


폰데로사 소나무 생육환경 연구

- 루즈벨트 국유림

작성자 : 이소림

< 목 차 >

1. 데이터 선정 이유 및 문제정의
2. 데이터 탐색 및 전처리
3. 모델 학습 및 성능확인
4. 모델 해석
5. 인사이트 도출 및 한계점



1. 데이터 선정 이유 및 문제정의

1. 데이터 선정 이유 및 문제정의



구상나무(Korea Fir)

- 2013년 한라산 집단고사 이후 현재 지리산에서 집단 고사 빠르게 진행 중
- 집단 고사 원인 : 기후 변화로 추정 (적설량 부족으로 인한 고산지역의 건조현상)
- 구상나무숲 회복을 위한 계획 추진 중 (금원산 복원시험지)

1. 데이터 선정 이유 및 문제정의

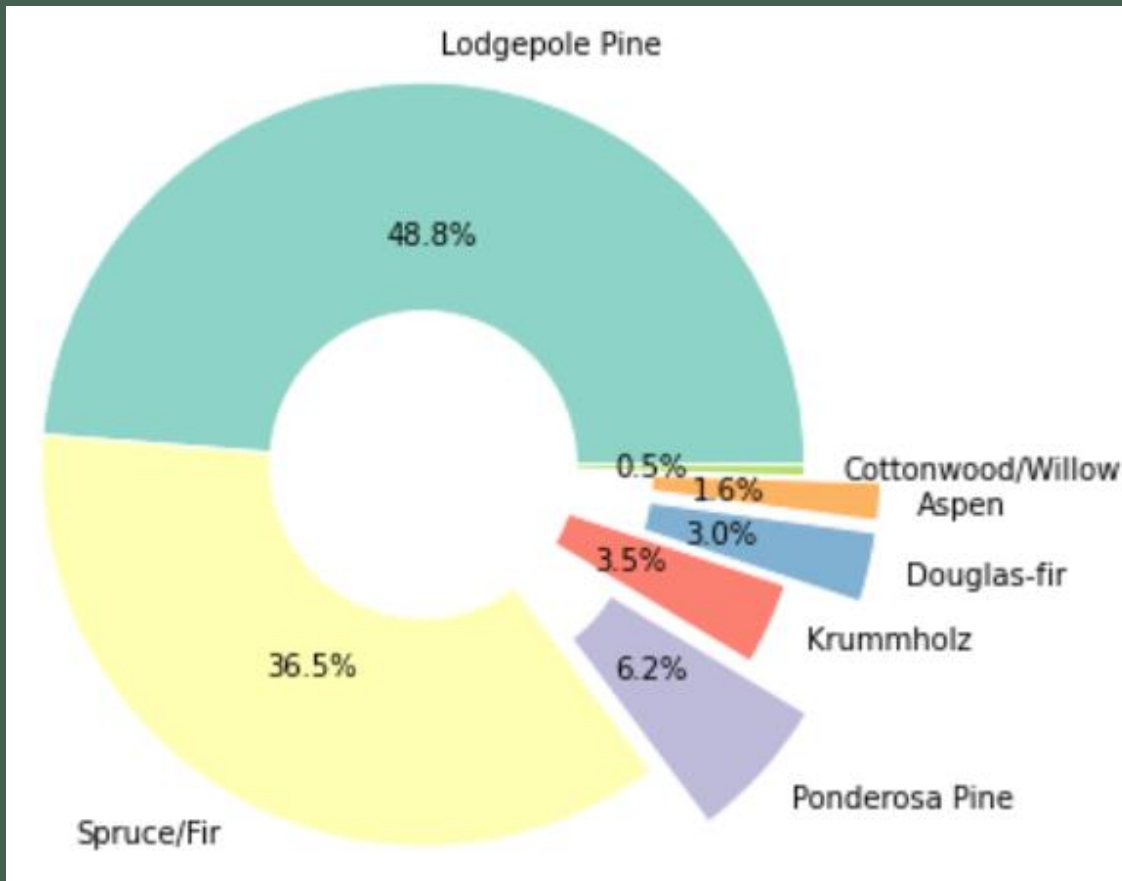
문제 정의

- 기후 변화 등으로 인하여 자연스럽게 자라던 나무들에게 인위적인 도움이 점점 필요해지고 있는 상황
- 계획 수립 및 의사결정을 위해서는 각 나무의 생육 환경에 대한 선행연구가 필요

데이터 선정

- 콜로라도 루즈벨트 국유림 4개 지역에서 생장하고 있는 나무 (Rawah/ Neota / Comanche Peak/ Cache la Poudre Wilderness Area)
- 30m x 30m 면적을 기준으로 나무 종류 식별
- 데이터 관측치 : 581,012개
- 데이터 출처 : 캐글(Forest Cover Type Dataset)

