

# Examensarbete 30 hp i data science

## *Using NLP methods to generate clinical variables from free text fields in patient records*

Region Västerbotten arbetar för att god hälsa och hållbar utveckling ska stärka varandra. Vi tar ansvar för en jämlik välfärd och för att forskning och innovation ger resultat. För att klara morgondagens utmaningar vill vi i Region Västerbotten möjliggöra mer tid till vård genom smarta servicelösningar och digitalisering. Centrum för informationsteknik och medicinsk teknik har som syfte att arbeta långsiktigt och modernt med rätt förmågor för att leverera och realisera regionens behov av utveckling, systemförvaltning, verksamhetsnära teknik och IT-infrastruktur. Vi är ca 300 anställda och finns representerade på länets tre sjukhus - i Lycksele, Skellefteå och Umeå. Vi bidrar med all form av teknik för att stötta hela regionen och i samverkan med dem bidrar vi till digitaliseringen av vården. Vi jobbar medvetet för att skapa ett högt värde för individen och en stolthet över det vi gör tillsammans.

## Bakgrund

AMHOS är ett paraplyprojekt på Region Västerbotten med målsättning att integrera **AI** och **Maskininläring** i **Hälsa- Och Sjukvård**. Delprojekten inom AMHOS innefattar att konstruera maskininlärningsmodeller med tillhörande pipelines för att prediktera vårdtid och -tyngd på postoperativ avdelning (PostOP) samt intensivvårdsavdelning (IVA). Prediktionerna görs med kliniska patientdata, vilka måste genomgå lämplig förbehandling för att reducera dimensionaliteten av datan och öka den *effektiva* komplexiteten.

En utmaning med patientdatan avser de fritextfält som finns i journalen. Fritextfälten kan delas in i två typer av kommentarsfält - *övergripande kommentarer* och *nivåspecifikation*. En övergripande kommentar avser övergripande och ograderad information rörande något kliniskt område. Ett exempel är ”övrig personal vid operation” där det kan finnas angivet huruvida en läkarstudent eller en särskild typ av specialist varit närvarande vid operationstillfället. En nivåspecifikation avser en kommentar med någon typ av gradering. Ett exempel är ”anestesivarning kommentar” där olika typer av varningar påverkar anestesin i olika grad.

Natural language processing (NLP) är ett samlingsnamn för metoder med syfte att extrahera information från fritext. Två huvudmetoder intressanta för ovan specificerade applikation avser *keyword extraction* [1] och *sentiment analysis* [2].

## Målsättning

Målet med examensarbetet är att använda olika NLP-metoder för att generera nya, informativa variabler för vårdtid på PostOP, baserat på tillgängliga fritextvariabler.

## Arbetsbeskrivning

För att identifiera viktiga, nya variabler baserat på fritextfälten skall arbetet inledas med en studie av tillgängliga data kombinerat med en undersökning av klinisk relevans. Baserat på detta skall förslag på nya variabler tas fram baserat på *minst* tre olika, befintliga fritextvariabler. Av de befintliga fritextvariablerna skall det identifieras vilka som utgör övergripande kommentarer och vilka som utgör nivåspecifikation. Krav på examensarbetet är att minst en av var typ används. Detta förväntas generera 5-10 nya variabler. Antalet nya variabler bestämmer studenten i samråd med handledare för att hålla lämplig nivå på arbetet.

För de förstnämnda skall lämpliga metoder av keyword extraction identifieras och appliceras. För de sistnämnda skall motsvarande för sentiment analysis undersökas. Var metod skall vara specifikt anpassad för den avsedda variabeln och där det är nödvändigt skall flera metoder utvärderas.

Detta arbete karaktäriseras av ett friare upplägg och bör justeras dynamiskt efter tillgänglig tid och de resultat som erhålles eftersom. Studenten bör vara ansvarstagande och ha god planeringsförmåga. Vidare behöver studenten ha en grundläggande förståelse för maskininlärning samt tidigare erfarenhet i analys av stora datamängder. Arbetet skall genomföras i Python.

## Referenser

[1] <https://www.analyticsvidhya.com/blog/2022/03/keyword-extraction-methods-from-documents-in-nlp/>

[2] <https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment-analysis/>

### Handledare vid IT (preliminärt)

Amanda Bertgren

CIMT, IT: Forskningsutveckling

Region Västerbotten

E-post: [amanda.bertgren@regionvasterbotten.se](mailto:amanda.bertgren@regionvasterbotten.se)