Bildegjenkjenning for menn og kvinner

[Jonas Skaslien, Christian Vidal/ DAT158 Førde], [14.11.2024]

Beskriv Problemet

Scope

Målet til dette prosjektet er å lage en modell som kan bestemme om en person er en mann eller en kvinne. Objektivet for prosjektet er å lage en gøy og morsom maskinlæringsmodell, for alle personer som er interessert om en person er en mann eller dame basert på et bilde. Produktet skal brukes for å se på bilder av enkelte personer, og basert på bildene skal den kunne bestemme om personen på bildet er en mann eller kvinne. Det finnes noen løsninger dette, som f.eks. Gender Detection Model using CNN (se referanse [3]), som er en maskinlæringsmodell som ser på bilder av personer, og basert på bildet skal den prøve å forutsi om personen er en mann eller dame, og hvor gammel personen er. Hvis dette prosjektet skal utføres manuelt, så kunne en person enkelt bare se på et bilde og bestemme selv om personen på bildet er en mann eller dame.

Ytelsen for prosjektet vil være en suksess hvis modellen klarer å få en treffsikkerhet på over 50%. Det vil være godt nok for dette korttids prosjektet. Prosjektet vil få fire ukers tid på å bli utført. Første uken vil bli brukt for å lete en passende prosjektoppgave, og planlegge hvordan prosjektet skal utføres. I andre uken skal modellen bli utviklet og trenes opp til å gjenkjenne kjønnet til en person på et bilde. Den tredje uken ble modellen koblet opp til en nettside, og testing av modellen og nettsiden ble utført den uken. I den fjerde og siste uken vil en rapport bli skrevet basert på resultat og arbeid som er gjort i de første tre ukene. For nå er det bare gruppen som er stakeholder for prosjektet, fordi gruppen ville lage en en gøy og morsom nettside som tar inn et bilde, og ut ifra bildet vil nettsiden fortelle om personen på bildet er en mann eller dame.

Behovet som prosjektet trenger er ikke så mye. Kun to personer er nødvendig for å kunne utvikle og trene opp modellen. Det vil trenges mange bilder av både menn og kvinner, for å trene opp modellen. Disse ressursene kommer fra Kaggle, se referanse [4]. Andre nødvendige ressurser er en server for nettsiden. Gruppen har valgt å bruke Gradio for å lage nettsiden.

Metrikker

For at prosjektet skal være en suksess, må modellen ha en treffsikkerhet på over 50% for at nettsiden skal være brukbart for folk som er interessert i å teste om en person er mann eller

dame. Det kvadratiske avviket burde være lavt for en høyere presisjon for modellen, og for prosjektet sin suksess burde den være lavere enn 50%. Nettsiden burde ha lav latency, og respondere fort tilbake til brukeren med resultatet. Throughput, eller gjennomstrømming, burde være høy for prosjektet. Det er ønskelig at nettsiden kan gjøre flere søk på kort tid. Med disse nedre grensene for prosjektets metrikker, vil gruppen anse prosjektet som en suksess hvis modellen og nettsiden kan møte disse metrikkene.

Modellen har fått en treffsikkerhet på rundt 70%, som betyr at vi har en høy nok treffsikkerhet for modellen, og kan gå videre til andre mål. Det kvadratiske avviket er rundt 30%, som gir oss et passende lavt avvik for prosjektet, men det er ønskelig med et lavere avvik for en mer presis modell. Latency er veldig lav, og nettsiden responderer nesten umiddelbart med resultatet sitt. Selvfølgelig, så kan latency bli høyere hvis et stort antall folk bruker nettsiden, som medfører at nettsiden får veldig høy trafikk, men ellers er latency godkjent for prosjektets metrikker. Gjennomstrømmingen for prosjektet er veldig høy. Det kan estimeres at en person bruker rundt 10-15 sekunder for å laste opp et bilde på nettsiden, trykke på “Submit” knappen, og få et resultat tilbake. Så hvis det tar rundt 10-15 sekunder per forsøk, så kan en person søke rundt 5-6 ganger i minuttet. Etter 10 minutter vil en person ha gjort rundt 50-60 forsøk, og etter en time kan en person ha gjort rundt 300-360 ulike forsøk. Dette vil gi oss en høy gjennomstrømming.