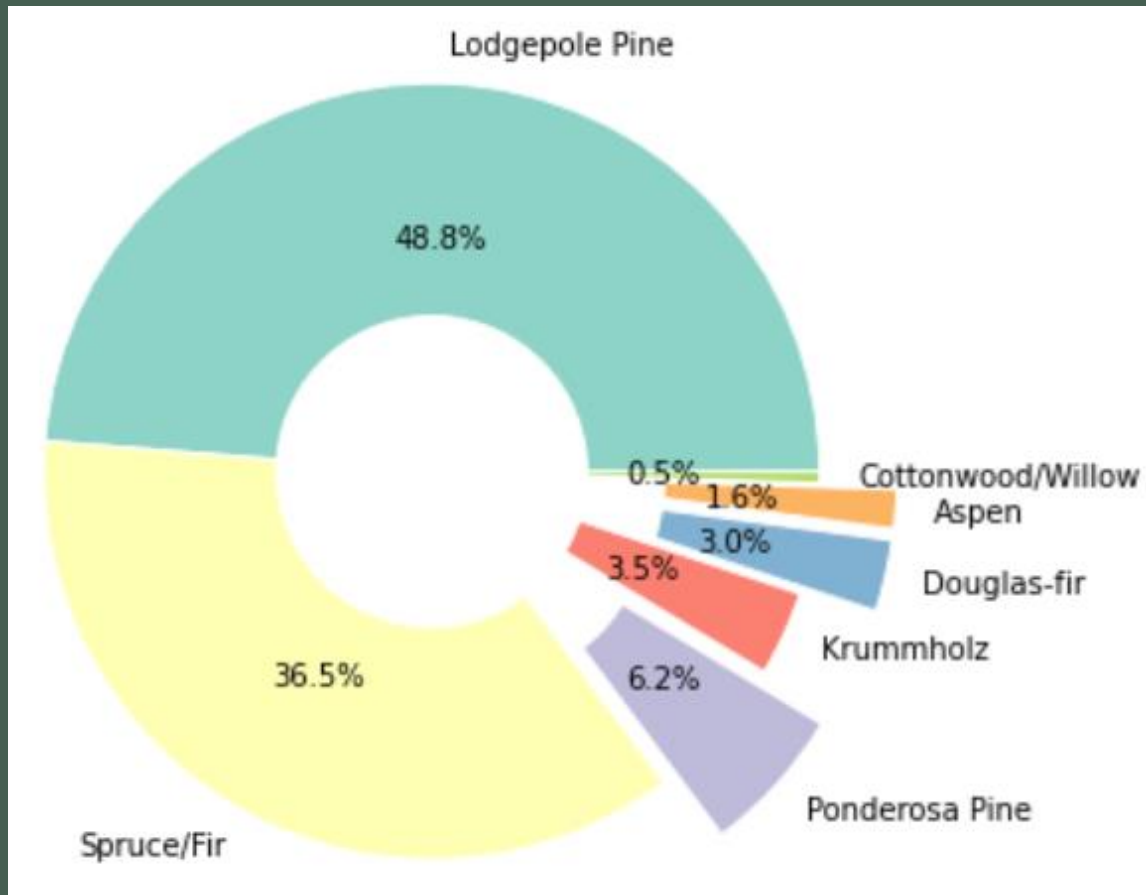
1. 데이터 선정 이유 및 문제정의



루즈벨트 국유림 현재상황

1. Lodgepole Pine(로지폴 소나무)
2. Spruce/Fir(가문비나무/전나무)
3. Ponderosa Pine(폰데로사 소나무)
4. Krummholz(작은 나무 숲)
5. Douglas-fir(미송)
6. Aspen(사시나무)
7. Cottonwood/Willow(미루나무/버드나무)

1. 데이터 선정 이유 및 문제정의



숲의 다양성

- 로지폴 소나무 + 가문비나무/전나무

⇒ 약 85% 차지

- 나무들의 생육환경 연구를 통해서 숲의 다양성을 유지할 수 있지 않을까?
- 이번 연구에서는 폰데로사 소나무의 생육 환경에 대한 연구를 진행

1. 데이터 선정 이유 및 문제정의

- 분류 문제

- 타겟 : 폰데로사 소나무(Yes, 1), 폰데로사 소나무 이외의 나무(No, 0)

- 가설설정

- 가설1. 폰데로사 소나무가 잘 자라는 높이(고도)가 있을 것이다.
- 가설2. 폰데로사 소나무가 잘 자라는 토양이 있을 것이다.



2. 데이터 탐색 및 전처리

2. 데이터 탐색 및 전처리

- 원본 데이터(Data Fields)

- **Elevation** - Elevation in meters
- **Aspect** - Aspect in degrees azimuth
- **Slope** - Slope in degrees
- **Horizontal_Distance_To_Hydrology** - Horz Dist to nearest surface water features
- **Vertical_Distance_To_Hydrology** - Vert Dist to nearest surface water features
- **Horizontal_Distance_To_Roadways** - Horz Dist to nearest roadway
- **Hillshade_9am** (0 to 255 index) - Hillshade index at 9am, summer solstice
- **Hillshade_Noon** (0 to 255 index) - Hillshade index at noon, summer solstice
- **Hillshade_3pm** (0 to 255 index) - Hillshade index at 3pm, summer solstice
- **Horizontal_Distance_To_Fire_Points** - Horz Dist to nearest wildfire ignition points
- **Wilderness_Area** (4 binary columns, 0 = absence or 1 = presence) - Wilderness area designation
- **Soil_Type** (40 binary columns, 0 = absence or 1 = presence) - Soil Type designation
- **Cover_Type** (7 types, integers 1 to 7) - Forest Cover Type designation

2. 데이터 탐색 및 전처리

- 원본 데이터(Data Fields)

- **The wilderness areas**

- 1 - Rawah Wilderness Area
 - 2 - Neota Wilderness Area
 - 3 - Comanche Peak Wilderness Area
 - 4 - Cache la Poudre Wilderness Area

- **Cover_Type**

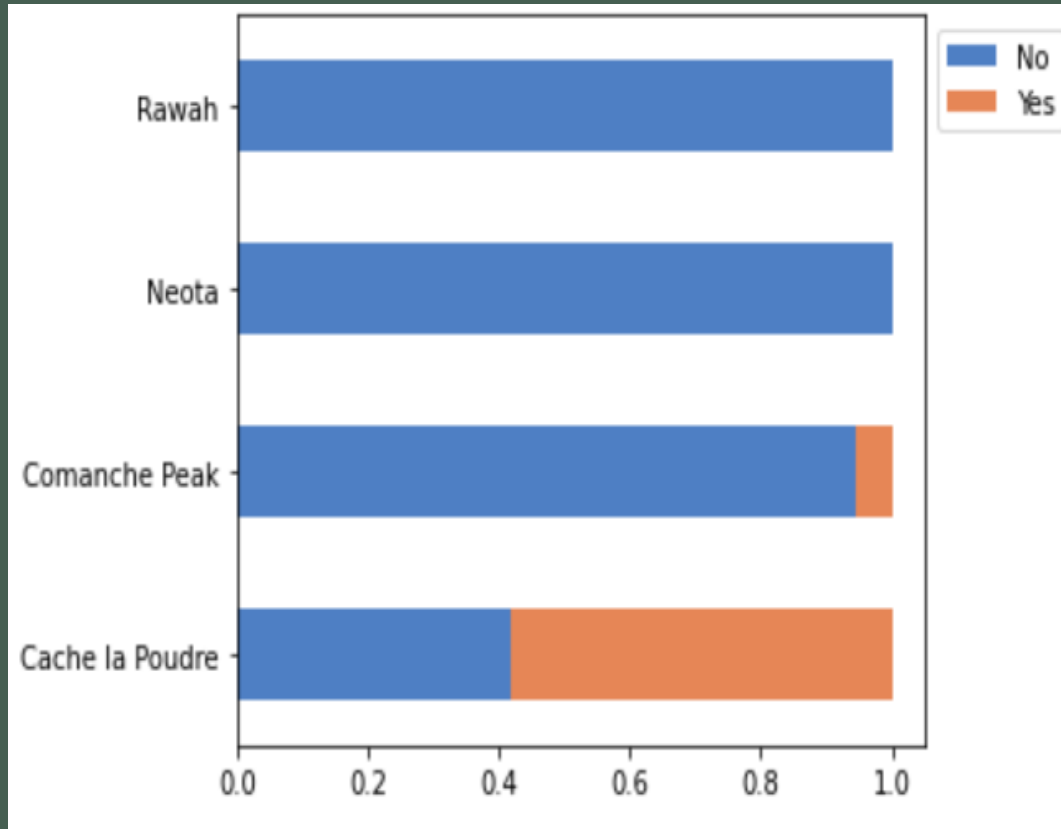
- 1 - Spruce/Fir
 - 2 - Lodgepole Pine
 - 3 - Ponderosa Pine
 - 4 - Cottonwood/Willow
 - 5 - Aspen
 - 6 - Douglas-fir
 - 7 - Krummholz

