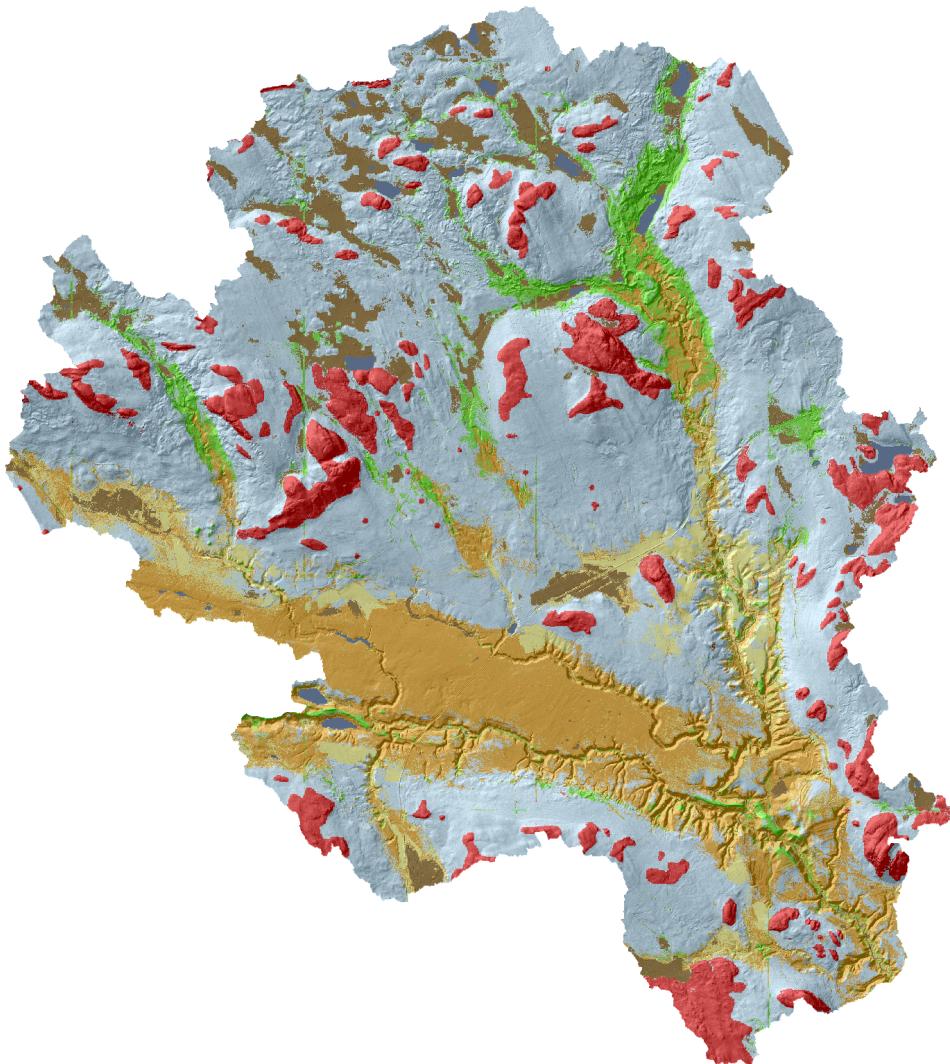


# Jordartskartering – maskininlärning

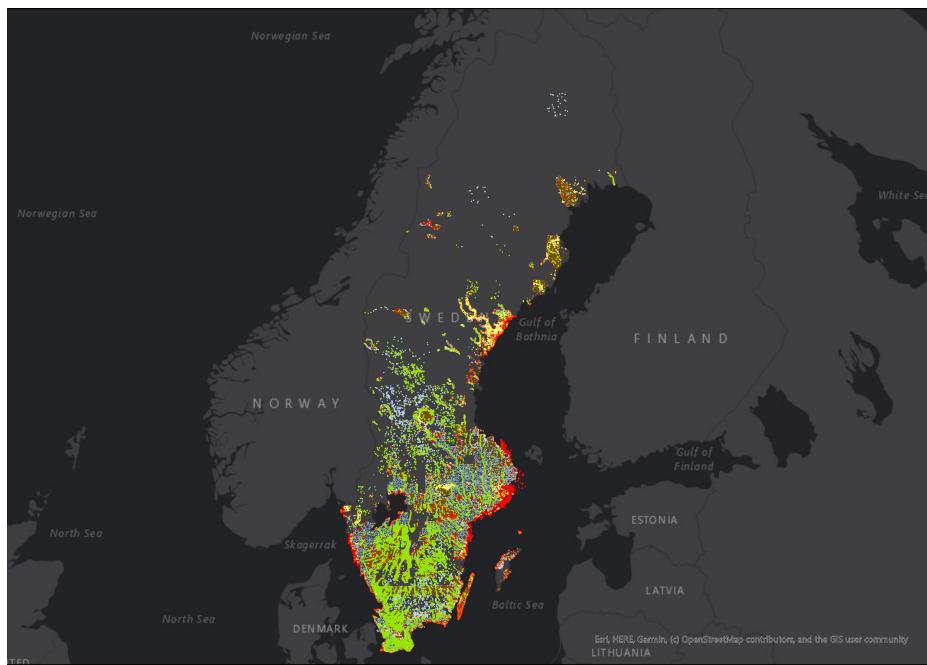


William Lidberg

## INTRODUKTION

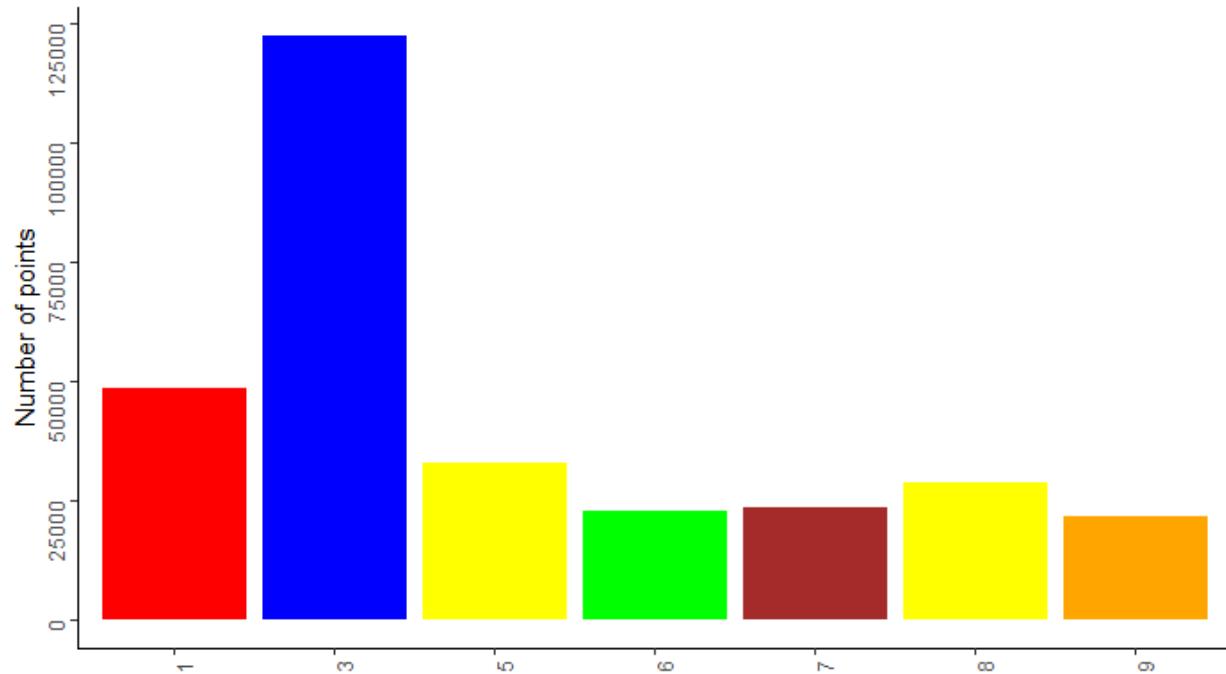
SGU jobbar med att uppdatera Sveriges jordartskarta i delar av landet där det saknas en detaljerad karta. Arbetet som presenteras här har gjorts i projektet ”Jordartskartering - maskininlärning” genom ett samarbete mellan SGU och SLU. Syftet med det här projektet var att hjälpa sätta upp en miljö och arbetsgång för att jobba med artificiell intelligens (AI) så att SGU kan ta fram ett underlag för jordartskarteringar.

