

기상예측을 통한
지역 내 호우/태풍 리스크 분석 및 제언

정희철 김민서 이주형 심수현

목차

1 분석 배경

2 분석 흐름

3 데이터 분석

3.1 데이터 개요

3.1.1 호우내역데이터 개요

3.1.2 재난복구비 데이터

3.1.3 기상 데이터

3.1.4 관측소 데이터

3.2. PCA

3.3 기상변수 예측

3.4 이상치 탐지

3.5 호우/태풍 이진분류

3.6 피해액 예측

4 리스크 분석

4.1 원인-연도-지역-시설별 피해액 데이터 개요

4.2 지역별 리스크 분석

5 보험산업의 대응 방안

5.1 풍수해 보험의 개선

5.2 정부 대상 호우/태풍 보험

1 분석 배경

세계기상기구 WMO는 2201년 '전 지구 기후현황 보고서' (State of the Global Climate 2021) 에서 기후변화의 4가지 주요 지표인 온실가스 농도, 해수면상승, 해수 온도, 해양 산성도 모두 2021년에 새로운 기록을 세웠다고 밝혔다. 기후 위기가 아닌 기후 재앙이라는 말이 사용되는 등 전 지구적인 기후 변화 정도가 심각하다. UNDRR이 2020년 발표한 '재난에 의한 인적 피해비용' (Human cost of disasters: an overview of the last 20 years)에 따르면 2000년에서 2019년까지 20년 동안 기후변화로 인해 발생한 재해 수는 6,681건으로 그 이전 20년(1980-1999)동안 발생한 3,656건에 비해 약 2배가량 상승하였으며, 이로 인한 피해액은 약 2조 9700억 달러로 추정된다.

우리나라 역시 기후 변화로 인한 위험에 직면하고 있다. 올해 8월 서울, 경기 등의 집중호우로 인한 피해 복구비로 7905억원이 투입되었고¹, 태풍 힌남노로 인한 포항의 재산 피해는 1조 7천억원으로 조사되었다². 산사태나 침수위험 지역 주민 2141세대 2906명이 마을회관·학교·숙박시설·친인척집 등으로 일시 대피했으며, 전국에서 6만6341가구의 전기가 끊기는 등 피해가 심각했다³. 지구 온난화로 인한 기후 변화로 이상기온 현상과 잦은 집중호우가 발생하고 있으며, 미래에는 그로 인한 재난과 재해의 빈도와 그에 따른 피해액 증가가 예상된다. 네이처 계열의 국제 학술지 네이처 기후변화에 실린 기후 변화에 따른 21세기 말의 홍수 피해 예측에 따르면, 산업화 이전 대비 기온이 1.5도, 2도, 3도 상승할 때마다 홍수에 의한 사망자는 1.83배, 2.3배, 3.7배 증가하고, 홍수에 의한 경제적 피해는 3.4배, 6.2배, 11배 증가할 것으로 예측된다고 한다. 따라서 기상 데이터 분석을 통해 앞으로 더 자주 발생할 호우와 태풍에 의한 시도별 피해액 예측과 리스크 분석을 실시하고, 이를 바탕으로 보험업계의 대응 방안에 대해 제안하고자 한다.

2 분석 흐름

가장 먼저 각 기상 변수에 대해 worst case scenario, 즉 상한값 또는 하한값을 예측했다. 그렇게 예측된 기상상황을 이용해 호우 또는 태풍이 있는 날을 이상치로 설정해 태풍/호우 이상치 탐지 모델, 즉 기상 상황에 따른 호우 혹은 태풍의 유무를 예측하는 모델을 만들었다. 이 태풍/호우 이상치 탐지 모델에 앞서 예측한 미래의 기상 예측 데이터를 넣어 태풍/호우의 유무 예측값 추출 후, 태풍/호우 이상치 탐지 모델로 추출한 이상치에서 호우와 태풍을 구분하기 위해 이상치로 분류된 날짜에 한해 기상 데이터에 따른 호우와 태풍 이진 분류 모델을 만들고, 미래 기상 상황에 따른 호우 또는 태풍 여부를 예측하였다. 기술된 과정들을 통해 완성된 데이터는 호우나 태풍이 일어날 때 각각의 상황에 따른

¹ 행정안전부, "8월 서울경기 등 집중호우 피해 복구비 7905억원 투입", 대한민국 정책브리핑, 2022년 9월 8일, <https://www.korea.kr/news/policyNewsView.do?newsId=148905747>

² 이준형, "포항시, '힌남노'로 인명피해 10명, 재산피해 1조7000억원 추정", 경북신문 2022년 9월 7일 <http://www.kbsm.net/news/view.php?idx=363128>

³ 김민욱, "[힌남노 피해집계] 인명피해 심각...12명 사망 또는 실종", 중앙일보, 2022년 9월 5일, <https://www.joongang.co.kr/article/25099521#home>

미래 기상 데이터인데, 이 미래 기상 예측 데이터를 피해액 예측 모델에 넣어 태풍이 있을 경우와 태풍 없이 호우만 있을 경우의 미래 기상 상황에 따른 피해액을 각각 예측하였다. 최종적으로, 피해액 예측값들을 연도-시도-시설별 호우/태풍 발 피해액 데이터에 포트폴리오 최적화를 적용해 지역별, 시설별 리스크를 분석하였다.

3 데이터 분석

3.1 데이터 개요

3.1.1 호우내역데이터 개요

행정안전부가 발행하는 2020년 재해연보에 의하면 최근 10년간 원인별 피해액 현황을 보았을 때 평균적으로 태풍과 호우가 다른 재해들에 비해 많은 금전적인 피해를 발생시켰음을 알 수 있다. 그렇기 때문에 다른 자연재해로 피해가 발생할 수 있으나 평균 피해액이 다른 재해의 약 10배에 달하는 태풍과 호우로 재해를 한정해 분석을 진행하기로 결정했다.

분석에는 기상청에서 제공하는 기상 빅데이터 분석 플랫폼 및 기상융합서비스인 기상청 날씨마루에서 제공하는 호우피해 이력정보 데이터를 이용했다.⁴ 1998년부터의 호우로 인해 발생한 피해 이력 정보를 담은 데이터로 전국 299개 지역구의 호우 피해 일자, 지속 기간, 피해액, 침수면적, 태풍 유무 및 종류, 강수 정보, 최대 풍속, 평균 기온에 대한 정보가 반영되어 있다. 해당 데이터 중 대구시 남부 데이터의 일부를 표 1로 첨부하였다.

번호	일자	침수면적 (1,000 m ²)	피해액 (백만 원)	총강수량(mm)	기간 (일)	태풍	강수계 속시간	최다강 수1H	일강수 량(mm)	누적강 수량 (mm)	최대풍 속(m/s)	평균기 온 (°C)
1	2002-8-30	1.4	56,400	240	3	없음	6	4	14	14	2.1	26.7
2	2002-8-31	1.4	56,400	240	3	루사	24	35.5	223	237	5.4	24
...
134	2017-09-18	0	0	0	3	없음	0		0	0		

표 1 : 대구시 남구 데이터

우선 기존 데이터는 각 시군구별로 데이터가 나누어져 있었기 때문에 전국에 대한 하나의 데이터로 만들어주는 과정이 필요했다. 이후에 시도별 특성을 반영하여 더욱 정확한 분석을 위해서는 시도별로 묶어서 분석을 진행해야 한다고 판단해 기존 데이터에는 존재하지 않는 시도와 시군구 변수를 추가해준 후 하나의 데이터로 병합해주었다. 그 후 태풍 변수의 경우에는 '없음'과 태풍명으로 나뉘어 표기되어 있었기 때문에 태풍 유무에 따른 이진변수를 추가해주었다. 또한, 연도와 월에 따른 피해 패턴을 파악하기 위해서 연, 월, 일 변수를 추가해주었다. 그리고 변수명을 아래 표 2와 같이 변경해주었다.

⁴ <https://bd.kma.go.kr/kma2020/fs/preventionSelect1.do?pageNum=5&menuCd=F050401000>

기존 변수명	변경한 변수명	기존 변수명	변경한 변수명
일자	date	누적강수량(mm)-2일	cum_rain_2
침수면적 (1,000m ²)	wet_area	누적강수량(mm)-3일	cum_rain_3
피해액(백만원)	damage	최대풍속(m/s)	max_wind
총강수량(mm)	total_rain	평균기온 (°C)	avg_temp
기간(일)	duration_day	시도	provin
태풍	typhoon	시군구	district
강수계속시간	duration_hour	일	day
최다강수1H	max_rain	월	month
일강수량(mm)	day_rain	년	year

표 2 : 변수명 변경

예측에 사용된 피해액은 모두 생산자 물가지수를 이용해 2021년 기준으로 변환된 피해액이다. 생산자 물가지수는 국내 생산자가 국내 내수 시장에 공급하는 상품의 가격 및 서비스의 요금을 경제에서 차지하는 중요도를 고려하여 평균한 종합적인 가격수준을 지수화한 통계이다. 생산자 물가지수는 각 연도의 경제 상황이 반영된 지수이므로 여러 연도의 피해액을 특정 연도 기준으로 변환시키는데 적합하다고 생각되어 이를 적용하였다. 데이터의 가장 최근 연도인 2021년을 기준 연도로 선택하고, 아래 수식 1을 이용하여 모든 연도의 피해액을 변환해주었다.

$$\text{각 연도의 피해액} \times \frac{\text{2021년 생산자 물가지수}}{\text{각 연도 생산자 물가지수}}$$

수식 1 : 2021년 기준 피해액 변환 공식

3.1.2 재난복구비 데이터

보다 자세한 분석에 앞서 실제 자연재해 피해비용과 관련하여 재난 복구비가 잘 쓰이고 있는지를 확인해보기 위해서 2020 재해연보 통계자료에서 제공한 최근 10년간 시도별/연도별 자연재난복구비 현황 데이터⁵를 활용하였다. 자세한 데이터의 형태는 다음 표 3과 같다. 연도는 2011년부터 2020년까지 10년으로 설정하였고, 서울/부산/대구 등 총 17개의 지역을 대상으로 한 재난복구비 데이터이다.

연도	서울	부산	대구	인천	광주	...	제주
2011	91934492	32397347	403500	9639594	94500	...	10229507
2012	1395718	4000	NA	180000	70500	...	102594690
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2020	658500	31913948	720525	965783	65113210	...	24965084

표 3 : 재난복구비 데이터

⁵ https://www.mois.go.kr/frt/bbs/type001/commonSelectBoardArticle.do?sessionId=JUA7i1tRMj-dNEOvPcYBqvMi.node50?bbsId=BBSMSTR_000000000014&nttId=89542

결측치 보간을 완료한 재난복구비 데이터를 2021년 화폐가치로 환산한 후, 호우내역데이터의 피해액과 비교하여 시각화한 결과는 다음과 같다. 시각화 결과, 피해복구비와 피해액의 흐름은 거의 유사하지만 피해액에 비해서 재난복구비의 예산이 훨씬 적게 사용되고 있음을 알 수 있었다. (그림 1)