2. 데이터 탐색 및 전처리

• Soil_Type

- 1 Cathedral family - Rock outcrop complex, extremely stony.
- 2 Vanet - Ratake families complex, very stony.
- 3 Haploborolis - Rock outcrop complex, rubbly.
- 4 Ratake family - Rock outcrop complex, rubbly.
- 5 Vanet family - Rock outcrop complex complex, rubbly.
- 6 Vanet - Wetmore families - Rock outcrop complex, stony.
- 7 Gothic family.
- 8 Supervisor - Limber families complex.
- 9 Troutville family, very stony.
- 10 Bullwark - Catamount families - Rock outcrop complex, rubbly.
- 11 Bullwark - Catamount families - Rock land complex, rubbly.
- 12 Legault family - Rock land complex, stony.
- 13 Catamount family - Rock land - Bullwark family complex, rubbly.
- 14 Pachic Argiborolis - Aquolis complex.
- 15 unspecified in the USFS Soil and ELU Survey.
- 16 Cryaquolis - Cryoborolis complex.
- 17 Gateview family - Cryaquolis complex.
- 18 Rogert family, very stony.
- 19 Typic Cryaquolis - Borohemists complex.
- 20 Typic Cryaquepts - Typic Cryaquolls complex.**Type**
- 21 Typic Cryaquolls - Leighcan family, till substratum complex.
- 22 Leighcan family, till substratum, extremely bouldery.
- 23 Leighcan family, till substratum - Typic Cryaquolls complex.
- 24 Leighcan family, extremely stony.
- 25 Leighcan family, warm, extremely stony.
- 26 Granile - Catamount families complex, very stony.
- 27 Leighcan family, warm - Rock outcrop complex, extremely stony.
- 28 Leighcan family - Rock outcrop complex, extremely stony.
- 29 Como - Legault families complex, extremely stony.
- 30 Como family - Rock land - Legault family complex, extremely stony.
- 31 Leighcan - Catamount families complex, extremely stony.
- 32 Catamount family - Rock outcrop - Leighcan family complex, extremely stony.
- 33 Leighcan - Catamount families - Rock outcrop complex, extremely stony.
- 34 Cryorthents - Rock land complex, extremely stony.
- 35 Cryumbrepts - Rock outcrop - Cryaquepts complex.
- 36 Bross family - Rock land - Cryumbrepts complex, extremely stony.
- 37 Rock outcrop - Cryumbrepts - Cryorthents complex, extremely stony.
- 38 Leighcan - Moran families - Cryaquolls complex, extremely stony.
- 39 Moran family - Cryorthents - Leighcan family complex, extremely stony.
- 40 Moran family - Cryorthents - Rock land complex, extremely stony.