Maskininlärning är en typ av AI där en dator lär sig att utföra en uppgift istället för att bli programmerad till det. I det här projektet lärde sig datorn att kartera jordarter i Sverige. Precis som människor behöver datorn gå igenom många olika exempel innan den lär sig att kartera jordarter. Eftersom SGU redan gjort en manuell uppdatering av jordartskartan i delar av landet så användes data därifrån för att träna datorn (figur 1).



Figur 1: Ca 300 000 exemplar på olika jordarter i Sverige.

Majoriteten av Sverige täcks av morän och inte helt överraskande så är majoriteten av datat exempel på morän (figur 2).



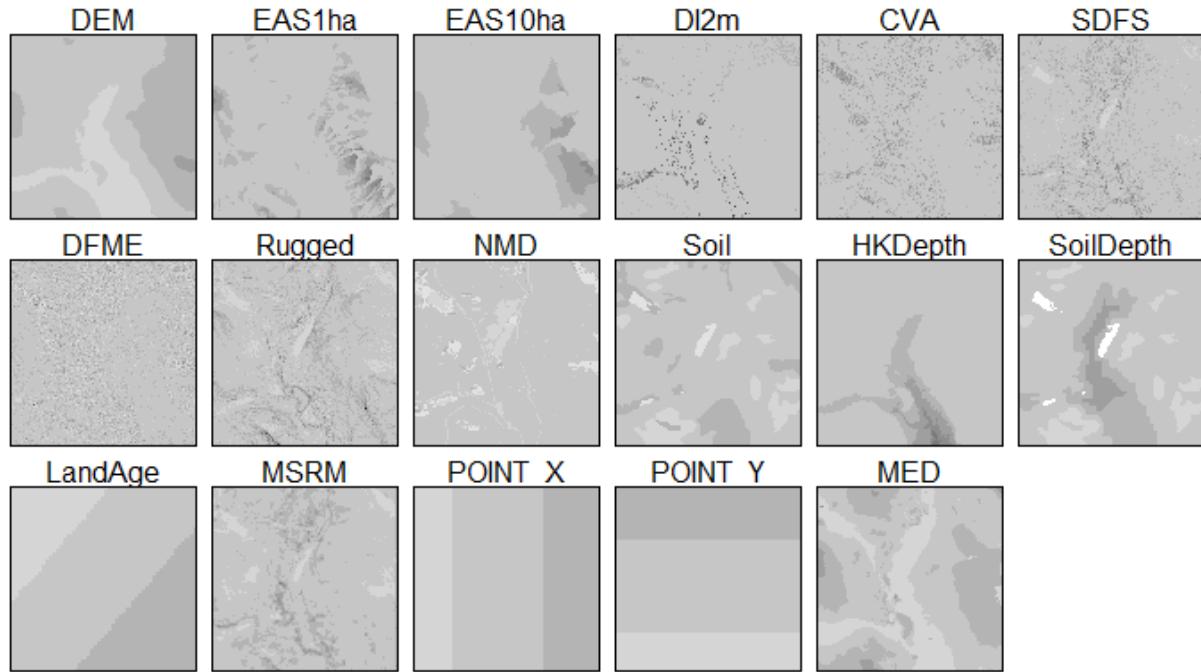
Figur 2: Fördelning av jordarter. 1 = berg, 3 = morän, 5 = glaciala finsediment, 6 = glacifluviala sediment, 7 = torv, 8 = postglaciala finsediment, 9 = postglaciala grovsediment.

