

| Term Project

| 산림화재 피해면적 예측 및
대응 단계별 이상적 방안 제시

1 팀

허건호	18102023
김도영	19101975
김시형	19101980
박지원	20100593

목 차

1
Proposal

2
프로젝트 진행 배경/목적/방법

3
Data Source

4
PreProcessing

5
Visualization

6
ReStructuring

7
Modeling & Result

8
Conclusion

9
Limitations & Improvement Measures

Term Project

Proposal

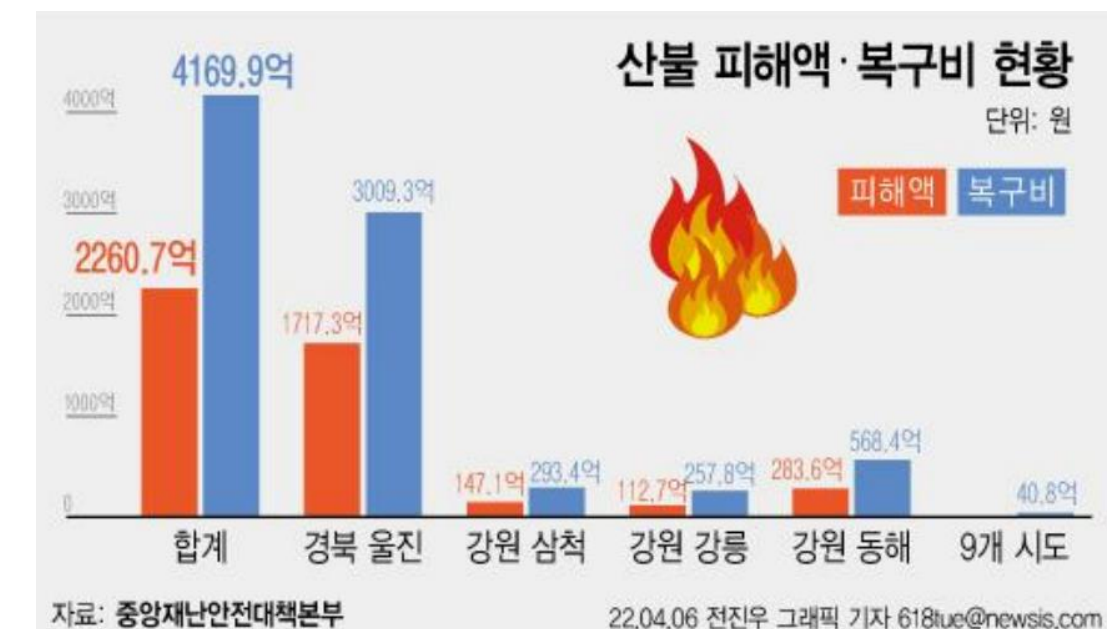
■ 예방적 차원

- 우리나라는 국토의 대부분이 산림으로 이루어져 산림 화재의 인명, 재산, 생태적 피해의 규모가 큼
- 부주의로 인한 화재가 3만 297건으로 전체 화재의 절반 이상(55.6%)을 차지
- 따라서 산림 화재 발생 위험을 줄이는 방향으로 사전에 예방해야 함



■ 대응적 차원

- 산림 화재는 기온, 습도, 풍향, 풍속 등이 연소 면적 및 피해 규모를 결정
- 따라서 산림 화재 상황 발생시, 적절한 대응을 통해 피해를 최소화 시켜야 함



Proposal

■ 목적

- 산림 화재 발생 가능성 및 피해 규모를 예측하는 모델을 생성
-> 예방 및 대응 차원에서 이점
- 발생률 및 피해 규모와 상관계수가 높은 위험요인 특정
-> 비용 및 시간 차원에서 새로운 관리 매뉴얼 제시
- 시간, 강수량과 같은 정량적 요인과 더불어
사회적 분위기, 등산객 행동 패턴 변화와 같은 정성적 요인
분석을 통한 다요인적 인과관계 확인
-> 현실적인 예방책 제시 가능

■ Formulation

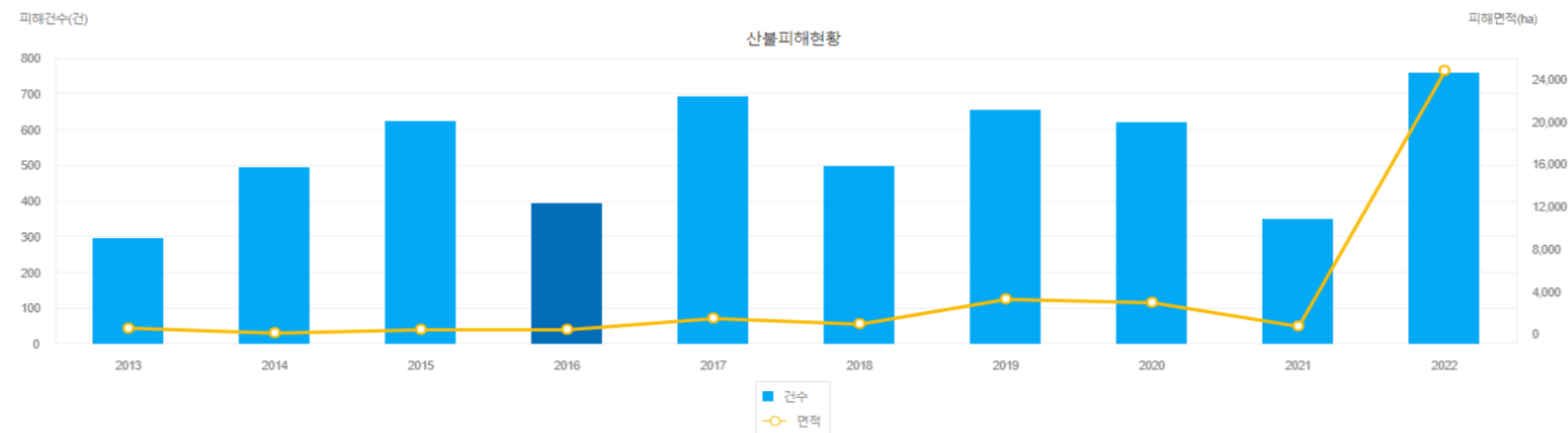
- Clustering을 활용하여 비슷한 추이를 가진 **산불 유형을 분류**
- SVM, Linear model, RF 등을 이용해서 **산불 피해 면적, 규모 등을 예측하는 모델** 생성

■ 피드백

- Formulation이 불명확
- 산불이 발생한 상황에 대한 데이터와
산불이 발생하지 않은 상황에 대한 데이터를 비교해야 함
-> 주제 축소 및 구체화

프로젝트 진행 배경 및 목적

- 국내 산림화재의 발생 및 피해가 증가하는 추세
- 우리나라는 국토의 대부분이 산림으로 이루어져 산림 화재의 인명, 재산, 생태적 피해의 규모가 큼
- 따라서 산림 화재 발생시, 피해 규모를 예측하고 그에 따라 적절한 대응을 통해 사회적, 경제적, 생태학적 측면에서 피해를 최소화해야 함



출처: 산림청 「산불통계연보」

- 산림 화재의 피해 면적은 발생 시기, 발생 원인, 진화 소요 시간, 기후 등 다양한 요인에 영향을 받아 결정될 것이라고 가정
- 본 프로젝트는 위의 가정을 전제로, 다양한 feature를 활용하여 산림 화재의 피해 면적을 예측하고자 함
- 또한 각 대응 단계(시기)별로 활용할 수 있는 feature의 종류가 다르다는 점을 집중하여, 해당 단계에서 피해 면적을 최소화할 수 있는 구체적 대응 방안을 제시하고자 함