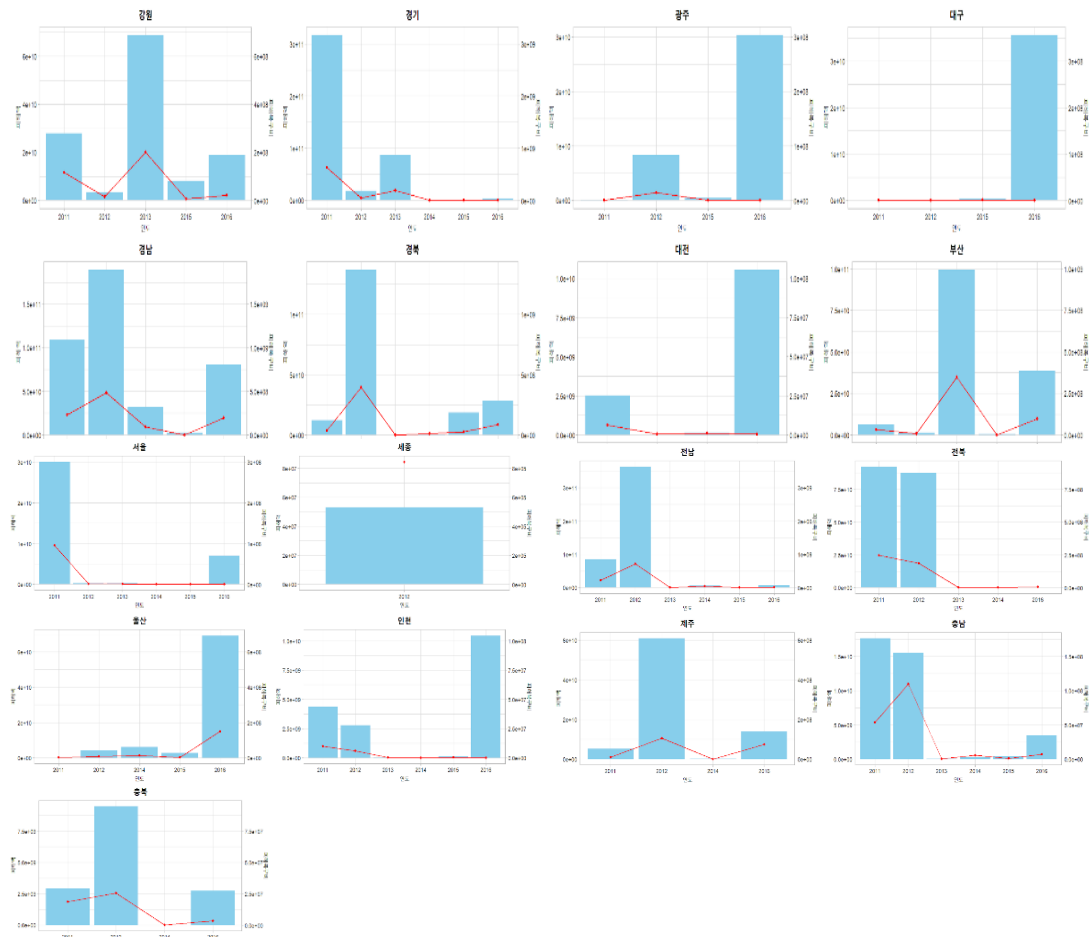


그림 1 : 재난복구비와 호우내역데이터상 피해액 비교

3.1.3 기상 데이터

기상청에서 운영 중인 기상 데이터 플랫폼 '날씨마루'⁶에서 2012년 1월 1일부터 2022년 10월 31일까지의 관측소-일별 평균 온도, 최고 온도, 최저 온도, 최다 풍향, 최고 풍속 풍향, 합성 풍향, 최고 순간 풍속 풍향, 평균 풍속, 합성 풍속, 최고 풍속, 최고 순간 풍속, 운량, 운고, 강수량, 시간당 최대 강수량을 추출하였다. 변수들은 동일한 계열의 변수끼리 묶여 각기 다른 데이터 테이블에 저장되어 있었고, 각 데이터 테이블에 어떤 변수들이 포함되어 있는지는 표 4에 나와 있다. 각 데이터 테이블 별로 따로 전처리를 진행하였다.

⁶ <https://bd.kma.go.kr/kma2020/svc/main.do>

테이블명	변수명
DB_SFC_CLOUD_DD	ca (운량), ch (운고)
DB_AWS_TA_DD	avg_ta (평균 온도), max_ta (최고 온도), min_ta (최저 온도)
DB_SFC_WIND	max_wd (최대 풍향), max_ws_wd (최고 풍속 풍향), mix_wd (합성 풍향), max_ins_ws_wd (최고 순간 풍속 풍향), avg_ws (평균 풍속), mix_ws (합성 풍속), max_ws (최고 풍속), max_ins_ws (최고 순간 풍속)
DB_AWS_RN_DD	sum_rn(강수량), hr1_max_rn (시간당 최대 강수량)

표 4 : 데이터 테이블별 변수명

3.1.4 관측소 데이터

각 데이터 테이블에는 'STN_ID'라는 변수가 공통적으로 있는데, 이는 각 기상 관측소 코드⁷를 의미하고, 각 시도에 속한 관측소의 개수가 모두 다르다. 먼저 해당 코드를 사용하여 기상청에서 제공하는 각 관측소의 이름과 주소⁸, 그리고 주소에 따른 위경도⁹를 모두 추출하였다. 완성된 데이터의 일부분으로 표 5에 참조하였다. 이를 이용하여 추후 데이터 테이블 병합과 시도별 기상 데이터 추출에 사용할 예정이다.

STN_ID(관측소 코드)	Lat(위도)	Long(경도)	add
1100000000	37.56	126.98	서울
1111056000	37.6	126.97	서울
1111061500	37.58	126.99	서울
1111067000	37.57	127.02	서울
1111058000	37.57	126.96	서울

표 5 : 관측소 코드별 위경도와 지역

이후, 병합을 위해 각 관측소의 주소에서 시도(서울, 부산, 대전 등 17개 지역)를 추출하여, 날짜와 시도가 동일한 행에 한해 병합을 해주었다. 이 때, 날짜-시도-변수별 대표 값으로는 중간값을 사용하였다. 세종의 경우 병합 이후에도 결측이 된 경우가 많았기에, 세종을 둘러싼 충북, 충남, 대전의 평균으로 대체하였다.

완성된 데이터는 17개의 모든 행정구역이 모든 변수에 대해 2012년 1월 1일부터 2022년 10월 31일까지의 모든 값들이 있어야 한다. 하지만 데이터를 확인해본 결과 데이터를 병합하는 과정에서 몇몇

⁷ <https://minwon.kma.go.kr/main/obvStn.do>

⁸ <https://www.code.go.kr/stdcode/regCodeL.do>

⁹ <https://www.google.com/maps/place/>

시도별로 몇몇 일자들이 없어졌음을 확인하였다. 병합 방식을 바꿔봐도 이러한 문제가 사라지지 않았고, 이러한 이유로 시도별로 결측된 일자에 한해서는 전날과 다음날의 값을 사용하여 선형보간법을 이용하여 결측치를 대체하였고, 결측값이 연속적인 날짜로 나타났다면 주변 지역의 중간값으로 대체하였다. (최종적으로 병합된 기상 데이터는 표 6을 참고)

date	add	ca	ch	avg_ws	mix_ws	...	avg_ta	max_ta	min_ta	sum_rn	hr1_max_rn
2012-12-01	강원	6.65	12.7	3.006	2.57	...	-12.19	-6.03	-17.67	0.04	0.115
2012-12-01	경기	3.5	10	2.246	2.009	...	-12.54	-6.76	-16.91	0	0
⋮											
2022-10-21	세종	세종	47.7	0.8	0.575	...	5.4	11.15	3.2	4.225	0.5

표 6 : 최종 기상데이터

3.2. PCA

PCA는 Principal Component Analysis의 약자로, 선형대수적 성질들을 이용하여 변수의 차원을 줄임과 동시에 변수 간 독립성을 만족시켜주는 통계적 분석 기법이다. 이를 이용하여 이후 시계열 예측에 활용할 기상변수를 선택하기 위해 PCA를 진행하였다. 본래 PCA를 통해 분석결과를 내기 위해서는 cross-sectional data이어야 한다. 하지만, 이 분석에서 PCA를 사용하는 목적은 결과를 이끌어내기 위함이 아닌 변수추출을 위한 변수 간의 관계를 분석하려는 것이기 때문에 시계열 데이터이지만 PCA를 사용해도 무리가 없다고 판단하였다. PCA를 진행하기 전에 데이터를 다음 3가지의 경우로 나누었다. 첫째는 호우와 태풍 모두 없는 날, 둘째는 호우는 있지만 태풍은 없는 날, 마지막으로 셋 째는 호우와 태풍 모두 있는 날이다.

먼저 호우와 태풍이 모두 없는 날의 결과를 살펴보겠다. 그림 2와 그림 3은 각 PC의 크기를 내림차순으로 정리한 그림과 각 PC에 기여하는 개별 변수들의 기여도를 나타낸 그림이다.

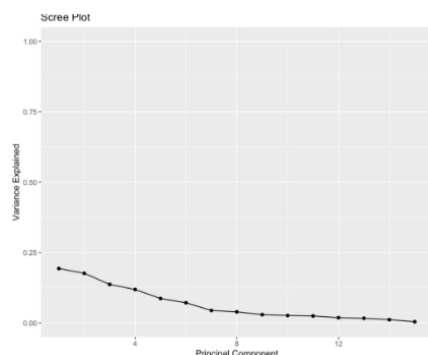


그림 2 : Scree Plot – 호우 X 태풍 X

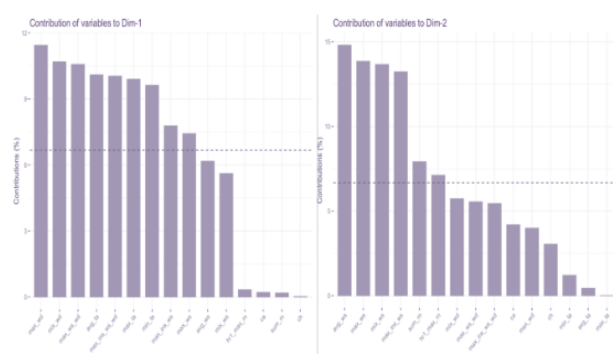


그림 3 : PC1, PC2 변수 기여도 - 호우 X 태풍 X

그림 2에서 확인할 수 있듯, 명확한 elbow point가 존재하지 않으므로, 각 PC들이 가지고 있는 정보량에 큰 차이가 없으며, 그림 3에서 PC1과 PC2에서의 개별 변수의 기여도를 보았을 때 역시 개별 변수들의 영향력이 큰 차이가 나지 않다는 것을 알 수 있다. 이 결과는 아래 그림에서 확인할 수 있듯이 호우는 있지만 태풍이 없는 날의 결과와 유사하다.

다음으로는 호우와 태풍 모두 있는 날의 결과를 살펴보겠다. 먼저 scree plot(그림 4)을 확인해보았을 때, PC4에서 elbow point를 발견할 수 있었다.

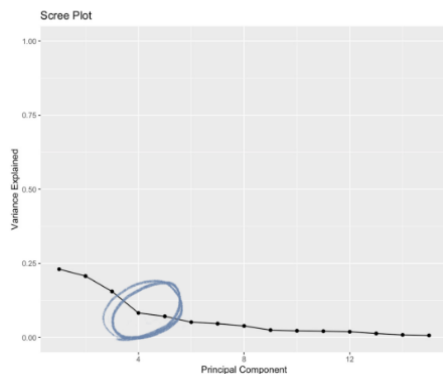


그림 4 : Scree Plot - 호우 O 태풍 O

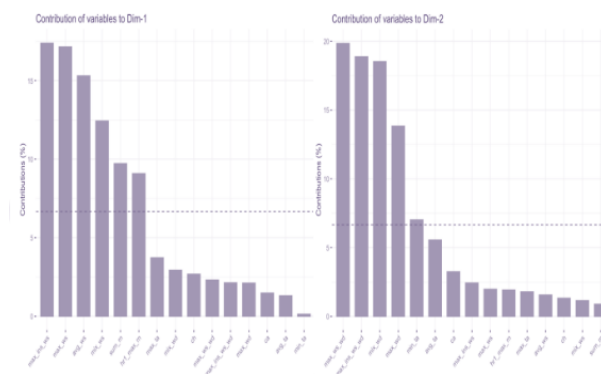


그림 5 : PC1, PC2 변수 기여도 - 호우 O 태풍 O

따라서 PC1, PC2의 Variable Contribution Plot을 그려본 결과, 강수량 관련 변수들이 가장 큰 기여도를 가지는 것으로 나타났다. 이는 분석을 진행하지 않더라도 충분히 예상될 수 있는 결과라고 생각되어 3가지 기상상황에 따라 나누어 진행하는 방안을 폐기하기로 하였다.

위와 같은 결과에 따라 3가지 기상상황에 따라 데이터를 나누는 것이 유의미한 분석결과를 도출하지 못할 가능성이 크다 판단하여, 동일한 계열의 변수끼리 묶기로 결정하였다. 먼저, 기온과 관련된 변수로는 최고기온, 평균기온, 최저기온으로 3가지 변수가 있는데, 높은 상관관계를 고려하여 평균기온만으로 충분한 정보량을 가진다 판단하여 최고기온과 최저기온은 폐기하기로 하였다. 다음으로, 풍향과 관련된 변수인 최다 풍향, 최고 풍속 풍향, 합성 풍향, 최고 순간 풍속 풍향에 대해 PCA를 따로 진행해주었고, 결과는 아래와 같다. (그림 6, 그림 7, 표 7)

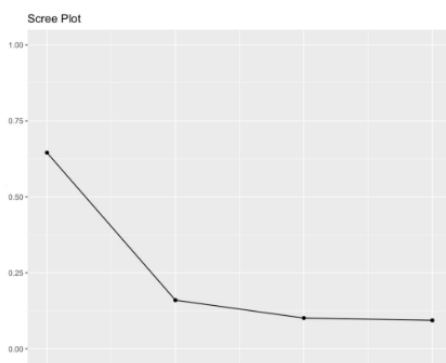


그림 6 : Scree Plot 풍향 관련 변수

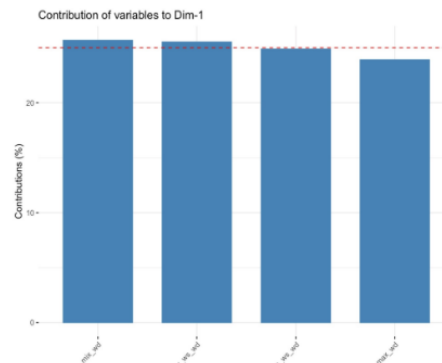


그림 7 : PC1 변수 기여도 - 풍향

합성 풍향	최고 풍속 풍향	최고 순간 풍속 풍향	최다 풍향
-0.507	-0.505	-0.5	-0.489

표 7 : 풍향 관련변수 선형변환 계수

Scree plot(그림 7)을 통해 PC2에서 elbow point가 존재함을 확인하였고, PC1의 변수 기여도(그림 8)를 살펴본 결과 각 변수들이 동일하다 싶은 기여도를 가졌음을 확인하였다. 각 변수 기여도 값을 사용하여 4개의 변수들을 선형결합한다면 4개의 변수를 단 하나의 변수로 표현가능하기에, 표 7에 나온 결과값을 이용하여 선형결합을 실시하였다. 마지막으로, 풍속과 관련된 변수인 평균 풍속, 합성 풍속, 최고 풍속, 최고 순간 풍속에 대해 PCA를 진행해주었고, 결과는 아래와 같다. (그림 8, 그림 9, 표 8)

