Sverige Brott Data



Gabriel Jara

EC Utbildning

2024/10/16

# Abstract

This project aimed to analyze and visualize crime data from across Sweden, gathered via the Polisen.se API.

The data was stored in a SQL server, cleaned, and processed for analysis.

We trained a T5 model to generate articles based on crime types and events, fine-tuning it with collected crime data and a columnists writing style from Aftonbladet.

The model was intended to produce cohesive, and insightful articles.

Additionally, we used Power BI to create interactive visualizations, including a heatmap, to highlight areas with high crime activity. These visual tools provide valuable insights that can help optimize police resource allocation.

Innehållsförteckning

[Abstract 2](#_Toc179987327)

[1 Inledning 1](#_Toc179987328)

[. 1](#_Toc179987329)

[2 Metod 2](#_Toc179987330)

[2.1 Agil arbetsmetodik 4](#_Toc179987331)

[3 Resultat 5](#_Toc179987332)

[4 Diskussion 6](#_Toc179987333)

[5 Slutsatser 7](#_Toc179987334)

[6 Självutvärdering 8](#_Toc179987335)

# Inledning

I dagens samhälle är brottsligheten ett ständig närvarande problem som påverkar säkerheten och tryggheten för invånare i hela landet. För att underlätta polisens arberte och förbättra resurdfördelningen, har vi genonm fört en datainsamling via en API från polisen.se. Genom denna insamling har vi fått en viss insyn i dom olika brott som rapporteras, allt från trafik brot till mer allvarliga våldsbrott i hela Sverige. Syftet med det här projektet är att analysera och visualisera på en karta vart flest brot begås och visa dessa resultat på en karta. Med hjälp av detta kan polisen lättare se var dom bör fokusera sina resurser för att förebygga eller bekämpa brott på ett effektivare sätt.

## .

# Metod

Vi har hämtat datan via polisens öppna API (https://polisen.se/api/events), som tillhandahåller rapporter om olika händelser och brott runtom i Sverige.

A computer screen with text

Description automatically generated

Den insamlade datan har lagrats på en SQL-server, där vi sedan rensade och bearbetade informationen för att extrahera relevanta brottstyper och händelser.

A group of people on a white background

Description automatically generated

För att skapa en djupare förståelse för brottsdata, har vi använt en T5-modell (Text-to-Text Transfer Transformer) för att generera artiklar baserade på de olika brottstyperna och händelserna från polisrapporterna. Modellen finjusterades med hjälp av den insamlade brottsdata och tränades för att kunna skriva sammanhängande artiklar i stil med en kolumnist från Aftonbladet.

Efter bearbetning av textdata, genomförde vi en analys och visualisering av brottsinformationen i Power BI. I den interaktiva visualiseringen skapade vi bland annat en heatmap, som visar var i Sverige brottsligheten är mest utbredd. Dessa visualiseringar ger en tydlig bild av områden där brott förekommer frekvent och kan användas som stöd för att allokera polisresurser mer effektivt.

## Agil arbetsmetodik

* Vi hade dagliga möten för att diskutera projektets framsteg och eventuella ändringar som behövde göras. Detta reflekterar den agila principen om frekvent och effektiv kommunikation inom teamet, vilket hjälper oss att snabbt identifiera förbättringsområden och justera arbetet.
* Teammedlemmarna arbetade utifrån sina styrkor, vilket stämmer överens med agila principer om att ha självorganiserande team där varje individ bidrar utifrån sina unika färdigheter och erfarenheter.
* Vi visade flexibilitet genom att anpassa oss till skolans tidschema och omfördela uppgifter när någon inte kunde fullfölja en task. Detta säkerställde att projektet fortskred utan hinder, vilket reflekterar den agila förmågan att snabbt reagera på förändringar.
* Projektet delades upp i mindre, hanterbara delar som möjliggjorde individuellt arbete samtidigt som vi arbetade mot ett gemensamt mål. Detta speglar principen om att leverera fungerande delar av projektet löpande.
* Efter varje större del av projektet reflekterade vi över vårt arbetssätt och insåg utmaningar, som till exempel svårigheten att bygga en chatbot baserad på en kolumnist. Genom dessa reflektioner strävade vi efter att ständigt förbättra vårt arbetssätt och processer.



# Resultat

**Datainsamling:** Vi lyckades samla in brottsdata från hela Sverige via Polisen.se API och lagrade denna data i en SQL-server. Efter datarensning hade vi en strukturerad dataset med brottstyper och händelser som täcker allt från trafikbrott till våldsbrott. Totalt samlade vi in [ange antal] datapunkter, vilket gav oss en omfattande överblick över brottsaktiviteten i olika delar av Sverige.