프로젝트 진행 방법

- 산림 화재의 발생부터
완전까지의 **대응 과정을
3단계로 분류**
- 각 단계가 진행됨에 따라
활용할 수 있는 feature를
활용해, **산림 화재 피해
면적 예측 모델을 개발**

산불 발생 단계

- **산불이 발생하여
신고가 접수된 시점**
- 화재발생 월, 주말여부,
온도, 습도, 발화원인 등

산불 진화준비 단계

- **산불이 발생하여 신고가
접수되어, 화재 진압
기관이 화재 현장으로
출동하는 시점**
- 화재발생 월, 주말여부,
온도, 습도, 발화원인,
**소방서명, 서센터명,
출동소요시간** 등

산불 진화 단계

- **산불이 발생하여 신고가
접수되고, 화재 진압
기관이 화재 현장으로
출동하여, 진압이 완료된
시점**
- 화재발생 월, 주말여부,
온도, 습도, 발화원인,
소방서명, 서센터명,
출동소요시간,
**전체인력수합계,
화재진압시간,
동원장비수** 등

Term Project

Data Source

소방안전 빅데이터 플랫폼

- 강원도 산불 현황 데이터 (2011 ~ 2021)

- 각 인스턴스는 산림화재 1건
- 일련번호, 조사서번호, 사망자수, 부상자수, 특보내용 등 총 70여개의 feature로 구성

화재발생일자	화재발생시간	시도명	시군구명	읍면동명	읍면동구분명	번지	도로명	발화지점	날씨	온도	습도	시간단위풍속	시간단위풍향	특보내용	발화원인	발화요인대분류명	발화요인소분류명	접수일자	접수시각	소방서명
20110122	233500	강원도	강릉시	죽헌동		108-41		산정상	맑음	-1	30	45	8		기타	방화의심	라이터불/성냥불	20110122	233511	강릉소방서
20110131	181300	강원도	양양군	현남면	상월천리	289		산아래	맑음	1	10	225	8	건조경보	기타	부주의	모닥불/연탄/숯	20110131	181311	양양소방서
20110201	210700	강원도	강릉시	대전동		147-3		산정상	맑음	1	550	135	4		기타	방화의심	기타	20110201	210710	강릉소방서
20110211	175000	강원도	홍천군	내촌면	물걸리	5		산중턱	구름맑음	-2	29	225	4		입산자실화	기타	기타	20110211	175032	홍천소방서
20110212	61800	강원도	원주시	귀래면	귀래리	1894-3		산중턱	구름맑음	-6	44	225	4		모닥불	부주의	모닥불/연탄/숯	20110212	61836	원주소방서
20110226	164200	강원도	원주시	태장동		363-33		산정상	맑음	14	23	180	4		담뱃불	부주의	담뱃불	20110226	164234	원주소방서
20110309	131800	강원도	영월군	영월읍	정양리			산아래	맑음	1	19	225	8		기타	부주의	불꽃/불티	20110309	131855	영월소방서
20110311	134200	강원도	횡성군	우천면	정금리	21		산정상	맑음	5	43	225	4		입산자실화	부주의	미상	20110311	134249	횡성소방서
20110314	193500	강원도	영월군	영월읍	정양리	80		산아래	맑음	4	67	45	4		미상	미상	미상	20110314	193519	영월소방서
20110318	115800	강원도	홍천군	홍천읍	연봉리			산정상	맑음	9	21	315	4		입산자실화	기타	기타	20110318	115807	홍천소방서

서센터명	소방지역대명	출동일자	출동시각	현장도착일자	현장도착시각	출동소요시간	현장소방서거리	현장안전센터거리	현장소방지역대거리	조진일자	조진시각	전체인력수합계	완진일자	완진시간	화재진압시간	동원장비수	인명피해수	재산피해금액
경포119안전센터		20110122	233615	20110122	235315	1700	9	9	0	20110123	121115	37	20110123	1115	1800	8	0	0
하조대119안전센터		20110131	181409	20110131	183024	1615	43	13	0	20110201	94538	1872	20110201	103016	155952	71	0	184330
경포119안전센터		20110201	210826	20110201	211500	634	7	6	0	20110201	102300	184	20110201	223800	12300	15	0	0
서석119안전센터	내촌119지역대	20110211	175220	20110211	181500	2240	37	37	7	20110211	74100	171	20110211	202000	20500	9	0	4864
홍업119안전센터	귀래119지역대	20110212	62010	20110212	64029	2019	20	20	8	20110212	82038	94	20110212	103038	35009	9	0	50
태장119안전센터		20110226	164353	20110226	164901	508	8	2	0	20110226	52601	46	20110226	172726	3825	8	0	440
영월119안전센터		20110309	131916	20110309	132916	1000	9	9	8	20110309	50016	204	20110309	200016	63100	26	0	600
우천119안전센터		20110311	134334	20110311	140409	2035	17	10	0	20110311	21746	214	20110311	151737	11328	10	0	0
영월119안전센터		20110314	193540	20110314	195222	1642	9	9	7	20110314	90022	116	20110314	230622	31400	5	0	64
홍천119안전센터		20110318	115900	20110318	120327	427	6	6	0	20110318	123300	88	20110318	124800	4433	15	0	3264

Term Project

Data Source

- 공공데이터포털
- 산림청_산불발생통계
 - 각 인스턴스는 산림화재 1건
 - 피해면적합계 등 총 16개의 피처로 구성

	피해면적 합계	진화종 료 일	진화종 료 월	진화종료 일 시	진화종료 연도	발생원 인	발생장소_ 지번	발생장소_ 동리	발생장소_ 시군구	발생장소_ 읍면	발생장소_ 시도	발생 일	발생 요일	발생 월	발생시 간	발생연 도
0	0.1	28	12	08:30:00	2011	입산자 실화	산202-2	장전평	홍천	홍천	강원	28	수요일	12	03:20:00	2011
1	0.5	24	12	16:30:00	2011	쓰레기 소각	산14-1	무계	청송	현서	경북	24	토요일	12	14:30:00	2011
2	0.01	22	12	18:20:00	2011	입산자 실화	산52	원	산청	시천	경남	22	목요일	12	17:15:00	2011
3	0.2	09	12	15:00:00	2011	기타	산27	용수	기장	정관	부산	09	금요일	12	14:11:00	2011
4	0.3	29	11	22:30:00	2011	담뱃불 실화	산4-1	거포	옥천	청성	충북	29	화요일	11	20:10:00	2011
...
10167	0.4	01	04	18:50:00	2021	쓰레기 소각	산102	쌍지	안성	고삼	경기	01	목요일	04	16:01:00	2021
10168	0.01	01	04	13:30:00	2021	입산자 실화	산251	양잠	태안	남	충남	01	목요일	04	12:14:00	2021
10169	0.1	01	04	15:30:00	2021	담뱃불 실화	산33-3	어천	공주	우성	충남	01	목요일	04	12:10:00	2021
10170	10.7	01	04	12:54:00	2021	기타	산146외1 필지	운치	정선	신동	강원	31	수요일	03	15:38:00	2021
10171	0.01	01	04	15:30:00	2021	기타	663-9	화전	고양	덕양	경기	01	목요일	04	15:03:00	2021