## INDICIER

Utöver koordinater på olika jordarter behöver datorn veta lite information om hur landskapet ser vid varje punkt. Följande geografiska information extraherades till varje punk i figur 1:

1. EAS 1ha = Höjd över 1ha vattendrag
2. EAS 10ha = Höjd över 10ha vattendrag
3. DI2m = Nedslutningsindex
4. DEM = Digital höjdmodell
5. MED = Maximal höjdavvikelse
6. CVA = Cirkulär variation av sluttning
7. SDFF = Standardavvikelse av lutning
8. DFME = Avvikelse från medelhöjd
9. Rugged = Terrain ojämnhetsindex
10. MSRM = Ojämnhet på olika skalor
11. CoordinateY = Y koordinater i Swereff 99 TM
12. CoordinateX = X koordinater of all cells in Swereff 99 <sup>TM</sup>
13. HKDepth = Höjdskillnar till HK
14. SoilDepth = Jorddjup
15. NMD = Nationella Marktäckedata
16. Soil = Jordartskartan
17. LandAge = Ålder sen isens tillbakadragande

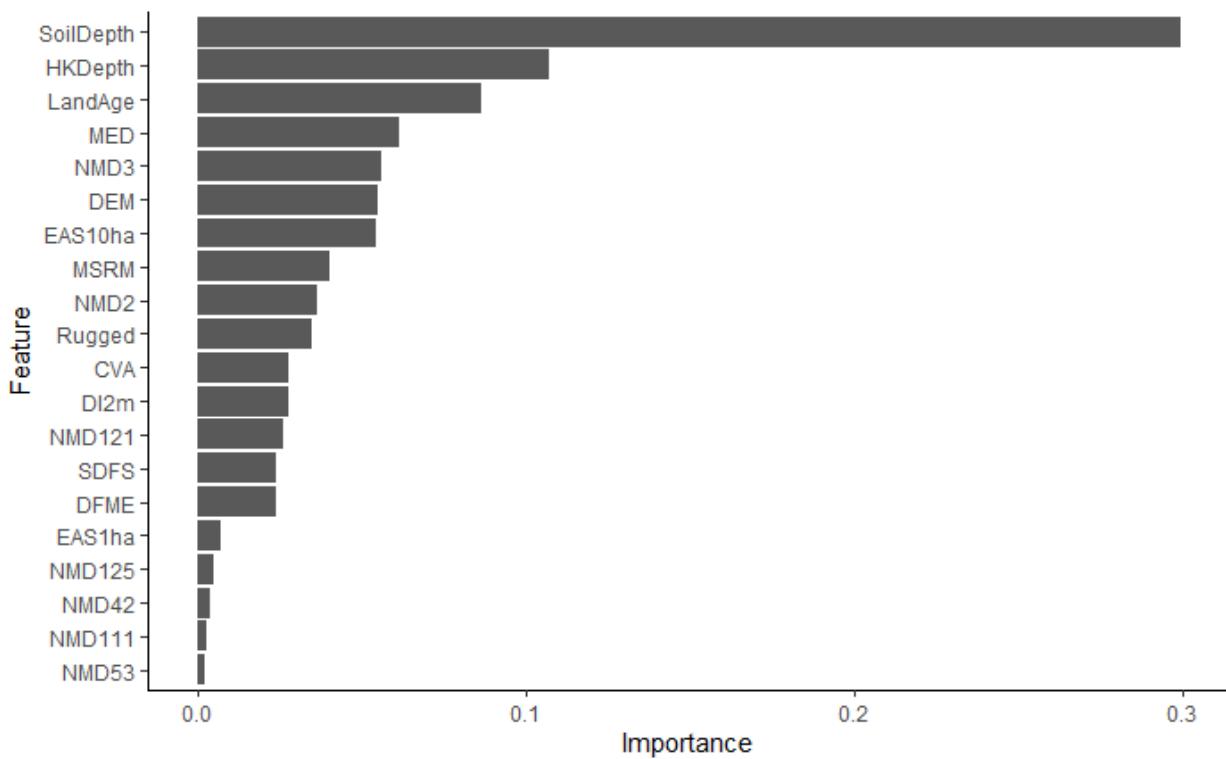
De flesta indicierna är beräknade utifrån lantmäteriets höjdmodell och fångar småskalig topografi och hydrologiska mönster (figur 3). Fler indicier kan läggas till i framtiden för att fånga storskaliga landskapsformationer.



Figur 3: Exempel på indicier som används för att träna datorn att kartera jordarter.

