

AI-sammanfattade klagomål på hälso- och sjukvård

Av David Aho och Leah Nhieu

Introduktion



Bakgrund

Detta examensarbete utfördes i samarbete med Patientnämnden Skåne, som är en av Region Skånes förvaltningar. En av Patientnämndens uppgifter är att sammanfatta patienters klagomål om hälso- och sjukvård. Patientnämnden står inför två utmaningar med denna uppgift. För det första är uppgiften tidskrävande. För det andra varierar kvaliteten på sammanfattningarna beroende på vem som är skribent. Patientnämnden vill undersöka ifall AI kan lösa dessa två utmaningar.

Syfte

Syftet med detta examensarbete är att undersöka vilka AI-modeller som kan generera bra sammanfattningar på klagomål.

Avgränsning

1. Ingen tillgång till Patientnämndens lagring av klagomål och mänskligt skriva sammanfattningar.
2. Internetbaserade modeller fick inte användas.
3. Begränsning av hårdvara

Terminologi

AI - Artificial intelligence. Datorns förmåga att utföra uppgifter som normalt kräver mänsklig intelligens.

Dataset - En kollektion av data som en modell använder för att utveckla sin förmåga i en uppgift.

Finjustering - När en förtränad modell tränas vidare på dataset för att lära sig eller för att förbättra sin förmåga i att utföra en specifik uppgift.

Modell - En matematiskt och beräkningsmässig representation av AI som är tränad på data för att utföra specifika uppgifter.

Extraktiv sammanfattning - En textsammanfattning som består av de beräknat viktigaste meningarna från originaltexten.

Abstrakt sammanfattning - En textsammanfattning bestående av meningar som inte finns i originaltexten.

Fabricerat klagomål - Ett verkligt klagomål, vilket har modifierats av Patientnämnden för att uppnå anonymitet.

Zero-shot - En modells förmåg att utföra uppgifter som den inte är specifikt finjusterad för.

Metod

Kriterier

Examensarbetarna diskuterade med Patientnämnden för att fastställa vilka kriterier som definierar en bra sammanfattning.

Dataset

Datasetet cnn_dailymail bedömdes vara mest lämpligt för sammanfattning av klagomål, jämfört med andra offentligt tillgängliga dataset.

Modeller

Sedan kartlades olika modeller som kan utföra textsammanfattning. För abstrakt sammanfattning valdes varianter av modellerna **BART**, **PEGASUS**, **T5**, **Flan-T5** och **GPT-Sw3**. För extraktiv sammanfattning valdes **TextRank**.

Till skillnad från andra modeller kunde PEGASUS och T5 inte hantera det svenska språket men de kunde hantera engelska. Därför användes två andra AI-modeller för att översätta klagomål respektive generad sammanfattning mellan svenska och engelska. Utav modellerna var BART, PEGASUS och TextRank specifikt designade för textsammanfattning.

Generering

För varje modell genererades sammanfattningar på de 29 fabricerade klagomålen erhållna från Patientnämnden.