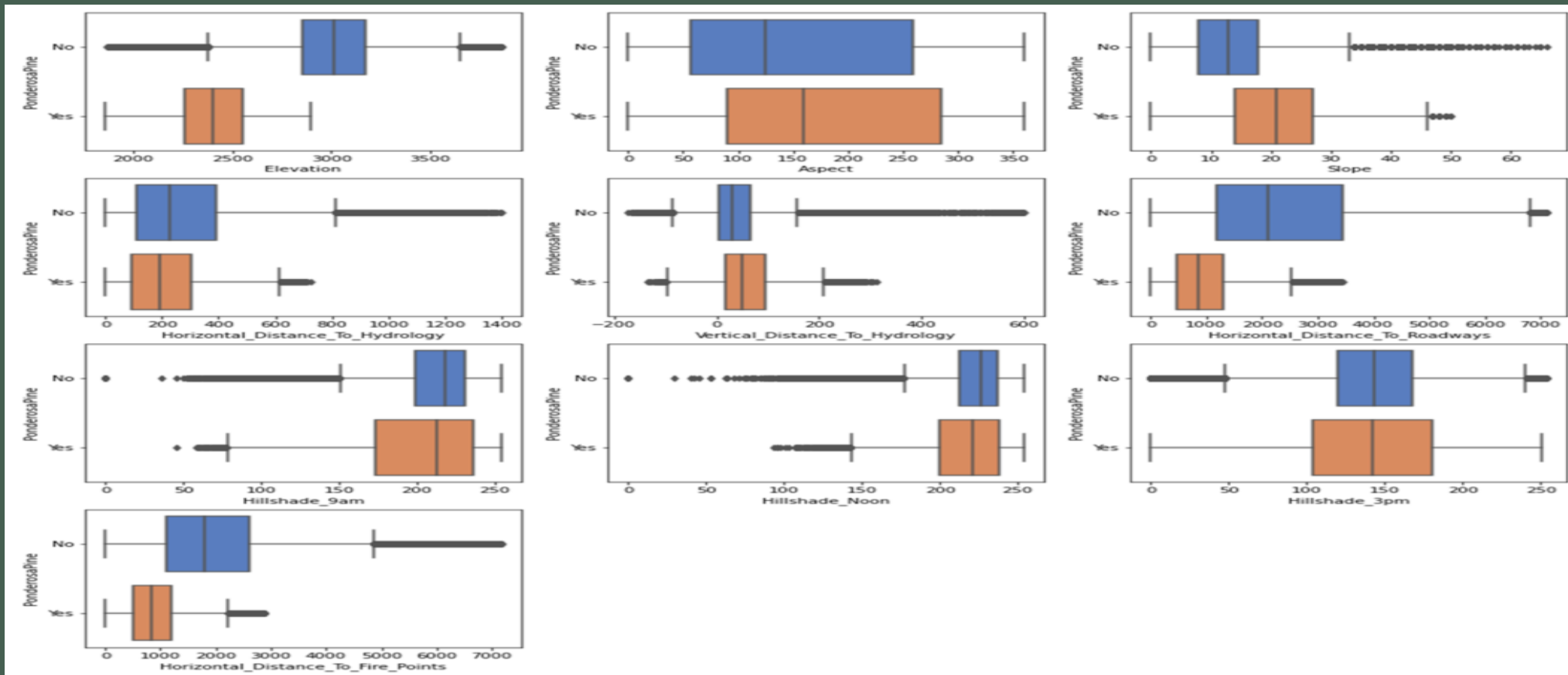
2. 데이터 탐색 및 전처리



표본 선정

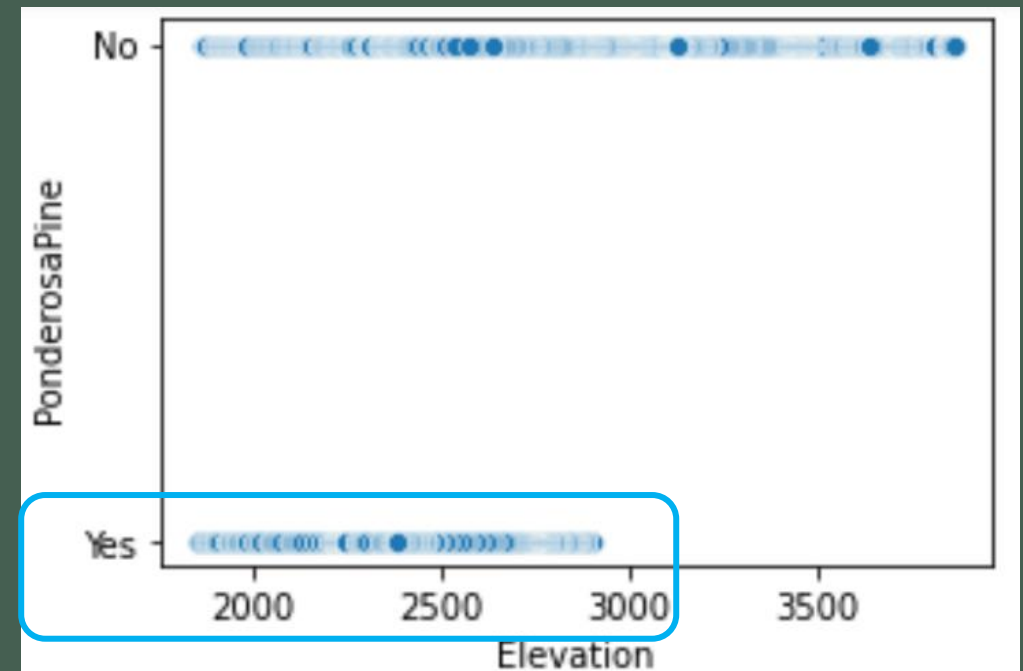
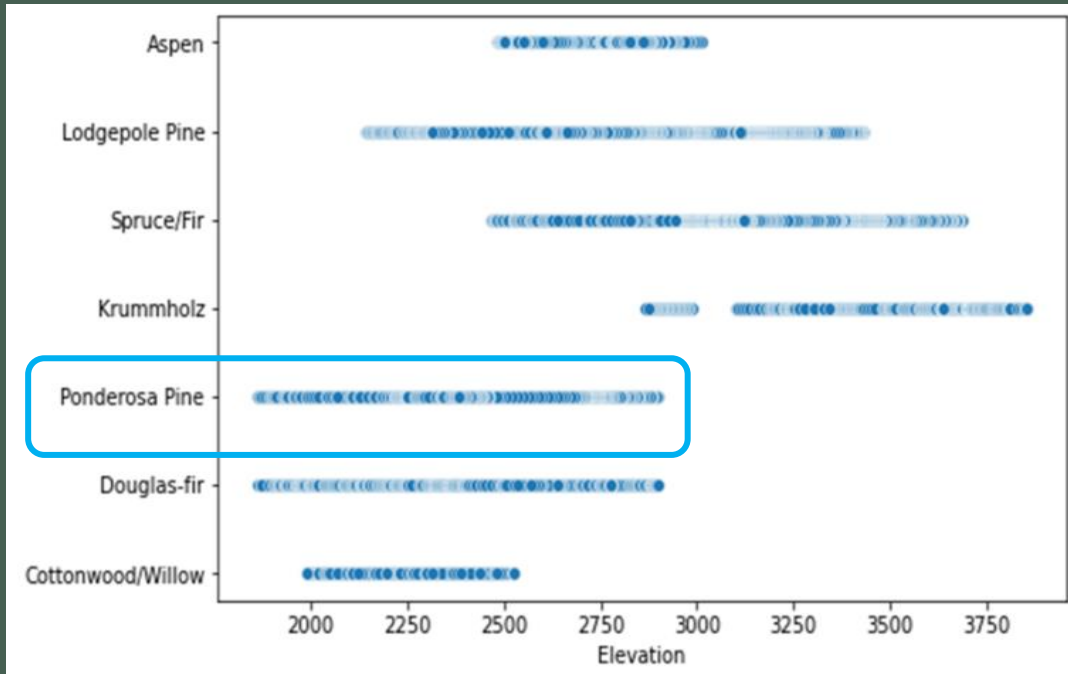
- 4개 지역 : Rawah/ Neota / Comanche Peak/ Cache la Poudre Wilderness Area
 - ⇒ 2개 지역 : Comanche Peak/ Cache la Poudre Wilderness Area
- 데이터 관측치 : 581,012개(원본)
 - ⇒ 데이터 관측치 : 290,332개(머신러닝)
 - ⇒ 중복치, 결측치 없음
 - ⇒ 이상치 제거하지 않음