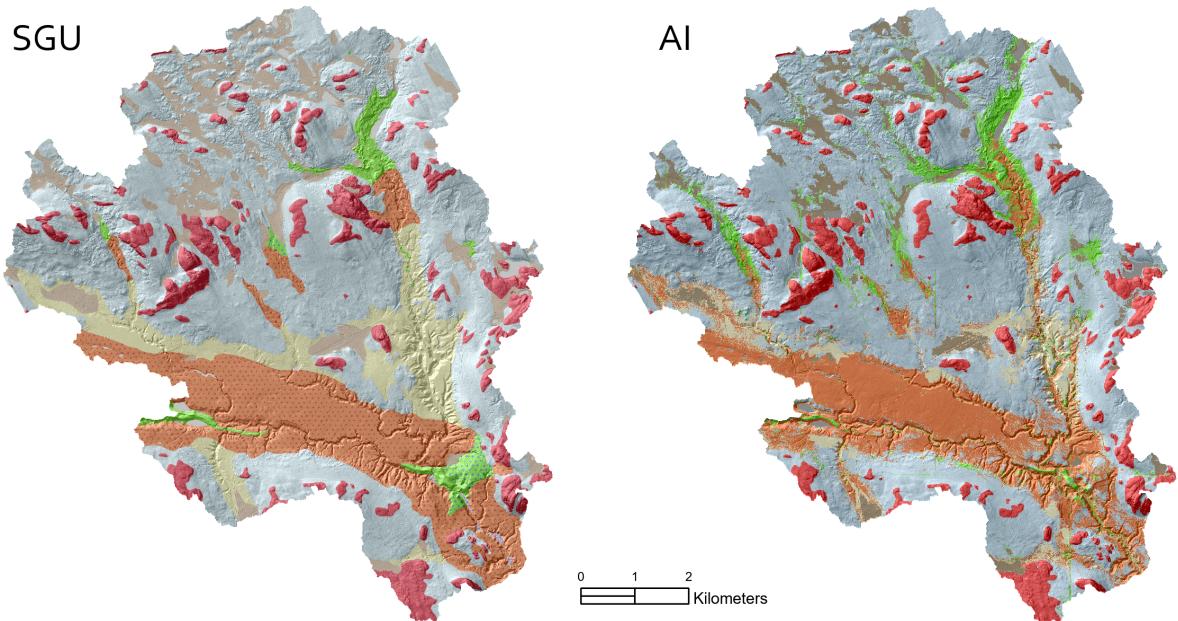
## MASKININLÄRNING

När data från alla indicerna slagits samman med koordinaterna så kan datorn börja träna. Det finns en rad olika modeller för att träna en dator och de är bra på att tackla lite olika problem. Därför testades följande modeller: Random Forest (RF), Support vector machine (SVM), Neural Network (NN), Stochastic Gradient Boosting (SGBoost) och Extreme Gradient Boosting (XGBoost). XGBoost visade sig något bättre än övriga och dessutom betydligt snabbare. Exempelpunkterna från figur 1 delades upp i träningsdata (70 %) och testdata (30 %). Träningsdatat användes för att träna datorn och testdatat användes för att utvärdera hur bra datorn är på att kartera jordarter. När datorn tränats så användes den för att klassa testdatat som den inte sett tidigare. Vid det testet klassade datorn 81 % av alla punkter rätt. Kappavärdet, ett mer balanserat mått på träffssäkerhet vid obalanserade dataset, var 0.75. Modellen uppskattar också hur viktig varje indice är för att kartera jordarter (figur 4)



