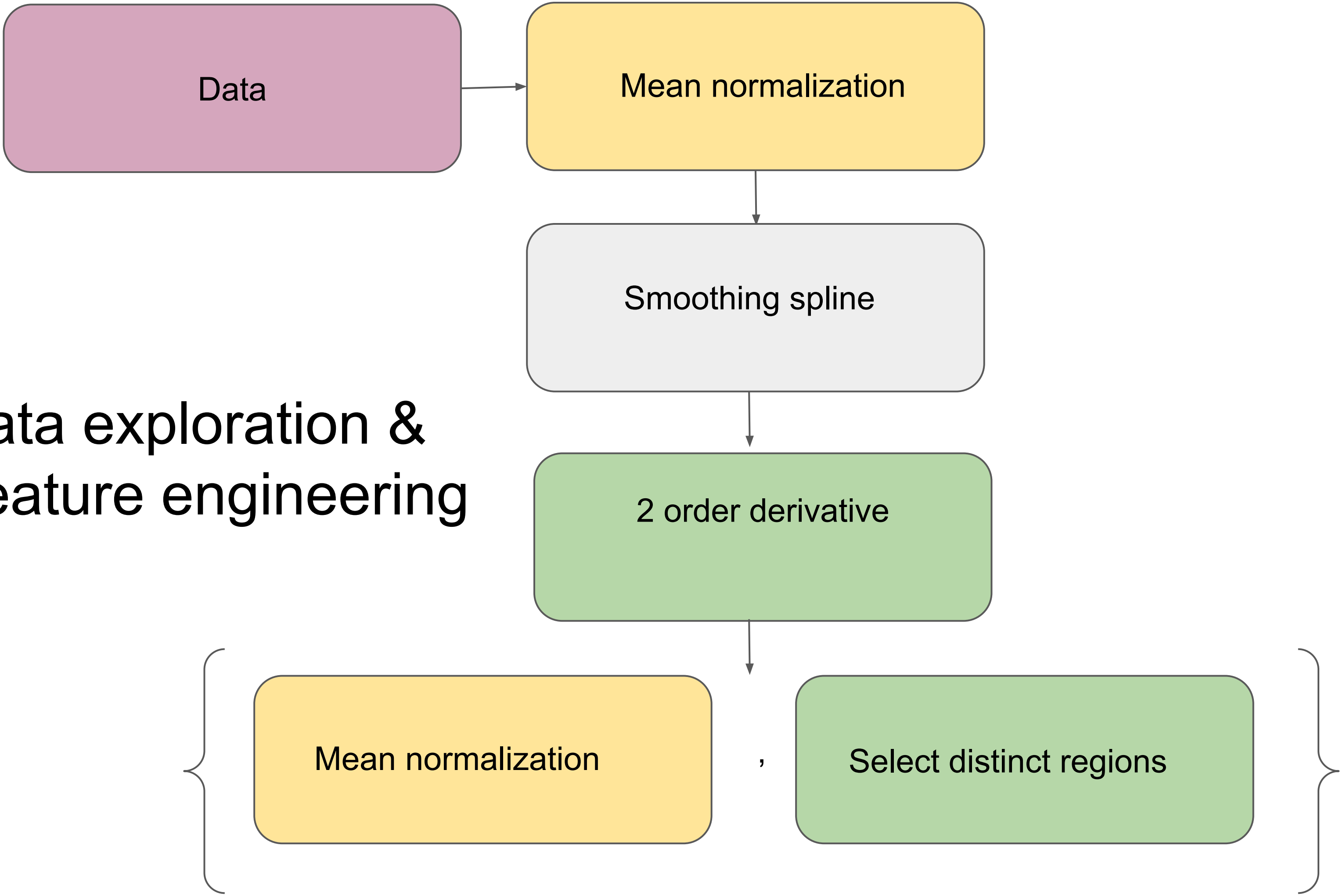
Wood classification

Flemming Morsch
Lys Sanz Moreta
Zhi Ye

Data exploration & Feature engineering



Data

Mean normalization

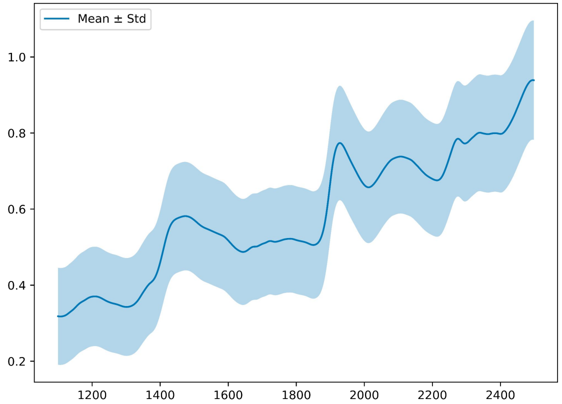
Smoothing spline

2 order derivative

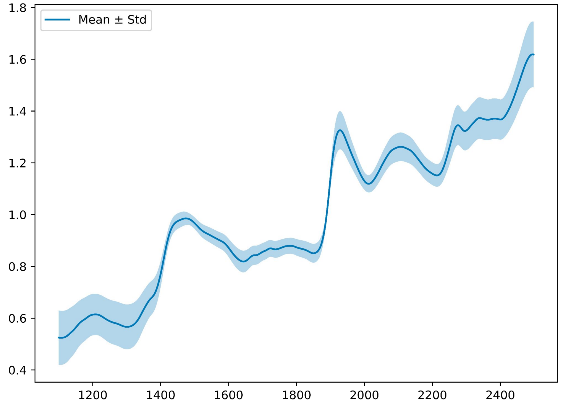
Mean normalization , Select distinct regions