Term Project

PreProcessing

Raw dataset 제작

- 강원도 산불 현황 데이터는 각 산림 화재에 대한 독립변수를 포함
- 산림청_산불발생통계는 각 산림화재에 대한 종속변수(산불 피해 면적)을 포함
- 두 원천 데이터를 종합하여, EDA(시각화 및 통계량 확인)에 사용할 dataset 확보

현장도착일자	현장도착시각	출동소요시간	현장소방서거리	현장안전센터거리	현장소방지역대거리	조진일자	조진시각	전체인력수합계	완진일자	완진시간	화재진압시간	동원장비수	인명피해수	재산피해금액	피해면적
20110122	235315	1700	9	9	0	20110123	121115	37	20110123	1115	1800	8	0	0	0.3
20110131	183024	1615	43	13	0	20110201	94538	1872	20110201	103016	155952	71	0	184330	30
20110201	211500	634	7	6	0	20110201	102300	184	20110201	223800	12300	15	0	0	0.2
20110211	181500	2240	37	37	7	20110211	74100	171	20110211	202000	20500	9	0	4864	0.3
20110212	64029	2019	20	20	8	20110212	82038	94	20110212	103038	35009	9	0	50	0.3
20110226	164901	508	8	2	0	20110226	52601	46	20110226	172726	3825	8	0	440	0.03
20110309	132916	1000	9	9	8	20110309	50016	204	20110309	200016	63100	26	0	600	2
20110311	140409	2035	17	10	0	20110311	21746	214	20110311	151737	11328	10	0	0	0.1
20110314	195222	1642	9	9	7	20110314	90022	116	20110314	230622	31400	5	0	64	0.1
20110318	120327	427	6	6	0	20110318	123300	88	20110318	124800	4433	15	0	3264	0.2
20110324	130321	600	46	6	0	20110324	15021	102	20110324	135021	4700	8	0	0	0.7
20110401	110300	1000	8	8	12	20110401	122020	160	20110401	133000	22700	19	0	0	0.3
20110405	112736	2200	40	25	0	20110405	120036	67	20110405	120036	3300	3	0	1485	0.1
20110406	144557	300	13	2	0	20110406	24857	294	20110406	155057	10500	11	0	0	0.2
20110406	164400	2000	18	17	0	20110406	45500	160	20110406	170000	1600	20	0	0	0.1
20110409	124438	1300	50	15	0	20110409	10128	68	20110409	130401	1923	5	0	0	0.04

PreProcessing

결측치 및 이상치 처리

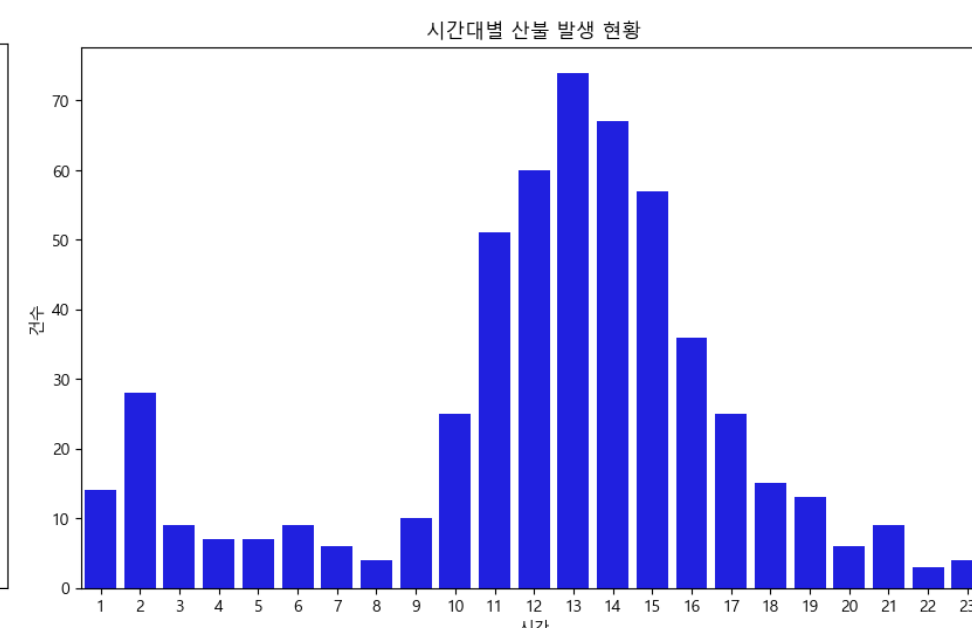
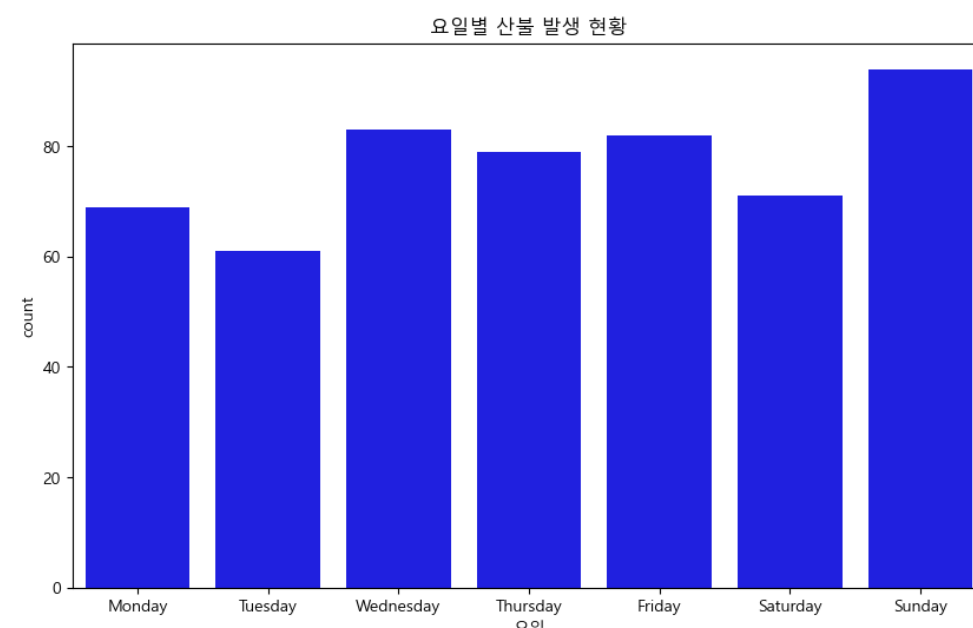
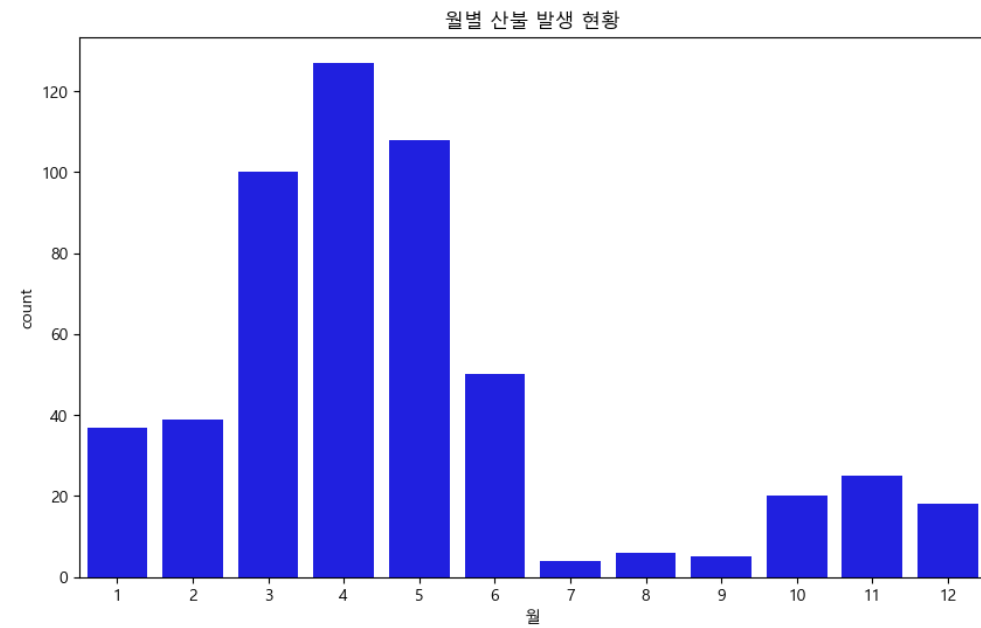
- Y값 (피해면적) == NULL인 인스턴스 제거
- 온도, 습도에서 결측치 또는 이상치를 포함하는 인스턴스 제거
- 풍향 및 풍속 feature 제거 (이상치)

