Examensarbete: AI-baserad klassificering och rensning av djupdata

Kan AI-baserade modeller användas för säker klassificering av störningar i punktmolnsbaserad djupdata och därmed bidra till ökad effektivitet?

Bakgrund

Vid sjömätning med multibeamekolod genereras stora mängder punktmolnsdata där varje datapunkt innehåller information om tid, position, djup, strålvinkel och svarsamplitud. Dessa data innehåller ofta störningar orsakade av exempelvis multipla ekon, luftbubblor, vegetation eller andra störningar.
I dagsläget kräver rensning av störningar en tidskrävande och till största delen manuell granskning, vilket ianspråktar betydande personella resurser.

Sjöfartsverket vill börja undersöka vilken potential det finns med att använda artificiell intelligens (AI) och maskininlärning som en del i denna process. Det är viktigt att upprätthålla en hög noggrannhet i identifieringen av sanna djup, för att säkerställa den nautiska säkerheten.

Syfte

Syftet med examensarbetet är att föreslå en, eller en sammansättning av, AI verktyg lämpade för tester. Test av en eller, om möjligt, flera föreslagna AI-baserade metoders förmåga att klassificera djuppunkter i data som troliga störningar eller godkända mätpunkter, om möjligt med en sannolikhetsskala för att en djuppunkt är en störning.

Förbättrade tekniska verktyg leder till en mindre grad av individberoende subjektiva bedömningar och på sikt till jämnare datakvalitet i vår djupdatabas. Riskerna för navigation är dock större om ett sant djup flaggats bort, än om en störning blir kvar i datamängden.

Det är i framtiden möjligt att en metod för rensning av data måste samköras med en metod som kan identifiera och klassificera bottenobjekt, för att ge ett utökat beslutsunderlag.

Mål

Examensarbetet ska resultera i:

1. En skriftlig Rapport som beskriver genomförande och resultat, samt rekommendationer om hur Sjöfartsverket bör fortsätta med tester med AI för bearbetning av djupdata.
2. Förslag till en eller flera sammansättningar av AI verktyg för fortsatta tester i form av enklare processbeskrivningar.

3. Test av ett eller flera föreslagna AI verktyg, eller hela sammansättningar av verktyg, som tränas på data set bestående av rådata och manuellt rensade data.
4. Om flera verktyg eller sammansättningar testas, en jämförelse av modellernas prestanda.

Möjlighet kan finnas att skriva en artikel för publicering i International Hydrographic Review (https://ihr.iho.int/) efter avslutat och bedömt arbete.

Tillgängliga data

Lämpligt data tas fram av SjöV då hänsyn måste tas till sekretesskrav.

- **Rådata (ASCII)**: innehåller tid, position, djup, strålvinkel, svarsamplitud.
- **Bearbetade (rensade och reducerade) data (XYZ ASCII)**: 1-25 punkter/m^2 i ursprungliga positioner, manuellt rensade från störningar (högre datatäthet i vissa områden och i rådata).
- **Metadata**: information om mätområde, bottentyp, djupintervall.

Metodik (förslag till arbetsgång)

1. **Datadriven analys och modellering**
    - Jämförelse av olika modelltyper:
        - Klassiska ML-metoder (t.ex. Random Forest, XGBoost).
        - Djupinlärningsmetoder för punktmoln (t.ex. PointNet++).
2. **Testning** (i den mån det ryms inom tidsramen)
    - Utvärdering av modell/modeller med hjälp av manuellt rensade data som "ground truth".
3. **Resultatanalys och slutsatser**
    - Sammanställning av projektets resultat.
    - Rekommendationer för framtida utveckling och tester.

Förväntat resultat

- När tester genomförts med AI modeller, eller sammansättningar av modeller vill Sjöfartsverket få tillgång till de använda körbara scripten, för fortsatt testning mm.
- En rapport som beskriver föreslagna metoder samt resultat av eventuella tester.
- En föredragning av resultat för personal på Sjöfartsverket.

Förkunskaper

Lämplig bakgrund:

- Kunskap i Engelska, både tal och skrift.
- Datateknik, maskininlärning, geomatik, marin mätteknik eller liknande.
- Programmeringskunskap (Python, eller annat liknande scriptspråk).
- Grundläggande förståelse för geografiska data/punktmoln.

Omfattning

- **Omfattning:** 30 hp.
- **Tidsåtgång:** ca 20 veckor.
- **Arbetsplats:** Datorbaserat arbete (kan till övervägande del utföras på distans).

Sjöfartsverket står för reskostnader för eventuella inrikes resor till Norrköping, samt vid behov hotellkostnader.

Förslag till titel

"AI-baserad identifiering och rensning av djupdata”