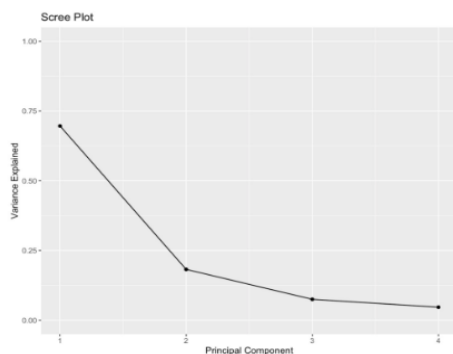


그림 8 : Scree Plot - 풍속 관련 변수

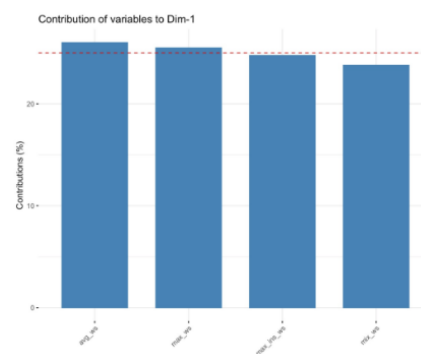


그림 9 : PC1 변수 기여도 - 풍속

평균 풍속	최고 풍속	최고 순간 풍속	합성 풍속
-0.51	-0.505	-0.497	-0.486

표 8 : 풍속 관련변수 선형변환 계수

Scree plot(그림 8)을 통해 PC2에서 elbow point가 존재함을 확인하였고, PC1의 변수 기여도(그림 9)를 살펴본 결과 각 변수들이 동일하다 싶은 기여도를 가졌음을 확인하였다. 각 변수 기여도 값을 사용하여 4개의 변수들을 선형결합한다면 4개의 변수를 단 하나의 변수로 표현가능하기에, 표 8에 나온 결과값을 이용하여 선형결합을 실시하였다.

나머지 변수들인 운량, 운고, 강수량, 시간당 최대 강수량 중, 운고는 PCA에서의 기여도뿐만 아니라 직관적인 생각을 해봐도 호우나 태풍 유무에 영향을 주지 않을 것으로 판단하여 폐기하였고, 시간당 최대 강수량은 강수량과의 높은 상관관계로 인해 폐기하였다. 따라서 시계열 모델링에 사용될 최종 기상 변수들은 표 9에 보이는 바와 같다.

개별 변수	평균기온, 운량, 강수량
풍향 합성 변수	-0.507(합성 풍향) - 0.505(최고 풍속 풍향) - 0.5(최고 순간 풍속 풍향) - 0.489(최다 풍향)
풍속 합성 변수	-0.51(평균 풍속) - 0.505(최고 풍속) - 0.497(최고 순간 풍속) - 0.486(합성 풍속)

표 9 : 최종 변수

3.3 기상변수 예측

주성분 분석을 통해 평균 기온, 운량, 운고, 총 강수량, 시간당 최대 강수량, 풍향 합성 변수, 풍속 합성 변수 기상 데이터를 최종 예측에 사용하기로 결정했었다.

기상 예측의 지역도 한정해주었다. 대한민국 국가지도집에 기록된 자연재해 피해 현황을 참고하여 지역을 선정했다¹⁰. 그림 11은 2005년부터 2014년까지의 지역별 자연재해 피해에 대한 시각화이다. 우선, 호우 피해가 자연재해의 100%를 차지하는 서울, 호우 피해의 비중이 높은 부산, 그리고 자연재해 피해액이 가장 큰 강원으로 기상 예측 지역을 제한했다. 그리고 서울을 기준으로 하여 2012년부터 2021년까지의 데이터를 train set으로 이용하여 모델을 구축하고 test set으로 2022년 10월 31일까지의 데이터를 사용한 후, 각 지역별로 2022년 11월 1일부터 365일의 데이터를 예측하기로 결정했다.

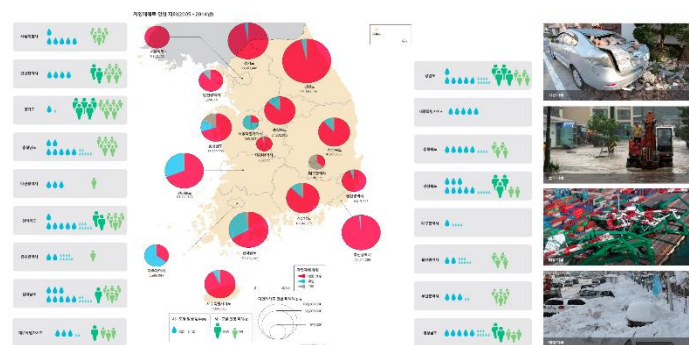


그림 10 : 자연재해 피해 현황

예측 전에 모든 변수들에 대해서 시계열 그래프를 그려보고 직관적으로 추세와 계절성이 존재하는지 파악한 후, 이를 참고하여 시계열 분해를 하였다. 그리고 정상성을 만족하는지 확인하기 위해 ADF 검정을 진행했다. ADF 검정은 시계열 데이터가 단위근을 포함하고 있는지를 확인하는 검정 방법이다. ADF 검정의 경우 귀무가설은 시계열 데이터가 단위근을 포함하고 있다, 즉 '정상성을 만족하지 않는다'이고 대립가설은 '정상성을 만족한다'이다. 또한 AR 모형과 MA 모형을 사용할 가능성을 염두하고 자기공분산함수(이하 ACF), 부분자기상관함수(이하 PACF)를 확인하여 모형 추정을 진행하고자 했다.

먼저 그림 11은 서울의 평균 기온 변수에 대한 시계열 분해 결과이다. 이하 모든 시계열 분해 결과의 첫번째 그래프는 데이터의 시계열 그래프, 두번째 그래프는 추세 그래프, 세번째 그래프는 계절성 그래프이다. 시계열 그래프를 살펴보면 추세는 보이지 않으나 1년을 주기로 계절성을 가지고 있음을 확인했다. 추세 그래프를 보면 추세는 상향하는 듯 하나 큰 경향성을 보이지 않는 것을 확인했고 계절성 그래프를 통해서 계절성이 1년 주기로 뚜렷하게 존재함을 확인했다. 정상성을 확인하기 위해 ADF 검정을 진행한 결과 p value가 0.000039로 정상성을 만족한다는 결과를 보였다. 또한 그림 12의 첫번째 그래프 ACF 그래프를 확인해본 결과 매우 느리게 0으로 수렴하고 있음을 확인했으며 두번째 그래프인 PACF 그래프의 경우 시차 1 이후로 절단됨을 확인했다.

¹⁰ http://nationalatlas.ngii.go.kr/pages/page_1348.php

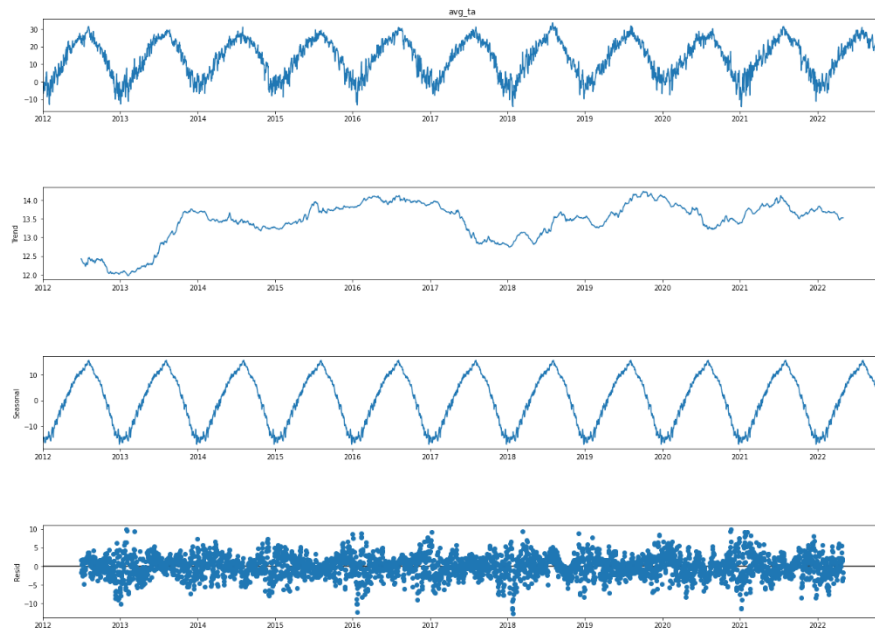


그림 11 : 평균 기온 변수 시계열 분해 결과

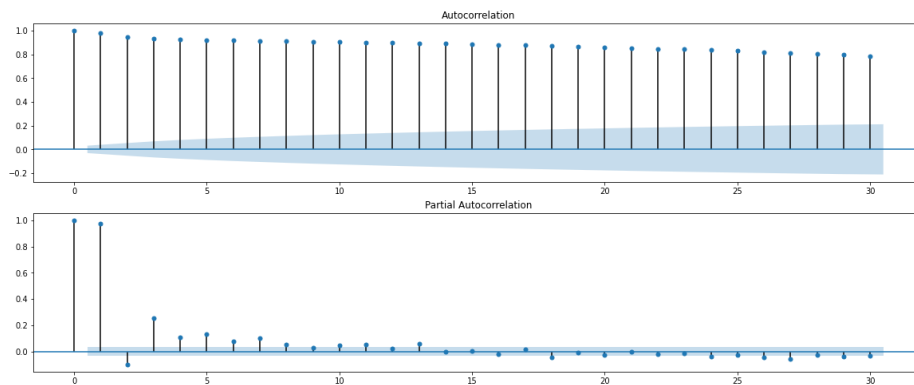


그림 12 : 평균 기온 변수 ACF, PACF 그래프

둘째로 서울의 풍향 합성 변수에 대해 살펴보았다. 그림 13은 풍향 합성 변수에 대한 시계열 분해의 결과이다. 시계열 그래프를 확인했을 때 매우 큰 분산을 가지고 특별한 추세는 보이지 않는 것으로 파악되었다. 그리고 추세 그래프와 계절성 그래프를 통해 추세는 존재하지 않으며 1년을 주기로 뚜렷한 계절성이 있음을 확인했다. 또한 ADF 검정을 진행한 결과 p value가 0.0으로 정상성을 만족한다는 결과가 나왔다. 그리고 그림 14의 첫번째 그래프인 ACF 그래프를 확인했을 때 시차 1 이후로, 두번째 그래프인 PACF 그래프를 확인했을 때 시차 1 이후로 눈에 띄게 0에 가까워짐을 확인했지만 PACF 그래프와 달리 ACF 그래프는 0에 수렴하지는 않는 것처럼 보인다.

