KENTECH 6월 과제

## “**AI 기반 태양광 발전량 예측: 선형회귀 및 비선형회귀 모델 탐구** ”

전북과학고등학교

**3107 오근정**

**3304 경규헌**

| **Abstract**  **전력망 사업에서 수요와 공급 예측은 매우 중요한 과제이다. 특히, 탄소중립을 목표로 재생 에너지의 비율을 높이기 위해서는 변동성이 큰 재생 에너지 발전량을 최대한 정확하게 예측해야 하는데, 인공지능(AI)을 활용하면 발전량을 비교적 정확히 예측할 수 있다. 본 연구에서는 기계학습 알고리즘 중 하나인 선형회귀의 원리를 이해하고 이를 통해 태양광 발전량 예측 모델을 만드는 방법을 소개한다.** |
| --- |

전력망은 크게 공급, 전송, 소비의 세 부분으로 나뉜다. 발전소에서 생산되는 전력량을 예측하고 소비자의 수요에 맞춰 에너지를 효율적으로 공급하기 위해 AI가 활용된다. 에너지 분야뿐만 아니라 여러 산업에서도 수요와 공급을 예측하는 데 AI가 사용된다. 탄소중립 목표를 위해 재생 에너지 비율이 확대되면서 에너지 분야에서 AI의 필요성이 커지고 있다. 기존의 에너지원과 달리 재생 에너지는 여러 요인의 영향을 받아 발전량이 일정하지 않다. 발전량의 변동성은 에너지 공급에 불확실성을 초래하므로 재생 에너지 발전량을 정확히 예측하는 모델을 만드는 것이 매우 중요하다. 이번 연구에서는 AI의 한 분야인 기계 학습의 기본 작동 원리를 익히고, 이를 에너지 분야에 적용한다. 국내 태양광 발전원의 실제 데이터를 사용해 태양광 발전량을 예측하는 모델을 개발하고 이를 통해 연구를 진행한다.

배경이론

기계학습은 인공지능의 한 분야로, 컴퓨터가 데이터를 기반으로 학습하고 이를 통해 다양한 작업을 수행할 수 있도록 하는 기술이다. 기계학습은 크게 세 가지 주요 유형으로 나뉜다: 지도학습, 비지도학습, 강화학습이다.

### 1. 지도학습 (Supervised Learning)

### 지도학습은 입력 데이터와 이에 대응하는 정답(레이블)을 사용하여 학습하는 방법으로, 분류와 회귀로 나뉜다. 분류는 이메일이 스팸인지 아닌지 분류 하거나, 이미지를 보고 사진 속에 있는 물체를 분류하는 작업을 예로 들 수 있다. 회귀는 주택 가격 예측이나 주식 가격 예측 등 연속적인 값을 예측하는 작업을 예로 들 수 있다. 지도학습의 예로는 선형회귀, 로지스틱 회귀, 서포트 벡터 머신 등이 있다.

### 2. 비지도학습 (Unsupervised Learning)

비지도학습은 입력 데이터에 대한 정답(레이블) 없이 학습하는 방법으로, 클러스터링과 차원축소를 예로 들 수 있다. 비지도 학습의 알고리즘은 데이터의 구조를 이해하고, 유사한 데이터끼리 그룹화하거나 데이터의 숨겨진 패턴을 찾는다. 클러스터링은 고객을 유사한 그룹으로 분류하거나, 뉴스 기사를 주제별로 그룹화하는 작업을 예로 들 수 있다. 차원 축소는 고차원의 데이터를 저차원으로 변환하여 시각화하거나, 데이터의 중요한 특징을 추출하는 작업이 해당한다. 비지도 학습의 대표 알고리즘으로는 K-평균 클러스터링 , 주성분 분석 등이 있다.

### 3. 강화학습 (Reinforcement Learning)

강화학습은 에이전트가 환경과 상호작용하며 최적의 행동을 학습하는 방법이다. 에이전트는 상태를 관찰하고, 행동을 취하며, 그 결과로 보상을 받는다. 보상을 최대화하는 방향으로 학습한다. 게임 플레이, 로봇 제어 등에 강화학습이 이용된다. Q-러닝 등이 대표적인 강화학습 알고리즘에 해당된다.