2. 데이터 탐색 및 전처리



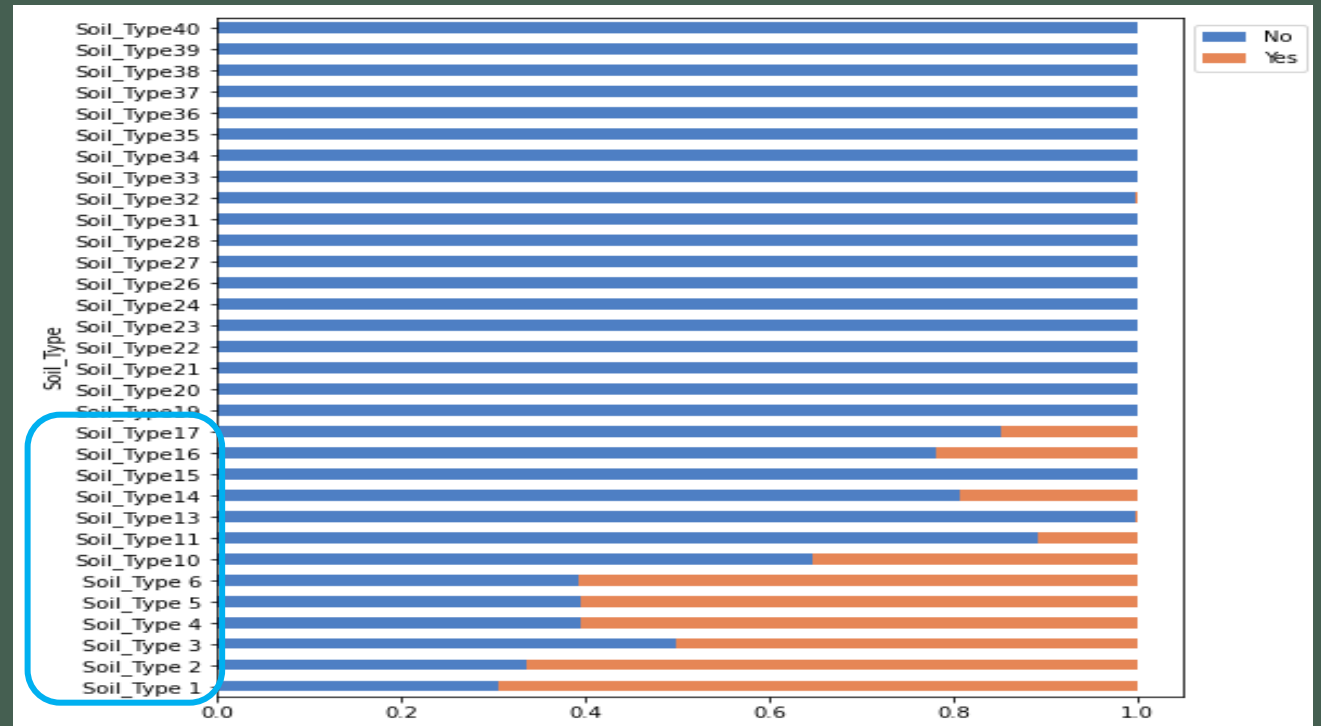
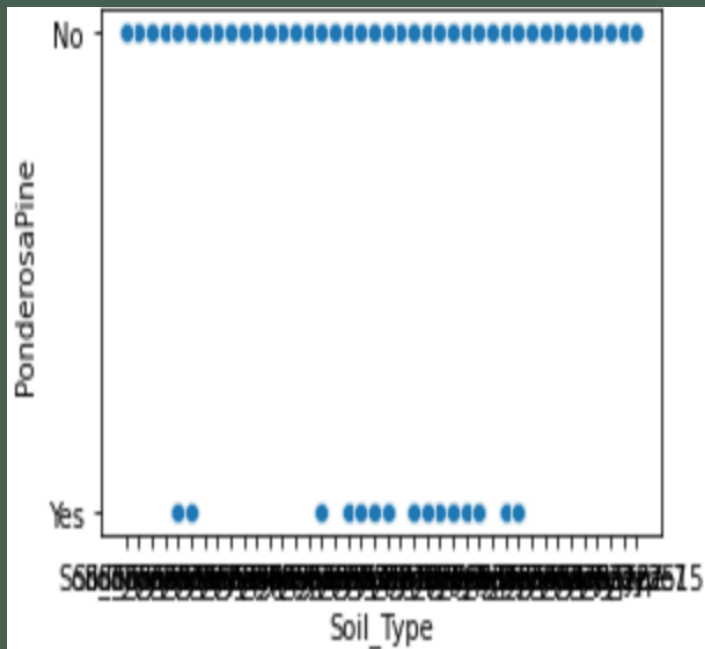
2. 데이터 탐색 및 전처리

가설1. 폰데로사 소나무가 잘 자라는 높이(고도)가 있을 것이다.



2. 데이터 탐색 및 전처리

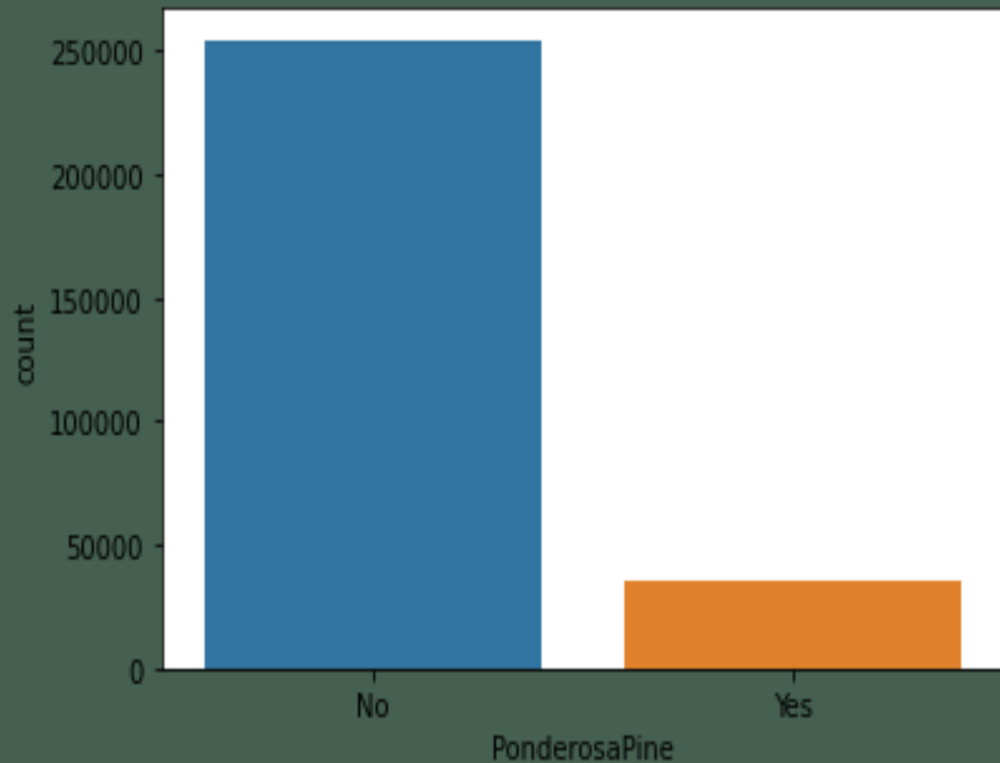
가설2. 폰데로사 소나무가 잘 자라는 토양이 있을 것이다.