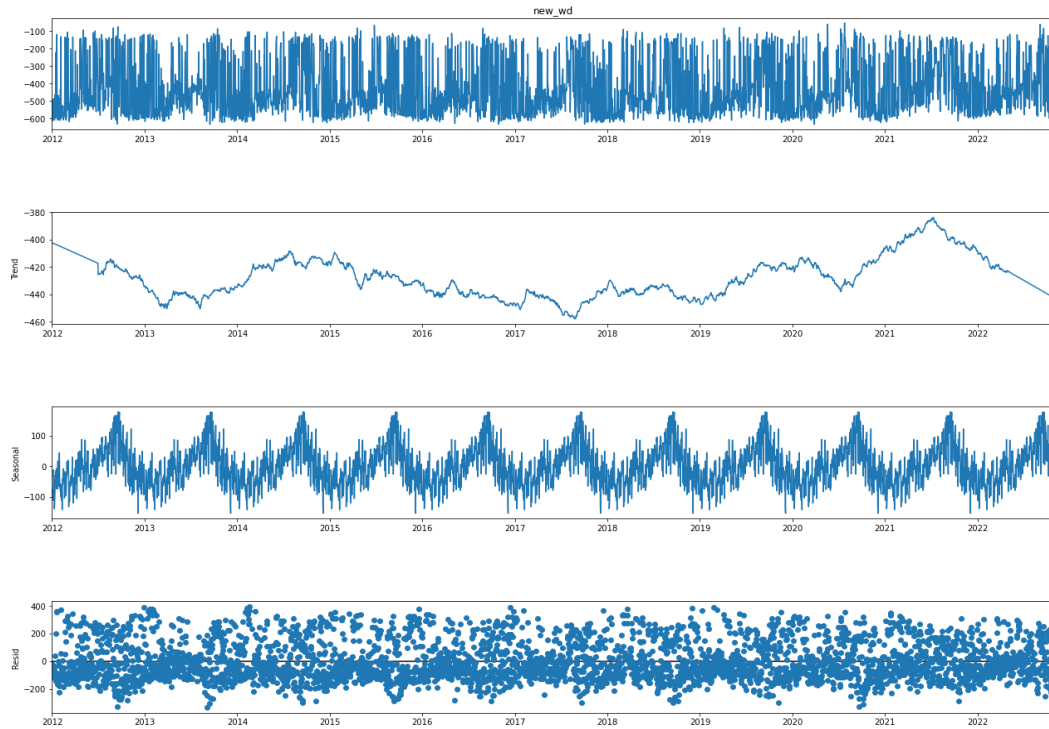


그림 13 : 풍향 합성 변수 시계열 분해

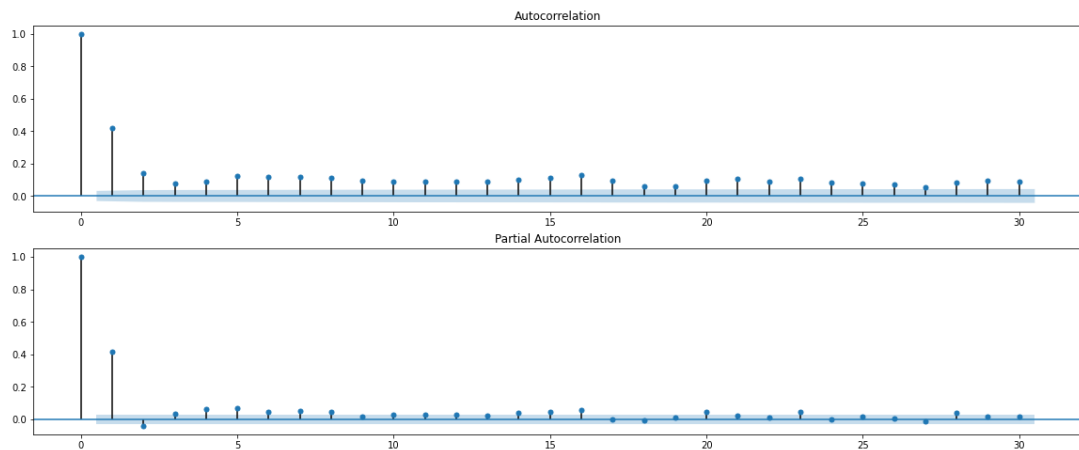


그림 14 : 풍향 합성 변수 ACF, PACF 그래프

다음으로 서울의 풍속 합성 변수에 대해 살펴보았다. 그림 15은 풍속 합성 변수의 시계열 분해 결과이다. 결과 속 시계열 그래프를 살펴보면 마찬가지로 추세는 보이지 않으나 1년을 주기로 계절성을 보이는 것처럼 보인다. 또한 추세 그래프와 계절성 그래프를 통해 확인 시 추세는 보이지 않았고 1년을 주기로 명확한 계절성이 있음을 보았다. 그리고 정상성을 확인하기 위해 ADF 검정을 진행한 결과 p value가 0.0으로 정상성을 만족한다는 결과를 보였으며 그림 16의 첫번째 그래프인 ACF 그래프의 경우 시차 1 이후로, 두번째 PACF 그래프의 경우도 시차 1 이후로 절단됨을 확인했다.

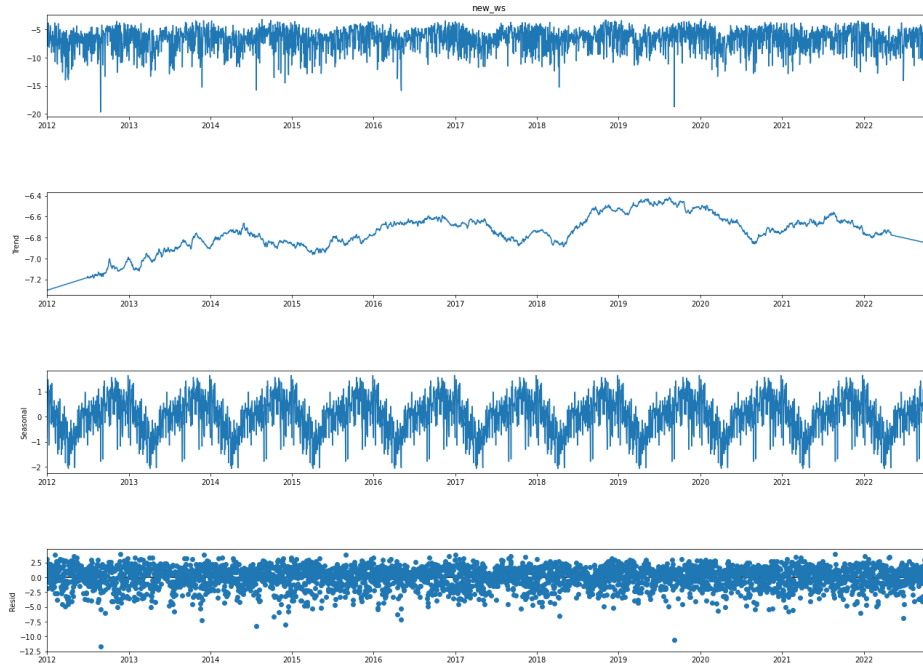


그림 15 : 풍속 합성 변수 시계열 분해

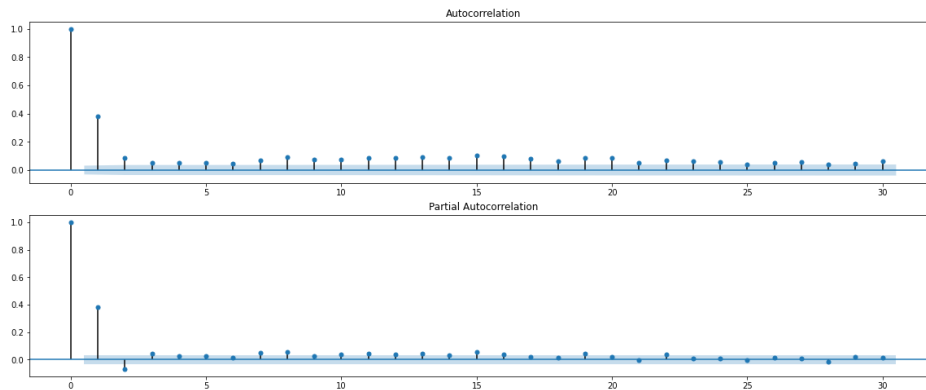


그림 16 : 풍속 합성 변수 ACF, PACF 그래프

다음으로 서울의 총 강수량 변수를 살펴보았다. 그림 17은 총 강수량 변수에 대한 시계열 분해 결과이다. 시계열 그래프를 확인했을 때 매우 불규칙적인 패턴을 보이거나 계절성은 존재하는 것으로 보였고 분산이 매우 큼을 확인했다. 추세 그래프와 계절성 그래프를 봤을 때 추세는 존재하지 않았고 명확한 계절성이 존재함을 확인했다. 그리고 ADF 검정 결과 p value가 0.0으로 정상성을 만족한다는 결과를 얻었으며 그림 18의 결과처럼 첫번째 그래프인 ACF 그래프는 시차 1 이후로, 두번째 그래프인 PACF 그래프는 시차 2 이후로 0에 가까워졌음을 확인했다.



그림 17 : 총 강수량 그래프 시계열 분해

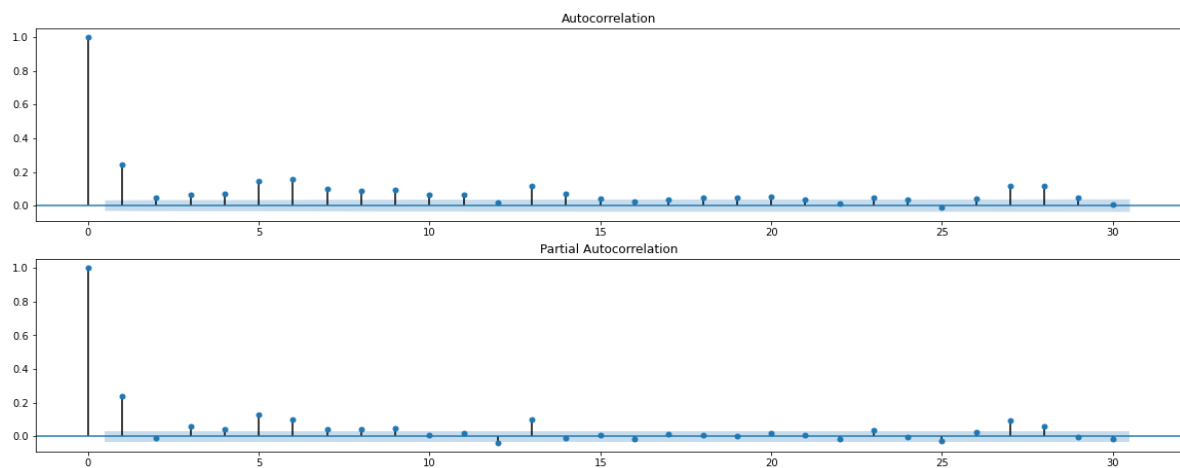


그림 18 : 총 강수량 그래프 ACF, PACF 그래프

마지막으로 서울의 운량 변수를 확인해 보았다. 그림 19는 운량 변수의 시계열 분해 결과이다. 시계열 그래프를 확인했을 때 추세는 존재하지 않는 것으로 보였고 분산이 크게 보였다. 추세 그래프와 계절성 그래프를 통해서 추세는 존재하지 않으며 계절성은 존재함을 확인했다. 그리고 ADF 검정 결과 p value가 0.0으로 정상성을 만족하는 것으로 나왔다. 마지막으로 그림 20과 같이 ACF 그래프와 PACF 그래프를 확인한 결과 첫번째 그래프인 ACF 그래프의 경우 시차 1 이후로 크게 감소하나 0에 수렴하지는 않는 것으로 보였고 두번째 그래프인 PACF 그래프를 확인한 결과 시차 1 이후로 크게 감소하는 것은 동일하나 시차가 커질 수록 0에 수렴하는 것으로 보였다.

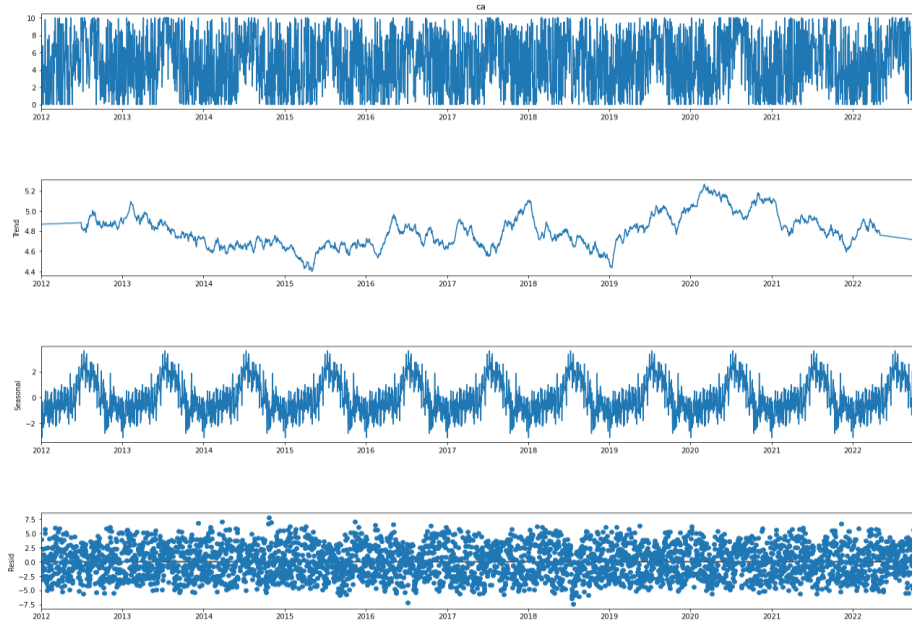


그림 19 : 운량 변수 시계열 분해

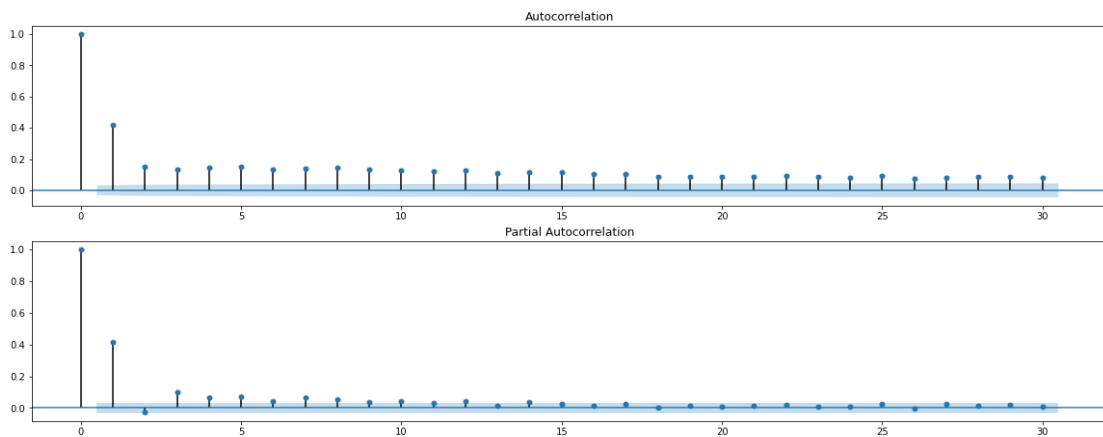


그림 20 : 운량 변수 ACF, PACF 그래프

모든 변수들이 1 년 주기의 명확한 계절성을 보임에도 불구하고 모든 변수에 대해서 ADF 검정을 실시했을 때 표 10 와 같이 정상성을 만족한다는 결과가 나왔다. 이는 ADF 검정이 가진 한계로 기인한 것이라 생각했다. Eric Zivot 과 Jiahui Wang 에 따르면 ADF 검정은 단위근에 대한 정상성 검정이기 때문에 계절성과 같은 다른 유형의 비정상성을 잘 찾아내지 못하는 한계를 가진다.¹¹ 이로 인해 뚜렷한 계절성을 가지는 우리 데이터에서도 모두 정상성을 만족한다는 결과가 나오는 것으로 판단했다.