기계학습은 이 외에도 반지도학습, 자기 지도 학습 등 다양한 하위 분야가 존재하며, 각 유형은 특정 문제를 해결하기 위해 고안되었다. 기계학습 알고리즘을 선택하고 적용하는 것은 문제의 성격, 데이터의 특성, 필요한 성능 등에 따라 달라진다.

4. 선형회귀

선형 회귀 문제의 해결을 위한 모델은 크게 가설 설정, 비용 함수, 경사 하강법의 세 단계로 나뉜다. 예제코드를 통해 선형회귀 실습을 진행헀다.

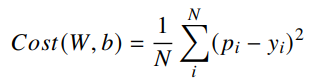
1) 가설 설정

첫 번째 단계는 가설을 세우는 것입니다. 선형 회귀 문제의 가설은 다음과 같은 형태를 갖는다. 여기서 W와 b는 초기에는 임의의 값으로 설정되고, 예제 코드에서는 W와 b를 0으로 초기화한다.



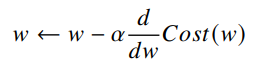
2) 비용 함수

현재 모델이 얼마나 잘못되었는지를 나타내는 지표가 필요한데, 이를 비용 함수라고 하며, 회귀 문제에서는 주로 평균 제곱 오차(Mean Square Error, MSE)를 사용합니다. 비용 함수는 다음과 같이 정의된다. 여기서 N은 데이터의 개수, p는 예측값 Wx+b, y는 실제 레이블 값이다.



3.)경사 하강법

모델이 가장 정답과 근접한 예측값을 가질 때는 비용 함수 W가 최솟값일 때이다. 이를 찾기 위해 경사 하강법 알고리즘이 사용된다. 경사 하강법은 비용 함수를 가중치에 대해 미분하여 최솟값을 찾아가는데, 수식으로 나타내면 다음과 같다.



탐구 제안

Q1. 태양광 발전량에 가장 큰 영향을 미치는 요인은 무엇인가?

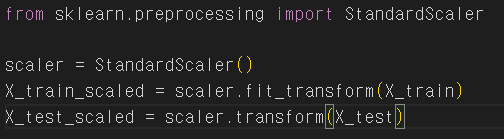
일사량(Solar radiation)은 태양으로부터 지구 표면에 도달하는 태양 에너지의 총량을 의미한다. 이는 태양광 패널이 실제로 흡수하여 전기 에너지로 변환할 수 있는 에너지의 양을 직접적으로 결정하게 된다. 높은 일사량은 더 많은 태양 에너지가 패널에 도달함을 의미하며, 이는 곧 더 높은 전기 생산량으로 이어진다. 태양광 패널은 일정한 효율성을 가지고 있으며, 더 많은 일사량을 받을수록 그 효율성에 따라 더 많은 전력을 생산할 수 있다. 따라서 일사량은 태양광 발전량의 핵심 결정 요인이다. 일조 시간은 하루 중 태양광 패널이 태양에 노출되는 시간을 의미한다. 일조 시간이 길수록 태양광 패널이 에너지를 흡수할 수 있는 시간이 많아지며, 이는 더 많은 전기를 생산할 수 있게 한다.태양광 발전은 태양이 떠 있는 동안에만 가능한다. 따라서 일조 시간이 길수록 태양광 발전 시스템이 작동하는 시간이 늘어나고, 이는 발전량의 증가로 이어진다.일사량과 일조 시간은 태양광 발전량에 가장 큰 영향을 미치는 요인이다. 일사량은 태양광 패널이 흡수할 수 있는 에너지의 양을 직접적으로 결정하고, 일조 시간은 패널이 태양에 노출되는 시간을 의미하기 때문에, 이 두 가지 요인은 태양광 발전의 효율성과 생산량을 결정하는 데 핵심적인 역할을 한다.

Q2. 예측의 속도나 정확도를 올리기 위해 어떤 방법을 사용할 수 있을까?

태양광 발전량 예측 모델의 예측 속도와 정확도를 높이기 위해 여러 가지 접근 방법을 고려할 수 있는데, 주요 방법은 데이터 전처리, 모델 개선, 하이퍼파라미터 튜닝, 피처 엔지니어링, 그리고 더 복잡한 모델 사용 등이 있다.