Data Exploration

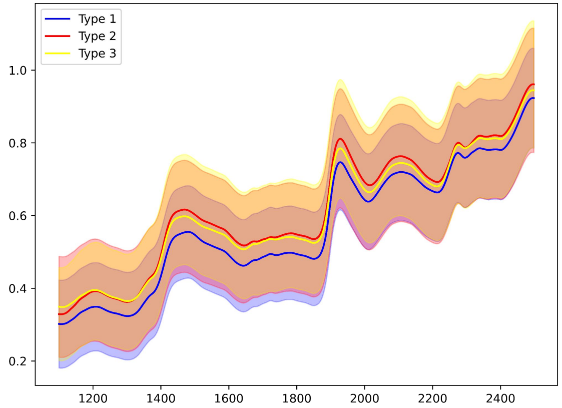
Overall

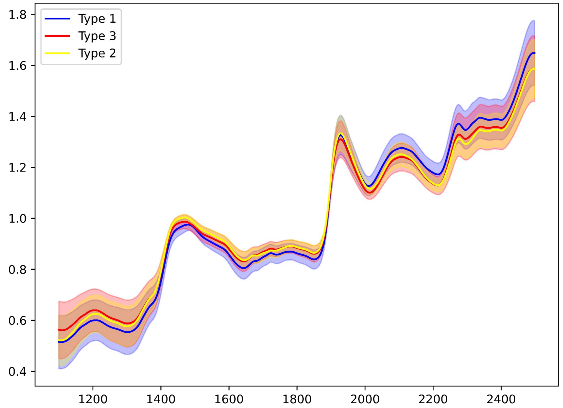


Mean Normalization

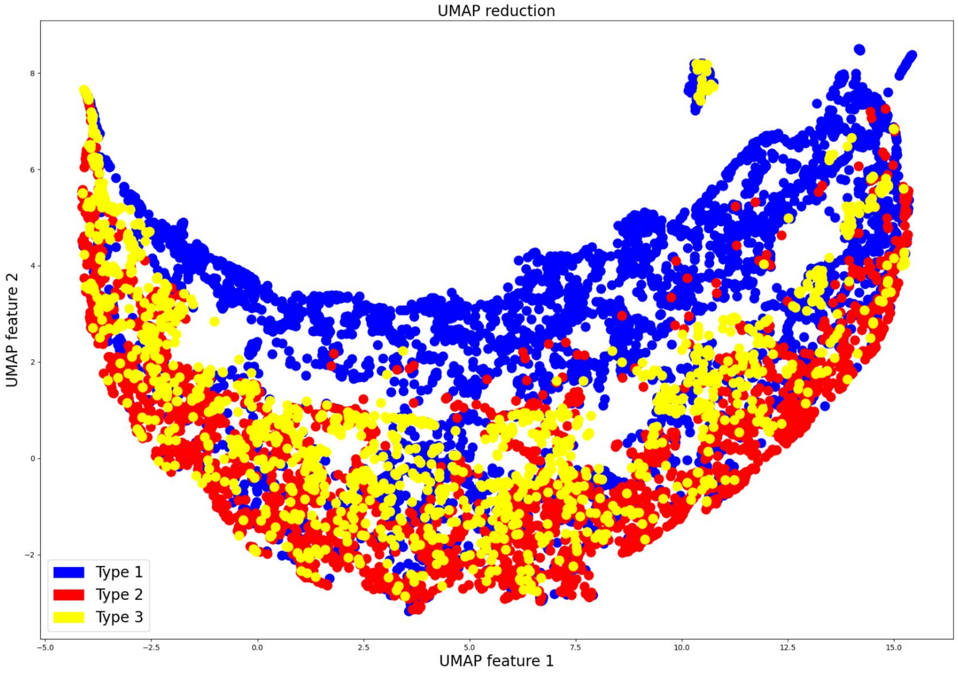


3 types