¹¹ 참고문헌 : 'Modeling Financial Time Series with S-PLUS'

변수	p-value
기온	0.000039
풍향 합성	0.0
풍속 합성	0.0
강수량	0.0
운량	0.0

표 10 : 변수별 ADF 검정 결과

따라서 ADF 검정 결과에 의존하기보다 시계열 분해와 ACF, PACF 그래프를 바탕으로 비정상 시계열로 판단하고 예측을 진행했다.

총 강수량 변수를 제외한 다른 변수들의 예측에는 페이스북에서 공개한 시계열 예측 라이브러리인 Prophet을 사용했다. Prophet은 추세와 계절성, 불규칙한 행사인 휴일 변수를 반영한 모형을 기반으로 데이터를 예측한다. 계절 효과가 강하고 여러 계절의 과거 데이터가 존재할 때 잘 작동한다는 특징을 가지며 추세의 변화에 강하고 결측치가 존재해도 무방하며 이상치를 잘 처리한다. Prophet 모델에는 매우 다양한 파라미터들이 존재하여 수정하며 모델을 만들기 용이하다. 우리 분석에서는 `nchangepts = 40`, `seasonality mode = additive`는 기본으로 두고 `changept prior scale`, `yearly seasonality`, `seasonality prior scale` 세 변수에 대해서 표에 나온 값으로 파라미터별로 정해둔 후보값들의 조합으로 최고의 성능을 내는 파라미터 조합을 찾는 `grid search`를 진행해주었다.

먼저 기본으로 설정한 파라미터에 대해서 소개해보자면 `nchangepts` 추세의 변경점의 개수에 관한 파라미터로 긴 기간의 데이터를 사용하는 우리 데이터의 특성상 기본으로 설정되어있는 25보다 큰 40으로 변경했다. `Seasonality_mode` 파라미터는 데이터의 계절성의 진폭의 변화에 따라 `additive`와 `multificative`로 설정해주며 우리 데이터에서는 크게 추세가 존재하지 않는 것으로 보고 `additive`로 설정했다. `change point prior scale`은 추세의 유연성을 조절해주는 파라미터로 높을 수록 유연하게 트렌드를 감지한다. `yearly seasonality`는 연 계절성에 관한 파라미터로 높을수록 연 주기 패턴을 유연하게 감지하며 `seasonality prior scale`은 계절성을 반영하는 강도이다. 모델의 평가지표로는 MAE를 활용했다. MAE란 실제 값과 예측 값의 절대 오차의 평균을 의미한다.

우리의 분석은 피해가 최대가 되는 상황을 고려하여 대비하는 것이 목적인 예측을 진행할 것이기 때문에 재난 피해와 관련된 기상 상황도 극한의 상황을 가정하고자 했다. 따라서 표 11와 같이 각 기상 변수에 대한 미래 값을 추출해주기로 했다.

변수	예측 값
기온	yhat
풍향 합성	yhat lower
풍속 합성	yhat lower
운량	yhat upper

표 11 : 기상변수별로 추출한 예측 값

Prophet 모델을 이용한 grid search 결과(표 12)를 바탕으로 test set과 비교하여 예측 성능을 평가했으며 결과 파라미터들을 기반으로 각 변수에 대한 1년 간의 예측값을 추출하여 주었다.

변수	change point prior scale	seasonality prior scale	yearly seasonality	MAE
기온	0.5	15	20	2.082
풍향 합성	0.1	8	20	113.16
풍속 합성	0.5	15	20	1.38
운량	0.5	8	20	2.42

표 12 : Prophet – Grid Search 결과

수많은 시계열 모델과 머신러닝 모델을 사용해보았으나 1년이라는 기간 동안의 큰 분산을 잡아내는 모델이 없었기에, 모델을 활용한 예측으로 분산을 잡아내기에 한계가 있다고 판단하여 새로운 관점이 필요하였다. 먼저 강수량이 자체가 어떠한 형태를 따르는지 대략적으로 파악할 필요가 있었다. 아래 그림 21은 강수량에 대한 Density Plot이다.

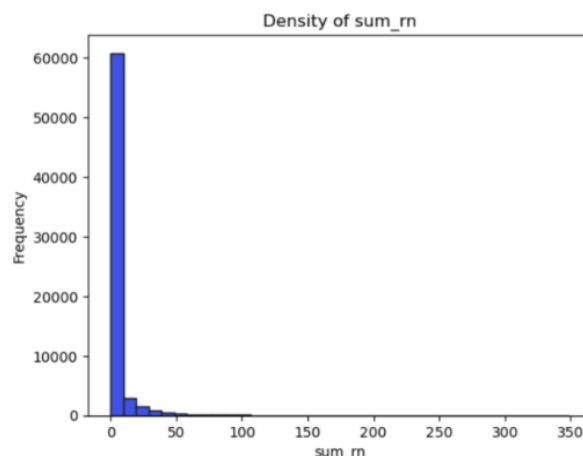


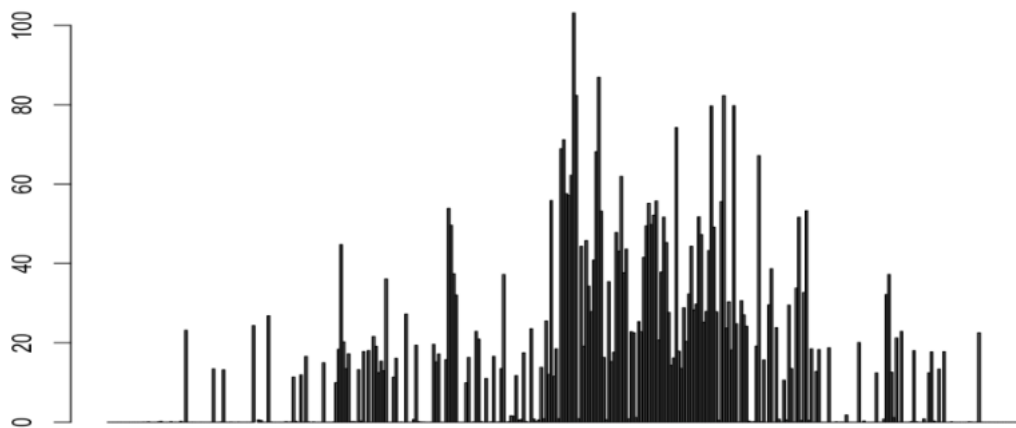
그림 21 : 강수량 분포

그림을 통해 주목해야 할 점은 강수량이 0일 확률이 압도적으로 많다는 것과, 만약 비가 오게 된다면 강수량 5~20 사이로 올 확률이 매우 컸고, 그 이상 올 확률은 매우 이례적인 일(Extreme case)이라고 가정하였고, 이를 토대로 Extreme Value Theory를 응용해보기로 하였다.

Training 데이터는 2012년부터 2022년까지 대한민국 17개 행정구역별 강수량이다. 이는 곧 1년의 각 365개의 일자들에 대해 11개의 관측치가 있다는 의미이다(이 때, 윤일은 제거해주었다). 이를 365개의 개별 cross-sectional data라고 가정하고, 각 일자들의 강수량이 Gumbel Distribution을 따른다고 가정하였다. 11개의 관측치와 Invariance Property of MLE를 사용하여 location parameter '에타'와 scale parameter '세타'를 추정하고, 추정된 모수를 통해 해당 일자의 95-분위수를 추출하여, 추출된 값을 예측값으로 사용하였다. 이 때, 단순히 모든 일자에 대해 모두 95-분위수를 사용하게 된다면, 극소수의

일자들을 제외하고는 모두 비가 오는 것으로 추정되고, 이는 아무리 최악의 기상상황이라 한들 현실성이 없다고 판단하였다. 이러한 이유로, 강수량이 4.5mm (물 웅덩이가 고이기 시작하고, 옷이 많이 젖기 시작하는 단계)이상의 값이 2개 미만일 경우, 해당 일자에 강수량이 4.5mm 이상으로 비가 올 확률은 10년에 한번 있을 만한 극단적인 경우라고 판단되어, 11개 관측치의 중간값을 사용하였다. 이 때, 평균보다 중간값을 사용한 이유는 혹시나 있을 이상치에 대처하기 위해서이다. 최종 예측된 강수량에 대해서는 그림 22와 같다. 위에서 다양한 모델들을 사용하여 예측한 값에 비해 분산을 상당히 잘 잡았음을 알 수 있다.

그림 22 : 일별 예측 강수량



3.4 이상치 탐지

태풍은 호우를 동반하기에 호우와 태풍에 의한 피해내역 데이터를 활용해 호우가 온 날짜들을 추출하여, 행정구역별로 호우가 온 날과 오지 않은 날을 구분하기 위해 Dummy Variable을 추가하였고, 0을 호우가 오지 않은 날, 즉 정상치라고 정의하고, 1을 호우가 온 날, 즉 이상치라고 정의한 뒤, 이상치 탐지 모델을 적용하여 미래 기상상황에 따른 호우 탐지 모델을 만들고자 하였다. 이상치 탐지에는 3가지 모델을 사용하고 성능비교를 하였다.

첫번째는 Stray Algorithm으로, Extreme Value Theory와 KNN distance with the maximum gap이라는 개념을 응용하여 이상치 탐지 및 Anomalous Score까지 제공해주는 모델¹²로, 1차원부터 고차원까지 다양한 차원의 데이터에 모두 효과적인 모델이다.

두번째는 트리 계열의 대표적인 이상치 탐지 모델인 Isolation Forest다. 비정상 데이터를 가장 얇은 깊이에 고립시키는 모델¹³로, 예측 모델로 뛰어난 성능을 보이는 트리 계열의 장점을 취하기 위해 사용하였다. 하이퍼 파라미터 튜닝을 위해서는 자동으로 하이퍼 파라미터를 최적화시켜주는 파이썬의

¹² 참고 문헌 : 'Anomaly Detection in High-Dimensional Data'

¹³ 참고 문헌 : 'Isolation Forest'

OPTUNA라이브러리를 사용하였다¹⁴.

세번째는 VAR을 사용하였다. 다차원 시계열 모델링의 일종인 VAR구조를 응용한 이상치 탐지 모델¹⁵로, Squared Errors의 평균과 표준편차를 더한 값을 Anomalous Threshold로 설정하고(표준편차에 scaler를 곱해 Threshold를 조정해줄 수 있다), 예측치의 Squared Error가 Threshold보다 크거나 같다면 이상치로 분류한다. 시계열 데이터를 사용하는 만큼 시계열 특성을 살리는 모델이 있어야 한다고 판단하였다.

예측의 목표가 단순 이진분류가 아닌 이상치 탐지이기에, 위 3개의 이상치 모델들의 성능 비교하기 위해 사용한 지표는 F1-Score이다. 이상치 탐지의 경우, 정상치의 개수가 이상치의 개수보다 월등히 많기 때문에 True Positive와 True Negative에만 의존하는 수식구조를 가진 Accuracy를 사용할 경우, 모두 0(정상치)로 예측하여도 높은 Accuracy를 가질 수 있기에 Confusion Matrix의 다양한 요인을 사용하는 F1-Score또는 Normalized Gini Index가 적절하다. 이 때, F1-Score가 속도면에서 우월하기 때문에 최종적으로 F1-Score를 사용하였다. F1-Score는 0부터 1사이의 값을 가지면, 1에 가까울수록 모델의 성능이 우수하다. 수식 2는 F1-Score를 계산하는 방법이다.

$$F1-Score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = 2 * \frac{precision * recall}{precision + recall} = \frac{2TP}{2TP + FN + FP}$$

수식 2 : F1 – Score 공식

Stray Algorithm은 모두 정상치로 분류해 의미가 없었고, Isolation Forest는 True Negative가 하나도 없는 방향으로 학습된 것을 확인하였다. 그림 23은 Isolation Forest를 이용해 강원, 부산, 서울에 대해 호우 탐지를 한 결과의 Confusion Matrix이며, True Negative가 하나도 없는 것을 확인할 수 있다.

