

# 머신러닝을 활용한 나노/탄소 소재 생산공정 최적화

00 개요





교육 일정

- 01 프로젝트 개요 및 데이터셋 이해
- 02 AI 적용을 위한 데이터셋 전처리
- 03 금속분말 생산공정 최적화를 위한 선형회귀 기법
- 04 금속분말 생산공정 최적화를 위한 비선형회귀 기법
- 05 금속분말 생산공정 최적화를 위한 딥러닝 기법 심화
- 06 AI 기법 성능 향상 방법론



# 머신러닝을 활용한 나노/탄소 소재 생산공정 최적화

01 프로젝트 개요 및 데이터셋 이해





프로젝트 개요 및 데이터셋 이해

- 01 프로젝트 개요
- 02 머신러닝과 데이터
- 03 금속분말 데이터셋 소개
- 04 머신러닝 라이브러리



四星型三개요



## ❷ 목표

- **나노/탄소 소재 산업**에서 발생한 생산공정 데이터를 활용하여 **생산공정을 최적화** 할 수 있는 조건에 대한 탐색 과정 이해를 함양
- AI 기술을 도입하여 실제 산업 내에서 어떠한 효율성을 개선할 수 있을지에 대한 가능성 인지
- 최적화 조건에 대한 이해를 통해 현장에서의 개선 방향을 시도하는 실무 역량을 강화

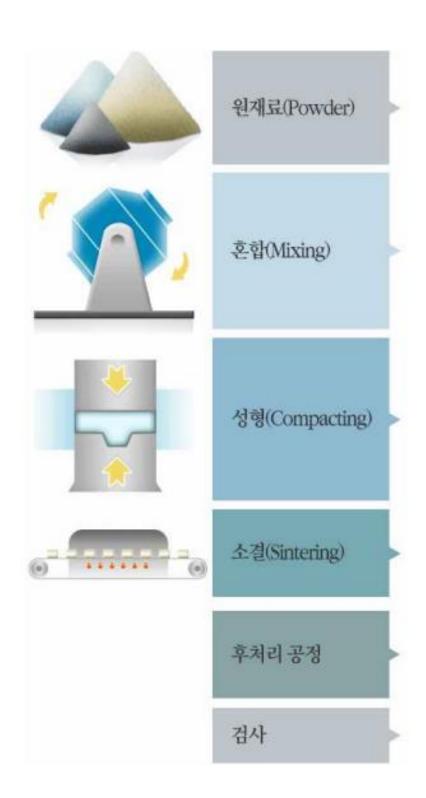


## ❷ 필요성

- 나노/탄소 소재산업은 다양한 산업에 적용되어 기존 제품의 성능향상 및 새로운 형태의 제품으로 융합되어 국내 산업발전에 크게 기여함
- •최근 나노/탄소 소재는 향후 대한민국을 이끌어 나갈 배터리, 미래자동차 (전기차, 자율주행) 등과 같은 첨단산업분야의 기술경쟁력 확보를 위해 중요한 역할을 할 것으로 예상됨
- •이러한 나노/탄소 소재분야의 확산을 위해서는 소재 성능 및 단가 안정화가 필수적이며 이를 위해소재 개발 및 제조 프로세스의 **디지털 전환** 및 **AI 적용**이 필수적임



# ❷ 금속분말 생산 공정

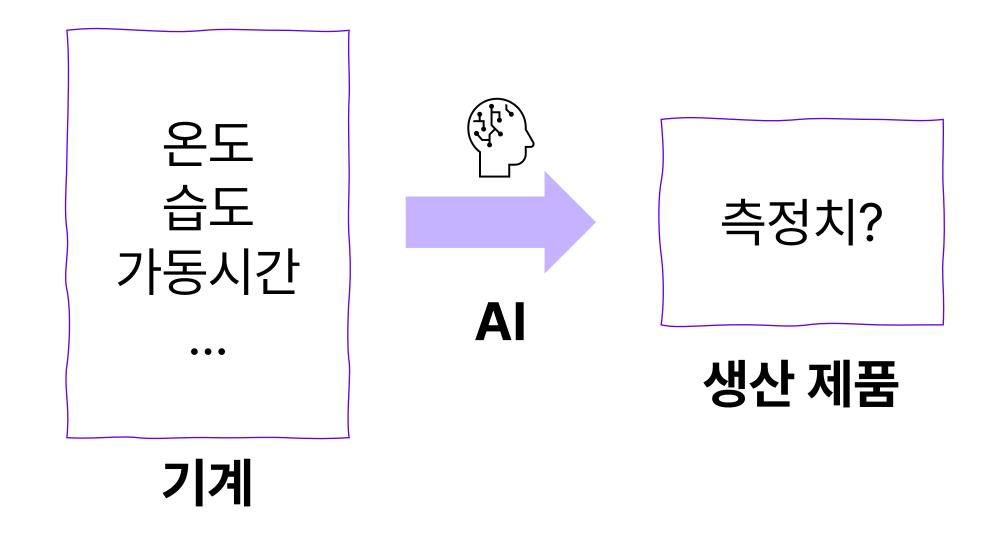


- 나노금속분말재료는 분말상태 또는 응용화 상태에 따라 고온구조재료, 공구재료, 전기·자기재료, 필터 및 센터 등의 응용이 기대되며, 다양한 산업에 사용되고 있음
- 분말의 혼합 시, 재료, 입자크기, 혼합기의 형태, 혼합기 크기, 분말 부피, 혼합속도, 혼합시간 등의 변수가 있음
- 혼합공정 원재료의 입자크기, 혼합기 등의 변수는 동일하여, 이외 환경적 요소인 습도, 온도, 설비가동시간에 따른 혼합속도 및 시간에 따라 혼합 후 분말의 밀도, 부피의 균일화가 중요함