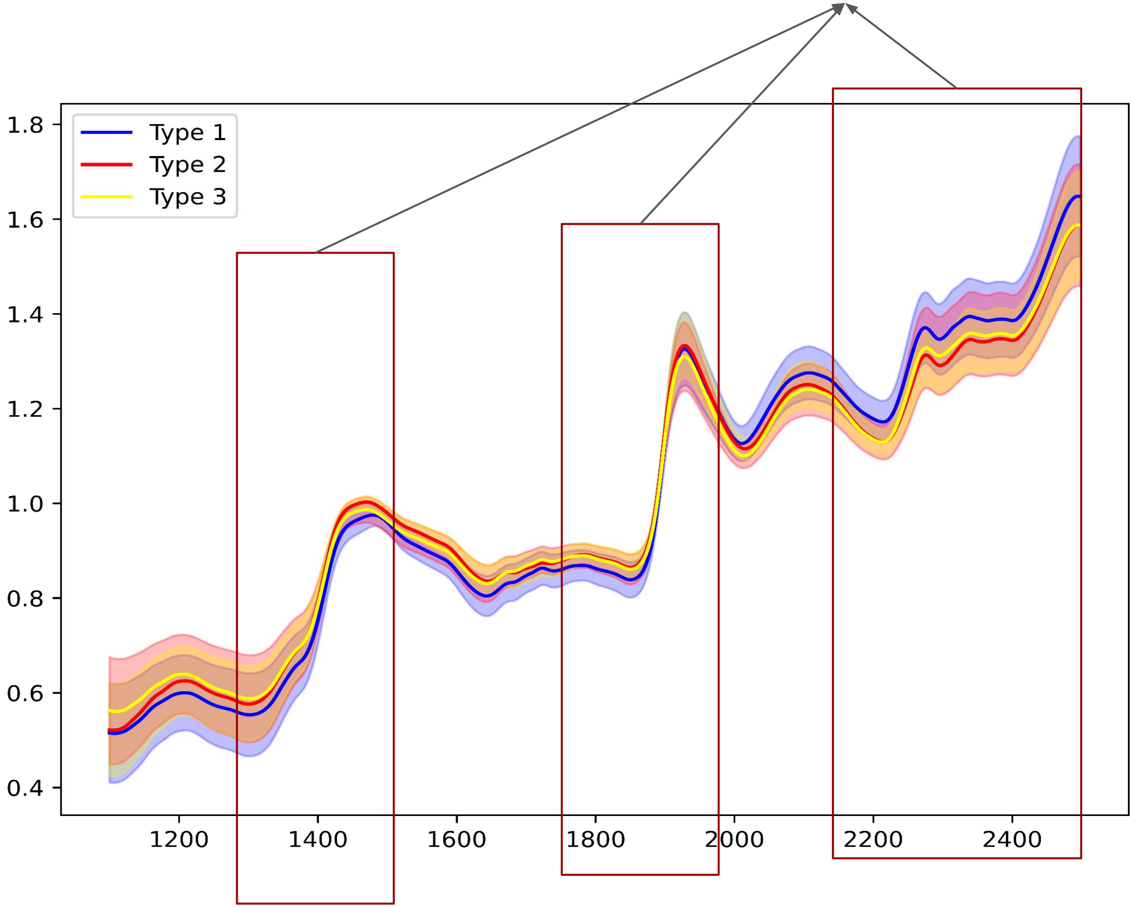


Data Exploration



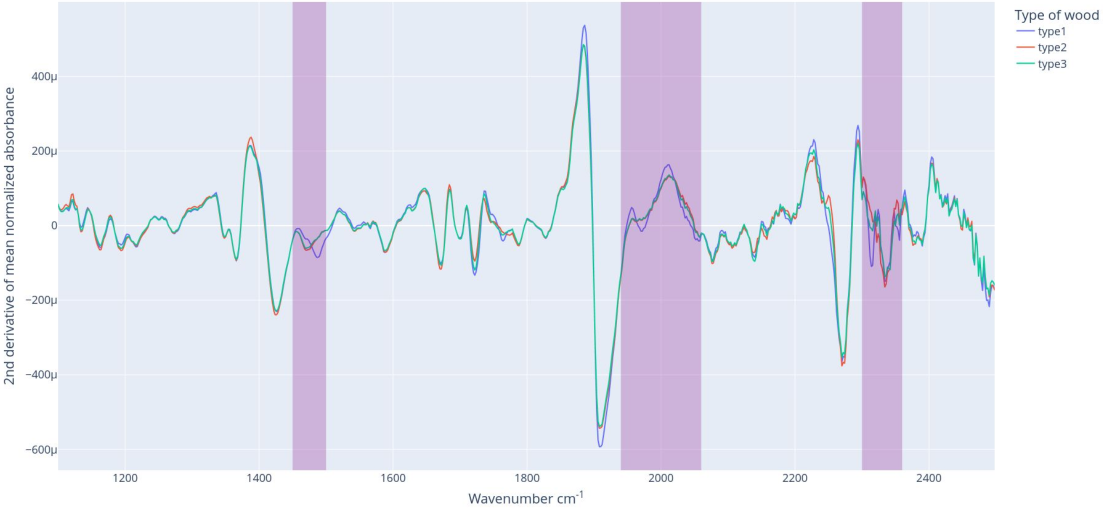
Data Exploration

Spectrums' trends are different for wood types in some regions