Utvärdering

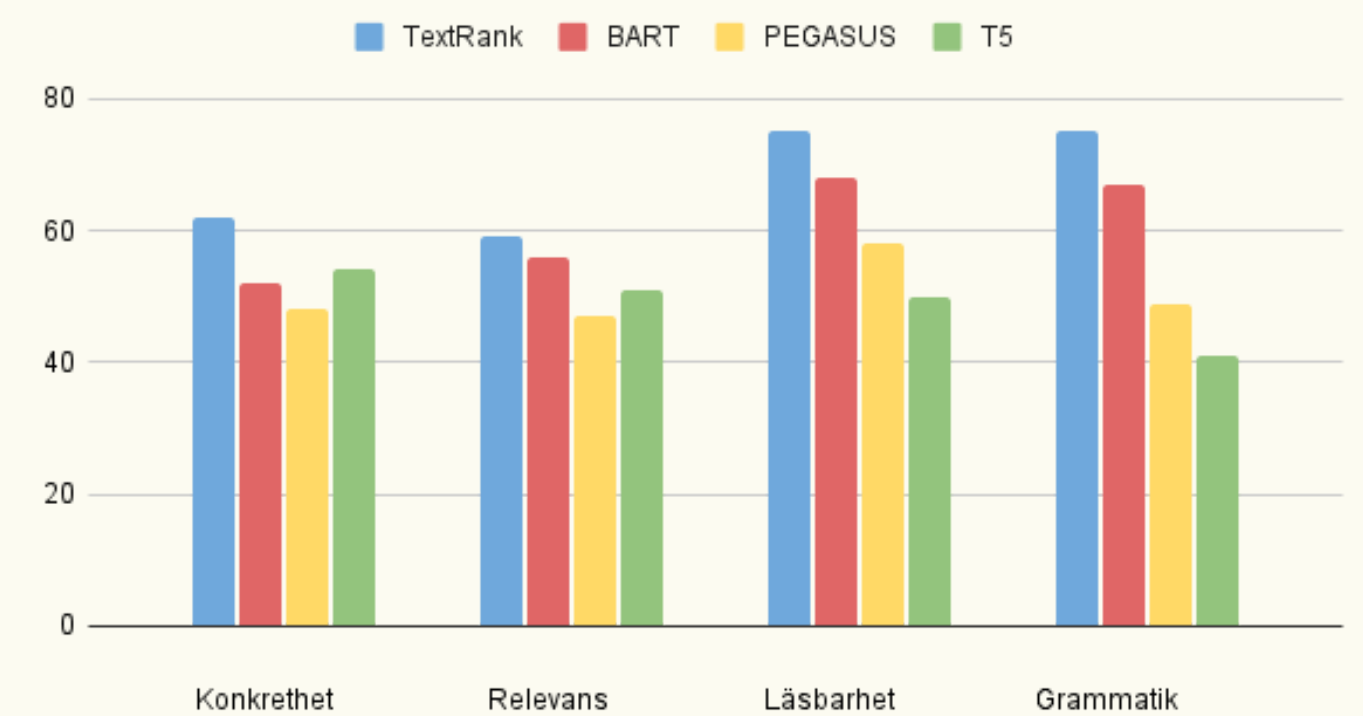
Sammanfattningarna av BART, PEGASUS, T5 och TextRank skickades in till Patientnämnden för utvärdering. Poängsystemet 1(låg) till 3(hög) användes för att utvärdera varje sammanfattning utifrån de fastställda kriterierna. Ju högre poäng en modell har, desto bättre var dess prestation.

Utmaningar

- På grund av avgränsning 1 kunde endast ett mindre lämpligt dataset användas.
- På grund av avgränsning 2 och 3 kunde endast nedladdningsbara modeller med färre parameter användas.
- Fåtal tillgängliga modeller som kan hantera det svenska språket.
- Gräns på hur lång respektive kort text en modell kan sammanfatta.



Modellers prestationer



Resultat

Resultatet var tydligt att de genererade sammanfattningarna från Flan-T5 och GPT-Sw3 inte uppnådde kriterierna för en bra sammanfattning. Därför valdes att Patientnämnden inte skulle utvärdera dessa modeller. Utav modellerna BART, PEGASUS, TextRank och T5 bedömde Patientnämnden att TextRank genererade bäst sammanfattningar på klagomål.

Diskussion

Problemformulering

Problem 1: Vilka kriterier definierar en bra sammanfattning?

- **Konkrethet:** sammanfattningen är kortfattad och innehållsrik.
- **Relevans:** sammanfattningen fokuserar på den mest relevanta informationen i originaltexten och undviker onödiga eller tangentiella detaljer.
- **Läsbarhet:** sammanfattningen är lätt att läsa och förstå för målgruppen.
- **Grammatisk korrekthet:** sammanfattningen har bra grammatik och välformade meningar.

Problem 2: Vilka modeller är lämpliga för Patientnämndens ändamål?

En modell som kan hantera det svenska språket och är finjusterad för abstrakt sammanfattning. Det ideala datasetet för finjustering borde innehålla par av klagomål och Patientnämndens mänskligt skrivna sammanfattningar.

Problem 3: Med hänvisning till kriterierna för en bra sammanfattning, vilken modell ger lämpligast resultat för Patientnämndens behov?

Utifrån resultatet från utvärderingen gav TextRank lämpligast resultat för Patientnämndens behov.

Framtida möjligheter

- Finjustera BART på det ideala datasetet som beskrivs i svaret för problem 2.
- För närvarande är GPT-Sw3 under utveckling. Om Patientnämnden inte vill finjustera en modell, finns möjligheten att använda framtida varianter av GPT-Sw3 på grund av dess förmåga i zero-shot scenarier.