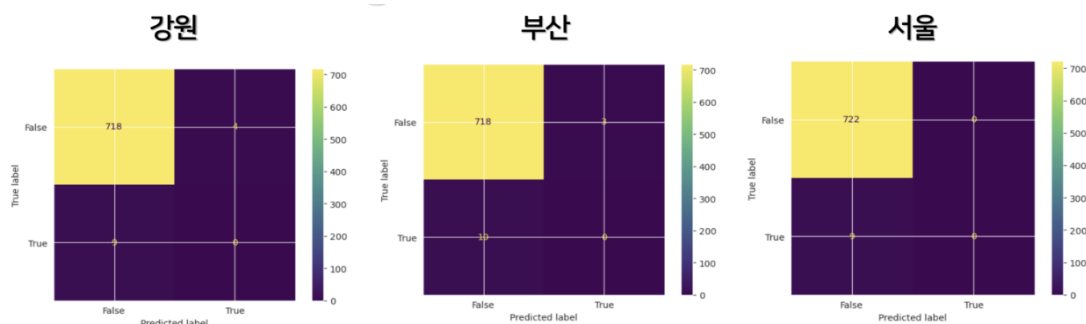


그림 23 : Isolation Forest Confusion Matrix

¹⁴ 참고 문헌 : 'Optuna: A Next-Generation Hyperparameter Optimization Framework'

¹⁵ <https://www.analyticsvidhya.com/blog/2021/08/multivariate-time-series-anomaly-detection-using-var-model/>

결론적으로는, 서울, 강원, 부산에 대해 각각 0.875, 0.823, 0.866의 높은 F1-score를 기록한 VAR모델을 사용하기로 하였다.

3.5 호우/태풍 이진분류

이상치 탐지 모델을 통해서 얻은 극한의 기상상황일 경우의 기상 데이터를 바탕으로 오직 호우만 존재할지, 또는 태풍을 동반할 호우가 존재할지 예측하는 이진분류 모델링을 진행하였다. 먼저 주요 기상 변수들이 존재하는 데이터셋에서 호우가 있는 행만 추출하고 태풍 유무에 따른 데이터의 개수를 파악하였다. 그 결과 호우만 존재하는 경우는 541행, 호우와 태풍이 모두 존재하는 경우는 165행으로 태풍 유무에 따른 클래스 차이가 존재한다는 것을 알 수 있었다. 따라서 태풍 유무의 비율을 맞춰서 train 데이터와 test 데이터로 분리하였다.

먼저 로지스틱 회귀 모델링을 진행한 결과 정확도는 0.7919, F1-score는 0.8782인 것으로 나타났다. 두번째로 Grid Search를 통한 Random Forest를 진행한 결과 파라미터 조합 별 F1-score를 구하였다. 그 후, F1-score가 가장 높은 파라미터로 mtry는 4, ntree는 300을 선정하여 다시 모델을 학습시킨 결과, 정확도는 0.8777, F1-score는 0.9258로 로지스틱 회귀보다 우수한 성능을 보였다. 따라서 태풍/호우 이진분류 모델에서는 Random Forest 모델을 채택하기로 결정하였다.

3.6 피해액 예측

대한민국 국가지도집에 수록된 자연재해 현황을 기준으로 호우 피해가 전체 피해의 100%를 차지했던 서울과 호우 피해의 비중이 매우 높은 부산과 강원 미래 호우 피해액을 예측하기로 결정했다. 피해액 예측에 설명 변수로 사용된 데이터는 평균 기온, 운량, 총 강수량, 풍속과 풍향이다. 여기서 풍속과 풍향은 PCA결과에 맞게 변환된 값을 사용하였다.

우선 전체 시도별 호우 피해액 데이터에서 서울, 부산, 강원 세 지역의 데이터만을 추출해 사용하였다. 세 지역의 데이터에서 태풍 혹은 호우가 있는 행만을 추출한 다음, 사용하기로 결정한 설명 변수를 추출했다. 8대 2의 비율로 train과 test set으로 나눈 후 각각의 데이터를 태풍 없이 호우만 있을 때와 태풍이 있을 때로 나눠주었다. 설명 변수에는 표준화를, 반응변수인 피해액은 먼저 로그변환을 시켜준 다음 표준화를 진행해 주었다. 전처리를 완료한 데이터에 각각 5가지 모델을 적합 시켜 가장 성능이 좋은 모델을 사용하였다. 피해액 예측 모델을 만들기 위해 사용된 방법은 선형회귀, XGBoost, Random Forest, Supporting Vector Machine, LightGBM이다. 5가지 모델 모두 서울, 부산, 강원 각 태풍이 있을 때와 태풍 없이 호우만 있을 때로 나누어 모델링을 진행한 후, test mse가 가장 작은 모델을 선택하였다.

서울의 태풍이 있을 경우의 5가지 모델의 mse와 태풍 없이 호우만 있을 경우의 5가지 모델의 test mse는 표 13과 같았다.

태풍이 있을 경우 (서울)		태풍 없이 호우만 있을 경우 (서울)	
모델	test mse	모델	test mse

선형회귀	0.5253	선형회귀	0.4802
XGBoost	0.3356	XGBoost	0.4372
Random Forest	0.4887	Random Forest	0.4178
Supporting Vector Machine	0.3350	Supporting Vector Machine	0.4364
LightGBM	0.5361	LightGBM	0.4342

표 13 : 태풍 유무에 따른 모델별 MSE

태풍이 있을 경우 SVM의 mse가 가장 작았고, 태풍 없이 호우만 있을 경우 Random Forest의 mse가 가장 작았다. 하지만 미래 기상 예측 데이터를 사용해 태풍이 있을 경우의 피해액 예측을 SVM으로 진행한 결과, 기상 상황이 모두 다름에도 불구하고 모든 피해액이 동일하게 예측되는 오류가 발생해 그 다음으로 mse가 작았던 XGBoost를 태풍이 있을 경우의 미래 피해액 예측 최종 모델로 선택하였다. 태풍 없이 호우만 있을 경우의 미래 피해액 예측은 Random Forest를 사용해 진행하였다. 아래는 서울의 XGBoost와 Random Forest의 산점도 그래프와 인덱스 순서대로 피해액과 예측값을 나열한 그래프이다.

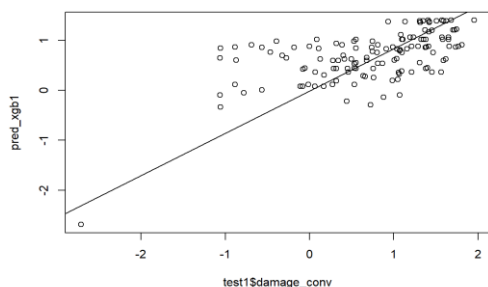


그림 24 : 예측/실제 산점도 - 태풍 O
(XGBoost)

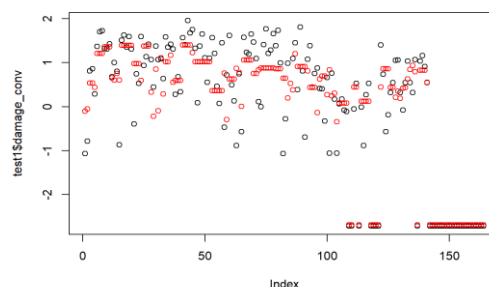


그림 25 : 인덱스별 예측/실제 비교
(XGBoost)

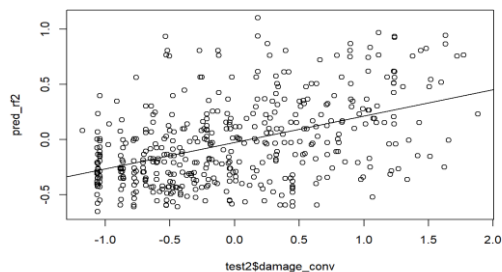


그림 26 : 예측/실제 산점도 - 태풍 X
(Random Forest)

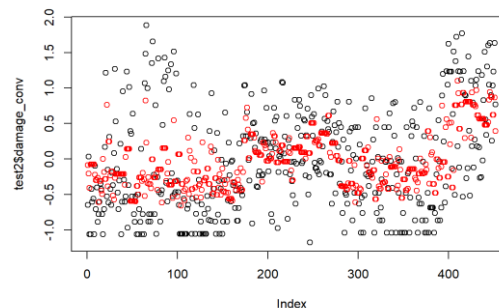


그림 27 : 인덱스별 예측/실제 비교
(Random Forest)

서울의 피해액을 예측하기 위한 모델링 과정을 부산과 강원에도 동일하게 적용해 피해액 예측에 사용할 모델을 결정하였다. 각각의 경우의 test mse는 모두 달랐지만 서울의 경우와 동일하게 부산과 강원도 태풍이 있을 경우에는 SVM, XGBoost, LightGBM, Random Forest, 선형회귀 순서로 성능이 좋았고 태풍 없이 호우만 있을 경우에는 Random Forest, LightGBM, SVM, XGBoost, 선형회귀의 순서로 성능이 좋았다. 서울의 태풍이 있을 경우 SVM을 사용해 미래 피해액을 예측할 때 발생했던 문제가 부산과 강원도 태풍이 있을 경우의 피해액을 예측할 때도 동일하게 발생했다. 따라서 태풍이 있을 경우에는 XGBoost를, 태풍 없이 호우만 있을 때는 Random Forest를 사용해 미래 피해액을 예측했다.

4 리스크 분석

4.1 원인-연도-지역-시설별 피해액 데이터 개요

미래의 기상 상황에 따라 호우와 태풍으로 인해 발생한 피해 중 시설별로 총 피해액에서 차지하는 비율이 어느 정도일지를 예측하기 위하여 사용한 데이터로 국가재난안전포털에서 제공하는 데이터를 사용하였다. 국내 17개의 시군구의 태풍과 호우에 의해 발생한 1998년부터 2020년 까지의 시설별 피해액을 다운받아 정리한 데이터이다. 피해를 입은 시설로는 사망 ·○실종, 이재민, 침수면적, 건물, 선박, 농경지, 공공시설 그리고 기타가 있다. 이 8가지의 시설 중 단위가 화폐가 아닌 사망 ·○실종과 이재민은 이 시설들의 피해를 화폐가치로 환산하기 어려워 총 피해액에서 어느 정도를 차지하는지를 파악하는 것이 불가능했기 때문에 시설별 피해액 예측에서 이들을 제외하였다. 따라서 시설별 피해액 예측에는 피해 단위가 화폐인 건물, 선박, 농경지, 공공시설 그리고 기타만을 사용해 호우와 태풍으로 인한 피해액 중 각 시설이 차지하는 비율을 판단하였다. 재해연보에 따르면 건물에 포함되는 구체적인 시설은 주거물, 공공시설에 포함되는 구체적인 시설로는 도로, 하천, 소하천, 수도, 항만, 어항, 학교, 철도, 수리시설, 사발, 순시설, 해양 등이 있고, 기타에 포함되는 구체적인 시설로는 축대, 담장, 가축, 축사, 담사, 수산증양식, 어망, 어구, 비닐하우스, 기타사유시설 등이 있다. 기존의 데이터는 2020년의 화폐가치를 기준으로 피해액이 작성되었기 때문에 2021년 생산자 물가지수를 사용하여 피해액을 2021년 기준으로 변경해 사용하였다. 2022년 기준으로 피해액을 변환하고 싶었으나 2022년 생산자 물가지수가 아직 공개되지 않았으므로 가장 최근인 2021년의 지수를 사용하였다. 피해액 변경 방법은 아래 수식 3과 같다.

$$2020 \text{년 기준 변환된 피해액} \times \frac{2021 \text{년 생산자 물가지수}}{2020 \text{년 생산자 물가지수}}$$

수식 3 : 2020년 기준 피해액 변환 공식

4.2 지역별 리스크 분석

마지막으로 분석의 최종 단계인 리스크 분석이 남았다. 리스크 분석을 실시하여, 호우나 태풍이 발생할 시 각 행정구역의 피해 유형들의 피해 비율들을 파악한 뒤, 앞서 기상 데이터로 예측된 피해액

을 대입하여, 각 시설별 예상 피해액을 추출할 것이다.

17개의 시도마다 각기 다른 지리적 특성과 기반시설이 있다. 이를 모두 고려하여 리스크 분석 모델링을 하기에는 매우 까다롭기에, 새로운 관점이 필요하다고 생각하였고, 포트폴리오 분석을 적용하기로 하였다. 먼저 리스크 분석에 사용될 데이터를 살펴보겠다.