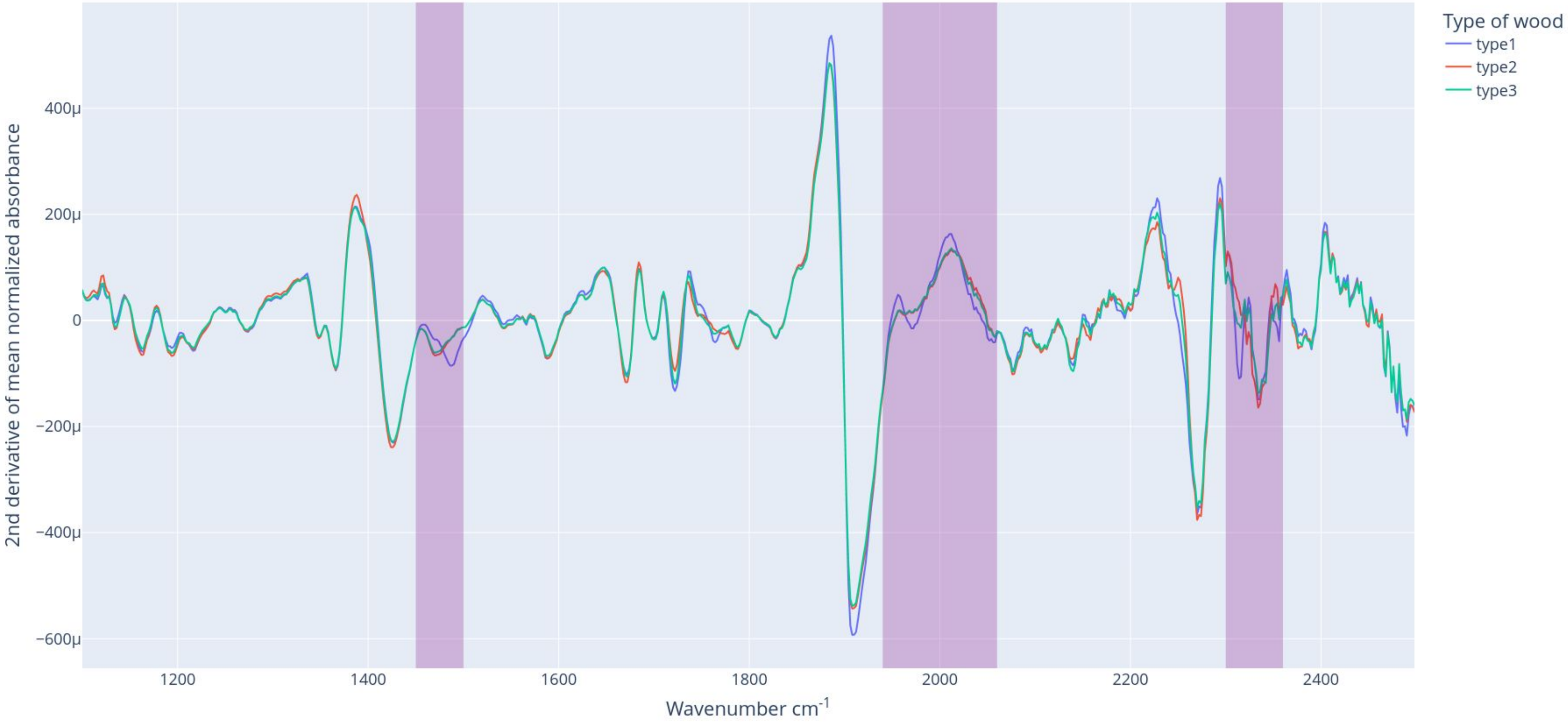
Feature Engineering

- Take the 2nd derivative of spectra.
- Type 1 is easy to be distinguished.
- Type 2 and 3 are very similar.

