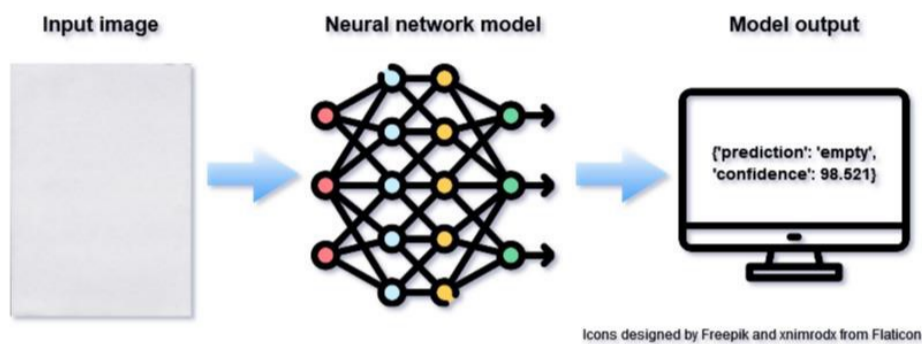


## Igenkänning av tomma sidor

### Bakgrund, fördelar och ändamål med komponenten

I processen för digitalisering av dokument används ofta dubbelsidig skanning. Om dokumenten är ensidiga skapas många tomma bilder vid digitaliseringen. En maskinell klassificering av tomma bilder och innehållsrika bilder underlättar digitaliserarens arbete. När igenkänningskvaliteten är tillräckligt hög, är det också möjligt att dölja tomma bilder eller till och med ta bort tomma bilder från förvaringen. Om dokumenten finns tillgängliga för kunden i tjänsten, kan möjligheten att dölja tomma bilder från vyn bidra till att förbättra kundupplevelsen. Om tomma bilder känns igen före senare faser i processen för behandling av dokument (bl.a. feligenkänning, textigenkänning, metadataigenkänning), kan hela kedjan för behandling av dokument också påskyndas.

### Vad gör komponenten?



Komponenten behandlar de bildfiler som den fått som indata en efter en och tar fram en prognos för bildens innehåll på grundval av de parametrar som den har lärt sig under utbildningen. Om komponenten känner igen innehåll i bilden med mer än 50 procent sannolikhet, klassificeras bilden som innehållsrik och i annat fall som tom. Komponentens återställer den förväntade klassen till användaren.

### Hur används komponenten?

Komponenten finns tillgänglig både som en del av Arkkiivi-gränssnittet (<http://www.arkkiivi.fi/>) och som en fristående app som tillhandahåller ett programmeringsgränssnitt (API) på DALAI-projektets GitHub-webbplats (<https://github.com/DALAI-project/EmptyAPI>). Närmare anvisningar om användningen av komponenten finns på den webbplatsen.

### Utbildning av komponenten

Bildfiler i .jpg-format har använts för att utbilda komponenten och deras storlek har harmoniserats till 224 x 224 pixlar. Komponentens har utbildats flera gånger och det finns ingen exakt information om det ursprungliga utbildningsmaterialets omfattning. Materialet som använts vid utbildningarna innehåller cirka 100 000 tomma bilder och 130 000 innehållsrika bilder.