3. 모델 학습 및 성능확인

3. 모델 학습 및 성능확인



타겟 데이터 비율

- Yes : 0.123, No : 0.877

⇒ 불균형 데이터

⇒ 불균형 데이터를 고려하여 모델 학습
(`class_weight`, `scale_pos_weight`)

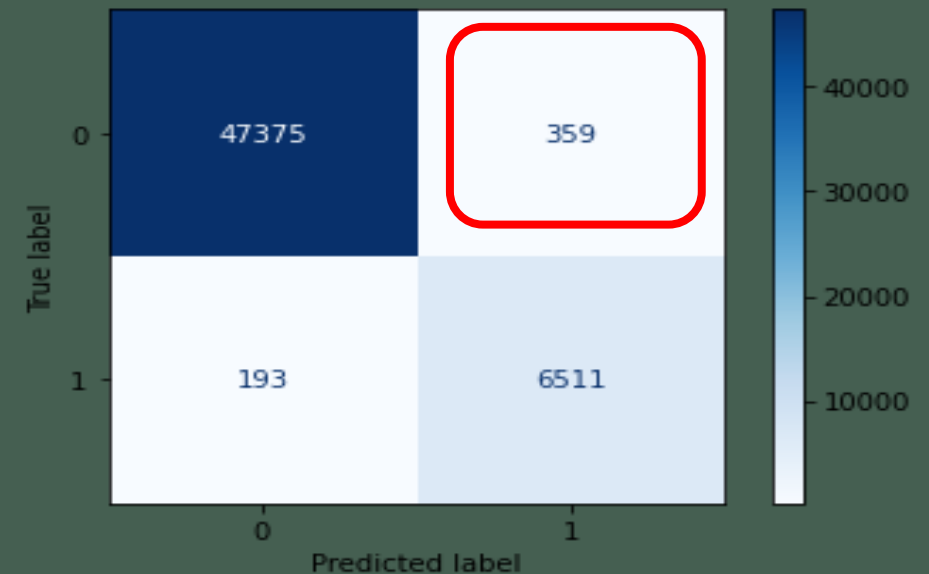
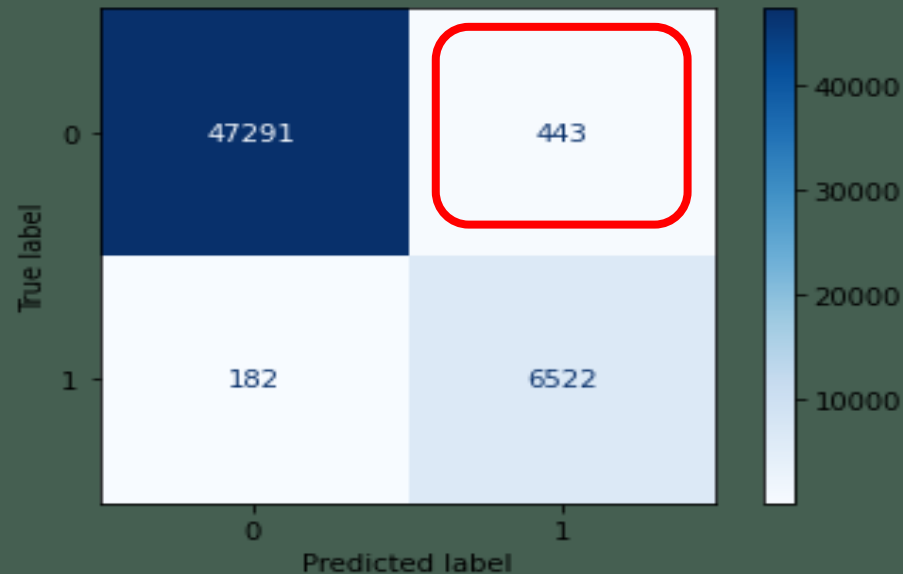
3. 모델 학습 및 성능확인

- 데이터 불균형 고려한 기초적 모델 분석 실시(max_depth =20)
- 여러 평가지표를 우수성을 나타낸 Gradient Boosting 모델 선택

분류 평가지표	기준 모델	Decision Tree	Random Forest	Gradient Boosting
Accuracy	0.88	0.97	0.98	0.99
Precision	0	0.87	0.90	0.94
Recall	0	0.93	0.96	0.97
F1-score	0	0.90	0.93	0.95

3. 모델 학습 및 성능확인

- 숲 조성 : 많은 시간과 비용 발생
- Yes라고 예측했는데 실제로는 No인 경우가 치명적인 오류라고 판단



3. 모델 학습 및 성능확인

- Gradient Boosting 모델의 성능 개선 (베이지안 튜닝)
- 일반화 성능확인

분류 평가지표	개선 전 (검증 데이터)	개선 후 (검증 데이터)	일반화 (테스트 데이터)
Accuracy	0.99	0.99	0.99
Precision	0.94	0.95	0.96
Recall	0.97	0.97	0.98
F1-score	0.95	0.96	0.97