## ਂ 프로젝트 목표

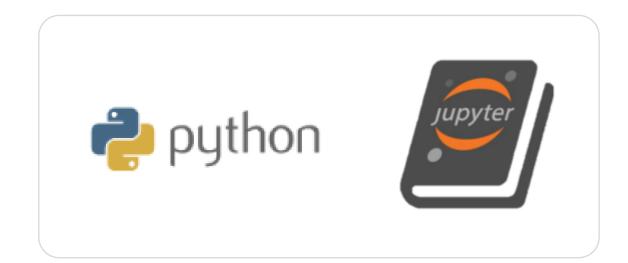
- 제품 생산 공정에서 **생산 제품의 측정치**는 **기계의 상태**에 따라 결정됨
- 기계가 임의의 상태를 가질 때, 생산 제품의 측정치를 AI 모델을 활용하여 예측





# ❷ 사용할 도구

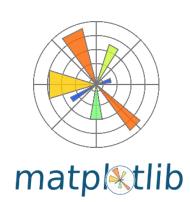
- Python & Jupyter notebook
- •머신러닝에 이용되는 다양한 python **라이브러리**













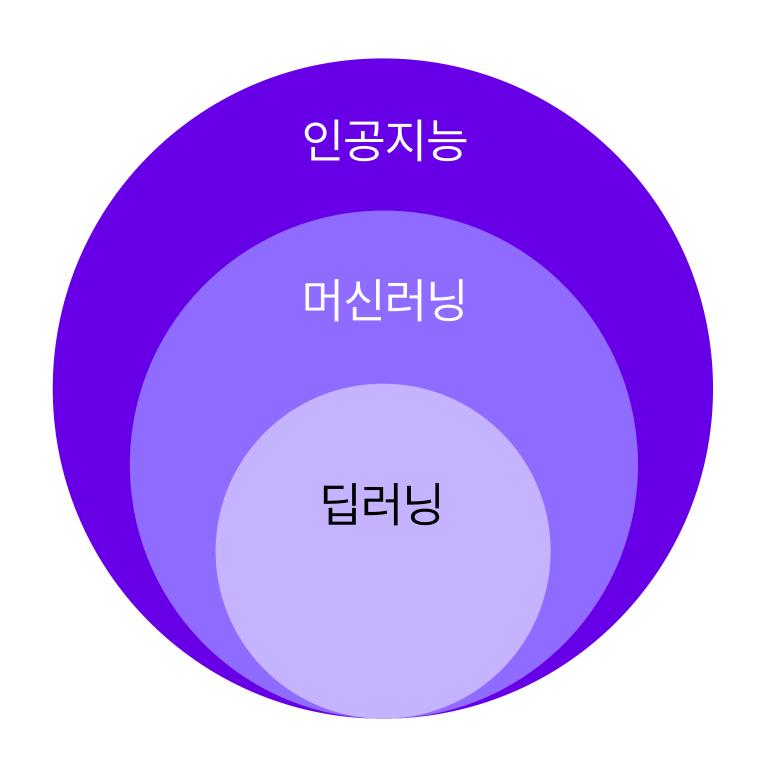




02 머신러닝과 데이터



# ❷ 인공지능, 머신러닝, 딥러닝

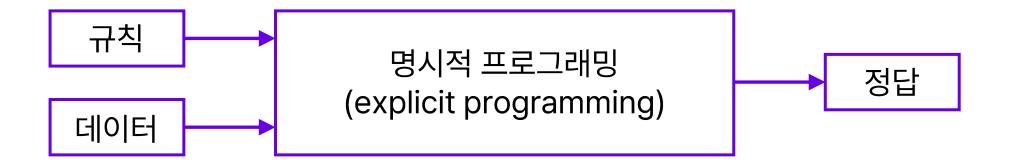


- 인공지능 인지, 추론, 행동, 적응할 수 있는 프로그램
- 머신러닝 데이터에 노출됨에 따라 성능이 향상하는 알고리즘
- 딥러닝 여러 레이어의 뉴럴 네트워크 (신경망)을 이용하여 데이터로부터 학습하는 알고리즘

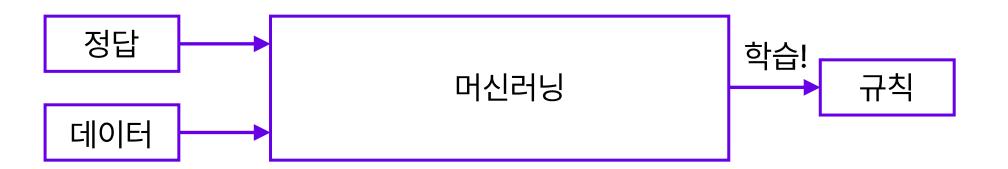


# ❷ 명시적 프로그래밍과 머신러닝의 차이

• 명시적 프로그래밍의 작동 방식



• 머신러닝의 작동 방식



#### 02 머신러닝과 데이터

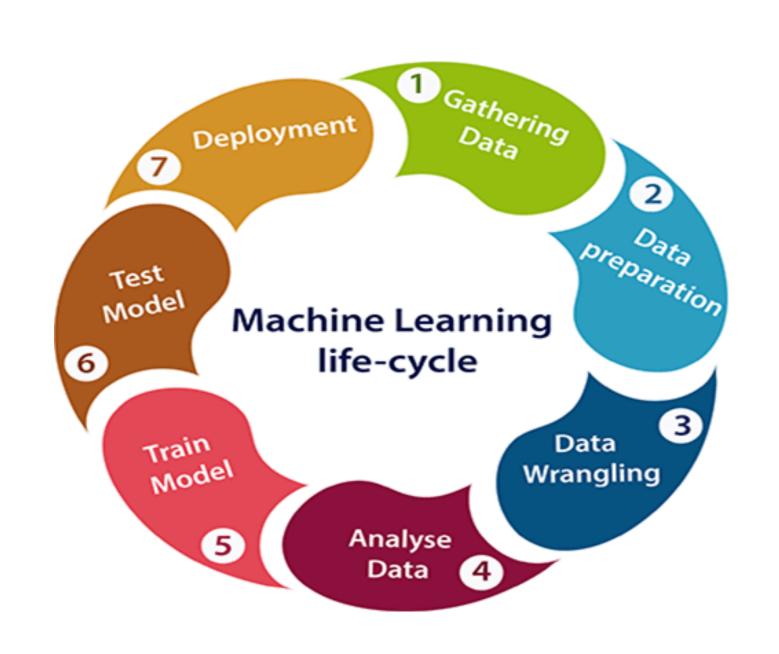


## ☑ 머신러닝의 정의

- Arthur Samuel 의 정의 (1959) 명시적으로 프로그래밍 하지 않고, 컴퓨터가 학습할 수 있는 능력을 가지게 하는 분야
- Tom Mitchell 의 정의 (1997) 어떠한 task T에 대해, 컴퓨터가 경험 E를 통해, 성능 P가 증가할 수 있다면 머신러닝이다.
- 스타크래프트 게임의 예시
  - E = 경험한 스타크래프트 경기 수
  - $T = \triangle$ 타크래프트
  - P = 다음 게임에서 이길 확률



