

Microsoft Fabric 을 이용한 한국 기후 변화 예측 분석

양소희¹, 이주리², 한지희³, 정홍주⁴
¹동덕여자대학교 세무회계학과 학부생
²동덕여자대학교 정보통계학과 학부생
³동덕여자대학교 데이터사이언스학과 학부생
⁴온스페이스 컨설팅사업부 이사

skg703@naver.com, jurilee54@gmail.com, jihee0965@naver.com, HongjuJung@hotmail.com

Predictive analytics for climate change in South Korea with Microsoft Fabric

So-Hee Yang¹, Ju-Ri Lee², Ji-Hee Han³, Hong-Ju Jung⁴
¹Dept. of Tax Accounting, Dongduk Women's University
²Dept. of Information and Statistics, Dongduk Women's University
³Dept. of Data Science, Dongduk Women's University
⁴Dept. of Consulting Service, ONSPACE

요 약

본 논문은 통합 데이터 분석 플랫폼인 Microsoft Fabric을 통해 기상청 데이터를 활용하여 한국의 지역별 기후 요소인 강수량, 기온을 예측하고자 한다. 앙상블 모델과 시계열 모델을 적용해 평균 제곱오차(MSE)를 기준으로 최적의 모델을 선정하였으며, 2024년부터 5년간의 기후 변화를 예측하였다. 연구 결과, 한국의 시간당 평균 최대 강수량과 평균 최고 기온이 지속적으로 상승함을 확인했다. 이는 기후 변화에 대응하기 위한 정책과 대비책 마련이 필요함을 시사한다.

1. 서론

지구온난화로 인한 폭우, 가뭄, 해수면 상승 등 이상 기후 현상이 전 세계적으로 빈번해지며, 한반도 역시 심각한 영향을 받고 있다. 2022년 서울에 기록적인 폭우로 큰 피해가 발생했으며, 기온 상승으로 농촌 지역의 생산량에도 변화가 나타나고 있다[1]. 환경부는 기후 변화에 대응하지 않을 경우 2100년까지 약 2,800조 원의 피해가 예상된다고 경고하고 있다[2]. 본 논문은 Microsoft Fabric을 이용하여 기계학습 기법 중 앙상블 및 시계열 모델을 통해 기후 데이터를 예측하여 기후 변화를 인식하고자 한다. 이를 통해 기후변화에 적극적으로 대응하기 위한 정책과 자발적 참여가 필요함을 알 수 있다.

2. 관련 연구

Facebook의 Prophet 모델을 사용하여 기후 데이터를 학습시키고 예측한 연구 결과, 예측 결과와 실제 관측 값 간의 상관관계 분석 결과 유의미한 유사성을 확인하였다[3]. 앙상블 모델로는 Random Forest, Gradient Boosting, XGBoost 등 있으며, 여러 모델을 결합하여 예측의 일반화 성능을 높일 수 있으며 과적합을 방지

한다[4]. Microsoft Fabric은 Data Factory, Data Engineering, Data Science, Power BI 등 구성요소를 통합한 환경으로 설계된 통합 데이터 분석 플랫폼이다[5].

3. 데이터 예측 분석 방법

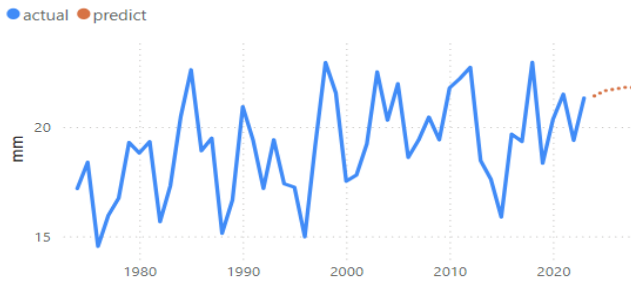
예측 분석을 위해 기상청 기상자료개방 포털의 데이터를 이용한다[6]. 제주 지역을 포함한 66개 지점에서 관측된 자료를 기반으로 1974년부터 2023년까지 데이터를 제공하는 지점만을 추출한 후 월간 평균, 최소, 최대값 데이터를 수집한다. 강수량, 기온을 중점으로 2024~2028년에 대한 예측 분석을 수행한다. 본 논문에서는 TDSP(Team Data Science Process)를[7] 기반으로 Microsoft Fabric 환경에서[5] 데이터 분석을 진행한다. 업무 이해 단계와 데이터 수집 및 이해 단계에서 필요 데이터를 수집하여 Notebook을 이용하여 전처리하고 Lakehouse에 데이터 테이블을 생성한다. Notebook에서 Python으로 여러 모델에 대한 평균제곱오차(MSE)로 모델을 평가하여 선정하고, 추가로 2023년 예측과 실제 데이터를 비교하여 모델을 평가한다. 전개 단계에서 추세 및 예측 분석 결과를 Power BI를 통해 시각화하여 기후변화 인사이트를 도출한다.