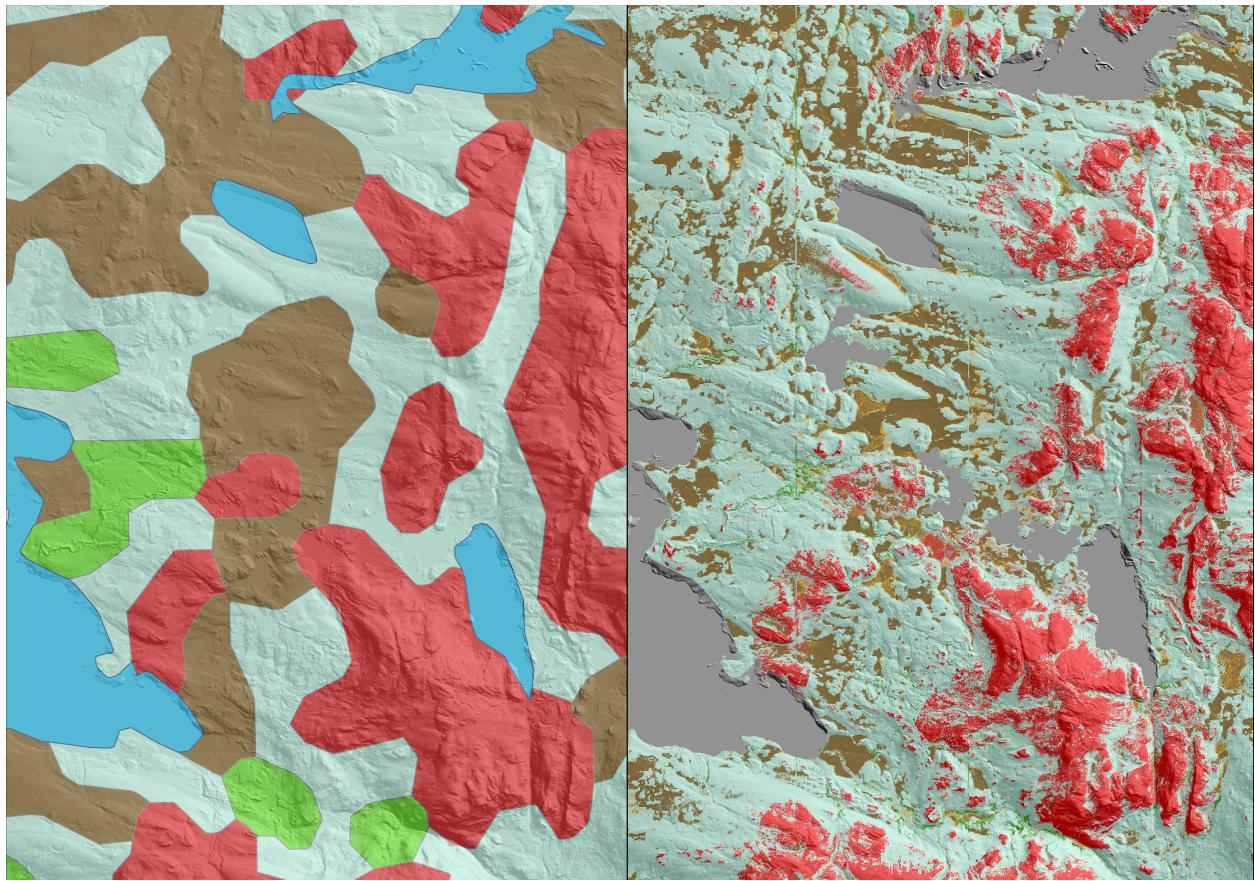
Figur 4: De översta indicierna är viktigaste medan de längst ner inte tillför så mycket information. Jorddjup och höjdskillnad från HK var alltså de viktigaste indicierna.

När datorn tränats klart användes den för att kartera jordarter i olika testområden. En visuell inspektion jämförelse gjordes på ett område med uppdaterad jordartskarta (figur 5) och ett område där den uppdaterade jordartskartan saknas i jämländ (figur 6).



*Figur 5: Jordartskartering i Krycklans avrinningsområde i västerbotten. Till vänster är en jordartskartering som gjordes av erfarna jordartskartering från SGU med höjdmodellen som hjälpmittel. Till höger är en jordartskartering gjord med AI.*

Datorn har svårast för att skilja på glacifluviala sediment och postglaciala grovsediment. Inte helt oväntat eftersom sediment ofta ser snarlika ut i höjdmodellen.



Figur 6: Ett område i Jämtland där det bara finns en översiktskarta. Till vänster är översiktskartan och till höger är jordartskarteringen gjord med AI.

Syftet med det här projektet var att sätta upp en arbetsgång så att SGU kan använda AI för att ta fram ett underlag till pågående uppdateringar av Sveriges jordartskarta. Fler indicier behövs och ett förslag är att lägga till några som fångar in storskalig topografi som glacifluviala formationer och deltaområden.