	화재발생일자	화재발생시간	시도명	시군구명	읍면동명	읍면동구분명	번지	도로명	발화지점	날씨	온도	습도	특보내용	발화원인	발화요인대분류명	발화요인소분류명	접수일자	접수시각	소방서명	서센터명
0	20110122	233500	강원도	강릉시	죽헌동		108-41		산정상	맑음	-1	30		기타	방화의심	라이터불/성냥불	20110122	233511	강릉소방서	경포119안전센터
1	20110131	181300	강원도	양양군	현남면	상월천리	289		산아래	맑음	1	10	건조정보	기타	부주의	모닥불/연탄/숯	20110131	181311	양양소방서	하조대119안전센터
2	20110211	175000	강원도	홍천군	내촌면	물걸리	5		산중턱	구름맑음	-2	29		입산자실화	기타	기타	20110211	175032	홍천소방서	서석119안전센터
3	20110212	61800	강원도	원주시	귀래면	귀래리	1894-3		산중턱	구름맑음	-6	44		모닥불	부주의	모닥불/연탄/숯	20110212	61836	원주소방서	홍업119안전센터
4	20110226	164200	강원도	원주시	태장동		363-33		산정상	맑음	14	23		담뱃불	부주의	담뱃불	20110226	164234	원주소방서	태장119안전센터
5	20110309	131800	강원도	영월군	영월읍	정양리			산아래	맑음	1	19		기타	부주의	불꽃/불티	20110309	131855	영월소방서	영월119안전센터
6	20110311	134200	강원도	횡성군	우천면	정금리	21		산정상	맑음	5	43		입산자실화	부주의	미상	20110311	134249	횡성소방서	우천119안전센터
7	20110314	193500	강원도	영월군	영월읍	정양리	80		산아래	맑음	4	67		미상	미상	미상	20110314	193519	영월소방서	영월119안전센터
8	20110318	115800	강원도	홍천군	홍천읍	연봉리			산정상	맑음	9	21		입산자실화	기타	기타	20110318	115807	홍천소방서	홍천119안전센터
9	20110324	125600	강원도	화천군	화천읍	풍산리	산240		산아래	맑음	4	47		논/밭두렁	부주의	쓰레기/논밭두렁	20110324	125625	춘천소방서	화천119안전센터

소방지역대명	출동일자	출동시각	현장도착일자	현장도착시각	출동소요시간	현장소방서거리	현장안전센터거리	현장소방지역대거리	초진일자	초진시각	전체인력수합계	완진일자	완진시간	화재진압시간	동원장비수	인명피해수	재산피해금액	피해면적
	20110122	233615	20110122	235315	1700	9	9	0	20110123	121115	37	20110123	1115	1800	8	0	0	0.3
	20110131	181409	20110131	183024	1615	43	13	0	20110201	94538	1872	20110201	103016	155952	71	0	184330	30
내촌119지역대	20110211	175220	20110211	181500	2240	37	37	7	20110211	74100	171	20110211	202000	20500	9	0	4864	0.3
귀래119지역대	20110212	62010	20110212	64029	2019	20	20	8	20110212	82038	94	20110212	103038	35009	9	0	50	0.3
	20110226	164353	20110226	164901	508	8	2	0	20110226	52601	46	20110226	172726	3825	8	0	440	0.03
	20110309	131916	20110309	132916	1000	9	9	8	20110309	50016	204	20110309	200016	63100	26	0	600	2
	20110311	134334	20110311	140409	2035	17	10	0	20110311	21746	214	20110311	151737	11328	10	0	0	0.1
	20110314	193540	20110314	195222	1642	9	9	7	20110314	90022	116	20110314	230622	31400	5	0	64	0.1
	20110318	115900	20110318	120327	427	6	6	0	20110318	123300	88	20110318	124800	4433	15	0	3264	0.2
	20110324	125721	20110324	130321	600	46	6	0	20110324	15021	102	20110324	135021	4700	8	0	0	0.7