## ❷ 머신러닝 모델의 사이클



- 1. 데이터 수집
- 2. 데이터 전처리
- 3. 데이터 청소 (이상치, 결측치 제거 등)
- 4. 데이터 분석
- 5. 머신러닝 모델 학습
- 6. 머신러닝 모델 평가
- 7. 머신러닝 모델 배포

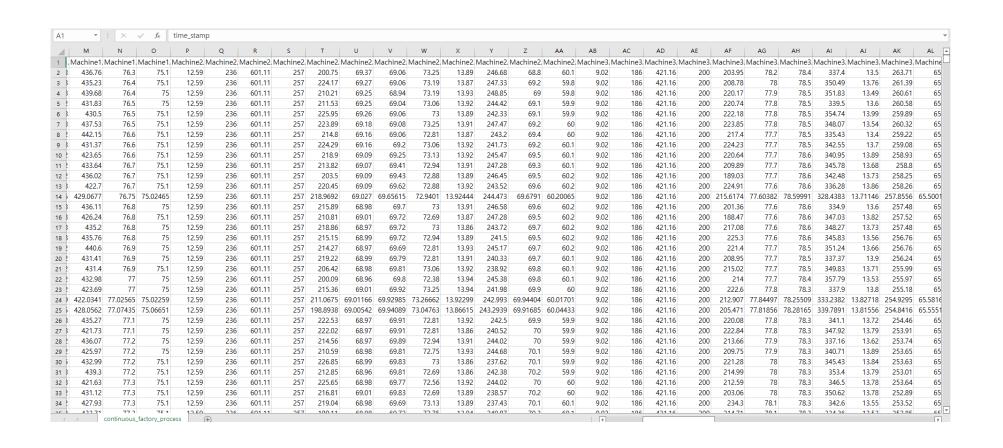


## ❷ 머신러닝의 구성 요소





# 정형 데이터와 비정형 데이터



- 구조화 된 데이터
- 날짜, 핸드폰 번호, 신용카드 번호, 온도 등



- 정해진 구조가 없는 데이터
- 이미지, 텍스트, 오디오, 비디오 등

#### 02 머신러닝과 데이터



## ❷ 정형 데이터의 분류

•범주형 데이터

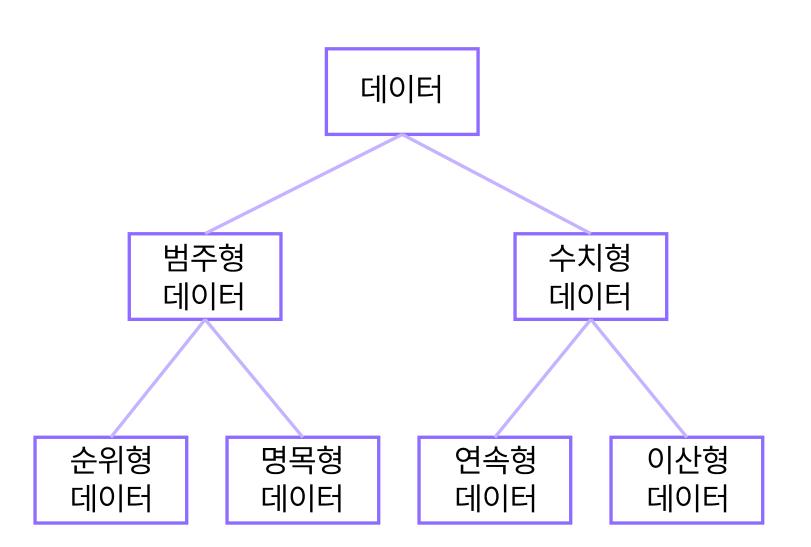
범주로 나타내어지는 데이터

- **순위형 데이터** 범주 간의 순서가 있는 데이터 (*e.g.,* 선호도, 성적)
- **명목형 데이터** 범주 간의 순서가 없는 데이터 (*e.g.,* 혈액형, 성별)

### • 수치형 데이터

수치로 측정되는 데이터

- **연속형 데이터** 실수로 표현 가능한 연속적인 데이터 (*e.g.,* 키, 몸무게, 온도, 습도)
- **이산형 데이터** 정수로 표현 가능한 셀 수 있는 데이터 (*e.g.,* 참여 인원, 나이)



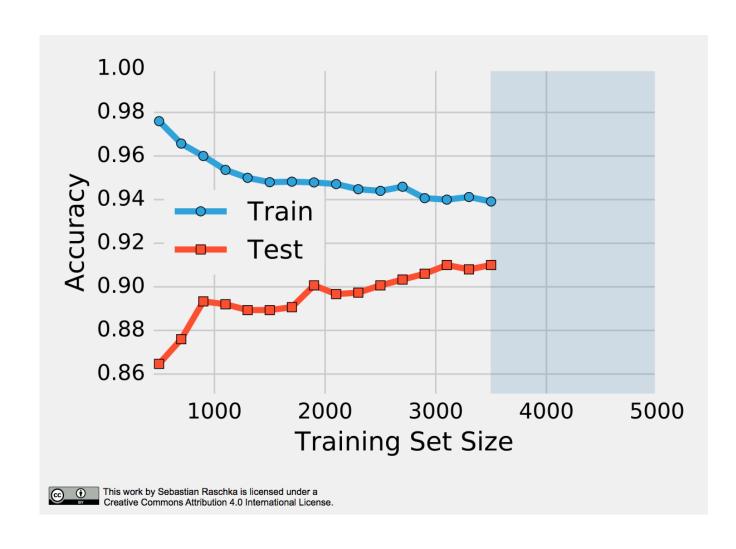


# ਂ 데이터의 중요성





#### • 더 많은 데이터는 훌륭한 알고리즘을 이긴다!



## • 하지만, 데이터만으로는 부족하다!

- 데이터에 "충분함" 이란 없다.
- 도메인 지식에 근거한 "가정" 을 머신러닝 알고리즘에 잘 반영하여야 한다.