**Textgenerering:** T5-modellen användes för att försöka generera artiklar baserade på brottstyperna, men vi stötte på betydande svårigheter. Modellen lyckades inte skapa sammanhängande eller meningsfulla texter, och resultaten var ofta osammanhängande och saknade den stil vi eftersträvade från kolumnisten. Vi tror att en av de främsta orsakerna till detta var att datamängden vi hade tillgänglig inte var tillräckligt stor eller varierad för att träna modellen på ett effektivt sätt. För att förbättra textgenereringen hade vi sannolikt behövt en större dataset och mer detaljerad finjustering av modellen.

**Visualisering:** Med hjälp av Power BI skapade vi interaktiva visualiseringar som en heatmap, som visade var brottsligheten var som mest utbredd. Heatmapen gav oss insikter om var polisen bör fokusera sina resurser, till exempel att vissa områden hade en hög koncentration av våldsbrott

medan andra hade fler trafikbrott.

# Diskussion

**Framgångar:** Genom vår datainsamling och visualisering fick vi fram en klar bild av var i landet brottsligheten var mest utbredd. Visualiseringarna visade sig vara särskilt användbara i att identifiera brotts-hotspots, vilket överensstämde med vårt initiala mål. Vi lyckades även samla in en omfattande mängd brottsdata och skapa meningsfulla visualiseringar som kan användas som beslutsunderlag för polisen.

**Utmaningar:** Den största utmaningen var att få T5-modellen att generera sammanhängande artiklar. Trots att vi tränade modellen med den tillgängliga datan lyckades vi inte uppnå önskade resultat. En av de främsta orsakerna var förmodligen bristen på tillräcklig mängd och variation i datasetet. Modellen hade även svårt att förstå kontexten i texten, vilket ledde till osammanhängande och icke-meningfulla resultat.

**Lärdomar:** Vi har insett att bygga en chatbot baserad på en kolumnist är svårare än förväntat, särskilt när mängden träningsdata är begränsad. Textmodeller som T5 kräver omfattande och varierad data för att producera sammanhängande och stilistiskt korrekta texter. Visualiseringarna var dock framgångsrika och gav oss tydliga insikter i brottsmönster, vilket bekräftar värdet av datadriven visualisering för brottsförebyggande arbete.

Förbättringsmöjligheter: För att förbättra textgenereringen skulle vi behöva mer data och en mer detaljerad finjustering av T5-modellen. Ett större dataset, som täcker fler brottstyper och mer varierad text, skulle sannolikt bidra till bättre resultat. För visualiseringarna skulle vi kunna gå vidare med att undersöka mer avancerade metoder, till exempel att analysera brottslighetens utveckling över tid eller skapa prediktiva modeller för att förutse var framtida brott kan ske.

# Slutsatser

I det här projektet har vi fått värdefulla insikter och resultat kring hur brottsligheten är geografiskt fördelad i Sverige och hur vi kan använda teknologi för att analysera och visualisera brottsmönster. Vi lyckades samla in och visualisera brottsdata från Polisen.se, och skapade en interaktiv heatmap som tydligt visar var brotten är mest koncentrerade. Denna visualisering kan hjälpa polisen att fokusera sina resurser där de behövs som mest.

När det gäller att generera artiklar baserade på brottstyper med hjälp av en T5-modell stötte vi dock på fler problem än väntat. Modellen kunde inte producera sammanhängande texter, och vi insåg att vi skulle ha behövt mer data och bättre justeringar för att få bättre resultat.

Sammanfattningsvis har vi visat hur kraftfull datadriven analys och AI kan vara för att förstå brottslighet, men vi har också sett att det krävs mer arbete, framför allt när det kommer till att använda AI för att skriva texter. I framtiden skulle mer data och bättre träning av modellerna kunna ge ännu bättre resultat både för textgenerering och för att analysera brottsligheten mer i detalj.

# Självutvärdering

1. Under arbetets gång stötte vi på flera utmaningar.. En av de största var att få T5-modellen att generera sammanhängande och meningsfulla texter baserade på brottsdata. Modellen kämpade med att producera text som höll en konsekvent stil och sammanhang, vilket vi tror beror på att vi inte hade tillräckligt med träningsdata och att finjusteringen av modellen inte var tillräcklig. För att hantera detta försökte vi justera modellen och testa olika dataset, men det blev tydligt att vi skulle behöva en större och mer varierad datamängd för att nå bättre resultat.
2. VG för jag uppfyller dom kriterierna satta