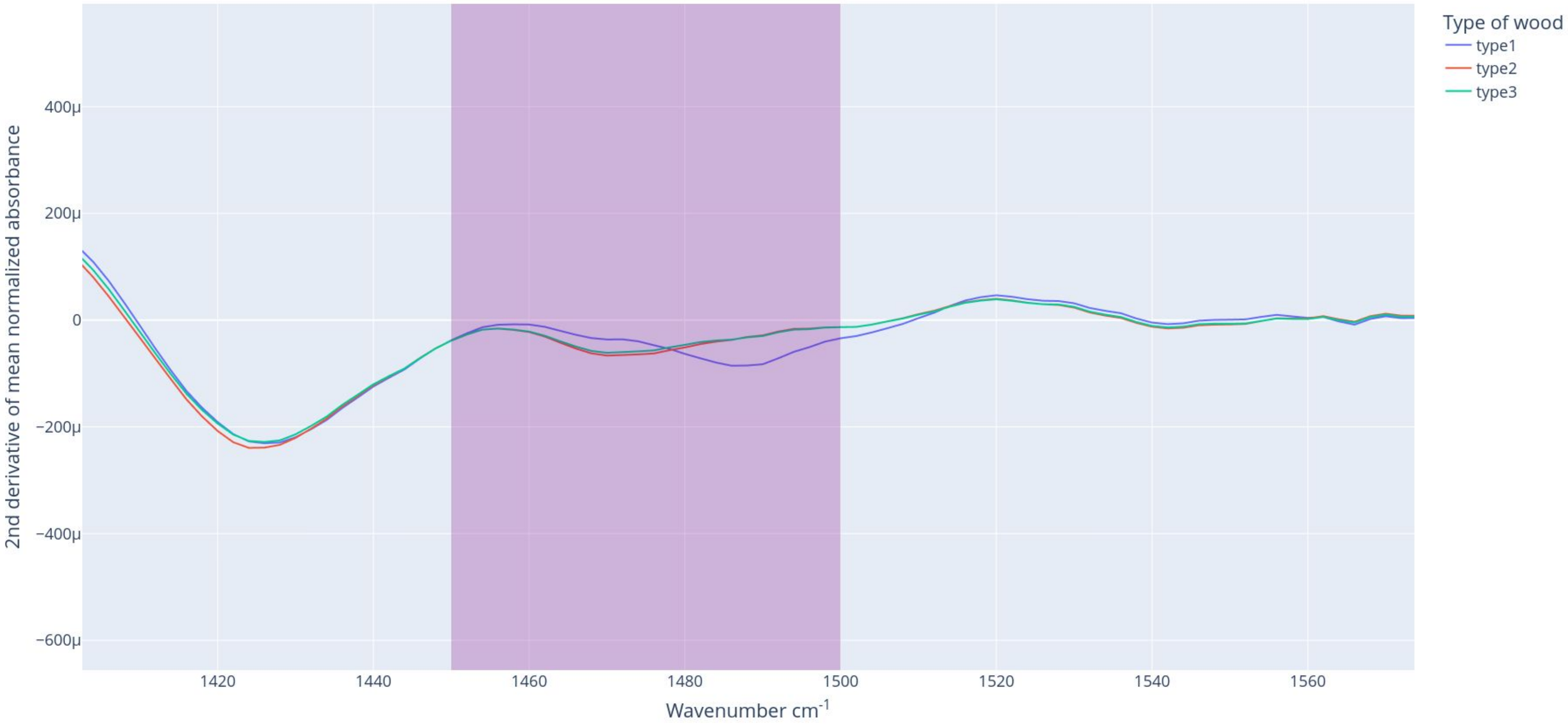


- The 2nd derivative implies the data difference clearly instead of just trends difference.