# ❷ 머신러닝의 분류

# 지도 학습 (Supervised learning)

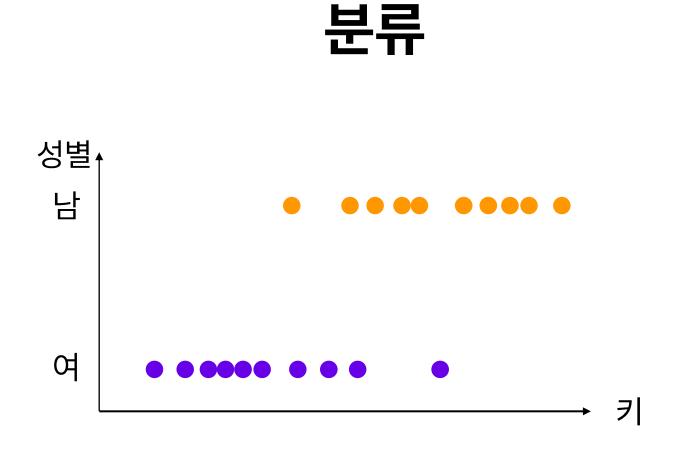
- 정답이 있는 데이터셋을 이용하여 학습
- 독립변수로부터 종속변수를 추론
- *e.g.,* 분류, 회귀

# 비지도 학습 (Unsupervised learning)

- 정답이 없는 데이터셋을 이용하여 학습
- *e.g.,* 군집화, 생성 모델

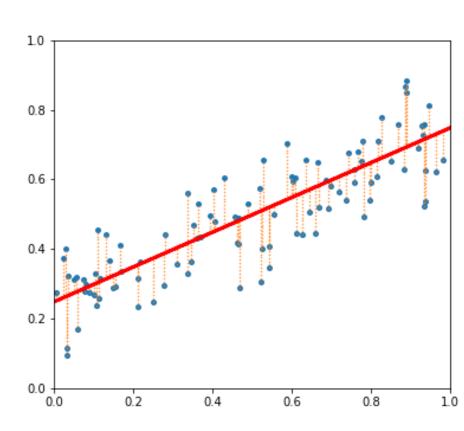


# ❷ 지도 학습



- **데이터**: 범주형 데이터
- X 축: 독립변수 (e.g., 신장)
- **Y 축**: 종속변수 (e.g., 성별)
- 분류 모델: 독립변수로부터 종속변수를 추론

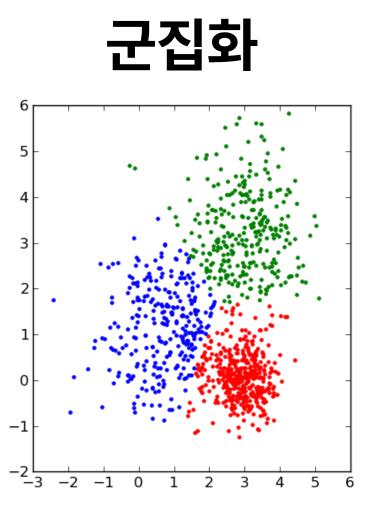
# 회귀



- 데이터: 수치형 데이터
- X 축: 독립변수 (e.g., 신장)
- **Y 축**: 종속변수 (e.g., 몸무게)
- 회귀 모델: 독립변수로부터 종속변수를 추론



# ❷ 비지도 학습



- 데이터의 분포를 분석하여, 데이터를 몇 개의 군집으로 나눔
- *e.g.,* 고객 세분화

생성



- 데이터의 분포를 학습하여, 데이터에 없는 새로운 데이터 를 생성
- *e.g.,* 딥페이크

## 02 머신러닝과 데이터



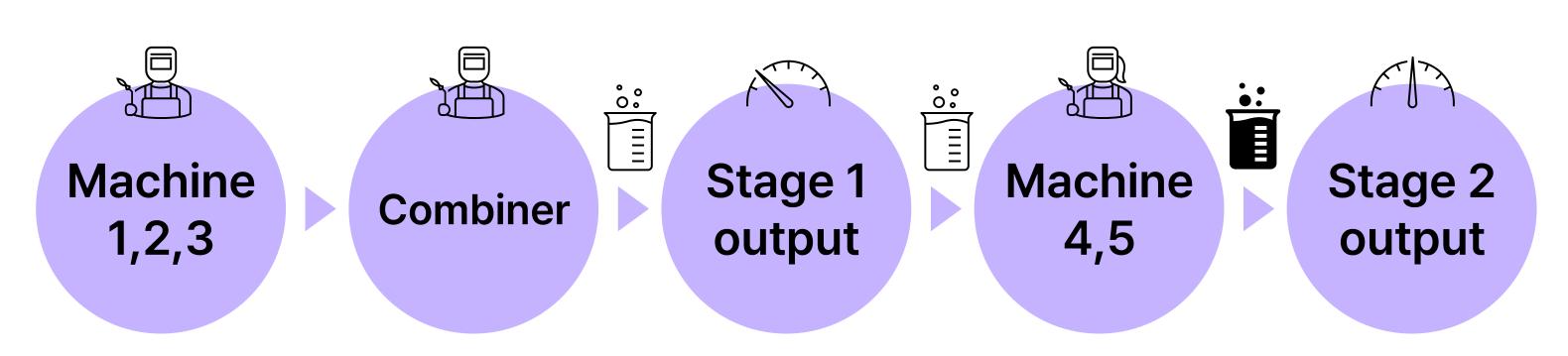
# **⊘** 프로젝트 목표 Recap

- •데이터셋
  - 수치형 데이터
  - 연속형 데이터
- •머신러닝 분류
  - 감독 학습
  - 회귀
    - 독립변수: 기계 및 작업 환경의 온도, 습도 등
    - 종속변수: 산출물의 출력값





# ❷ 생산 공정



- 각각의 기계는병렬적으로 동작합니다.
- 출력물을 Combiner로 보냅니다.
- Combiner는
   첫번째 산출물
   을 생산합니다.
- Combiner에
   서 생산된 산출
   물을 15개 위치
   에서 측정한 값
   입니다.
- 각각의 기계는**직렬적**으로 동작합니다.
- 출력물은 최종
   산출물이 됩니다.
- Stage 1에서의 산출물과 마찬 가지로, 15개의 위치에서 측정 한 값입니다.



