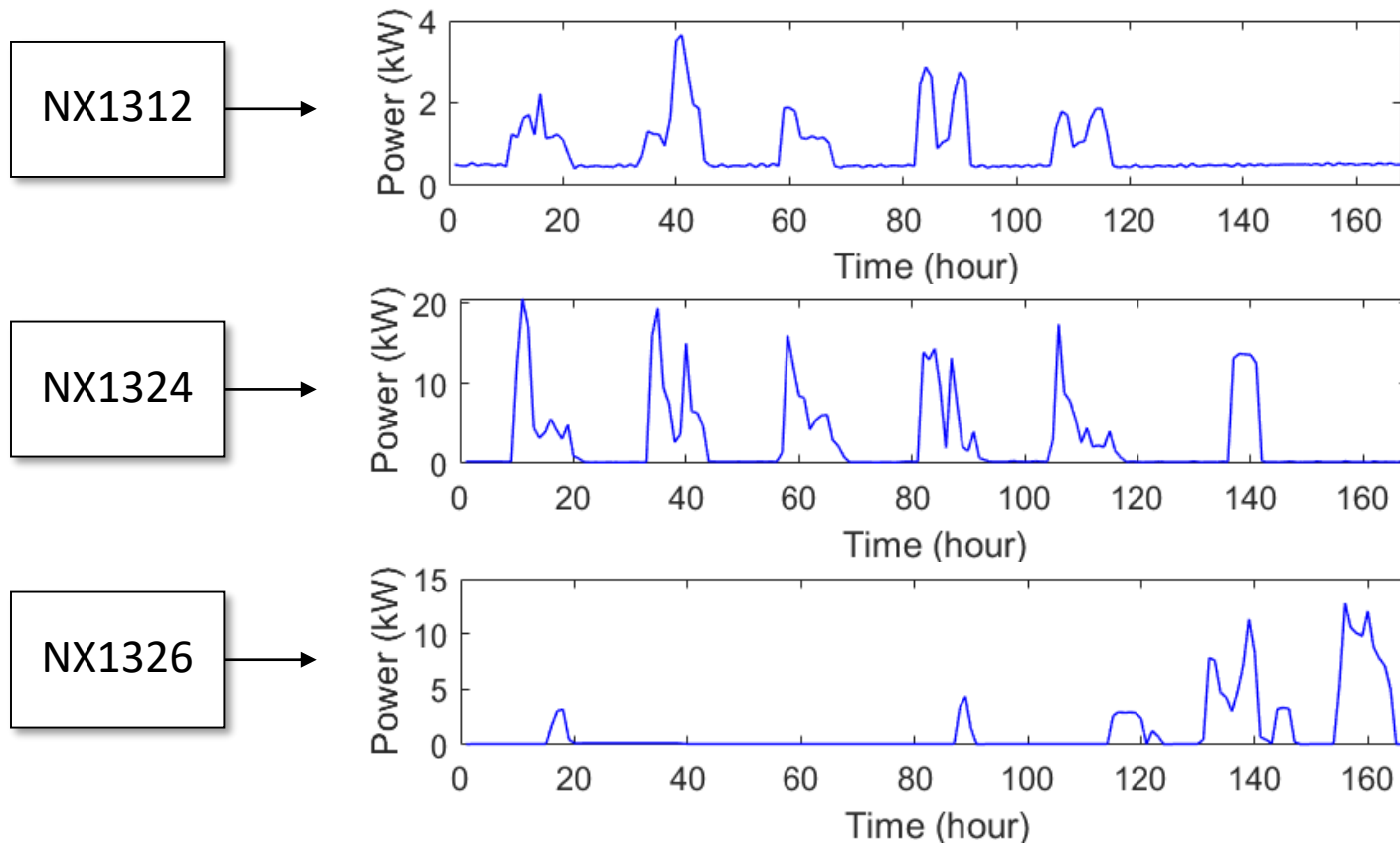
- [1]: <https://newsroom.koscom.co.kr/wp-content/uploads/2020/06/koscomnews-report-0622.pdf>
- [2]: Statista, Volume of data/information created worldwide from 2010 to 2025 (in zetabytes), 2018.12
- [3]: <https://hotelrestaurant.co.kr/mobile/article.html?no=6658>
- [4]: <https://yanolja.in/ko/> 야놀자, IoT 기술역량 강화 위해 커누스에 투자
- [5]: 김에덴, 이병탁. "전력 데이터의 특징 추출 및 XGBoost를 이용한 숙박 업소 재실 여부 판단", 한국정보처리학회 2020년 춘계학술발표대회, VOL 27 NO. 01 PP. 0458 ~ 0460 2020. 05
- [6]: Fan, H., I. F. MacGill, and A. B. Sproul. "Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia." *Energy and Buildings* 105 (2015): 9-25.
- [7]: Yan, Siqing, et al. "Time–Frequency Feature Combination Based Household Characteristic Identification Approach Using Smart Meter Data." *IEEE Transactions on Industry Applications* 56.3 (2020): 2251-2262.
- [8]: <https://www.hospitalitynet.org/opinion/4098555.html>