4. 예측 분석 결과

4-1 강수량

시간당 최대 강수량은 특정 지역에서 1시간 동안 내린 비의 최대량을 뜻하며 집중호우나 폭우의 강도를 평가하는 중요한 지표로 활용된다. 예측 변수로는 일 최대 강수량, 평균 습도, 평균 기온, 최고 기온, 최대 풍속을 이용하였다. 예측 모델은 최종적으로 가장 낮은 MSE를 보인 Random Forest를 선정하였다.

Average Maximum hourly rainfall



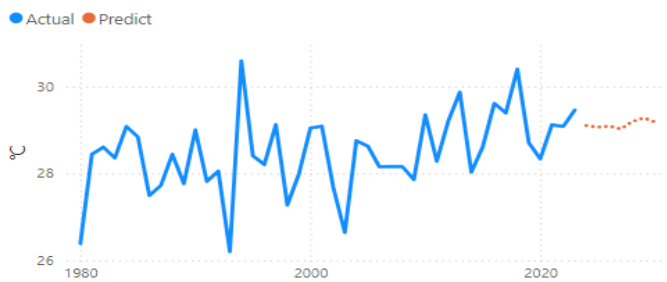
(그림 1) 시간당 최대 강수량

2024년부터 시간당 평균 최대 강수량은 지속적으로 증가할 것으로 예측되어, 집중호우와 폭우로 인한 자연재해 발생 위험이 커질 것으로 예상된다.

4-2 여름 기후

여름 기후 요소인 여름 평균 최고 기온과 여름 평균 강수량의 1980년부터 2023년 지역별 데이터를 LSTM, ARIMA, Prophet 모델을 사용하여 예측하였다. 지역별로 예측을 진행하였기 때문에, 지역에 따라 MSE가 가장 낮은 최적 모델을 선정하였다.

Average Summer Highest Temperature



(그림 2) 평균 여름 최고 기온

예측 결과, 1980년부터 2023년까지 여름 평균 최고 기온은 전반적으로 상승하였으며, 2024년부터 2028년 또한 지속적인 상승이 예상되어 1980년 대비 2028년에는 약 2.8°C 상승할 것으로 예측된다.

4-3 기온

1976년부터 2023년까지의 평균 월별 최고기온을 Prophet 모델을 사용하여 예측하였다. 지역별로 평균 절대 오차(MAE)가 가장 작은 모델을 선택한 결과, 1976년부터 2023년까지 평균 월별 최고기온이 전반적으로 증가하는 추세를 확인하였다.

Max. Average monthly Highest temperature



(그림 3) 평균 월별 최고 기온

2024년부터 2028년까지 지속적으로 30°C 이상의 기온을 유지할 것으로 보이며 2028년에는 1976년보다 평균 월별 최고기온이 약 1.2°C 상승할 것으로 예측된다.

5. 결론

본 논문의 예측 결과를 바탕으로 향후 5년 동안 시간당 평균 최대 강수량은 2028년에 1974년보다 4.61mm 증가할 것으로 예측된다. 여름 기후 또한 1980년 대비 2028년에 약 2.8°C 상승하고, 평균 월별 최고기온은 2028년에는 1976년보다 약 1.2°C 상승할 것으로 예측된다. 지속적으로 증가하는 강수량에 대비하여 방재성능 수립 시 참고 자료로 본 논문을 활용할 수 있다. 또한 여름 기후와 기온의 지속적인 상승은 지역 농업과 생태계가 파괴될 수 있으며, 지구 온난화로 인한 인명피해와 경제적 손실이 우려된다. 이에 따라 정부와 기업이 탄소중립 등 정책을 더 활성화하여 기후 위기의 사회적 인식을 제고하고 자발적 참여를 촉진하는 등 적극적인 대응이 필요하다.

본 논문은 과학기술정보통신부 대학디지털교육역량강화 사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

참고문헌

- [1]이덕배, 심교문, 기후변화가 농업에 미치는 영향과 대책, 한국농촌경제연구원 기타연구보고 농업전망 2011:농업·농촌과 농식품산업: 새로운 시장과 기회, 319-344, 2011
- [2]환경부 한국환경정책평가연구원, 우리나라 기후변화 영향에 따른 경제적 손실 2,800조원, 환경부, 2011, p.11
- [3]김준석, 김성희, 윤주상, 강재환, Prophet모델을 사용한 기상데이터 예측, 제1회 한국 인공지능 학술대회, 한국, 2020, p.110
- [4]최종석, 앙상블 러닝 기반 동적 가중치 할당 모델을 통한 보험금 예측 인공지능 연구, 한국정보전자통신기술학회논문지, 17, 4, 221-228, 2024
- [5]Microsoft Fabric, <https://www.microsoft.com/ko-kr/microsoft-fabric>
- [6]기상자료개방포털, <https://data.kma.go.kr/>
- [7]TDSP, <https://learn.microsoft.com/ko-kr/azure/architecture/data-science-process/overview>