# ⊙ 데이터 구성



# continuous\_factory\_process.csv

#### 11,270 rows $\times$ 116 columns

1 .M	lachine1 Ma				Q	R	S	T	U	V	W	X	Υ	Z	AA	AB	AC	AD	AE	AF	AG	AH	Al	AJ	AK	AL 📥
2	iaci iii le i , ivia	achine1. M	lachine1.I	Machine2. N	1achine2.N	Machine2. N	Machine2.	Machine2.N	/lachine2.l	Machine2.1	Machine2.	Machine2.	Machine2.	Machine2.	Machine2.N	/lachine3. N	Machine3. N	/lachine3.I	Machine3.I	Machine3.	Machine3.	Machine3.I	Machine3.	Machine3.N	1achine3.	Machine
2 )	436.76	76.3	75.1	12.59	236	601.11	257	200.75	69.37	69.06	73.25	13.89	246.68	68.8	60.1	9.02	186	421.16	200	203.95	78.2	78.4	337.4	13.5	263.71	65
3 }	435.23	76.4	75.1	12.59	236	601.11	257	224.17	69.27	69.06	73.19	13.87	247.33	69.2	59.8	9.02	186	421.16	200	208.78	78	78.5	350.49	13.76	261.39	65
4 }	439.68	76.4	75	12.59	236	601.11	257	210.21	69.25	68.94	73.19	13.93	248.85	69	59.8	9.02	186	421.16	200	220.17	77.9	78.5	351.83	13.49	260.61	65
5 ?	431.83	76.5	75	12.59	236	601.11	257	211.53	69.25	69.04	73.06	13.92	244.42	69.1	59.9	9.02	186	421.16	200	220.74	77.8	78.5	339.5	13.6	260.58	65
6 }	430.5	76.5	75.1	12.59	236	601.11	257	225.95	69.26	69.06	73	13.89	242.33	69.1	59.9	9.02	186	421.16	200	222.18	77.8	78.5	354.74	13.99	259.89	65
7 }	437.53	76.5	75.1	12.59	236	601.11	257	223.89	69.18	69.08	73.25	13.91	247.47	69.2	60	9.02	186	421.16	200	223.85	77.8	78.5	348.07	13.54	260.32	65
8	442.15	76.6	75.1	12.59	236	601.11	257	214.8	69.16	69.06	72.81	13.87	243.2	69.4	60	9.02	186	421.16	200	217.4	77.7	78.5	335.43	13.4	259.22	65
9 }	431.37	76.6	75.1	12.59	236	601.11	257	224.29	69.16	69.2	73.06	13.92	241.73	69.2	60.1	9.02	186	421.16	200	224.23	77.7	78.5	342.55	13.7	259.08	65
10	423.65	76.6	75.1	12.59	236	601.11	257	218.9	69.09	69.25	73.13	13.92	245.47	69.5	60.1	9.02	186	421.16	200	220.64	77.7	78.6	340.95	13.89	258.93	65
11	433.64	76.7	75.1	12.59	236	601.11	257	213.82	69.07	69.41	72.94	13.91	247.28	69.3	60.1	9.02	186	421.16	200	209.89	77.7	78.6	345.78	13.68	258.8	65
12 2	436.02	76.7	75.1	12.59	236	601.11	257	203.5	69.09	69.43	72.88	13.89	246.45	69.5	60.2	9.02	186	421.16	200	189.03	77.7	78.6	342.48	13.73	258.25	65
13 }	422.7	76.7	75.1	12.59	236	601.11	257	220.45	69.09	69.62	72.88	13.92	243.52	69.6	60.2	9.02	186	421.16	200	224.91	77.6	78.6	336.28	13.86	258.26	65
14 5 4	429.0677	76.75	75.02465	12.59	236	601.11	257	218.9692	69.027	69.65615	72.9401	13.92444	244.473	69.6791	60.20065	9.02	186	421.16	200	215.6174	77.60382	78.59991	328.4383	13.71146	257.8556	65.5001
15 }	436.11	76.8	75	12.59	236	601.11	257	215.89	68.98	69.7	73	13.91	246.58	69.6	60.2	9.02	186	421.16	200	201.36	77.6	78.6	334.9	13.6	257.48	65
16 }	426.24	76.8	75.1	12.59	236	601.11	257	210.81	69.01	69.72	72.69	13.87	247.28	69.5	60.2	9.02	186	421.16	200	188.47	77.6	78.6	347.03	13.82	257.52	65
17 }	435.2	76.8	75	12.59	236	601.11	257	218.86	68.97	69.72	73	13.86	243.72	69.7	60.2	9.02	186	421.16	200	217.08	77.6	78.6	348.27	13.73	257.48	65
18 }	435.76	76.8	75	12.59	236	601.11	257	215.15	68.99	69.72	72.94	13.89	241.5	69.5	60.2	9.02	186	421.16	200	225.3	77.6	78.6	345.83	13.56	256.76	65
19 ?	440.6	76.9	75	12.59	236	601.11	257	214.27	68.97	69.69	72.81	13.93	245.17	69.7	60.2	9.02	186	421.16	200	221.4	77.7	78.5	351.24	13.66	256.76	65
20 2	431.41	76.9	75	12.59	236	601.11	257	219.22	68.99	69.79	72.81	13.91	240.33	69.7	60.1	9.02	186	421.16	200	208.95	77.7	78.5	337.37	13.9	256.24	65
21 2	431.4	76.9	75.1	12.59	236	601.11	257	206.42	68.98	69.81	73.06	13.92	238.92	69.8	60.1	9.02	186	421.16	200	215.02	77.7	78.5	349.83	13.71	255.99	65
22 ?	432.98	77	75	12.59	236	601.11	257	200.09	68.96	69.8	72.38	13.94	245.38	69.8	60.1	9.02	186	421.16	200	214	77.7	78.4	357.79	13.53	255.97	65
23 ?	423.69	77	75	12.59	236	601.11	257	215.36	69.01	69.92	73.25	13.94	241.98	69.9	60	9.02	186	421.16	200	222.6	77.8	78.3	337.9	13.8	255.18	65
24 ) 4	422.0341 77	7.02565	75.02259	12.59	236	601.11	257	211.0675	69.01166	69.92985	73.26662	13.92299	242.993	69.94404	60.01701	9.02	186	421.16	200	212.907	77.84497	78.25509	333.2382	13.82718	254.9295	65.5816
25 5 4	428.0562 77	7.07435	75.06651	12.59	236	601.11	257	198.8938	69.00542	69.94089	73.04763	13.86615	243.2939	69.91685	60.04433	9.02	186	421.16	200	205.471	77.81856	78.28165	339.7891	13.81556	254.8416	65.5551
26 }	435.27	77.1	75	12.59	236	601.11	257	222.53	68.97	69.91	72.81	13.92	242.5	69.9	59.9	9.02	186	421.16	200	220.08	77.8	78.3	341.1	13.72	254.46	65
27 }	421.73	77.1	75	12.59	236	601.11	257	222.02	68.97	69.91	72.81	13.86	240.52	70	59.9	9.02	186	421.16	200	222.84	77.8	78.3	347.92	13.79	253.91	65
28	436.07	77.2	75	12.59	236	601.11	257	214.56	68.97	69.89	72.94	13.91	244.02	70	59.9	9.02	186	421.16	200	213.66	77.9	78.3	337.16	13.62	253.74	65
29 ?	425.97	77.2	75	12.59	236	601.11	257	210.59	68.98	69.81	72.75	13.93	244.68	70.1	59.9	9.02	186	421.16	200	209.75	77.9	78.3	340.71	13.89	253.65	65
30	432.99	77.2	75.1	12.59	236	601.11	257	226.85	68.99	69.83	73	13.86	237.62	70.1	59.9	9.02	186	421.16	200	221.28	78	78.3	345.43	13.84	253.63	65
31 }	439.3	77.2	75.1	12.59	236	601.11	257	212.85	68.96	69.81	72.69	13.86	242.38	70.2	59.9	9.02	186	421.16	200	214.99	78	78.3	353.4	13.79	253.01	65
32 }	421.63	77.3	75.1	12.59	236	601.11	257	225.65	68.98	69.77	72.56	13.92	244.02	70	60	9.02	186	421.16	200	212.59	78	78.3	346.5	13.78	253.64	65
33	431.12	77.3	75.1	12.59	236	601.11	257	216.81	69.01	69.83	72.69	13.89	238.57	70.2	60	9.02	186	421.16	200	203.06	78	78.3	350.62	13.78	252.89	65
34	427.93	77.3	75.1	12.59	236	601.11	257	219.04	68.98	69.69	73.13	13.89	237.43	70.1	60.1	9.02	186	421.16	200	234.3	78.1	78.3	342.6	13.55	253.52	65
2E-)	122.71 con	ר דד ntinuous_fa	ctory proc	1250	226	601 11	257	100 11	60.00	60.72	72.75	12.04	240.07	70.2	60.1	: [1]	106	121.16	200	21/71	70.1	70.2	224.26	12.52	252.05	



