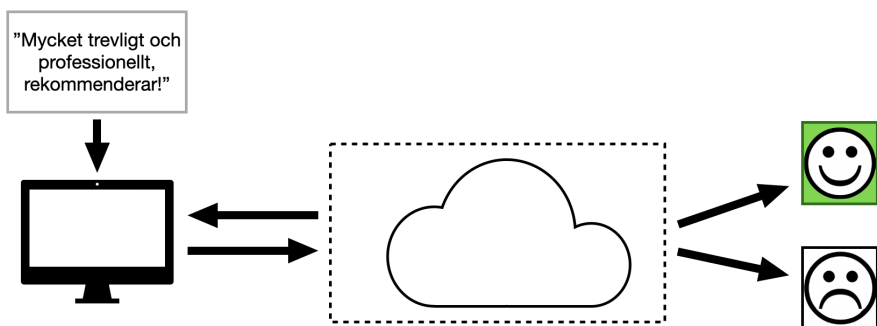


Maskininlärning för analys av svensk text - en jämförande studie av modeller och molntjänster

Hur kan organisationer använda maskininlärning för att analysera skriven text? Det här arbetet undersöker olika tillämpningar av maskininlärning och molntjänster för sentimentanalys, och utforskar hur företag kan utnyttja dessa teknologier för att få en djupare förståelse för kundernas åsikter och preferenser.

Att förstå den underliggande meningen eller attityden i en skriven text är vad som utgör en så kallad sentimentanalys. I takt med den ökande mängden textdata som idag finns att tillgå har automatiserad sentimentanalys blivit ett värdefullt verktyg när företag ska förstå sina kunder och deras åsikter. Till exempel kan det appliceras på stora mängder kundrecensioner eller foruminlägg, som illustrerat i Figur 1 nedan, för att ge företag insyn i vad kunderna egentligen tycker om företagets produkter eller tjänster. För att kunna nyttja de här värdefulla funktionerna inom en organisation används de ofta i samband med molntjänster som gör det möjligt att bygga, distribuera och lagra de verktyg som används. Arbetet går igenom några av dagens mest populära hjälpmedel, och genom jämförelser undersökt vilka avvägningar som bör göras när det appliceras på svenska texter.



Figur 1 Sentimentanalys för att upptäcka positiva eller negativa attityder i en text. Användaren matar in den text som ska analyseras och får automatiskt tillbaka ett svar från modellen.

Det finns en mängd verktyg för sentimentanalys som anpassats för olika förkunskapsnivåer. För att snabbt komma igång utan krav på att kunna bygga egna modeller finns populära alternativ som erbjuder färdigställda lösningar. Om användaren dock vill bygga en egen lösning finns istället olika plattformsalternativ. Dessa är tjänster som, till skillnad från färdiga lösningar, låter användaren själv bygga, träna och distribuera sina modeller. I jämförelsen mellan de två alternativen är det viktigt att ta hänsyn till specifika avvägningar. Om resurser är begränsade, eller om det behövs en enkel och okomplicerad molnlösning, kan en färdig molntjänst vara det bästa valet. Men om det behövs mer anpassning, kontroll och skalbarhet, kan egenbyggda molnlösningar vara det bättre alternativet på lång sikt.

I arbetet jämförs därför de färdiga tjänsterna mot en rad egenbyggda modeller som sträcker sig från enkla regressionsmetoder till komplexa språkmodeller. Där själva jämförelsen baserades på ett fiktivt användarscenario, som genom ett dataset bestående av verkliga recensioner (skrivna på svenska) simulerade hur modellerna skulle fungera om de sattes upp i produktion.

Efter jämförelsen står det klart att färdiga lösningar (som inte är anpassad för svenska) når ett förhållandevis bra resultat, men når inte upp till de nivåer som nås med de egenbyggda modellerna. Att använda de färdiga metoderna kunde även, i de flesta fallen, kopplas till en högre kostnad. Resultatet visar dock på att det finns övervägningar som bör göras då vissa färdiga metoder kan vara mer kostnadseffektiva sålänge de används på en mindre mängd data. Intressant nog visade det sig också att det inte var de största och mest komplexa av de egenbyggda modellerna som visade allra mest lovande resultat. Utan att enklare modeller både kunde visa på bättre prestanda samt lägre kostnader.

Av arbetet framgår att sentimentanalys är ett komplext men användbart forskningsområde för att hjälpa företag att identifiera trender och effektivisera sitt agerande. När fler organisationer vänder sig till maskininlärning för sina behov av sentimentanalys är det troligt att vi kommer att fortsätta se framsteg på området, med ännu mer sofistikerade metoder och verktyg som utvecklas för att hjälpa företag att bättre förstå sina kunder.