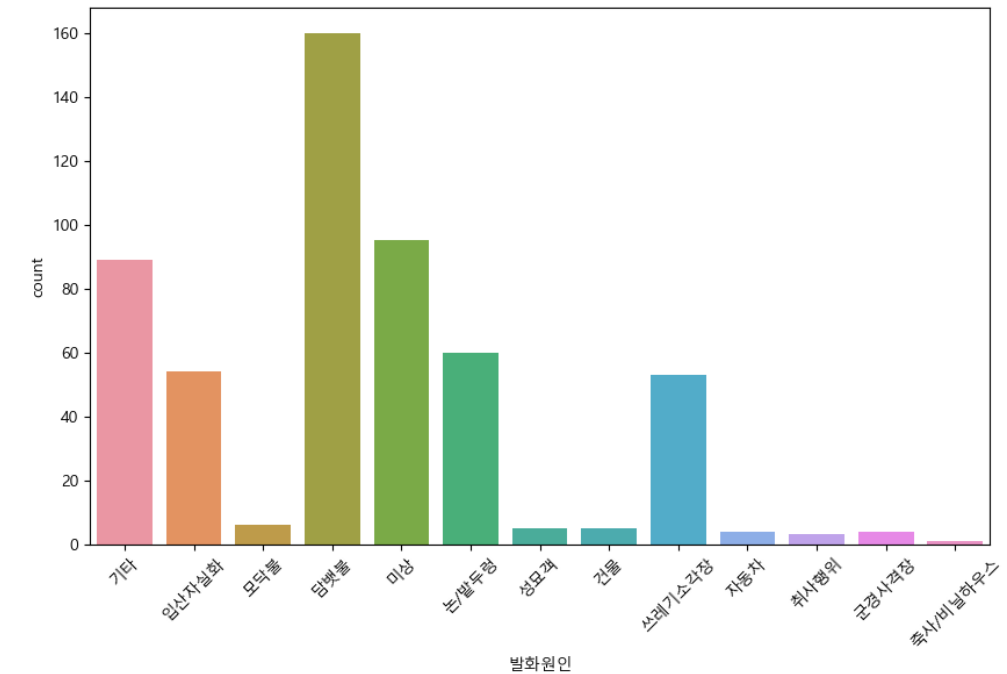
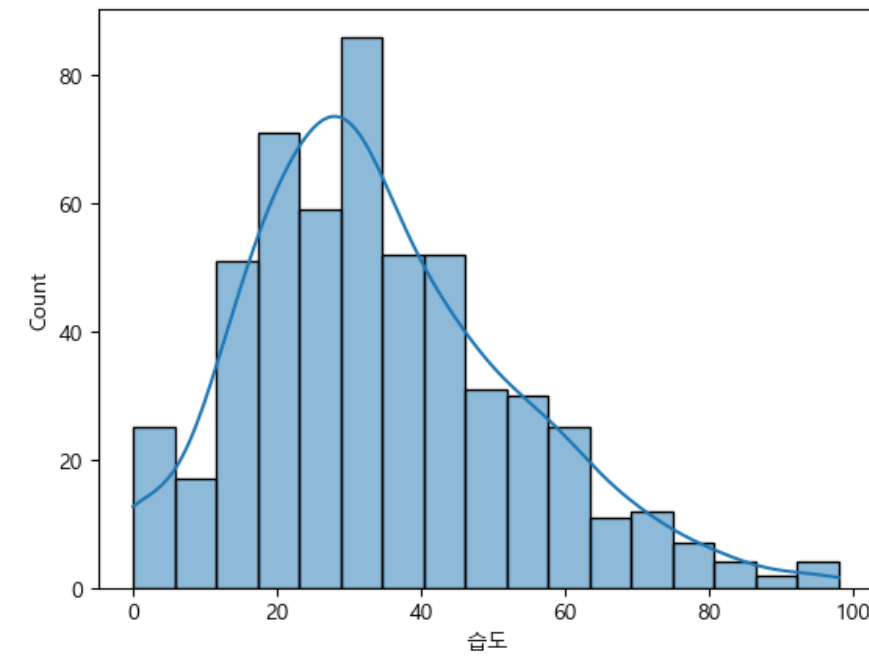
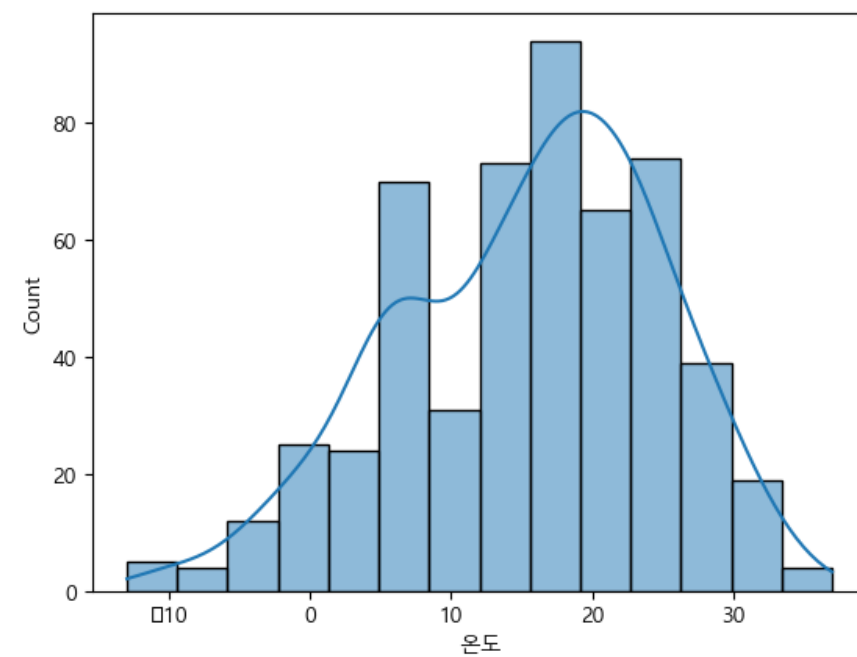
Term Project

Visualization



- 봄철, 일요일, 낮시간에 산불 많이 발생
- 온도 20 °C, 습도 30% 내외일 때 많이 발생
- 담뱃불이 압도적으로 많은 비율 차지

- 시각화를 통해, 각 피처의 구성과 분포를 파악, 이후 데이터셋 재구조화에 해당 정보 활용
- 발화 원인 분포 확인, 온습도와 같은 수치형 데이터의 분포 확인



Term Project

ReStructuring

- 화재발생일자 바탕으로 주말 여부 컬럼 생성
- 화재발생시간 카테고리화
- 지역범위 수정
- 데이터 분석 및 학습에 사용하지 않을 feature 제거

화재발생일시	화재발생_연	화재발생_월	화재발생_일	주말여부	화재발생시간대	시군구명	읍면동명	발화지점	날씨	온도	습도	특보내용	발화원인종합	소방서명
2011-01-22 23:35:00	2011	1	22	1	21~24	강릉시	죽현동	산정상	맑음	-1	30		방화 및 방화의심	강릉소방서
2011-01-31 18:13:00	2011	1	31	0	18~21	양양군	현남면	산아래	맑음	1	10	건조경보	인간의 부주의	양양소방서
2011-02-11 17:50:00	2011	2	11	0	15~18	홍천군	내촌면	산중턱	구름많음	-2	29		인간의 부주의	홍천소방서
2011-02-12 06:18:00	2011	2	12	1	3~6	원주시	귀래면	산중턱	구름많음	-6	44		인간의 부주의	원주소방서
2011-02-26 16:42:00	2011	2	26	1	15~18	원주시	태장동	산정상	맑음	14	23		인간의 부주의	원주소방서
2011-03-09 13:18:00	2011	3	9	0	12~15	영월군	영월읍	산아래	맑음	1	19		인간의 부주의	영월소방서
2011-03-11 13:42:00	2011	3	11	0	12~15	횡성군	우천면	산정상	맑음	5	43		인간의 부주의	횡성소방서

서센터명	소방지역대명	출동소요시간	현장소방서거리	현장안전센터거리	현장소방지역대거리	전체인력수합계	화재진압시간	동원장비수	인명피해수	재산피해금액	피해면적
경포119안전센터		1700	9	9	0.0	37	1800	8	0	0	0.3
하조대119안전센터		1615	43	13	0.0	1872	155952	71	0	184330	30.0
서석119안전센터	내촌119지역대	2240	37	37	7.0	171	20500	9	0	4864	0.3
홍업119안전센터	귀래119지역대	2019	20	20	8.0	94	35009	9	0	50	0.3
태장119안전센터		508	8	2	0.0	46	3825	8	0	440	0.03
영월119안전센터		1000	9	9	8.0	204	63100	26	0	600	2.0
우천119안전센터		2035	17	10	0.0	214	11328	10	0	0	0.1

Modeling & Result

Model_01 at 산불 발생 단계

- Used features
 - 화재발생_월, 주말여부, 화재발생시간대, 시군구명, 발화지점, 날씨, 온도, 습도, 발화원인 종합

```
data = df.drop(['화재발생일시', '화재발생_연', '화재발생_일', '읍면동명', '특보내용', '소방서명', '서센터명', '소방지역대명', '출동소요시간'])
```

```
#분석에 필요한 변수만 남김
data
```