# 

time_stamp	• 데이터를 측정한 시간
<b>Ambient Conditions</b>	<ul><li>주변 상태</li><li>2개 열</li></ul>
Machine1	• Machine 1의 상태 • 12개 열
Machine2	• Machine 2의 상태 • 12개 열
Machine3	• Machine 3의 상태 • 12개 열



# 

FirstStage	• Combiner의 상태 • 3개 열
Stage1	<ul> <li>첫번째 산출물의 측정값</li> <li>30개 열</li> </ul>
Machine4	• Machine 4의 상태 • 7개 열
Machine5	• Machine 5의 상태 • 7개 열
Stage2	<ul> <li>최종 산출물의 측정값</li> <li>30개 열</li> </ul>



# 

#### time\_stamp

- 0 2019-03-06 10:52:33
- 1 2019-03-06 10:52:38
- 2 2019-03-06 10:52:40
- 3 2019-03-06 10:52:42
- 4 2019-03-06 10:52:43

...

11265 2019-03-06 14:47:13

11266 2019-03-06 14:47:16

**11267** 2019-03-06 14:47:17

11268 2019-03-06 14:47:18

**11269** 2019-03-06 14:47:20

- 1개 열
- 데이터가 기록된 시간
- 약 4시간 가량의 데이터
- 데이터 사이의 시간 간격은 가변적



## AmbientConditions

	AmbientConditions.AmbientHumidity.U.Actual	AmbientConditions.AmbientTemperature.U.Actual
0	17.24	23.53
1	17.24	23.53
2	17.24	23.53
3	17.24	23.53
4	17.24	23.53
	** •	
11265	13.84	24.43
11266	13.84	24.43
11267	13.84	24.43
11268	13.84	24.43
11269	13.84	24.43

- 2개 열
- 각각 작업 환경의 습도와 온도를 의미하 는 것으로 유추됨



```
l'Machinel.RawMaterial.Propertyl',
 'Machine1.RawMaterial.Property2',
 'Machine1.RawMaterial.Property3',
 'Machine1.RawMaterial.Property4',
 'Machine1.RawMaterialFeederParameter.U.Actual'.
 'Machine1.Zone1Temperature.C.Actual',
 'Machine1.Zone2Temperature.C.Actual',
 'Machine1.MotorAmperage.U.Actual',
 'Machine1.MotorRPM.C.Actual'.
 'Machine1.MaterialPressure.U.Actual',
 'Machine1.MaterialTemperature.U.Actual',
 'Machine1.ExitZoneTemperature.C.Actual']
```