연도	건물	선박	농경지	공공시설	기타	합계	시도
1998	508018.286	0	9573492.45	79228775.2	652306.692	89962592.6	강원
1998	36828357.3	33966.107	32979395.3	435509999	119355371	624707089	경남
1998	1828380.75	6681.52577	7645584.52	92070225.2	2986368.63	104537241	경북
1998	1156903.9	5479.46812	32509197.1	334528896	14827630.1	383028107	광주
1998	257833.92	0	22826.3302	373632.496	837.185286	655129.931	대구
1998	234492.727	0	0	3428232.26	1455657.78	5118382.76	대전
1998	255412.785	0	239222.238	11083980	191370.77	11769985.8	부산
1998	0	0	0	47496.1701	0	47496.1701	서울
1998	1221318.23	0	71851.1347	36676968.2	31769275.3	69739413	울산
1998	0	0	0	438169.163	0	438169.163	인천
1998	2840310.12	0	4076297.71	27605471	2673124.11	37195203	전남
1998	693606.414	2713.67175	3043149.37	63004873.2	1301601.86	68045944.6	전북
1998	285426.994	0	540768.507	19176459.8	583003.281	20585658.6	제주
1998	0	0	32608.7392	31210.9483	0	63819.6875	충남
1998	3276032.61	1685.00825	13148733.2	93291601.9	6314250.35	116032303	충북

연도	건물	선박	농경지	공공시설	기타	합계	시도
1998	10855.7508	0	0	697797.659	1257594.85	1966248.26	강원
1998	124099.156	186413.612	3141393.65	69218552.9	2622119.63	75292578.9	경기
1998	969656.295	13296.0342	12136849.6	191144174	33287649.7	237551626	경남
1998	67851.3676	0	0	511898.911	26654.8306	606405.109	경북
1998	27139.9088	0	15942.6885	4743981.17	1979135.8	6766199.56	광주
1998	0	0	0	0	0	0	대구
1998	0	79958.1093	0	262284.725	62694.2211	404937.055	대전
1998	0	0	0	0	0	0	부산
1998	13570.4863	0	45123.9678	5813100.82	419062.829	6290858.1	서울
1998	0	0	0	0	0	0	울산
1998	753295.491	176667.371	791050.739	20915817	12811679.4	35448510	인천
1998	36255.3354	46004.7676	28315.3722	4357445.17	1926616.52	6394637.16	전남
1998	0	426585.795	0	66724.8375	558377.055	1051687.69	전북
1998	116718.734	0	0	55157.426	507447.043	679323.203	제주
1998	173.394157	0	0	553313.521	0	553486.915	충남
1999	244599.585	32643.8436	27970.7115	8559692.99	6704616.98	15569524.1	충북

표 14 : 연도-시도-시설별 피해액 데이터

표 14의 왼쪽 표는 1998년부터 2020년까지 '호우에 의한' 시설별 피해액을 나타낸 데이터고, 오른쪽 표는 같은 기간동안 '태풍에 의한' 시설별 피해액을 나타낸 데이터이다. 해당 데이터를 시도별로 분류하고, 포트폴리오 최적화를 실시하였다. 각 시설(건물, 선박, 농경지, 공공시설, 기타)을 주식종목, 그리고 총 피해액(합계)을 주식수익으로 간주하고 포트폴리오 최적화를 실시하였다. 이 때, 포트폴리오 최적화는 2가지 방법이 있는데, 하나는 리스크(분산)가 최저로 되는 최적화 방법이고, 다른 하나는 수익이 최대화되는 최적화 방법이다.

보고서에서의 분석 목표는, 기상 조건이 최악으로 주어졌을 때의 지역별 리스크 분석이다. 최악의 기상 조건이 주어진다면, 그만큼 호우 또는 태풍의 연간 발생률이 올라갈 것이고, 그로 인해 피해액 역시 증가할 것이라는 가정 하에 최종 리스크 분석에 사용한 포트폴리오 최적화 방법은 두 번째 방법이다. (기대 주식수익(기대 피해액)이 극대화되었을 때의 각 주식종목(각 시설)이 차지하는 수익 비율(피해액 비율))

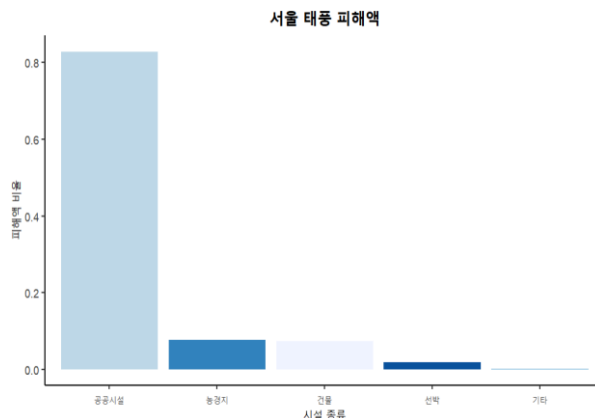


그림 28 : 태풍이 동반된 경우의 서울 시설별 예상 피해액 비율

건물	선박	농경지	공공시설	기타
0.073	0.018	0.076	0.828	0.002

표 15 : 태풍이 동반된 경우의 서울 시설별 예상 피해액 비율

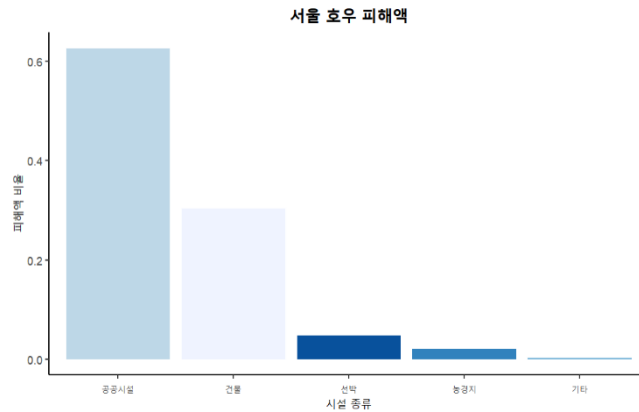


그림 29 : 호우만 있는 경우의 서울 시설별 예상 피해액 비율

건물	선박	농경지	공공시설	기타
0.303	0.047	0.021	0.625	0.002

표 16 : 호우만 있는 경우의 서울 시설별 예상 피해액 비율

1) 서울

포트폴리오 최적화를 통해서 얻은 서울의 시설 별 피해액 비율은 위와 같다. (그림 28, 그림 29, 표 15, 표 16) 서울의 경우, 태풍의 유무와 관련없이 공공시설의 피해액 비율이 가장 높기 때문에 추후 공공시설을 중심으로 재난 복구 비용을 투자해야 한다고 판단된다.

2) 부산

포트폴리오 최적화를 통해서 얻은 부산의 시설 별 피해액 비율은 다음과 같다. (그림 30, 그림 31, 표 17, 표 18)

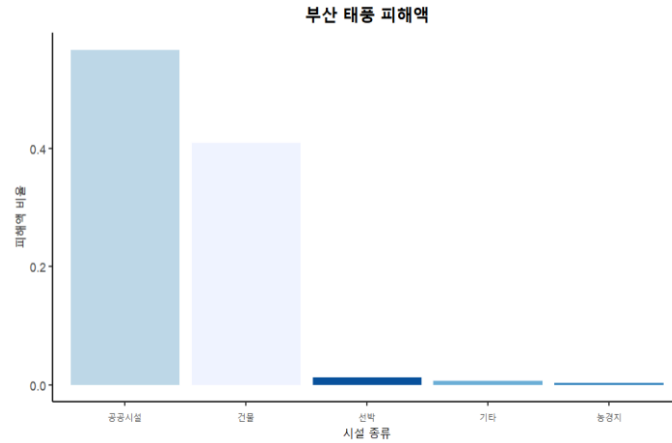


그림 30 : 태풍이 동반된 경우의 부산 시설별 예상 피해액 비율

건물	선박	농경지	공공시설	기타
0.408	0.013	0.003	0.566	0.007

표 17 : 태풍이 동반된 경우의 부산 시설별 피해액 비율

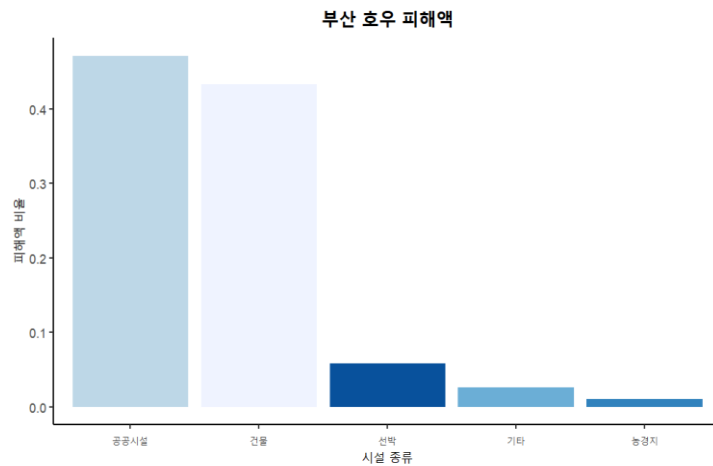


그림 31 : 호우만 있는 경우의 부산 시설별 예상 피해액 비율

건물	선박	농경지	공공시설	기타
0.432	0.058	0.011	0.471	0.026

표 18 : 호우만 있는 경우의 부산 시설별 예상 피해액 비율

즉 부산의 경우, 태풍의 유무와 관련없이 공공시설과 건물의 피해액 비율이 가장 높기 때문에 추후 공공시설과 건물을 중심으로 재난 복구 비용을 투자해야 한다고 판단된다.

3) 강원

포트폴리오 최적화를 통해서 얻은 강원의 시설 별 피해액 비율은 다음과 같다. (그림 32, 그림 33, 표 19, 표 20)

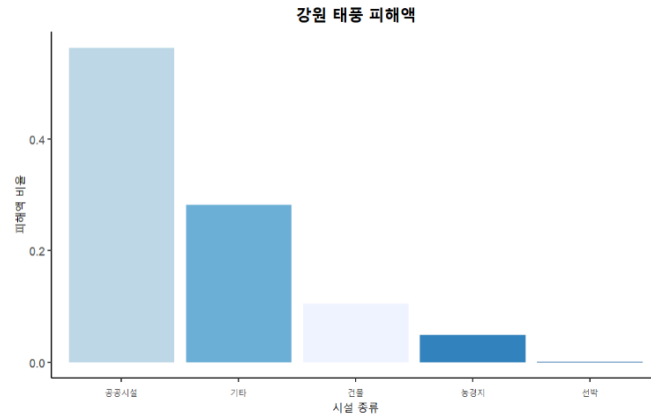


그림 32 : 태풍이 동반된 경우의 강원 시설별 예상 피해액 비율

건물	선박	농경지	공공시설	기타
0.104	0.001	0.048	0.563	0.281

표 19 : 태풍이 동반된 경우의 강원 시설별 예상 피해액 비율

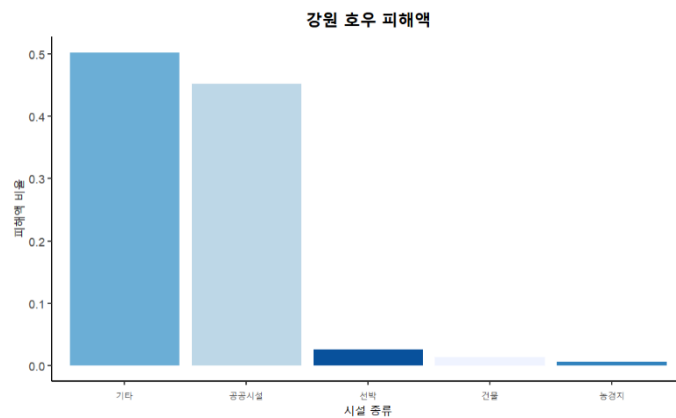


그림 33 : 호우만 있는 경우의 강원 시설별 예상 피해액 비율

건물	선박	농경지	공공시설	기타
0.013	0.025	0.006	0.451	0.502

표 20 : 호우만 있는 경우의 강원 시설별 예상 피해액 비율

즉 위 그림과 같이 강원의 경우, 태풍이 있을 때는 공공시설이, 호우가 있을 때는 기타 시설의 피해 비율이 가장 높은 것으로 보인다. 따라서 태풍이 올 때에는 공공시설의 피해를 예상하고 추가적으로 기타시설에 대한 재난 복구 비용 투자를 늘려야 할 것으로 판단된다.

5 보험산업의 대응 방안

5.1 풍수해 보험의 개선

현재 재난에 대비하기 위한 보험으로 많은 보험사들이 풍수해 보험 상품을 판매하고 있다. 이때 대부분의 보험사들은 특수건물과 일반건물, 주거용 건물과 상업용 건물, 공동 주택과 단독 주택과 같이 단편적으로 구분된 건물의 면적과 연식 등을 기준으로 풍수해 보험을 판매한다. 하지만 각 건물 구분의 경우에도 아래 단계의 보다 세부적인 구분이 존재할 것이다. 예를 들어보자면 공동 주택의 경우에도 아파트, 연립 주택, 기숙사 등으로 분류할 수 있고 상업용 건물에도 상가, 백화점, 대형마트 등으로 분류가 가능하는 등 주 구분 아래의 세부적인 구분이 존재한다. 이러한 세부 구분에 따라서 만들어지는 건물의 일반적인 구조나 설계방식, 원재료가 상이할 것이다. 그렇기 때문에 건물 각각의 유형과 특성에 따라서 재난 발생시 건물에 따른 철거 비용과 복구 비용이 천차만별일 것이므로 재난 상황 시 투입되는 비용의 분산이 매우 클 수밖에 없다. 하지만 현재의 보험 상품 대부분은 앞서 언급한 세부 구분을 고려하지 않고 보험금을 책정한다. 이 말은 곧 주 구분에만 국한하여 보험금을 책정한다는 것인데, 이는 보험사, 그리고 풍수해 보험금의 대부분을 지원하는 정부의 손실로 이어질 수밖에 없다. 따라서 주 구분의 면적을 넘어서서 보다 자세한 보험 목적물의 분류를 제언한다.