	화재발생_월	주말여부	화재발생시간대	시군구명	발화지점	날씨	온도	습도	발화원인종합	피해면적
0	1	1	21~24	강릉시	산정상	맑음	-1	30	방화 및 방화의심	0.30
1	1	0	18~21	양양군	산아래	맑음	1	10	인간의 부주의	30.00
2	2	0	15~18	홍천군	산중턱	구름많음	-2	29	인간의 부주의	0.30
3	2	1	3~6	원주시	산중턱	구름많음	-6	44	인간의 부주의	0.30
4	2	1	15~18	원주시	산정상	맑음	14	23	인간의 부주의	0.03
...
534	5	1	12~15	철원군	산중턱	맑음	19	26	인간의 부주의	0.03
535	5	1	15~18	철원군	산정상	맑음	18	22	인간의 부주의	0.01
536	5	0	15~18	태백시	산중턱	맑음	23	43	기계/전기/화학 요인	1.20
537	6	0	15~18	강릉시	산중턱	맑음	25	62	인간의 부주의	0.01
538	6	0	12~15	홍천군	산정상	맑음	18	78	자연적 요인	0.04

539 rows × 10 columns

```
# 문자열 데이터를 숫자형 데이터로 변환
```

```
features = pd.get_dummies(data = features, columns = ['화재발생시간대'], prefix = '화재발생시간대')
```

```
features = pd.get_dummies(data = features, columns = ['시군구명'], prefix = '시군구명')
```

```
features = pd.get_dummies(data = features, columns = ['발화지점'], prefix = '발화지점')
```

```
features = pd.get_dummies(data = features, columns = ['날씨'], prefix = '날씨')
```

```
features = pd.get_dummies(data = features, columns = ['발화원인종합'], prefix = '발화원인종합')
```

	화재발생_월	주말여부	온도	습도	화재발생시간대_0~3	화재발생시간대_12~15	화재발생시간대_15~18	화재발생시간대_18~21	화재발생시간대_21~24	화재발생시간대_3~6	...	날씨_구름조금	날씨_맑음	날씨_비	날씨_소나기	날씨_흐림	발화원인종합_기계/전기/화학요인	발화원인종합_미상	발화원인종합_방화및방화의심	발화원인종합_인간의부주의	발화원인종합_자연적요인
0	1	1	-1	30	0	0	0	0	1	0	...	0	1	0	0	0	0	0	1	0	0
1	1	0	1	10	0	0	0	1	0	0	...	0	1	0	0	0	0	0	0	1	0
2	2	0	-2	29	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	1	0
3	2	1	-6	44	0	0	0	0	0	1	...	0	0	0	0	0	0	0	0	1	0
4	2	1	14	23	0	0	1	0	0	0	...	0	1	0	0	0	0	0	0	1	0
...
534	5	1	19	26	0	1	0	0	0	0	...	0	1	0	0	0	0	0	0	1	0

Term Project

Modeling & Result

Model_01 at 산불 발생 단계

- **Data split:** test_size = 0.2
- **Used Model:** DecisionTreeRegressor
- **Result**
 - RMSE(오차의 크기) on test set: **26.51**
 - R^2 (모델의 설명력) on test set: **-0.04**

```
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error

model = DecisionTreeRegressor(random_state=22)

model.fit(X_train, y_train)
```

```
DecisionTreeRegressor(random_state=22)
```

```
y_train_hat = model.predict(X_train)
y_test_hat = model.predict(X_test)
```

```
# 모델 성능 평가 결과 training set에 대해서는 성능이 우수하나 test set에 대해서는 성능이 떨어지는 overfitting의 양상을 보임
from sklearn.metrics import mean_squared_error
import numpy as np

rmse_train = np.sqrt(mean_squared_error(y_train, y_train_hat))
rmse_test = np.sqrt(mean_squared_error(y_test, y_test_hat))

print("RMSE on training set:", rmse_train)
print("RMSE on test set:", rmse_test)
```

RMSE on training set: 2.855703028628603e-18

RMSE on test set: 26.519992441063664

```
from sklearn.metrics import r2_score
r2 = r2_score(y_test, y_test_hat)
print("R2 Score:", r2)
```

R^2 Score: -0.042342255942229734

Modeling & Result

Model_01 at 산불 발생 단계

```
#Data split: Train, test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
    features, target, test_size=0.2, random_state = 22)
```

```
print("X_train shape:", X_train.shape)
print("y_train shape:", y_train.shape)
```

```
X_train shape: (431, 46)
y_train shape: (431,)
```

```
print("X_test shape:", X_test.shape)
print("y_test shape:", y_test.shape)
```

```
X_test shape: (108, 46)
y_test shape: (108,)
```

- **Data split:** test_size = 0.2
- **Used Model:** DecisionTreeRegressor + GridSearchCV
- **Result**
 - RMSE(오차의 크기) on test set: **36.58**
 - R²(모델의 설명력) on test set: **-0.98**

```
#GridSearchCV 를 통한 적절한 hyperparameter 찾기
from sklearn.model_selection import GridSearchCV
```

```
# 모델 생성
model = DecisionTreeRegressor(random_state=22)
```

```
# GridSearchCV 수행을 위한 파라미터 후보군 정의
param_grid = {
    'max_depth': [2, 3, 4, 5, 10, 20],
    'min_samples_split': [2, 3, 4, 5, 12, 13, 14, 15],
    'min_samples_leaf': [1, 2, 3, 4, 5],
}
```

```
# GridSearchCV 수행
grid_search = GridSearchCV(model, param_grid=param_grid, cv=5, scoring='neg_mean_squared_error')
grid_search.fit(X_train, y_train)
```

```
# 최적의 하이퍼파라미터와 점수 출력
print("Best hyperparameters: ", grid_search.best_params_)
print("Best score: ", -grid_search.best_score_)
```

```
Best hyperparameters: {'max_depth': 2, 'min_samples_leaf': 5, 'min_samples_split': 2}
Best score: 4647.860131072654
```

```
model = DecisionTreeRegressor(max_depth=2, min_samples_leaf=5, min_samples_split=2, random_state=22)
model.fit(X_train, y_train)
```

```
DecisionTreeRegressor(max_depth=2, min_samples_leaf=5, random_state=22)
```

Term Project

Modeling & Result

Model_02 at 산불 진화준비 단계

Used Model

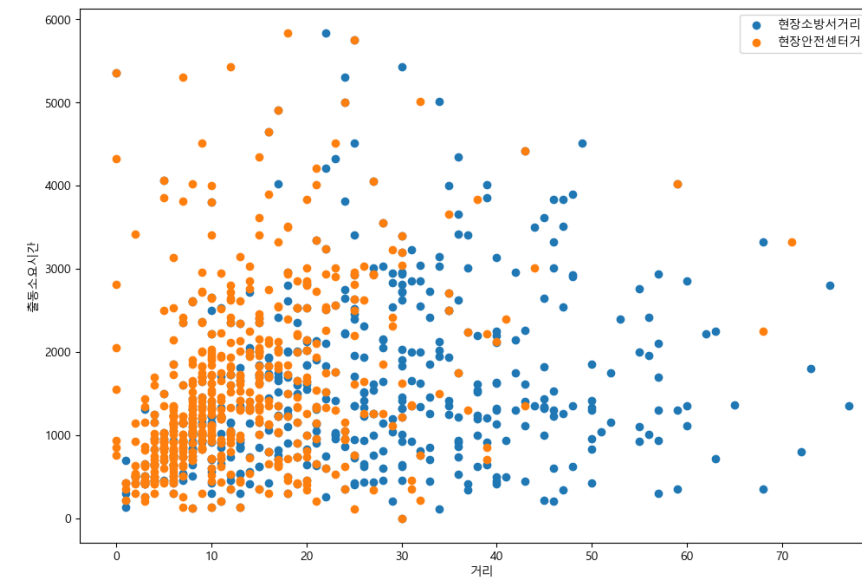
- DecisionTreeRegressor

Used features

- 화재발생_월, 주말여부, 화재발생시간대, 시군구명, 발화지점, 날씨, 온도, 습도, 발화원인 종합
- 소방서명, 서센터명, 출동소요시간, 현장소방서거리, 현장안전센터거리
- 유동적으로 feature 조정