**THANK
YOU**

팀 명: FE Lab.

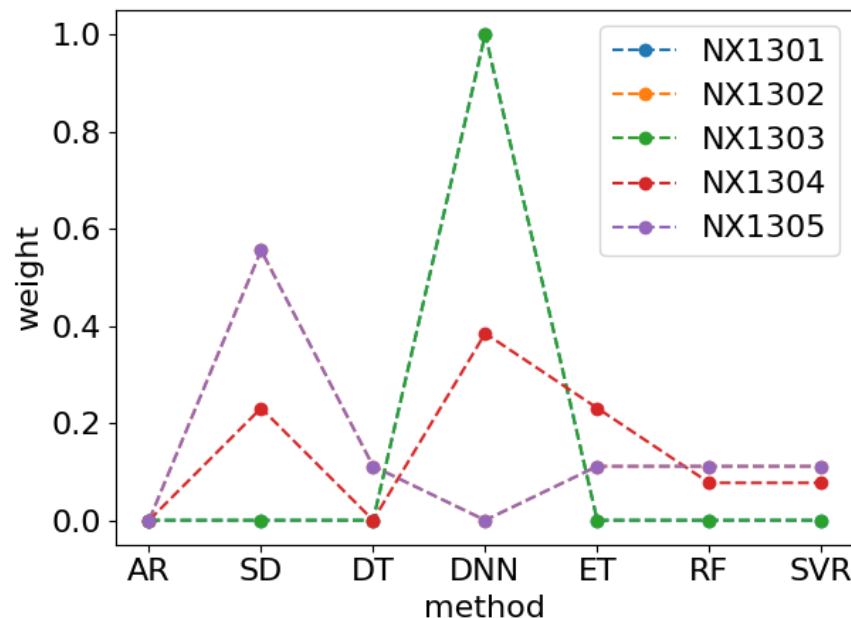
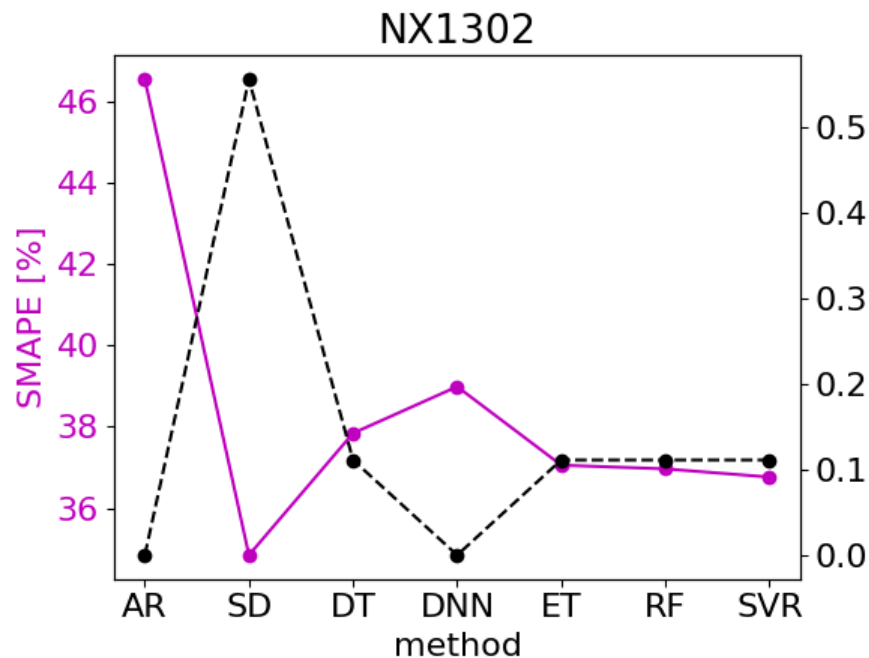
발표자: 송준호

