

Projektinlämning

Uppgift

Du ska välja en av 5 uppgifter och försöka lösa problemet med hjälp av djupa neurala nätverk. Varje uppgift innehåller ett dataset och en problemformulering. Du ska själva ta fram vilken typ av modell som är lämplig för projektet, träna den, optimera hyperparametrar och utvärdera.

Du får inte lägga till egen data i projektet, men du ska städa datan. Det är tillåtet med data agumentering för att utöka ditt dataset.

Du ska skapa en modell från grunden, transfer learning får inte användas. Det finns inga krav på modellprestanda (finns heller inget svar på hur bra din modell kan bli). Det är godkänt att ha flera modeller som används, men alla ska vara djupa neurala nätverk.

Datan finns på onedrive.

Inlämningen och presentationen är individuell.

För betyg G

Skapa en modell från grunden. Du ansvarar själv för städning av datan och uppdelning i träning, validering och testning.

Frågor:

- Motivera din modell, dess arkitektur och val av relevanta hyperparametrar.
- Har du använt någon databehandling? Varför?
- Utvärdera modellens prestanda.
- Är prestandan bra? Varför/ varför inte?
- Vad hade du kunnat göra för att förbättra den ytterligare?

För betyg VG

För VG ska du utvärdera modellen noggrannare och systematiskt optimera hyperparametrar.

Frågor:

- Testa olika hyperparametrar systematiskt.
- Vilken typ av fel gör modellen?
 - Gör den fel på vissa klasser? Gör den fel specifik typ av data? Annat?
 - Kan du göra något för att hantera vanliga fel modellen gör? (Med den data du har, att träna på mer data är inte ett godkänt svar)
- Finns det några begränsningar med modellen?
 - Exempelvis på grund av prestanda eller hur träningsdatan ser ut.

Presentation

Förutom att lämna in en notebook ska du skapa en presentation (exempelvis en power point) där du ska presentera erat problem, vad för modell du skapat och hur den presterar. Ni kommer presentera i smågrupper.

Presentationen ska vara 10-15 minuter samt 5-10 minuter till frågor. Det är obligatoriskt att presentera och vara med under alla presentationer i din grupp. Du får vara med och lyssna på andra grupper om du vill.

Tillåtna hjälpmedel

API dokumentation, google, böcker och föreläsningar.

Ej tillåtet att använda chatGPT eller annan generativ AI.

Inlämning

Inlämning av koden ska ske senast 26-01-2024 23:59 via läroplattformen.

Presentationerna kommer ske 24-01-2024

Lämna in en jupyter notebook och en requirements.txt fil som anger vad för paket ni använder er av.

Om ni bifogar material så som bilder i notebooken, glöm inte att även skicka med materialet. Datasetet behöver ni inte bifoga.

Problem

Algblooming

Algblooming drabbar svenska kuster årligen. Främst rapporterar vaniga människor algblooming i olika områden och så måste det kontrolleras manuellt. Genom att skapa en modell som kan hitta algblooming i bilder kan problemet kartläggas bättre och mer effektivt.

Uppgift: Skapa en modell som givet en bild kan säga om det finns algblooming eller inte i bilden.

Data: ca 300 bilder på vatten med algblooming och 100 bilder med vatten utan algblooming.

Sjuk fisk

Klimatförändringarna påverkar djurlivet runt om i världen. En sak som upptäckts att fler och fler fiskar drar på sig sjukdomar. Genom att bättre kunna se om en fisk är sjuk eller inte kan vi få bättre förståelse för hur utbredd problemet är och vidta åtgärder om det behövs.

Uppgift: Skapa en modell som kan avgöra om fisken i bilden är sjuk eller inte. Vilken typ av sjukdom behövs inte identifieras.

Data: ca 200 bilder på frisk fisk och ca 300 bilder på sjuk fisk (fördelat på 6 olika sjukdomar).

Skogsbränder

Att kontrollera om delar av en skog brinner är svårt och dyrt, speciellt om delen av skogen ligger lång ifrån civilisation. Med hjälp av bilder från drönare kan detta effektiviseras då drönarna kan "patrullera" skogen med jämna mellanrum samt skickas ut vid behov.

Uppgift: Skapa en modell som kan avgöra om det är en skogsbrand eller inte.

Data: ca 5000 bilder, jämnt fördelade mellan bränder och icke bränder.

Covid fake news

Under corona så exploderade det med nyheter och rubriker som sedan visat sig att de inte är sanna. Detta har fortsatt efteråt och blir ett större och större problem. För att undvika att stora falska informationsflöden får fäste så behövs modeller som kan avgöra vad som är sant och inte.

Uppgift: Skapa en modell som kan avgöra om en rubrik är kopplad till en falsk nyhet eller inte.

Data: Ca 10 000 rubriker med flagga om den är kopplad till en falsk nyhet eller inte.

Motverka hatretorik

Denna uppgift innehåller hatiska texter riktad mot flera olika grupper och kan därmed vara jobbig för vissa att genomföra.

På sociala medier och forum skrivs idag mycket hatiska texter, både mot personer och grupper. Oftast bevakas sociala medier och forum av moderatorer, men det är ett tidskrävande arbete och hat sprids snabbare än vad de hinner jobba. Genom att ha en modell som kan identifiera denna typ av text så kan vi snabba på moderatorerna arbete och minska att hat sprids.

Uppgift: Skapa en modell som kan identifiera om en text är hatisk mot en specifik grupp (och vilken grupp). Det är några typer av hat som ligger i fokus, exempelvis hat baserat på sexualitet eller kön.

Hat mot exempelvis en yrkesgrupp räknas inte som en hatisk text.

Dataset: ca 4000 texter, om den är hatisk och i så fall mot vilken grupp.