4. 모델 해석

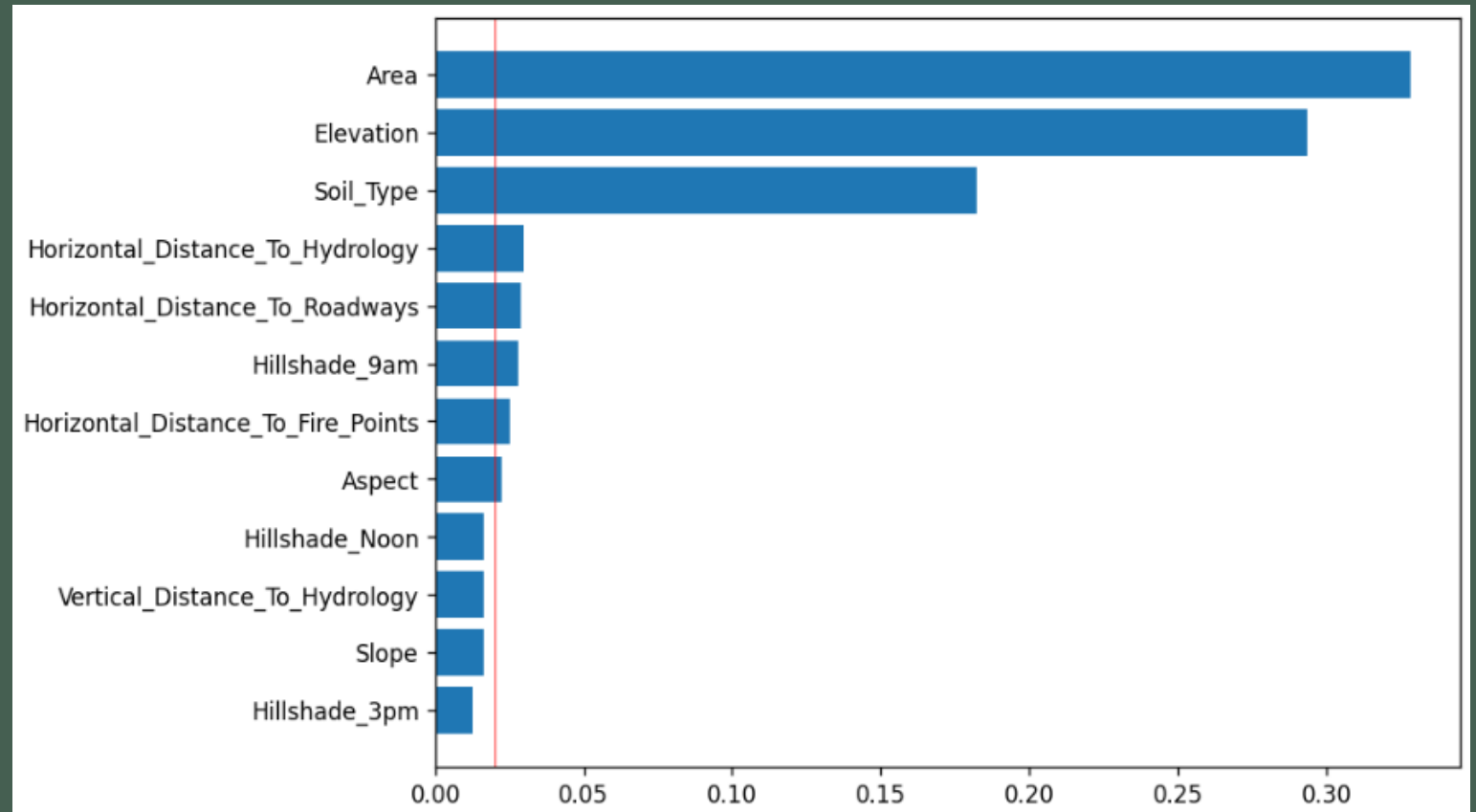
4. 모델 해석

- 폰데로사 소나무
생육환경에 대한
특성 중요도

⇒ Area

⇒ Elevation

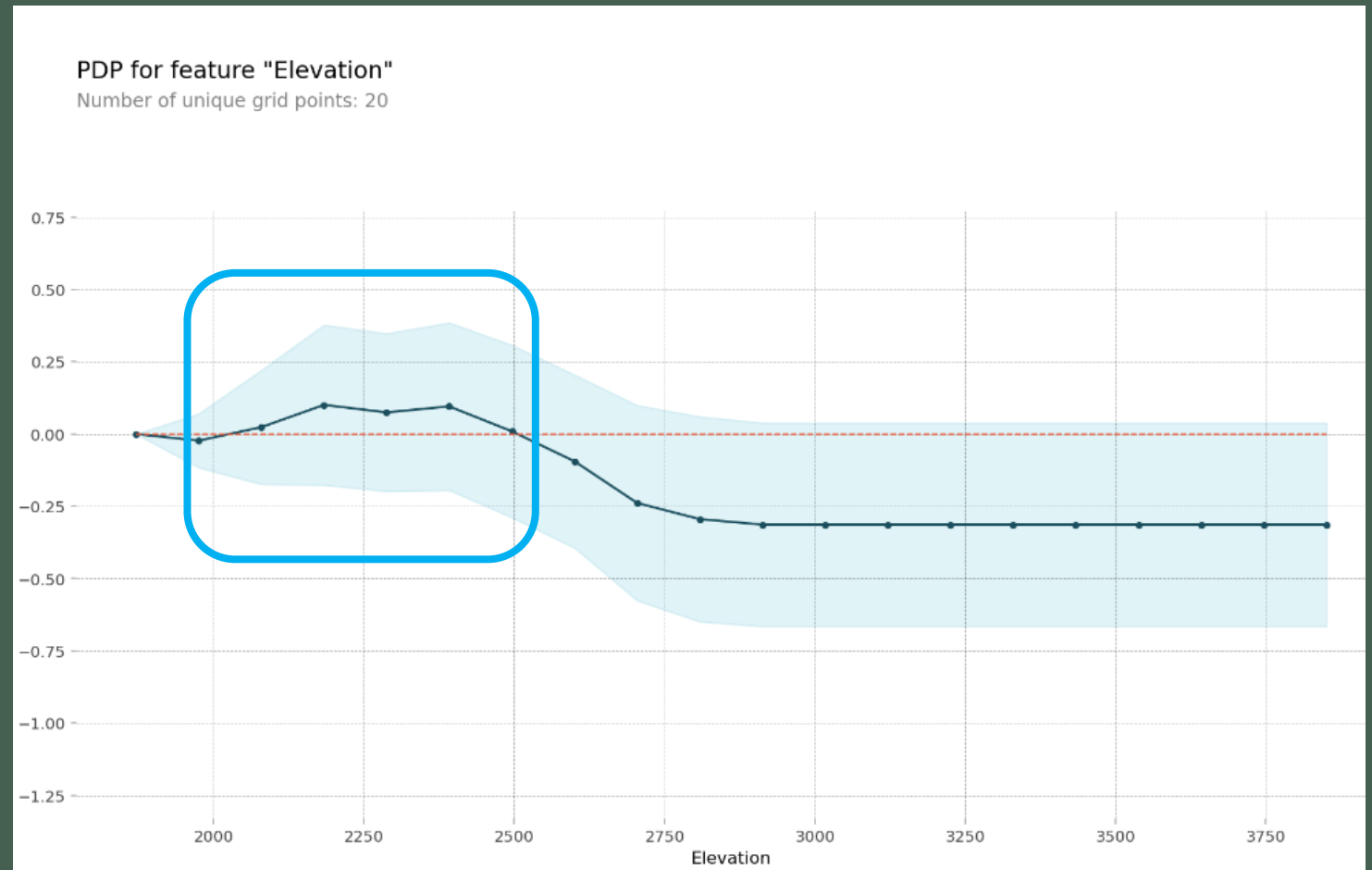
⇒ Soil_Type



4. 모델 해석

- 가설1. 폰데로사 소나무가 잘 자라는 높이(고도)가 있을 것이다.

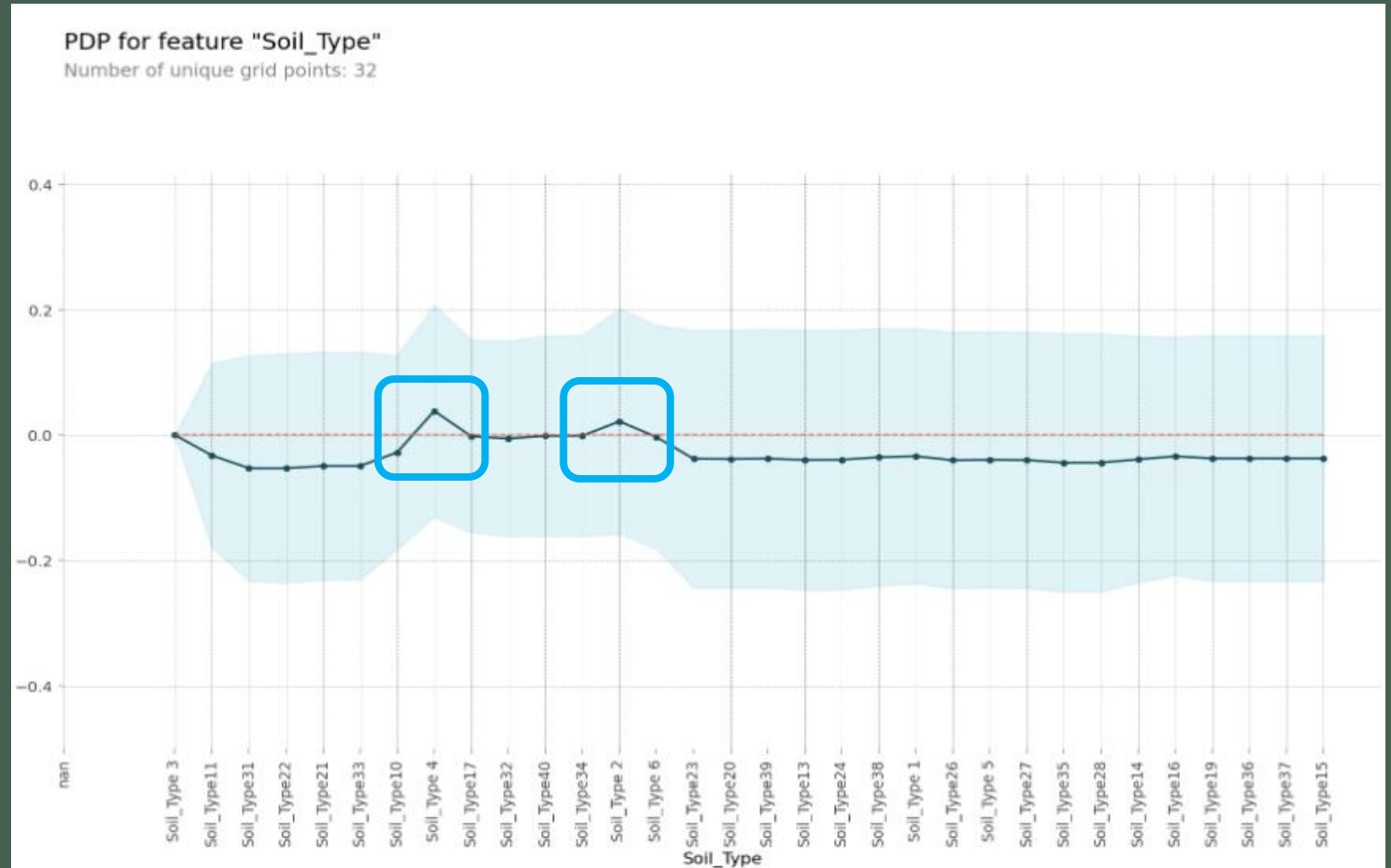
⇒ 2100m ~ 2500m
까지의 높이




4. 모델 해석

- 가설2. 폰데로사 소나무가 잘 자라는 토양이 있을 것이다.

⇒ Type 4(Ratake family), Type2(Vanet - Ratake families)의 토양





5. 인사이트 도출 및 후속 연구

5. 인사이트 도출 및 한계점

- (Q) 나무들의 생육환경 연구를 통해서 숲의 다양성을 유지할 수 있지 않을까?
- 폰데로사 소나무의 생육환경에 대한 연구를 진행
 - ❖ 루즈벨트 국유림 : Comanche Peak/ Cache la Poudre Wilderness Area
 - ❖ 2000m~2800m이하까지 생존가능
 - ⇒ 2100m ~ 2500m까지의 높이를 가장 선호
 - ❖ Soil_Type(1, 2, 3, 4, 5, 6, 10, 11, 12, 14, 16, 17)에서 생존 가능
 - ⇒ Type2(Vanet - Ratake families), Type 4(Ratake family), 의 토양 선호

5. 인사이트 도출 및 한계점

- 프로젝트의 한계점

- ❖ 머신 러닝의 성능이 높음 : 정보 누수 여부 확인 필요
- ❖ 해외 데이터이므로 우리나라 구상나무에 적용하기에는 한계가 있음
- ❖ 미래의 생육환경까지 고려하여 산림조성 계획 수립 및 의사결정에 활용하기 위해서는 기후 변화 예측 데이터까지 고려가 필요



감사합니다.