- 12개 열
- Machine 1 의 상태



```
['Machine2.RawMaterial.Property1',
 'Machine2.RawMaterial.Property2',
 'Machine2.RawMaterial.Property3',
 'Machine2.RawMaterial.Property4',
 'Machine2.RawMaterialFeederParameter.U.Actual'.
 'Machine2.Zone1Temperature.C.Actual',
 'Machine2.Zone2Temperature.C.Actual',
 'Machine2.MotorAmperage.U.Actual',
 'Machine2.MotorRPM.C.Actual',
 'Machine2.MaterialPressure.U.Actual',
 'Machine2.MaterialTemperature.U.Actual',
 'Machine2.ExitZoneTemperature.C.Actual']
```

- 12개 열
- Machine 2 의 상태



```
['Machine3.RawMaterial.Property1',
 'Machine3.RawMaterial.Property2',
 'Machine3.RawMaterial.Property3',
 'Machine3.RawMaterial.Property4',
 'Machine3.RawMaterialFeederParameter.U.Actual',
 'Machine3.Zone1Temperature.C.Actual',
 'Machine3.Zone2Temperature.C.Actual',
 'Machine3.MotorAmperage.U.Actual',
 'Machine3.MotorRPM.C.Actual',
 'Machine3.MaterialPressure.U.Actual'.
 'Machine3.MaterialTemperature.U.Actual',
 'Machine3.ExitZoneTemperature.C.Actual']
```

- 12개 열
- Machine 3 의 상태



# **⊘** FirstStage

```
['FirstStage.CombinerOperation.Temperature1.U.Actual', 'FirstStage.CombinerOperation.Temperature2.U.Actual', 'FirstStage.CombinerOperation.Temperature3.C.Actual']
```

- 3개 열
- Combiner의 상태



# Stage1

```
['Stage1.Output.MeasurementO.U.Actual',
 'Stage1.Output.MeasurementO.U.Setpoint',
 'Stage1.Output.Measurement1.U.Actual',
 'Stage1.Output.Measurement1.U.Setpoint',
 'Stage1.Output.Measurement2.U.Actual',
 'Stage1.Output.Measurement2.U.Setpoint',
 'Stage1.Output.Measurement3.U.Actual',
 'Stage1.Output.Measurement3.U.Setpoint',
 'Stage1.Output.Measurement4.U.Actual',
 'Stage1.Output.Measurement4.U.Setpoint',
 'Stage1.Output.Measurement5.U.Actual'.
 'Stage1.Output.Measurement5.U.Setpoint',
 'Stage1.Output.Measurement6.U.Actual',
 'Stage1.Output.Measurement6.U.Setpoint',
 'Stage1.Output.Measurement7.U.Actual'.
 'Stage1.Output.Measurement7.U.Setpoint'.
 'Stage1.Output.Measurement8.U.Actual',
 'Stage1.Output.Measurement8.U.Setpoint',
 'Stage1.Output.Measurement9.U.Actual',
 'Stage1.Output.Measurement9.U.Setpoint',
 'Stage1.Output.Measurement10.U.Actual',
 'Stage1.Output.Measurement10.U.Setpoint',
 'Stage1.Output.Measurement11.U.Actual',
 'Stage1.Output.Measurement11.U.Setpoint',
 'Stage1.Output.Measurement12.U.Actual',
'Stage1.Output.Measurement12.U.Setpoint',
 'Stage1.Output.Measurement13.U.Actual',
 'Stage1.Output.Measurement13.U.Setpoint',
 'Stage1.Output.Measurement14.U.Actual',
'Stage1.Output.Measurement14.U.Setpoint']
```

- 30개 열
- Machine 1,2,3과 Combiner를 거쳐 나
   온 첫번째 산출물을 15개 위치에서 측정
   한 값
- 각 위치에서 2개의 값 씩 측정하여 총 30 개의 값



```
['Machine4.Temperature1.C.Actual',
'Machine4.Temperature2.C.Actual',
'Machine4.Pressure.C.Actual',
'Machine4.Temperature3.C.Actual',
'Machine4.Temperature4.C.Actual',
'Machine4.Temperature5.C.Actual',
'Machine4.ExitTemperature.U.Actual']
```

- 7개 열
- Machine 4 의 상태



## Machine5

```
['Machine5.Temperature1.C.Actual',
'Machine5.Temperature2.C.Actual',
'Machine5.Temperature3.C.Actual',
'Machine5.Temperature4.C.Actual',
'Machine5.Temperature5.C.Actual',
'Machine5.Temperature6.C.Actual',
'Machine5.ExitTemperature.U.Actual']
```

- 7개 열
- Machine 5 의 상태



## **⊘** Stage2

```
['Stage2.Output.Measurement0.U.Actual',
 'Stage2.Output.MeasurementO.U.Setpoint',
 'Stage2.Output.Measurement1.U.Actual',
 'Stage2.Output.Measurement1.U.Setpoint',
 'Stage2.Output.Measurement2.U.Actual',
 'Stage2.Output.Measurement2.U.Setpoint',
 'Stage2.Output.Measurement3.U.Actual',
 'Stage2.Output.Measurement3.U.Setpoint'.
 'Stage2.Output.Measurement4.U.Actual'.
 'Stage2.Output.Measurement4.U.Setpoint',
 'Stage2.Output.Measurement5.U.Actual'.
 'Stage2.Output.Measurement5.U.Setpoint',
 'Stage2.Output.Measurement6.U.Actual',
 'Stage2.Output.Measurement6.U.Setpoint',
 'Stage2.Output.Measurement7.U.Actual',
 'Stage2.Output.Measurement7.U.Setpoint'.
 'Stage2.Output.Measurement8.U.Actual'.
 'Stage2.Output.Measurement8.U.Setpoint',
 'Stage2.Output.Measurement9.U.Actual',
 'Stage2.Output.Measurement9.U.Setpoint',
 'Stage2.Output.Measurement10.U.Actual',
 'Stage2.Output.Measurement10.U.Setpoint',
 'Stage2.Output.Measurement11.U.Actual',
 'Stage2.Output.Measurement11.U.Setpoint',
 'Stage2.Output.Measurement12.U.Actual',
 'Stage2.Output.Measurement12.U.Setpoint',
 'Stage2.Output.Measurement13.U.Actual',
 'Stage2.Output.Measurement13.U.Setpoint',
 'Stage2.Output.Measurement14.U.Actual',
 'Stage2.Output.Measurement14.U.Setpoint']
```

