

최재완

- · 세인관세법인 수입통관팀장 (관세무역분야 10년 근무)
- Python Data Science 과정 이수
- Skills
 - ✓ Python Programming
 - ✓ Microsoft Office
 - ✓ Machine Learning
 - ✓ Deep Learning
 - ✓ MySQL
 - ✓ MongoDB



국민들의 동파에 대한 불안감을 해소하고 시기적절한 사전 대비를 유도하기 위해 동파 관련 데이터의 빅데이터 분석을 통해 동파 영향인자 도출하여, 영향인자를 바탕으로 지역별 동파위험지수를 개발, 대국민 동파 예보 시행





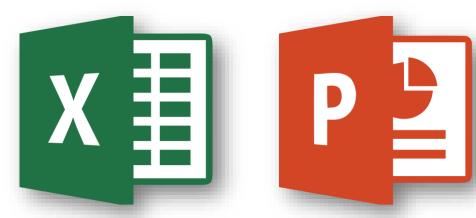


Process



X Tools







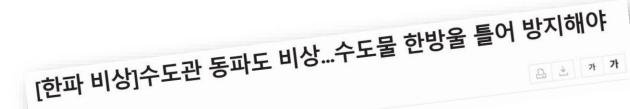




1. 문제 인식 및 가설수립

물은생명이다





[뉴시스] 입력 2016.11.23 12:00





수도관 동파신고 폭주...겨울철 동

수도관 동파신고 폭주...겨울철 동파방지 대비요령은?

원성심 headlinejeju@headlinejeju.co.kr 🖂 승인 2016.01.26 22:17:00

댓글 0 📑 💟 N 💇 🧿 🗾 💻

32년만에 찾아온 최강한파가 제주도를 엄습한 지난 23일부터 25일 사이 수도계량기 동파관련 신고가 폭주했던 것으로 나타났다.

제주특별자치도 재난관리대책본부가 26일 오후 4시 기준으로 집계한 자료에 따르면 지난 사흘 사이 접수된 수도계량기 동파신고는 1256건에 달한다. 이중 1147건은 처리가 완료됐고, 나머 지 109건은 현재까지 응급복구가 진행 중이라고 밝혔다.

또 수도관이 갑자기 얼어붙어 물이 안나오는 동결 급수불편 민원도 2561건 접수됐다.

이번 한파 내습 기간 중 제주도의 최저기온은 영하 5도에서 7도.

이 정도 기온이 불과 이틀정도 지속됐는데도 이처럼 동파 및 동결 관련 신고가 폭주한 것은 한파 날씨가 극히 적은 제주도의 날씨 특성상 수도관 동파에 대비한 준비가 부족했던 때문으로 풀이 된다.

大ちに

"이번주말 한파, 수도계량기 동파 조심"...주요 원인은?

"이번주말 한파, 수도계량기 동파 조심"…주요 원인은?

[뉴시스] 입력 2017-01-12 15:54 | 수정 2017-01-12 16:05

안내▶JTBC 뉴스는 여러분의 생생한 제보를 기다리고 있습니다.





豆は大川 短れて十十十日 豆をトワルトル?





수도미터 동파의 가장 큰 원인은 "한파"이다.

동파에 영향을 미치는 기상요소는 여러 가지가 있을 것이며,

이러한 기상 요소들의 조합은 동파와 일정한 상관관계가 있을 것이다



파주시 동파데이터

✓ 내용: 파주시 수도 미터 및 민원에 대한 데이터

✓ 대상기간: 2010년 ~ 2016년

✓ 파일형식 : Excel

✓ 출처 : 공모전 주최 측

구분	동파 발생 데이터	수용가 데이터	수도미터 데이터	급수관로 데이터
내용	파주시 수도계량기 민원 정보	수용가 단위 월별 수도 사용량	수용가 단위 수도미터 정보	급수관로 위치 및 세부정보
열	5	86	13	4
행	6,582	53,099	54,589	53,889

기상데이터

✓ 내용: 종관기상관측장비(ASOS)로 부터 관측된 기상자료

✓ 대상기간: 2010년 ~ 2016년

✓ 파일형식: csv

✓ 출처: 기상청 국가기후데이터센터(https://data.kma.go.kr/)

평균기온(°C)	최저기온(°C)	최저기온 시각 (hhmi)	최고기온(°C)	최고기온 시각 (hhmi)	강수 계속시간(hr)	10분 최다 강수량 (mm)	10분 최다강수량 시각(hhmi)	1시간 최다강수량 (mm)	1시간 최다 강수 량 시각(hhmi)
일강수량(mm)	최대 순간 풍속 (m/s)	최대 순간 풍속 풍 향(16방위)	최대 순간풍속 시 각(hhmi)	최대 풍속(m/s)	최대 풍속 풍향 (16방위)	최대 풍속 시각 (hhmi)	평균 풍속(m/s)	풍정합(100m)	평균 이슬점온도 (°C)
최소 상대습도(%)	최소 상대습도 시 각(hhmi)	평균 상대습도(%)	평균 증기압(hPa)	평균 현지기압 (hPa)	최고 해면기압 (hPa)	최고 해면기압 시 각(hhmi)	최저 해면기압 (hPa)	최저 해면기압 시 각(hhmi)	평균 해면기압 (hPa)
가조시간(hr)	합계 일조 시간 (hr)	1시간 최다일사 시각(hhmi)	1시간 최다일사량 (MJ/m2)	합계 일사(MJ/m 2)	일 최심신적설 (cm)	일 최심신적설 시 각(hhmi)	일 최심적설(cm)	일 최심적설 시각 (hhmi)	합계 3시간 신적 설(cm)
평균 전운량 (1/10)	평균 중하층운량 (1/10)	평균 지면온도(°C)	최저 초상온도(°C)	평균 5cm 지중온 도(°C)	평균 10cm 지중 온도(°C)	평균 20cm 지중 온도(°C)	평균 30cm 지중 온도(°C)	0.5m 지중온도(° C)	1,0m 지중온도(° C)