Result

- RMSE(오차의 크기) on test set: **27.34**
- R²(모델의 설명력) on test set: **-0.11**



- 출동소요시간과 현장소방서거리, 현장안전센터거리 사이의 관계 파악

- 피어슨 상관계수

	현장소방서거리	현장안전센터거리	출동소요시간
현장소방서거리	1.000000	0.475144	0.245447
현장안전센터거리	0.475144	1.000000	0.340478
출동소요시간	0.245447	0.340478	1.000000

화재발생_월	주말여부	온도	습도	출동소요시간	현장소방서거리	현장안전센터거리	화재발생시간대_0~3	화재발생시간대_12~15	화재발생시간대_15~18	...	서센터명_하조대119안전센터	서센터명_학성119안전센터	서센터명_혁신119안전센터	서센터명_홍천119안전센터	서센터명_화천119안전센터	서센터명_화성119안전센터	서센터명_횡성119안전센터	서센터명_자평119안전센터	서센터명_후평119안전센터	서센터명_흥업119안전센터
0	1	1	-1	30	1700	9	9	0	0	0	...	0	0	0	0	0	0	0	0	0
1	1	0	1	10	1615	43	13	0	0	0	...	1	0	0	0	0	0	0	0	0
2	2	0	-2	29	2240	37	37	0	0	1	...	0	0	0	0	0	0	0	0	0
3	2	1	-6	44	2019	20	20	0	0	0	...	0	0	0	0	0	0	0	0	1
4	2	1	14	23	508	8	2	0	0	1	...	0	0	0	0	0	0	0	0	0

Term Project

Modeling & Result

Model_03 at 산불 진화 단계

- Used Model

- DecisionTreeRegressor

- Used features

- 한달내의 시기, 계절, 주말여부, 화재발생시간대, 시군구명, 발화지점, 날씨, 온도, 습도, 특보내용, 발화원인종합, 출동소요시간, 현장소방서거리, 현장안전센터거리, 전체인력수합계, 화재진압시간, 동원장비수, 피해면적
 - 출동소요시간, 현장소방서거리, 현장안전센터거리
 - 전체인력수합계, 화재진압시간, 동원장비수, 피해면적
 - 유동적으로 feature 조정

	한달내 의 시기	계 절	주말 여부	화재발생 시간대	시군 구명	발화 지점	날씨	온 도	습 도	특보내 용	발화원인종합	출동소요 시간	현장소방 서거리	현장안전 센터거리	전체인력 수합계	화재진압 시간	동원장 비수	피해면 적
0	월말	겨울	1	21~24	강릉 시	산정 상	맑음	-1	30	None	방화 및 방 화의심	1700	9	9	37	1800	8	0.30
1	월말	겨울	0	18~21	양양 군	산아 래	맑음	1	10	특보 건조	인간의 부 주의	1615	43	13	1872	155952	71	30.00
2	월중순	겨울	0	15~18	홍천 군	산중 턱	구름 많음	-2	29	None	인간의 부 주의	2240	37	37	171	20500	9	0.30
3	월중순	겨울	1	3~6	원주 시	산중 턱	구름 많음	-6	44	None	인간의 부 주의	2019	20	20	94	35009	9	0.30
4	월말	겨울	1	15~18	원주 시	산정 상	맑음	14	23	None	인간의 부 주의	508	8	2	46	3825	8	0.03

주말 여부	온 도	습 도	출 동 소 요 시 간	현 장 소 방 서 거 리	현 장 안 전 센 터 거 리	화재진 압시간	한달 내의 시기 _월말	한달 내의 시기 _월중순	한달 내의 시기 _월초	계절 _겨울	계절 _봄	계절 _여름	화재 발생 시간 대 _0~3	화재발 생시간 대 _12~15	화재발 생시간 대 _15~18	화재발 생시간 대 _18~21	화재발 생시간 대 _21~24	화재발 생시간 대 _3~6	화재발 생시간 대 _6~9	화재발 생시간 대 _9~12	시군 구명 _강릉 시	시군 구명 _고성 군	시군 구명 _동해 시	시군 구명 _삼척 시	시군 구명 _속초 시	시군 구명 _양구 군	시군 구명 _양양 군	시군 구명 _영월 군
----------	--------	--------	----------------------------	---------------------------------	--------------------------------------	------------	-----------------------	------------------------	-----------------------	-----------	----------	-----------	-----------------------------	---------------------------	---------------------------	---------------------------	---------------------------	-------------------------	-------------------------	--------------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

- Result

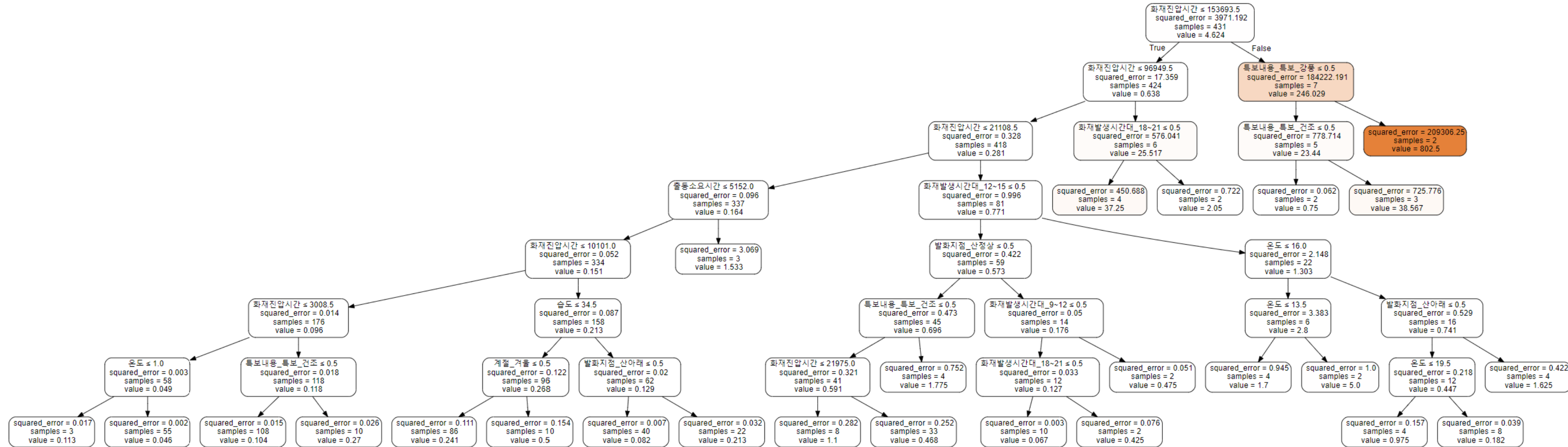
- RMSE(오차의 크기) on test set: **23.07**
 - R²(모델의 설명력) on test set: **0.21**