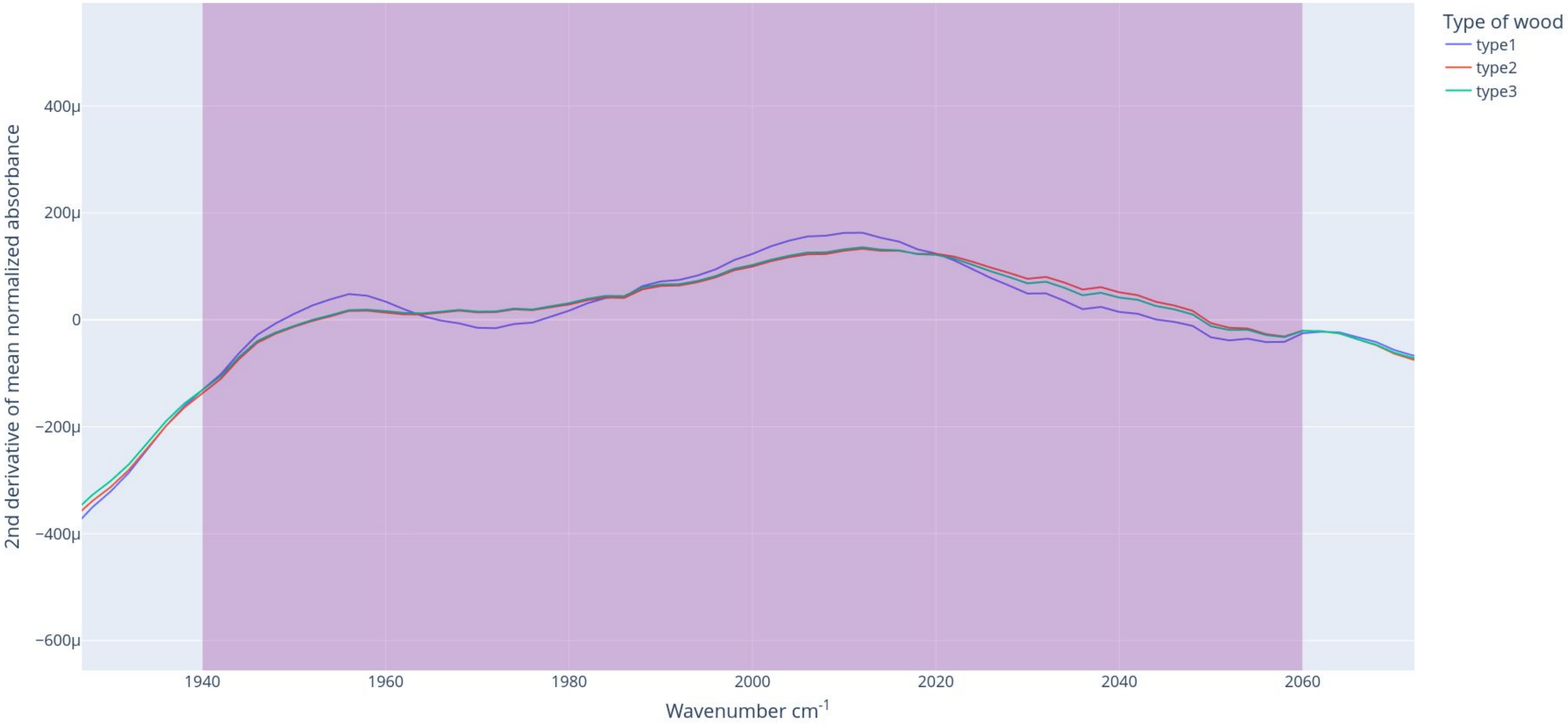
Feature Engineering - 2nd derivative



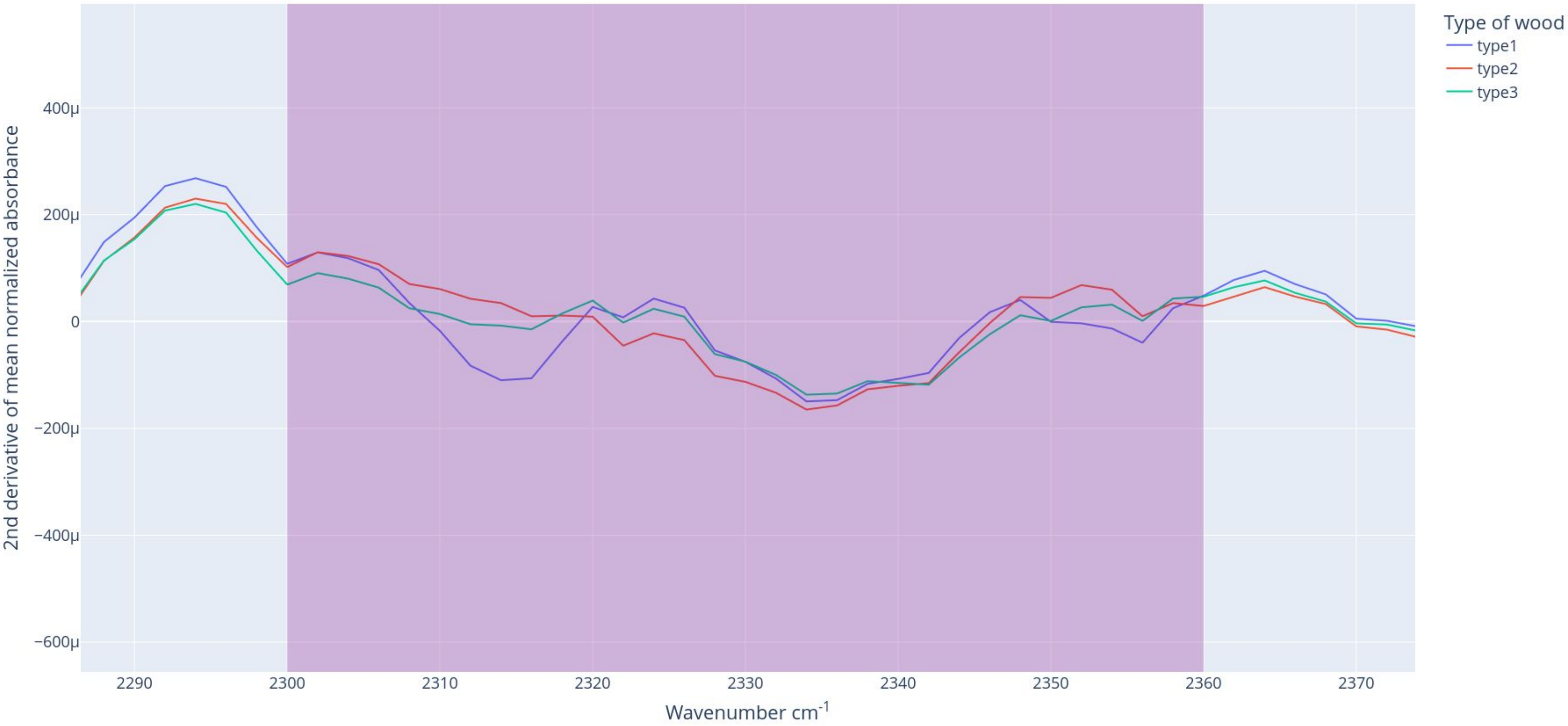
Feature Engineering - 2nd derivative



Feature Engineering - 2nd derivative



Feature Engineering - 2nd derivative



Unbalanced data

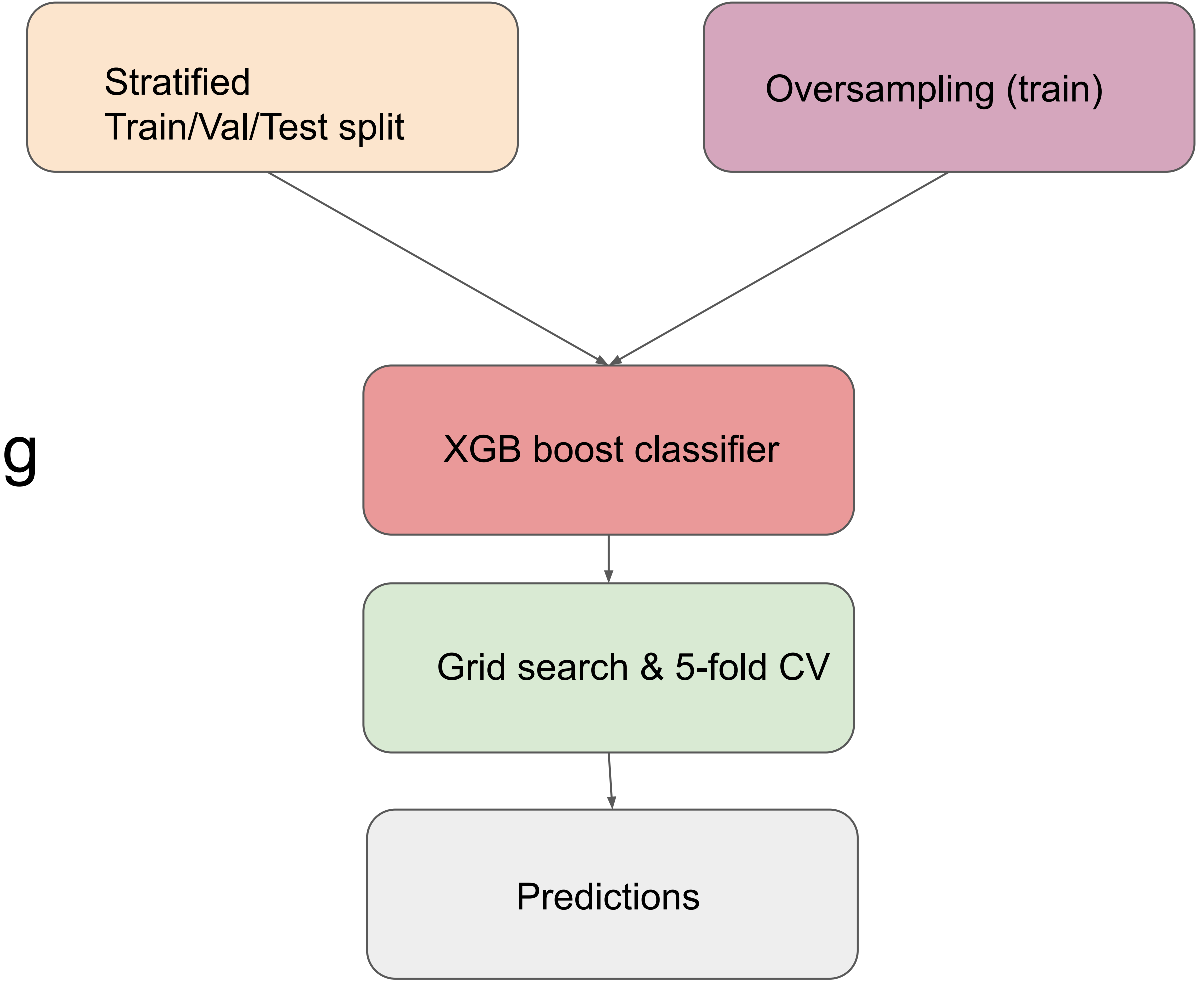
- Unbalanced data set:

Type 1	Type 2	Type 3
4416	2784	1344

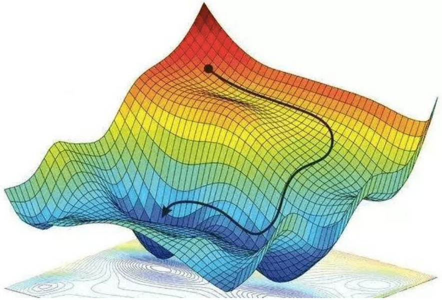
- Solution - Upsampling!
  - Synthesize new samples for the minority classes to obtain a balanced data set.
  - SMOTE (Synthetic Minority Oversampling Technique).
  - Choose a random sample from the minority class and compute its 5 nearest neighbors.
  - Randomly selected a neighbor from 5-nearest neighbors and generate a synthetic sample between these two samples in feature space.
- After upsampling we obtain a balanced data set for classification task:

Type 1	Type 2	Type 3
4416	4416	4416

Modelling

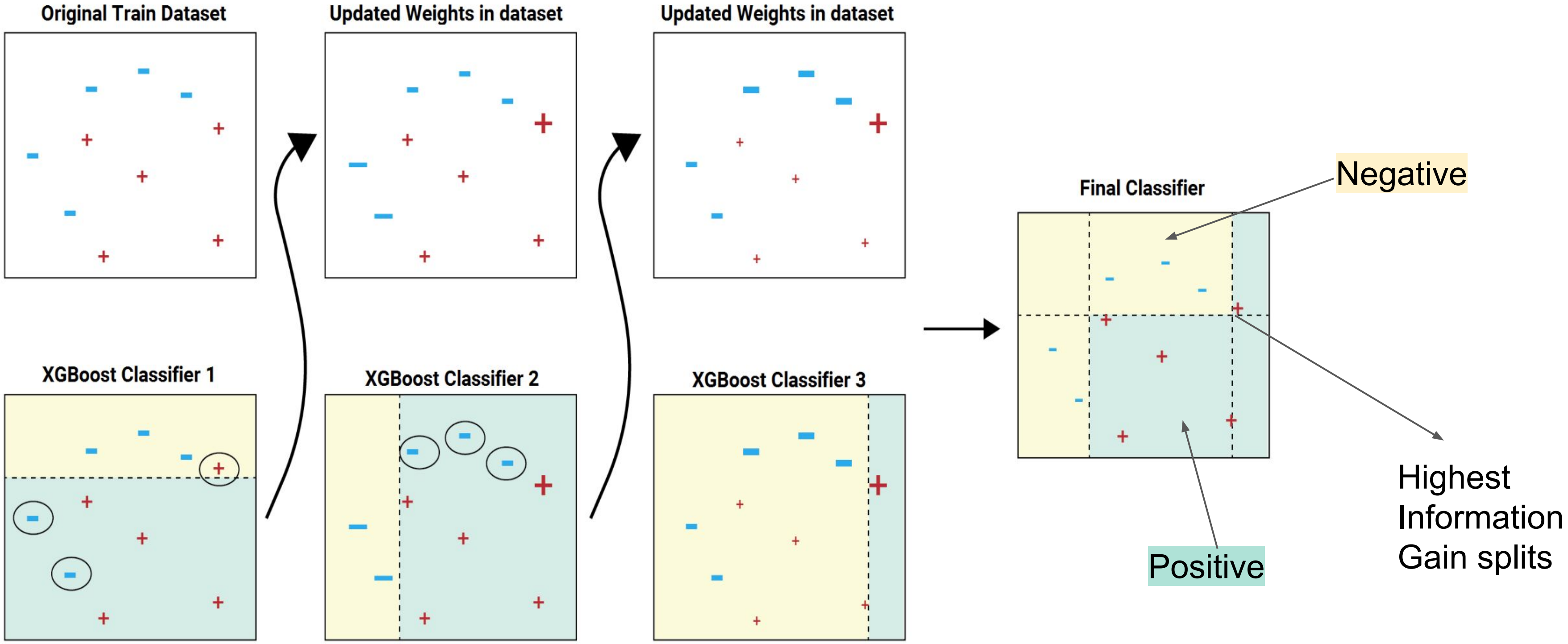

Stratified
Train/Val/Test split
Oversampling (train)
XGB boost classifier
Grid search & 5-fold CV
Predictions

XGB Boost Classifier



- **Extreme Gradient Boosting Decision tree.**
- Minimizes error loss function using **gradient descent**
- Makes use of **gradient boosting.** Trees are trained sequentially
- **Parallelized tree building** for extreme computational performance.
- Many hyperparameters: tree depth, number of trees, regularization,...

XGB Boost Classifier



Results

Class	Precision	Recall	F1-score	Support
1	0.90	0.78	0.84	36
2	0.73	1.00	0.84	27
3	0.94	0.71	0.81	21
average	**0.86**	**0.83**	**0.83**	
accuracy				**0.83**

Future work

- Consistency: Model not very consistent across scans/replicates within same sample

- Try different model types

- Time limitation: Comprehensive Hyperparameter search

tack
kiitos
mulţumesc
nandri
tack
gracias
köszönöm
obrigado
paldies
thank you
grazie
köszönöm
dank u
ďakujem
tack
obrigado
paldies
ačiū
asante
mulţumesc
kiitos
děkuji
gracias
nandri
misaotra
ngiyabonga
THANK YOU
mulţumesc
terima kasih
děkuji
ačiū
takk
takk
grazie
dziękuję
grazie
danke
misaotra
obrigado
salamat
ngiyabonga
köszönöm
takk
thank you
ďakujem
kiitos
terima kasih
dziękuję
ďakujem
salamat
paldies
asante
merci
ngiyabonga
gracias
misaotra
terima kasih
ačiū
danke
děkuji
dank u
nandri
dziękuję
merci
hvala
salamat
dank u
hvala
danke
asante
merci
hvala