1.5m 지중온도(°

3,0m 지중온도(° C) 5,0m 지중온도(° C)

기상데이터

평균기온(°C)

최저기온(°C)

평균기온이 영하로 지속된 날 수 최저기온이 영하로 지속된 날 수

기준일 전날의 평균기온 기준일 전날의 최저기온 기준일 전날의 최저기온과 기준일 최저기온의 차이

최종 분석용 데이터의 독립변수

기상정보

평균기온(°C)

최저기온(°C)

평균기온이 영하로 지속된 날 수 최저기온이 영하로 지속된 날 수

기준일 전날의 평균기온 기준일 전날의 최저기온 기준일 전날의 최저기온과 기준일 최저기온의 차이

수도미터 정보

급수상태

업종

구경

용도



기계학습 알고리즘

01

로지스틱 회귀모델 (Logistic Regression)

종속 변수와 독립 변수간의 관계를 구체적인 함수로 나타내어 향후 예측 모델에 사용하는 것으로, 데이터의 결과가 특정 분류로 나뉘기 때문에 일종의 분류 (classification) 기법으로도 볼 수 있다.

02

K-최근접 이웃 알고리즘 (K-Nearest Neighbors: KNN)

가장 간단한 분류 알고리즘 중의 하나로써 새로운 데이터가 들어왔을 때 학습 데이터셋 중 그 데이터와 가장 가까운 K개의 데이터의 클래스를 보고 가장 많은 클래스로 새 데이터의 클래스를 예측하는 알고리즘이다.

03

서포트 백터 머신 (Support Vector Machine: SVM)

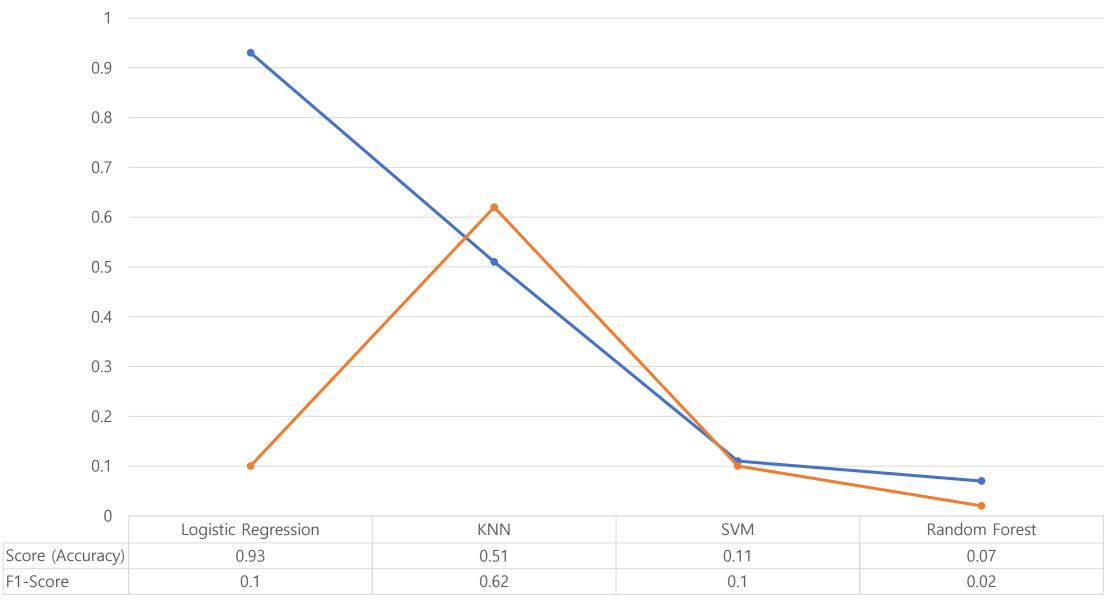
두 카테고리 중 어느 하나에 속한 데이터의 집합이 주어졌을 때, SVM 알고리즘은 주어진 데이터 집합을 바탕으로 하여 새로운 데이터가 어느 카테고리에 속할지 판단하는 비확률적 이진 선형 분류 모델을 만든다.

04

랜덤 포레스트 (Random Forest)

분류, 회귀 분석 등에 사용되는 앙상블 학습 방법의 일종으로, 훈련 과정에서 구성한 다수의 결정 트리로부터 부류(분류) 또는 평균 예측치(회귀 분석)를 출력함으로써 동작한다.

예측 모델별 정확도



→ Score (Accuracy) → F1-Score

F1-Score

Best Model

K-최근접 이웃 알고리즘 (K-Nearest Neighbors: KNN)

수집

분석/예측

동파예보

기상 데이터 KNN 동파 확률

재난 문자 SNS