데이터 분석



- 건물에 따라 **다양한 전력 사용량 프로파일**을 보임
- 전력 사용량이 측정되지 못하여 **nan 값**으로 표시되는 부분들이 있음

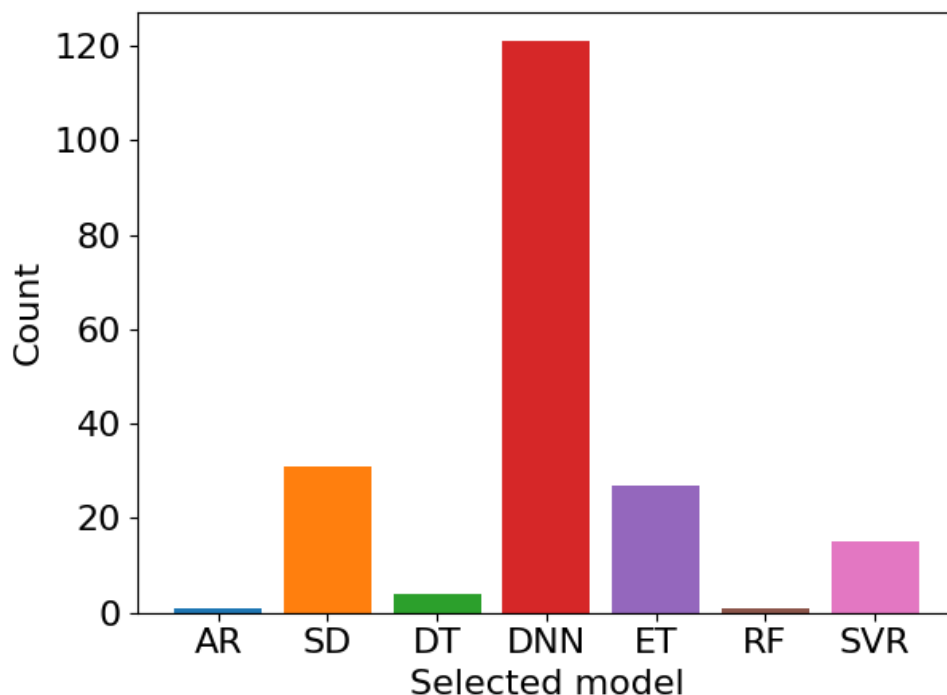
Ensemble model



- DNN 모델은 multi-layer perceptron (MLP)과 long short term memory (LSTM) 중 성능이 좋은 모델을 선택함
- 각 모델의 SMAPE 결과에 따라 **weight**를 할당하여 **ensemble** 수행
- 성능이 지나치게 낮은 경우에는 해당 모델을 선택하지 않음
- 각 건물 부하마다 **개별적으로 다른 weight**를 적용함

