

데이터사이언스

PROPOSAL

TEAM 2

201700638 김비아

201700658 김서희

201701387 박보성

201501257 박세인

INDEX



1

주제 소개



2

데이터 선정



3

분석 목표



4

프로젝트 진행계획

센서 데이터를 활용한 공정 이상 예측

Order Pick Time

00:05:15

Order Pick Time

002:09:27

Order Pick Time

15:14:03

공정 모니터링 시스템

- 화학 및 환경 공정의 이상을 감지하고 진단하는 시스템
- 제조 회사에서 불량품을 판별하는 방법 중 하나
- 생산 설비의 중요 처리 부위에 센서를 부착하여
품질에 영향을 미칠 수 있는 데이터를 실시간으로 수집
- 데이터 처리 기술이 더해져 이상을 사전 예측 가능

공정 이상 예측의 효과

- 사람으로 인한 오류 발생 감소
- 인력과 비용을 절감, 기업의 이윤 극대화
- 단순히 발생한 불량품을 찾는 것 뿐 아니라, 공정 과정에서 발생하는 데이터를 통해 문제를 미연에 방지

데이터 선정

데이터: uni-secom.csv

- 공정 센서와 관련된 590개의 변수
- 0부터 589는 센서 번호를 의미함
- 열 정보 : Time, 0, 1, ..., 589, Pass/Fail

A1																		
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	
2	2008-07-19 11:55	3030.93	2564	2187.733	1411.127	1.3602	100	97.6133	0.1242	1.5005	0.0162	-0.0034	0.9455	202.4396	0	7.9558	414.87	
3	2008-07-19 12:32	3095.78	2465.14	2230.422	1463.661	0.8294	100	102.3433	0.1247	1.4966	-0.0005	-0.0148	0.9627	200.547	0	10.1548	414.734	
4	2008-07-19 13:17	2932.61	2559.94	2186.411	1698.017	1.5102	100	95.4878	0.1241	1.4436	0.0041	0.0013	0.9615	202.0179	0	9.5157	416.707	
5	2008-07-19 14:43	2988.72	2479.9	2199.033	909.7926	1.3204	100	104.2367	0.1217	1.4882	-0.0124	-0.0033	0.9629	201.8482	0	9.6052	422.289	
6	2008-07-19 15:22	3032.24	2502.87	2233.367	1326.52	1.5334	100	100.3967	0.1235	1.5031	-0.0031	-0.0072	0.9569	201.9424	0	10.5661	420.592	
7	2008-07-19 17:53	2946.25	2432.84	2233.367	1326.52	1.5334	100	100.3967	0.1235	1.5287	0.0167	0.0055	0.9699	200.472	0	8.6617	414.242	
8	2008-07-19 19:44	3030.27	2430.12	2230.422	1463.661	0.8294	100	102.3433	0.1247	1.5816	-0.027	0.0105	0.9591	202.0901	0	9.035	415.885	
9	2008-07-19 19:45	3058.88	2690.15	2248.9	1004.469	0.7884	100	106.24	0.1185	1.5153	0.0157	0.0007	0.9481	202.417	0	13.6872	408.401	
10	2008-07-19 20:24	2967.68	2600.47	2248.9	1004.469	0.7884	100	106.24	0.1185	1.5358	0.0111	-0.0066	0.9494	202.4544	0	12.6837	417.600	
11	2008-07-19 21:35	3016.11	2428.37	2248.9	1004.469	0.7884	100	106.24	0.1185	1.5381	0.0159	0.0049	0.944	202.5999	0	12.4278	413.367	
12	2008-07-19 21:57	2994.05	2548.21	2195.122	1046.147	1.3204	100	103.34	0.1223	1.5144	-0.019	0.0013	0.9433	201.7125	0	11.8566	411.957	
13	2008-07-19 22:52	2928.84	2479.4	2196.211	1605.758	0.9959	100	97.9156	0.1257	1.469	0.017	-0.0154	0.9445	202.1264	0	9.1084	419.901	
14	2008-07-20 3:35	2920.07	2507.4	2195.122	1046.147	1.3204	100	103.34	0.1223	1.531	-0.0259	0.0216	0.9595	202.1269	0	8.4828	415.518	
15	2008-07-21 8:21	3051.44	2529.27	2184.433	877.6266	1.4668	100	107.8711	0.124	1.5236	-0.0209	-0.0031	0.9441	226.0086	0	9.7686	409.888	
16	2008-07-21 11:53	2963.97	2629.48	2224.622	947.7739	1.2924	100	104.8489	0.1197	1.4474	0.0144	-0.0119	0.9582	195.3787	0	9.7561	422.367	
17	2008-07-22 0:03	2988.31	2546.26	2224.622	947.7739	1.2924	100	104.8489	0.1197	1.5465	0.025	-0.0024	0.9616	192.9787	0	12.4364	424.753	
18	2008-07-22 2:59	3028.02	2560.87	2270.256	1258.456	1.395	100	104.8078	0.1207	1.4368	0.015	-0.0037	0.9623	195.1742	0	12.1805	428.982	
	uci-secom																	

분석 목표

- “센서 데이터를 활용한 공정 이상 예측”
- 어떤 센서가 이상을 발생시키는가?
=> 위험에 미리 대응하여 피해 손실 감소
- 만든 모델이 Pass/ Fail의 여부를 잘 예측하는가?
=> 가장 좋은 성능을 가진 모델링 기법 탐색

프로젝트 진행계획

A light blue chevron arrow pointing to the right, containing the number 1 in white.

1

데이터 읽기

- 파이썬 (주피터)
- Numpy, pandas

A medium blue chevron arrow pointing to the right, containing the number 2 in white.

2

데이터 정제

- 결측치 확인
- 결측치 처리
- 0으로 대체

A dark blue chevron arrow pointing to the right, containing the number 3 in white.

3

데이터 시각화

- pass/fail 시각화
- 막대그래프
- 의미있는 센서 시각화

프로젝트 진행계획



4

데이터 전처리

- feature : label
- train : test
- 정규화
- StandardScaler



5

데이터 모델링

- Logistic Regression
- 다른 모델링
Decision Tree, Random Forest,
SVC, XGB 등



6

모델 평가

- 평가 지표 설정
Accuracy, Precision, Recall

감사합니다