1)데이터 전처리

데이터에 결측지가 있는지 확인하고, 이를 적절히 처리한다. 또는 입력 피처를 스케일링하여 모델의 성능을 향상시킬 수 있다. 아래 코드는 StandardScaler를 사용하여 피처를 정규화한 코드이다.



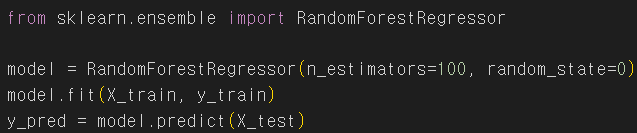
2) 피처 엔지니어링

태양광 발전량은 일사 시간 외에도 다른 피처들에 영향을 받을 수 있기 때문에, 이러한 피처들을 추가하여 모델을 개선할 수 있다.

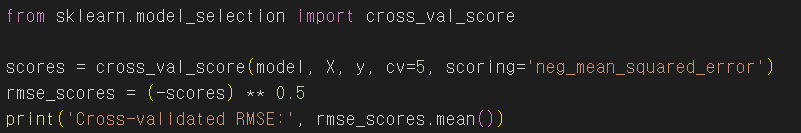
3)다양한 모델 시도

단순 선형 회귀 외에도 결정 트리, 랜덤 포레스트, 그라디언트 부스팅, 또는 딥러닝 모델을 활용하면 모델 성능을 향상시킬 수 있을 것이다.

ex) 랜덤 포레스트

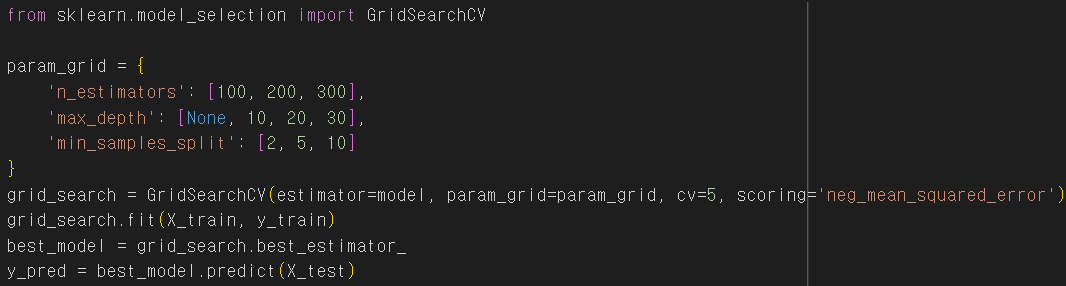


또는 train\_test\_split 대신 cross\_val\_score를 사용하여 더 안정적인 성능 평가를 할 수 있다.



4) 하이퍼파라미터 튜닝

‘GridSearchCV’를 사용하여 최적의 하이퍼파라미터를 찾을 수 있다.



Q3. 제시한 방법을 활용하여 모델을 제작해보고, 결과를 분석한다.

단순 선형 회귀 대신 Light GBM를 이용해 학습을 진행했고, 회귀 모델 성능 평가 지표로 이용되는 RMSE를 이용했다.

-단순 선형 회귀 RMSE



-LightGBM RMSE



Q4. x, y가 비선형 관계를 가질 때 , 회귀분석에 대한 아이디어를 제시한다.

비선형 관계를 가지는 x와 y의 경우, 선형 회귀 모델은 적합하지 않을 수 있다. 이 경우에는 다음과 같은 아이디어를 고려해 볼 수 있다.

1)다항 회귀

다항회귀는 x와 y의 비선형 관계를 설명하기 위해 x의 다항식 형태로 모델을 확장하는 방법이다. 예를 들어, x를 x, x^2, x^3 등의 다항식으로 변환하여 선형 회귀 모델을 적용할 수 있다.

2) 비선형 변환

x나 y를 로그 변환, 지수 변환 등의 비선형 함수를 적용하여 선형 관계를 강화할 수 있다. 이를 통해 데이터의 분포를 변환하거나 특정 패턴을 효과적으로 확인할 수 있을 것이다.

3)비선형 회귀 모델

의사 결정 트리 기반의 DecisionTreeRegressor나 RandomForestRegressor과 같이 비선형 관계를 직접 모델링할 수 있는 다양한 비선형 회귀 모델을 이용한다.