첫째로 건물의 원재료를 고려한 구분을 제언한다. 예시를 들어보자면 단독주택의 경우에도 목조 주택, 콘크리트주택, 벽돌주택 등의 종류가 다양하고 종류별로 사용하는 자재가 상이하기 때문에 공사 비용과 보수 비용이 다를 수밖에 없다. 더불어 현재는 러시아, 우크라이나 사태로 인해 국제적인 원자재 수급 위기의 상황이다. 대한건설협회에 따르면 지난해에 비해 철근은 49.3%, 시멘트의 원재료인 유연탄은 256%, 원유는 66%가량 상승했다고 한다.¹⁶ 천연자원을 자력으로 공급하지 못하는 우리나라의 특성상 앞서 언급한 건축 자재들은 전적으로 수입에 의존해야 하는데, 이러한 재료들의 공급 리스크는 우리나라에서 충분히 사용 가능한 나무나 벽돌과 같은 재료들의 공급과 비교했을 때 상대적으로 높을 수밖에 없다. 현재 건물가액은 건물 건축 시 사용된 비용을 현재를 기준으로 계산하여 건물 가치를 다시 산정하나, 건축 기술이 발달되었고 앞서 언급한 바와 같이 원자재 시장에서의 대한민국의 위치를 고려했을 때 당시 사용된 비용이 아닌 건축에 사용된 주 원자재를 기준으로 계산을 해줄 것을 제언한다.

둘째로 노후 건물에 대한 추가적 지표 생성을 통한 인센티브 제도를 제언한다. 국토교통부에 의하면 2021년 기준 연면적 단위로 계산했을 때 10년 미만의 건축물은 25.3%, 10~20년 미만의 건축물은 27.0%, 20~30년 미만의 건축물은 28.4%, 30년 이상의 건축물은 17.2%이다.¹⁷ 노후화된 건축물일수록 재난 발생시 피해의 위험도가 높고 피해 발생시 복구 비용과 철거 비용이 더 많이 요구될 것이다. 또한 노후 건물의 경우에는 철거 작업의 도중에 붕괴 사고와 같이 추가적인 인명 피해의 가능성도 덜 노

¹⁶ 성재용, "올라도 너무 오른 원자재價"...공사비 올려라 무력시위에 건설사 '전전긍긍', 뉴데일리경제, 2022년 4월 20일, <https://biz.newdaily.co.kr/site/data/html/2022/04/20/2022042000058.html>

¹⁷ 210310(조간)전국 건축물 총 7275266동(건축정책과), 10쪽, 《용도별 노후 건축물 현황》

후화된 건물에 비해서 높을 수밖에 없다. 건물의 연식에 따라 건물 가액이 산정되기 때문에 건물의 연식이 높다면 앞서 언급했던 바와 같이 기후 재난 시의 리스크가 크므로 더 높은 보험금이 산정된다. 하지만 현재 우리나라의 경우 풍수해 보험과 화재보험의 가입률이 매우 저조하다. 2022년 기준 손해보험사들의 소상공인 대상의 풍수해보험 가입률은 7.1%에 그친다고 한다.¹⁸ 노후화된 건물일수록 오히려 재난 관련 보험 가입이 장려되어야 하나 건물 연식을 고려했을 때 노후도에 따라 높아지는 보험금이 가입률의 저해로 이어진 요인 중 하나라고 보았다. 따라서 예비 노후 건물, 즉 20~30년 미만의 건축물을 대상으로 건물 관리 정도에 따라 평가를 진행하여 인센티브를 줄 것을 제안한다. 승강기 점검, 화재 경보기 점검, 건축물 균열 확인과 같이 붕괴, 화재, 지진 등 재난 상황에 대한 대비를 얼마나 철저히 하는지 점검 빈도 및 점검 사항에 대한 개선 정도 등을 반영하여 노후 건축물의 관리도에 대해서 보험사가 자체적인 지표를 만들 수 있을 것이다. 이러한 지표를 기반으로 노후 건축물에 대한 안전 점검 및 재난 대비도에 대해서 등급을 매겨 차등적으로 보험금을 지급하는 제도를 제안한다. 이는 단순히 보험 가입률의 제고에만 국한된 것이 아니라 더 나아가 노후 건축물 자체의 보강으로 이어져 재난에 대한 노후 건축물의 잠재적인 위험도를 낮추는 데에 기여할 수 있을 것이다.

5.2 정부 대상 호우/태풍 보험

이상기후에 대한 보험산업의 대응의 두 번째로는 정부를 대상으로 한 시설별 호우/태풍 보험이다. 앞서 기술하였듯이, 재난으로 인한 피해금액의 규모에 비하여 정부의 지원금이 턱없이 부족한 실정이다. 하지만 동시에 한정된 재원에서 무조건적으로 지원금을 늘릴 수는 없는 노릇이기에, 이러한 어려움을 극복하기 위해선 정부(지자체) 역시 보험이 필요하다.

앞서 진행된 일련의 분석 결과들로부터, 호우 또는 태풍이 발생할 시 대부분의 피해액은 공공시설¹⁹에서 발생할 것이라는 걸 유추할 수 있었다. 이러한 점을 고려하여, 각 시도별 지자체마다 특정 공공시설에 대해 호우 또는 태풍에 의한 피해액이 일정 금액을 넘을 시 보험금을 지급하는 방식의 정부 대상 보험 상품을 만드는 것을 제안한다. 분석의 대상이었던 서울과 부산에 대해 차례로 예시를 들어보겠다. 첫째, 서울은 도로, 철도, 수도와 같은 기반시설들이 매우 높은 밀도로 형성되어 있는데, 이러한 이유로 호우 또는 태풍 발생 시 천문학적인 피해규모가 동반된다. 실제로, 올해 8월 집중호우로 인해 지반이 낮은 지역에서 수재민 수가 매우 컸고, 이 외에도 기록적인 차량침수 건수와 소상공인 피해가 야기됐다. 이번 집중호우로 인해 특히 서울의 배수시설이 참사의 원인으로 지목된 만큼, 서울의 경우 배수시설과 관련되어 계약을 맺는 것이 적절해 보인다. 둘째, 부산은 대한민국의 대표적인 해안 도시임에서 알 수 있듯이, 호우 또는 태풍 발생 시 파도의 위협을 누구보다 크게 받는다. 그렇기 때문에 부산은 항만과 그 주변시설들과 관련되어 계약을 맺는 것이 적절해 보인다.

¹⁸ 유희곤, "손보사 소상공인 풍수해 보험 가입 소극적...가입률 7.1%", 경향신문, 2022년 9월 18일, <https://m.khan.co.kr/economy/finance/article/202209181104001>

¹⁹ 공공시설 : 도로, 하천, 소하천, 수도, 항만, 어항, 학교, 철도, 수리시설, 사발, 순시설, 소규모, 해양

이 때, 지자체에서 지급해야 할 보험료는 직접적인 현금지급보다는, 정부가 법인세 감면을 통해 각 지자체의 보험료를 대납하는 형식이 된다면, 지자체 역시 직접적으로 지출하게 되는 재원이 없게 되고, 기업 역시 세금감면 혜택을 보면서 새로운 사업활동을 전개할 수 있다는 상부상조의 방법이 될 것이다. 2022년 9월을 기준으로 삼성화재의 3분기 당기순이익은 2827억²⁰으로, 해당 순이익에 유효세율 26.72%가 적용된다면 약 750억을 정부에서 법인세로 징수한다. 이 때, 만약 유효세율을 단순히 20% 정도로만 조정되어도, 약 200억가량의 유동성 현금이 확보된다. 이 중 책임준비금에 적립할 금액을 제외하고는 운용할 자산으로 활용할 수 있을 것으로 예상된다.

이러한 정부 대상 보험 상품도 건강보험에서의 정기적인 건강검진에 대한 보상과 같이 적극적인 피해 방지책 갈구 유도과 도덕적 해이 방지를 위해 적절한 보상 시스템이 필요하다. 만약 보험 계약 후 지자체 차원에서 특정 공공시설에 대해 호우/태풍 피해 방지를 위한 보수공사를 실시할 시, 심사를 통해 보험료 감면(법인세 추가징수) 혜택이 주어지고, 이렇게 추가로 징수된 법인세에 대해서 일부를 다시 지자체 지원금으로 제공하는 것이다. 이러한 보상 시스템이 체계적으로 갖춰진다면, 더 이상 매년 말에 정부 지원금을 탕진하기 위해 실시되는 소모적인 도로 보수공사보다는 국민의 생명과 안전을 지키는 선순환 시스템이 갖춰지게 될 것이라 예상된다. 지자체의 적극적인 피해 방지책 갈구의 한 예시로 남해군의 전동식 파도막이가 있다.

그럼 왜 굳이 정부를 대상으로 보험 상품을 만들어야 하는 것인가? 답은 간단하다. 보험 상품은 특허를 낼 수 없기 때문에 경쟁사에서 모방 상품이 나오기 쉬운 구조이다. 이로 인해 사실상 상품 차별화가 전무하다고 봐도 무방할 정도로 보험산업에서의 상품 차별화란 매우 어려운 과제이다. 지자체 별로 특정 공공시설에 대해 '독점계약'을 맺게 된다면, 경쟁사에서 나올 모방 상품에 대한 리스크가 매우 낮아지기에, 사업 안정성을 확보하기 수월할 것으로 예상된다.

이와 더불어, 세금감면 형식으로 징수하게 될 보험료에 한해서는 최소금액을 제외한 나머지는 기후변화에 적극적으로 대응하는 스타트업에 투자하거나 관련 상장기업의 주식을 매입하는 등의 자산 운용 전략을 취해야 한다. 또한, 관련 재난 발생 시 지자체를 도와 피해현장 뒷수습을 전문으로 하는 자회사를 설립하는 등 재난 발생 시 일방적으로 보험금 지불을 통해 손해만 보는 수익 구조를 재편해야 한다. 이러한 방식으로 기후변화에 능동적으로 대처할 수 있는 친환경적 자산운용전략을 갖추게 된다면, 보험사는 자신들의 위험관리 전문성을 살리면서 사회적 책임을 이행함과 동시에 수익성을 높일 수 있는 효과를 가질 수 있다 ('기후변화 위험과 보험사의 대응방안', 이승준).

²⁰ <https://www.google.com/finance/quote/000810:KRX?sa=X&ved=2ahUKewi5ut74vtf7AhWMk1YBH aZCBSEQ3ecFegQIJRAi>

참고문헌

- 김민욱, [한남노 피해집계] 인명피해 심각...12명 사망 또는 실종, 중앙일보, 2022년 9월 5일, <https://www.joongang.co.kr/article/25099521#home>
- 국토교통부, *어린이를 위한 대한민국 국가지도집*, 세종, 2020.
- 국토교통부, 210310(조간)전국 건축물 총 7275266동(건축정책과), 세종, 2021
- 성재용, "올라도 너무 오른 원자재價"...공사비 올려라 무력시위에 건설사 '전전공공', 뉴데일리경제, 2022년 4월 20일, <https://biz.newdaily.co.kr/site/data/html/2022/04/20/2022042000058.html>.
- 이승준. (2019). *기후변화 위험과 보험사의 대응방안*. 보험연구원: 경성문화사(2019).
- 이준형, "포항시, '한남노'로 인명피해 10명, 재산피해 1조7000억원 추정", 경북신문 2022년 9월 7일 <http://www.kbsm.net/news/view.php?idx=363128>
- 유희곤, "손보사 소상공인 풍수해 보험 가입 소극적...가입률 7.1%", 경향신문, 2022년 9월 18일, <https://m.khan.co.kr/economy/finance/article/202209181104001>.
- 행정안전부, 2020 재해연보, 세종, 2021.
- 행정안전부, "8월 서울경기 등 집중호우 피해 복구비 7905억원 투입", 대한민국 정책브리핑, 2022년 9월 8일, <https://www.korea.kr/news/policyNewsView.do?newsId=148905747>
- Akiba, Takuya, et al. "Optuna: A Next-Generation Hyperparameter Optimization Framework." ACM Digital Library, 25 July 2019, <https://dl.acm.org/doi/10.1145/3292500.3330701>.
- Eric Zivot, Jiahui Wang, *Modeling Financial Time Series with S-PLUS* (2003), NewYork : Springer New York
- F. T. Liu, K. M. Ting and Z. -H. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 413-422, doi: 10.1109/ 2008 Eighth IEEE International Conference on Data Mining
- Francesco Dottori. "Increased Human and Economic Losses from River Flooding with Anthropogenic Warming." Nature Climate Change 8 (2018): 781-786
- Priyanga Dilini Talagala, Rob J. Hyndman & Kate Smith-Miles (2021) *Anomaly Detection in High-Dimensional Data*, Journal of Computational and Graphical Statistics, 30:2, 360-374, DOI: [10.1080/10618600.2020.1807997](https://doi.org/10.1080/10618600.2020.1807997)
- UNDRR. *The Human Cost of Disaster: An Overview of the Last 20 years (2000-2019)*. 2020.
- WMO. *State of the Global Climate 2021*. 2022.