0	1	-1	30	1700	9	9	1800	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
1	0	1	10	1615	43	13	155952	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
2	0	-2	29	2240	37	37	20500	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
3	1	-6	44	2019	20	20	35009	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	1	14	23	508	8	2	3825	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Term Project

Modeling & Result

Model_03 at 산불 진화 단계



Term Project

Conclusion

진행 전 예상결과

- 산림 화재 피해면적은, 확보한 데이터셋에서 가공한 feature를 통해 설명 가능할 것
- 단계가 올라감에 따라 더 많은 feature를 활용할 수 있고, 이에 따라 모델의 설명력, 즉 성능이 향상될 것

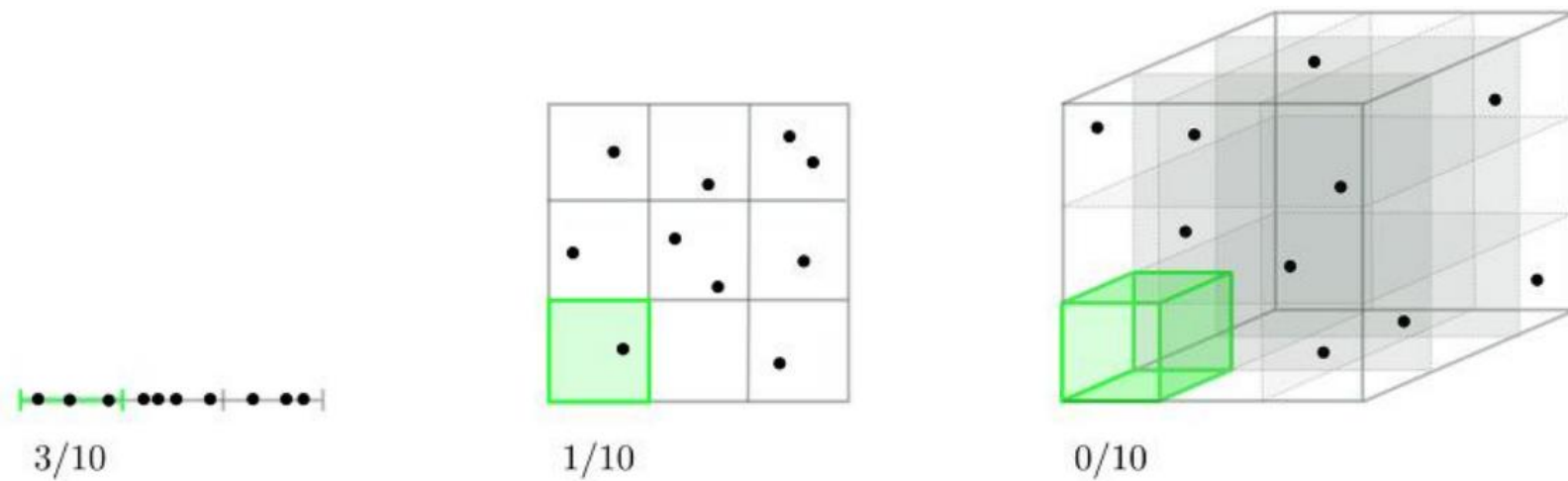
결론 및 해석

- Model_02의 성능이 Model_01의 성능보다 낮게 도출되었음
- 낮은 성능으로 인해, 초기에 기획했던 ‘다음 단계에서의 대응 방안을 제시’는 어려울 것
- 최종성능(R^2): 0.21 with Model_03

Limitations & Improvement Measures

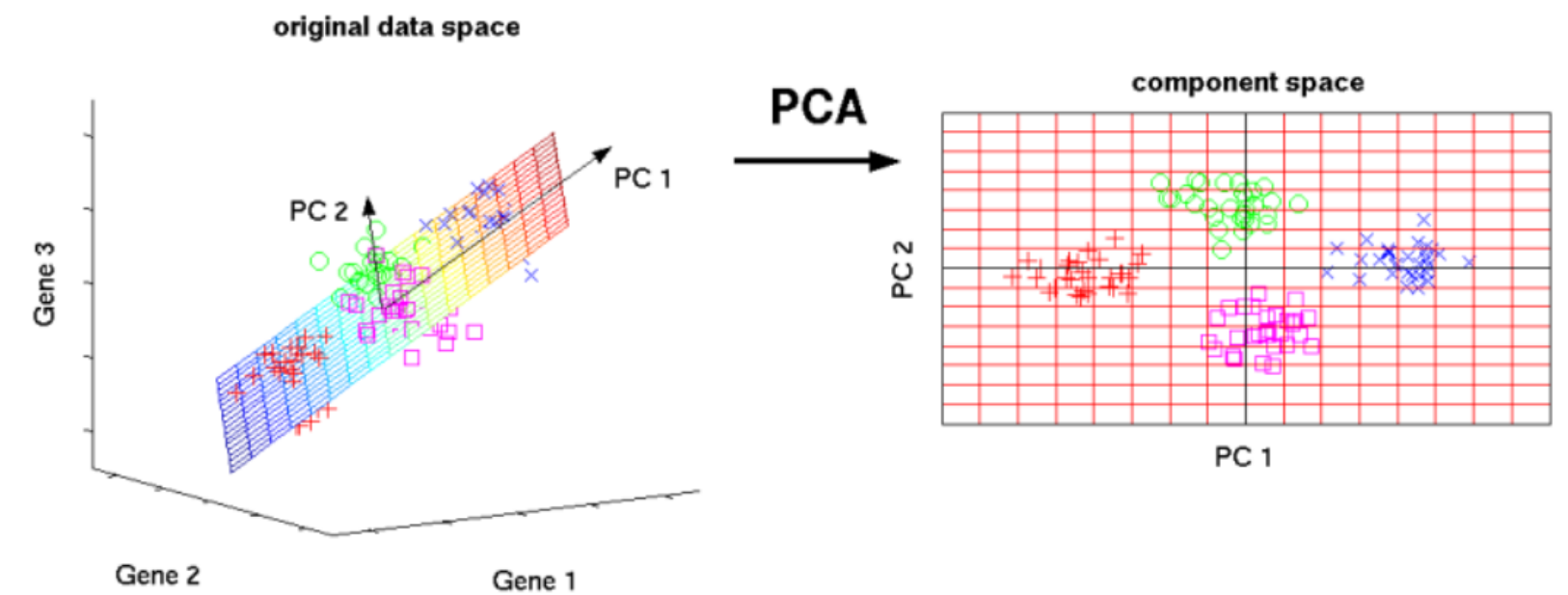
한계

- 산림 화재 피해면적은 확보한 데이터셋만으로 충분히 설명할 수 없음
- Feature가 늘어남에 따라, 필연적으로 **Curse of dimensionality** 발생



개선방안

- 산불의 확산속도와 같은, 더 직접적이고 유효한 feature 획득
- PCA 등 차원축소기법



Term Project

출처

- 소방안전 빅데이터 플랫폼

<https://www.bigdata-119.kr/>

- 공공데이터포털

<https://www.data.go.kr/data/3070842/openapi.do>

- e-나라지표 산불피해현황

https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=1309

- 전라남도 빅데이터허브 산불발생분석

https://data.jeonnam.go.kr/index.do?menuCd=DOM_000000108001004000

- Folium을 사용한 데이터 시각화

<https://blog.naver.com/PostView.nhn?blogId=kcchang61&logNo=221350672356>