예측 결과

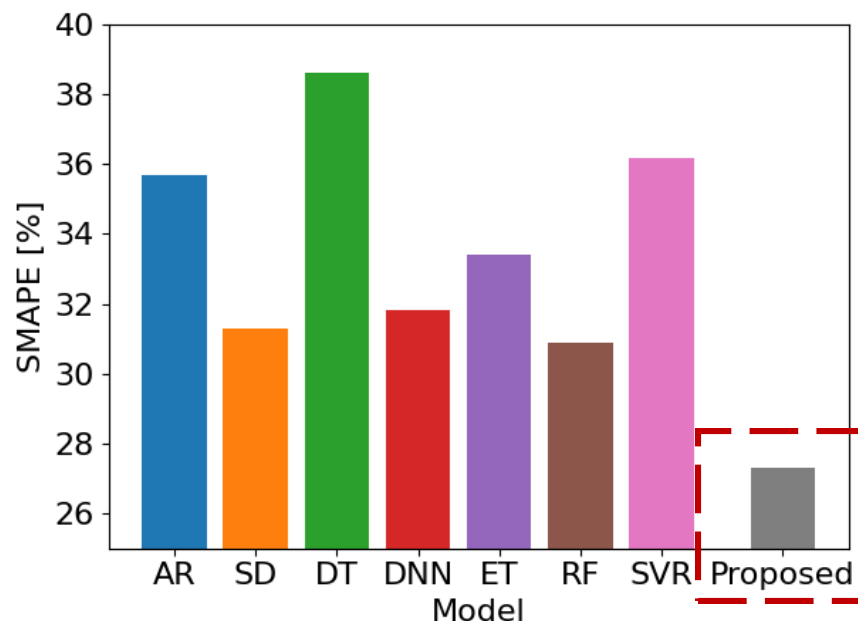
- 개별 건물 부하에 대한 모델 선택 횟수



- Test.csv 데이터의 200개 건물 부하에 대하여 예측 모델들을 평가하였을 때 **DNN이 가장 많이 선택됨**
- SD, ET, SVR이 비슷한 정도로 선택됨
- AR, RF는 거의 선택되지 않음

예측 결과

- 제안한 앙상블 기반 예측 방법과 단일 예측 모델과의 예측 성능 비교



- Test.csv 데이터의 200개 건물 부하에 대한 평균 SMAPE 비교
- 단일 예측 모델에서 가장 예측 정확도가 높은 것은 RF이며 SMAPE는 30.8%로 나타남
- 제안한 **앙상블 기반 예측의 SMAPE는 27.3%**이며 단일 예측 RF 보다 약 **11.6% 개선됨**

제안한 방법에 사용된 예측 모델

- **Auto-regressive (AR) prediction model**
 - 데이터 간의 선형 조합을 이용하여 데이터를 예측하는 모델
 - Pseudo-inverse method를 이용하여 AR model의 coefficient 계산

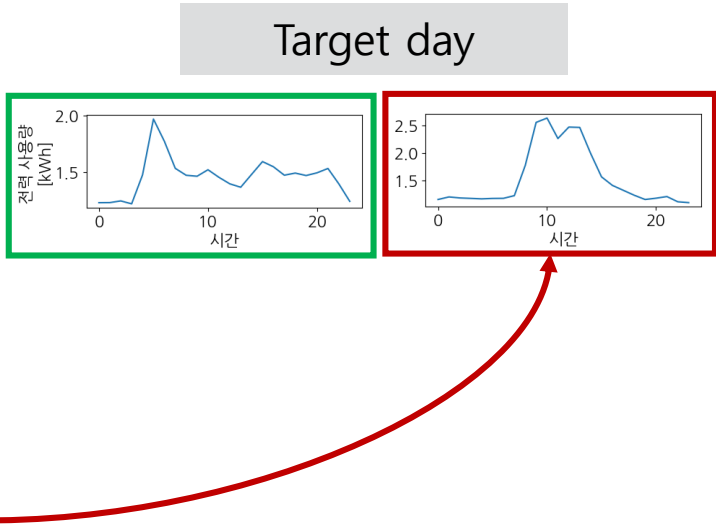
Equation	Parameter
$P_t = c + \sum_{i=1}^n P_{t-i} * W_i + \epsilon_t$	<div><div>[P_t]</div>Power</div> <div><div>[W_i]</div>Coefficient</div> <div><div>[c]</div>Constant</div> <div><div>[ϵ_t]</div>White noise</div>