- 30개 열
- 첫번째 산출물을 Machine4,5를 차례대로 통과시켜 나온 최종 산출물을 15개 위치에서 측정한 값
- 각 위치에서 2개의 값 씩 측정하여 총 30 개의 값



## ❷ 독립 변수와 종속 변수

# 독립 변수 종속 변수

#### •독립 변수

- 원인이 되는 변수
- 연구자가 조정할 수 있는 값
- 머신러닝 모델의 입력값

#### •종속 변수

- 독립 변수에 따른 결과
- 연구자가 알고 싶어하는 값
- 머신러닝 모델의 출력값



## 독립 변수와 종속 변수

#### Stage 1

- 독립변수
  - Machine 1,2,3 의 상태
  - Combiner의 상태
- 종속변수
  - Stage 1 Output

#### Stage 2

- 독립변수
  - Stage 1 Output
  - Machine 4,5 의 상태
- 종속변수
  - Stage 2 Output



- 각각의 기계 는 **병렬적**으 로 동작합니 다.
- 출력물을 Combiner 로 보냅니다.



Combiner

 Combiner 는 첫번째 산 **출물**을 생산 합니다.



Stage 1 output

Combiner

에서 생산된

산출물을 15

**개** 위치에서

**측정**한 값입

니다.

Machine 4,5



Stage 2 output

- 각각의 기계 는 **직렬적**으 로 동작합니
- 출력물은 **최** 종 산출물이 됩니다.

• Stage 1에서 의 산출물과 마찬가지로, **15개**의 위치 에서 **측정**한 값입니다.



## ❷ 결측치와 이상치

#### time\_stamp

- o 2019-03-06 10:52:33
- 1

- NaN
- 2 2019-03-06 10:52:40
- 3 NaN
- 4 NaN
- 5 2019-03-06 10:52:44

### • 결측치

- 수집 과정에서 오류로 인해 누락되어 비어 있는 값
- NaN (Not a Number) 으로 표시 됨
- 정제 과정을 통해 제거해주어야 함



## ❷ 결측치와 이상치

AmbientConditions.AmbientHumidity.U.Actual	
17	.24
16	.73
16	.64
16	.54
16	.44
1563	.00
15	.02
14	.42
13	.94
13	.84

#### •이상치

- 수집 과정에서 오류로 인해 정상 범주에서 크게 벗어난 값
  - 극단적으로 크거나 작은 값 (*e.g.,* 몸무게: 230 kg)
  - 논리적으로 있을 수 없는 값 (*e.g.,* 키: -30 cm)
- 정제 과정을 통해 제거해주어야 함
  - 이상치가 포함된 데이터를 제거
  - 이상치를 특정 값 (*e.g.,* 평균)으로 대체



04 머신러닝라이브러리



## ❷ 머신러닝 및 데이터분석에 활용할 수 있는 프로그래밍 언어

R, Matlab, 그리고 Python은 모두 활용성 높은 라이브러리들이 잘 구축되어 있어 머신러닝 및 데이터분석에 적합!



- 오픈 소스 (무료)
- 일반적으로 매트랩보다 빠르고, 파이썬보다 느림
- 통계 분석에 특화



- 상업 프로그램 (유료)
- 수식 계산 및 시각화에 특화
- 사용하기 쉬움



- 오픈 소스 (무료)
- 다양한 목적에 적합
- 일반적으로 가장 빠름
- 신경망 모델을 비롯한 머신러닝 모델을 사용하기에 가장 적합



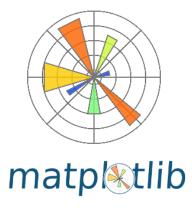
## ਂ 프로젝트에 사용할 오픈소스 파이썬 라이브러리



- Pandas
- txt, csv, excel, json, sql, json 등의 tabular data을 읽고 처리하기에 적합
- 2차원 데이터를 dataframe 데이터 구조를 활용하여 효율적으로 처리



- Numpy
- 고성능 수치 데이터를 처리하기에 적합
- C언어 기반으로 작성되어 빠르며, Pandas는 Numpy를 기반으로 작성됨
- Pandas로 읽어온 데이터를 numpy ndarray로 변환하여 머신러닝 등 이후 task 진행



- Matplotlib
- 데이터를 쉽고 빠르게 그래프로 시각화 가능



- Keras
- Tensorflow를 기반으로 작성
- 단순한 신경망 생성에 유리



- Scikit-learn
- 분류, 회귀, 클러스터링, 차원 축소 등과 관련한 다양한 알고리즘을 코드 몇 줄로 생성 가능



## **Quiz**

머신러닝의 7가지 사이클에 포함되지 않는 것은 무엇인가요?

- ① 데이터 수집
- ② 데이터 분석
- ③ 명시적 프로그램 작성
- ④ 머신러닝 모델 평가



## **Quiz**

데이터는 \_(1)\_ 데이터와 \_(2)\_ 데이터로 구분할 수 있으며, \_(1)\_ 데이터는 순위형 데이터와 명목형데이터로, \_(2)\_ 데이터는 \_(3)\_ 데이터와 \_(4)\_ 데이터로 구분됩니다.

- ① 범주형, 수치형, 연속형, 이산형
- ② 수치형, 연속형, 수치형, 이산형
- ③ 연속형, 이산형, 범주형, 수치형
- ④ 범주형, 이산형, 수치형, 연속형



## **Quiz**

데이터 수집 중 오류로 인해 손실된 값으로, 일반적으로 NAN 으로 표시되는 데이터는 무엇인가요?

- ① 이상치
- ② 결측치
- ③ 분산값
- ④ 추정값