Data

Dataen som skal bli brukt i prosjektet vil være bilder av personer der ansikt er synlig, helst av kun én person per bilde. Labels som skal brukes vil være de binære tallene 0 for menn og 1 for kvinner. For å samle inn data er det brukt Human Images Dataset - Men and Women (se ref. [4]) og det finnes andre større datasett som CelebA Dataset (se ref. [5]) og UTKFace | Large Scale Face Dataset (se ref. [6]). Siden modellen ikke skal benytte seg av dyp læring og kompleksiteten av problemet ikke er veldig komplisert kan det bli brukt et mindre datasett. Men det er fortsatt viktig og få et stort nok datasett til å kunne fange opp forskjeller i menn og kvinner som kan forandre seg basert på alder, etnisitet og unike personlige trekk. For å få tak i labels som trengs er det i de fleste datasett vi har sett på allerede markert om det er en mann eller kvinne, men hvis det ikke var det kunne flere personer ha sett på bildet og markert om det var en dame eller mann. Hvis det skulle bli uenighet om personene var en mann eller dame, så kunne det bli brukt et stemmesystem eller bare forkastet dataen. Siden prosjektet er bygget på personopplysninger “Et bilde regnes som en *personopplysning* dersom personer kan gjenkjennes” [1], så må man under norsk lov ha samtykke “En virksomhet kan behandle personopplysninger dersom den har innhentet gyldig samtykke fra personen eller personene det gjelder” [2]. Dette kan bli et stort problem for innhenting av data, løsningen vil være å bruke datasett som allerede har samtykke. Dataene eller bildene maskinlæringsmodellen skal bruke må først sørges for at det er tydelige bilder av én person der ansiktet synes. Så skal bildene skaleres til samme størrelse for så å bli gjort om til en tabell med en kollonne der en verdi er en piksel verdi. Å representere bildet som en dimensjonal tabell er noe som ofte blir gjort for å forenkle bildet og fordi mange modeller krever data som ikke har flere dimensjoner.

Modellering

I prosjektet har gruppen valgt å ikke bruke nevrale nettverksteknikk, men heller gå for en simpel modell basert på det vi har lært. Derfor skal modellering heller prøve å gjøres med modeller som logistisk regresjon, støttevektormaskiner (SVM), random forests og K-nærmeste naboer. Siden problemet er enkelt for et menneske å løse kan modellen måles opp mot hvor bra mennesker kan løse problemet. Gruppen gjorde en liten test på tyve personer og klarte å gjette riktig på alle bildene. Ettersom det kan være personer som kan være vanskeligere å gjette seg fram til enn på de bildene det ble testet på, så blir det lagt på en feilmargin på fem prosent, noe som gir en standard på 95% som modellen skal prøve å nå. Fordi det ikke blir brukt nevrale nettverk i modellen som er bedre enn standard maskinlæringsmodeller på å gjenkjenne bilder velger gruppen å sette baseline-ytelse til 65-70%. Som er bedre enn vill gjetning, men fortsatt ikke for høyt. For å finne ut hvor ting går feil skal gruppen teste modellen på bilder av forskjellige kategorier som menn,damer,eldre,yngre,etnisitet etc. å se på hvem kategori modellen feiler oftest. For feature importance skal det gjennom LIME bli lagd “heatmaps” på bildene som indikerer hvilke features som er de viktigste for modellen i bildet. Gjennom finne ut hvem kategorier modellen sliter med kan det bli lagt til mer data for de kategoriene og med heatmaps kan en se om modellen f.eks ser på ting som er irrelevant for kjønn som bakgrunn.

Deployment

Modellen skal settes i drift ved å kjøre Gradio kode på en server, dette vil generere en link som skal kunne åpnes så lenge serveren kjører. Prediksjonene skal brukes for moro skyld, læring for gruppen om hvordan man utvikler en maskinlæringsmodell basert på bildegjenkjenning og hvordan en kan koble opp en maskinlæringsmodell opp mot en nettside. For å forbedre systemet etter det er satt i drift kan det samles inn bilder og svar om modellen hadde riktig når noen bruker den. Dette kan hjelpe med å få mer data til modellen, men det krever også at de som bruker den gir ordentlige bilder og svarer riktig på om de faktisk var mann eller kvinne. Derfor vil det kreve mye mer vedlikehold for å gå gjennom bildene brukere lastet opp og sørge for at de svarte ordentlig. En annen forbedringsløsning vil være å bare oppdatere modellen med flere bilder fra et større datasett. Som siste forbedring kunne modellen også ha blitt trent på bilder som er rotert, speilvendt eller redigert. Dette kunne blitt gjort ved å rotere, speile og redigere bilder vi allerede har og mate dem inn i modellen.

Referanser

[1] https://www.datatilsynet.no/rettigheter-og-plikter/personopplysninger/er | Datatilsynet
hentet: 14.11.2024

[2] https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/om-behandlingsgrunnlag/samtykke/ hentet: 14.11.2024

[3] https://medium.com/@skillcate/gender-detection-model-using-cnn-a-complete-guide-279706e94fdb hentet 14.11.2024

[4] https://www.kaggle.com/datasets/snmahsa/human-images-dataset-men-and-women/data hentet 14.11.2024

[5] https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html hentet 14.11.2024

[6] https://susanqq.github.io/UTKFace/ hentet 14.11.2024