제안한 방법에 사용된 예측 모델

Similar day approach (SD)

- 과거 데이터 셋을 Day type에 따라 전력 프로파일을 분류
- 예측 전날의 전력 프로파일과 가장 유사한 프로파일을 distance 계산하여 추출
- 비슷한 날의 데이터를 여러 개 추출하여 전력 프로파일을 평균하여 예측

Class	전날	예측날
W-N		
W-W		
N-W		
N-N		



제안한 방법에 사용된 예측 모델

▪ Multilayer perceptron (MLP)

- MAE loss 기반의 학습
- Adaptive Moment Estimation Algorithm (Adam) optimizer 사용
- Activation function: ReLU 사용
- Hidden layer의 개수와 unit의 개수는 예측하고자 하는 smart meter마다 개별적으로 설정

▪ Long short-term memory (LSTM)

- 첫 번째 hidden layer에 LSTM을 사용
- 모든 smart meter에 대해 동일한 hyper parameter 적용

Layer type	Unit	Activation
Input	24	
Hidden layer 1	432	ReLU
Hidden layer 2	168	ReLU
Hidden layer 3	432	ReLU
output	24	Linear

제안한 방법에 사용된 예측 모델

- **Random forest (RF)**

- 다수의 결정 트리를 통해 예측을 수행함
- MAE (mean absolute error) loss를 이용한 학습

- **Extra tree (ET)**

- Random forest보다 random성이 높음

- **Decision tree (DT)**

- 간단한 계산을 통해 어떤 항목에 대한 관측 값과 목표 값을 예측

- **Support vector regression (SVR)**

- Support Vector Machine이 가진 예측 능력을 바탕으로, 회귀문제 영역을 해결하기 위해 제안된 모델

시뮬레이션

- 사용 데이터

	주거용 전력 데이터 [1]	호텔 이용객 수 [2]
Period	2009-09-01 - 2010-12-31	2015-09-01 - 2016-12-30
Sampling rate	30분	1일
Location	아일랜드	포르투갈

- 호텔의 각 객실마다 이용객 수와 동일한 수의 사람이 주거하는 전력 데이터를 통해 가상의 전력 데이터를 생성
- 학습에 365일, 검증에 108일을 사용

- 성능 검증 지표

- Symmetric mean absolute percentage error (SMAPE)

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|X_t - \widehat{X}_t|}{(|X_t| + |\widehat{X}_t|)/2}$$

[1] <https://www.ucd.ie/issda/data/commissionforenergyregulationcer/>

[2] Antonio, Nuno, Ana de Almeida, and Luis Nunes. "Hotel booking demand datasets." Data in brief 22 (2019): 41-49.

수요 자원(Demand response) 거래시장



출처: 전력거래소(KPX)

- 전기 소비자가 아낀 전기를 전력 시장에 판매하고 금전으로 보상 받는 제도

데이터 솔루션 시장현황

- 인공지능 기반 예측 솔루션 회사

업체명	제공 서비스	고객사
Megaputer	<ul style="list-style-type: none">빅데이터 분석 및 예측 솔루션컨설팅 및 Biz 모델 제공	제조업, 요식업, 금융기관
데이타솔루션	<ul style="list-style-type: none">빅데이터 분석 및 예측 솔루션컨설팅 및 Biz 모델 제공	금융기관, 공공기관, 통신사
데일리인텔리전스	<ul style="list-style-type: none">빅데이터 분석 및 예측 솔루션챗봇 상담 서비스 제공	금융기관, 공공기관